In silico design of novel inhibitors of dengue virus replication

A thesis submitted in accordance with the conditions governing candidates for the degree of Philosophiae Doctor in Cardiff University

by

Iuni Margaret Laura Trist

December 2014
Cardiff School of Pharmacy and Pharmaceutical Science
Cardiff University
“Fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza”

(“Ye were not made to live like unto brutes,
But for pursuit of virtue and of knowledge.”
vv. 119-120, Canto XXVI, Inferno, The Divine Comedy, D. Alighieri)
DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed [Signature] (candidate) Date 17/12/2014

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD

Signed [Signature] (candidate) Date 17/12/2014

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed [Signature] (candidate) Date 17/12/2014

STATEMENT 3: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University’s Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed [Signature] (candidate) Date 17/12/2014
ACKNOWLEDGEMENTS

So, here I am at the end of my PhD. I almost cannot believe it! It went so quickly! Maybe that’s because “time passes quickly when you are having fun!!”. It’s now time to collect all my work (and my stuff from the lab: there will be a lot of space!!!!) and to thank all the people that somehow helped me to get here. But before starting, I must apologise with the reader! I am bilingual and I am used to change language in the middle of the conversation at home (anyone who has been there or who has seen me with my parents can confirm it). For this, I have decided to write the acknowledgements in line with my style: mixing Italian and English together!

First of all I would like to thank my supervisor, Dr Andrea Brancale for giving me the opportunity to join his research group, patiently encouraging me to find the way through the problems. A big thank you also for involving me in many different things and giving me a lot of trust! Grazie anche, Dr Brande, per le risate durante le pause caffè, per avermi fatto conoscere Darren Brown, per il bowling e per aver cercato (mi dispiace, inutilmente) di insegnarmi a giocare a biliardo.

I would like to thank the Marie Curie Action FP7 program for sponsoring my PhD as well as all the EUVIRNA consortium members and organisers (Prof Frank van Kuppeveld and Dr Clasien Oomen in particular) for giving me the great opportunity to get such broad and multidisciplinary background with scientific and complementary training. It has been tough and particularly full of paperwork, but it was absolutely worthwhile. Huge thanks also to my peers for the great time spent together at the network meetings, making them really precious occasions and for all the encouragement.

An enormous thank you goes to our collaborators at Marseille University, KU Leuven and Heidelberg University, without which the PhD study could have not been concluded. In particular, I would like to thank Prof. Johan Nytes, Dr Pieter Leyssen and Dr Suzanne Kaptein for performing the cell-based assays at the Rega Institute for Medical Research (KU Leuven) and Pietro Scaturro and Prof. Ralf Bartenschlager at Hidelberg University for the stimulating collaboration on the ST-148 mode of action study. A big thank you to all the people who have helped me at AFMB Marseille University during and after my secondment. Above all, I would like to thank Dr Bruno Coutard for the great coordination of the collaboration on the NS3hel and NS5 RdRp projects and for all the support and help he has given me during my secondment period. Thank you to Julie for preparing all the protein samples I needed, Prof. Etienne Decroly for the precious advices, Axelle Collet for patiently teaching me the FP assay, Changqing Li for helping me with the ATPase assay and for performing some of the biochemical assays on NS3hel and Cécilia Eydoux and Prof. Jean-Claude Guillemot for carrying
out the NS5 RdRp enzymatic assays. A big “grazie” also to Lorenzo Subissi for making me feel at home also in Marseille. I would also like to thank a lot Dr Karine Barral at AFMB, Marseille University for involving me in the interdisciplinary fragment-based drug discovery study on NS5 MTase.

The success of the work of a person does not depend only on the amount of effort, of hours spent working (mainly in front of the computer in my case), or on the quality of the collaboration teams that that person has. I strongly believe that the environment in which that person is working is just as important. For this reason I really want to thank all the people of the lab! Independently from the time spent in Cardiff at the School of Pharmacy, every one of you has been precious for me.

Ευχαριστώ Thanasis! For the chats over the longest coffees/teas ever, for introducing me to GPUs and for the surf lesson. It was real fun!! A huge thank you, Samia for you friendship, all the kind words and the support you are still giving me also now that you are back in Egypt. I promise I’ll come to visit you one day, you promise me you will keep being so curious about the World. A big thank you also to all the fantastic visitors that have passed by in these three years Roberta, Simona (Rabito), Andrea, Alberto, Valeria, Daniel, Maria. Every one of you has somehow personalised the lab and has taught me something.

Non posso non ringraziare Willi Wonka e tutti gli Oompa Loompa per il miglior travestimento di sempre! Peccato non ci sia stato un sequel. Ringrazio tanto la mitica compagnia del pranzo Salvo, Marcella, Michela, Cinzia e Cecilia per le ore passate di fronte al microonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista! Un grazie allo zio Davidino, per le risate, le avventure vagabonde e per avermi fatto scoprire che se andasse male con il modelling potrei sempre provare a fare la barista!

Anche se sono a Cardiff da ormai quasi quattro anni, in Italia mi sono rimaste molte persone care che mi fanno sentire un po’ come se non fossi mai partita. Vorrei ringraziarle tutte, in particolare, la mitica Francesca, storica compagna di avventure, ed il suo Roberto per “rubarmi” sempre a mia mamma (lei forse ringrazia un po’ meno...) quando torno a San Bonifacio e per il continuo incoraggiamento.
Ora sono in difficoltà. What language should I use to thank my family? Il grazie più grande va sicuramente a chi mi sopporta da ormai quasi 28 anni, cioè alla mia famiglia: Sandra, David e Riccardo. Non ci sono parole per descrivere quanto prezioso sia il vostro supporto. Voi ci siete sempre: dalle risate per un caffè macchiato senza latte ai crampi al braccio per le lunghissime telefonate; dal “pulling my leg” all’ incoraggiarmi ad “andare lontana da casa” nonostante il distacco sia molto difficile e al sopportarmi nei miei momenti di nervosismo e scontrosità. Grazie soprattutto per avermi insegnato l’amore per la conoscenza e per avermi trasmesso la curiosità, qualità senza le quali non sarei mai arrivata qui. Simply, thank you!! Un grazie speciale anche a Smoky per tutti i giochi riportati!

Last, but not least... Un immenso grazie ad Antonio, mia metà di due (cit.)! Grazie di essere così pazzerello, per tutti i sorrisi che mi fai tornare, per le chiacchierate chimiche e quantistiche, per il continuo supporto e forza che mi dai giorno per giorno, rendendo il mio mondo così speciale.

In summary: DIOLCH YN FAWR!
SUMMARY

Dengue virus (DENV) is a health burden responsible of 50-100 million new cases and 22,000 deaths per year and its four serotypes are worryingly spreading out of the endemic regions. Current therapy is symptomatic, making antiviral research on DENV an unmet need. Vaccine development is more challenging than expected, so the development of anti-DENV drugs is particularly important for infection management. DENV is a positive sense single stranded RNA virus that replicates within cells exploiting both host and viral enzymes to replicate.

Based on the hypothesis that DENV infection can be stopped with the inhibition of one or more of the enzymes that are fundamental for its replication, the aims of the studies reported in this PhD thesis were to: identify novel targets to combat DENV infection, generate new basic knowledge and discover potential novel chemical leads exploiting those targets. Novel approaches combining molecular modelling techniques, classical Medicinal Chemistry approaches, chemical synthesis and in vitro assays were applied to four essential viral-encoded proteins: the capsid (C), the NS3 NTPase/helicase (NS3hel), the NS5 methyltransferase (NS5 MTase) and the NS5 RNA-dependent RNA polymerase (NS5 RdRp).

Novel understanding of the 3′-5′ translocation mechanism of NS3hel along the RNA has been hypothesised, increasing awareness about DENV-encoded proteins.

Important knowledge on the mode of action of promising antiviral compounds has been acquired, for example that ST-148 stabilises C protein-protein interactions and that published N-sulphonylanthranilic acid RdRp inhibitors bind to a unique allosteric site.

Novel promising DENV inhibitor scaffolds have also been developed and the chemical synthesis of one of them has been described, showing that the adopted drug discovery approaches are suitable starting points for the development of anti-DENV medicines.

The results obtained represent a significant contribution to DENV research in increasing basic knowledge and in identifying good chemical leads for future work.
LIST OF CONTENTS

DECLARATION II
ACKNOWLEDGEMENTS III
SUMMARY VI
LIST OF CONTENTS VII
ABBREVIATIONS XI
GENERAL ABBREVIATIONS XI
AMINO ACIDS ABBREVIATIONS XVI
NUCLEOTIDES ABBREVIATIONS XVI

SECTION 1: INTRODUCTION 1

1.1 DENGUE VIRUS 2
EPIDEMIOLOGY 2
TRANSMISSION 4
TAXONOMY 6
CLINICAL FEATURES 8
BIOLOGY 9
DIAGNOSIS AND CURRENT TREATMENT 15
PREVENTION 16
FUTURE POTENTIAL TREATEMENTS 17

1.2 MOLECULAR MODELLING 29
MOLECULAR MODELING AND DRUG DISCOVERY 29
QUANTUM MECHANICS 32
MOLECULAR MECHANICS 33
STRUCTURE-BASED DRUG DESIGN 36
LIGAND-BASED DRUG DESIGN 40
MOLECULAR DYNAMICS SIMULATION 42

1.3 INTRODUCTION TO THE BIOLOGICAL ASSAYS 44
CYTOPATHIC EFFECT (CPE) INHIBITION ASSAY 44
THERMAL SHIFT ASSAY (TSA) 45
FLUORESCENCE POLARISATION (FP) ASSAY 46
HELICASE ATPASE ACTIVITY INHIBITION ASSAY 49
HELICASE UNWINDING ACTIVITY INHIBITION ASSAY 50
POLYMERASE ACTIVITY INHIBITION ASSAY USING RADIOACTIVE PROBES 51
POLYMERASE ACTIVITY INHIBITION ASSAY USING PICOGREEN 52
METHYLTRANSFERASE 2’-O AND N-7 METHYLATION ACTIVITY INHIBITION ASSAY 53

1.4 AIMS AND OBJECTIVES 54

1.5 BIBLIOGRAPHY 55
### SECTION 4: NS3 HELICASE MECHANISM OF TRANSLOCATION ALONG RNA

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 INTRODUCTION</td>
<td>141</td>
</tr>
<tr>
<td>NS3 HELICASE MECHANISM OF TRANSLOCATION</td>
<td>141</td>
</tr>
<tr>
<td>AIMS AND OBJECTIVES</td>
<td>142</td>
</tr>
<tr>
<td>4.2 RESULTS AND DISCUSSION</td>
<td>144</td>
</tr>
<tr>
<td>CLASSICAL MOLECULAR DYNAMICS SIMULATIONS</td>
<td>144</td>
</tr>
<tr>
<td>ACCELERATED MOLECULAR DYNAMICS SIMULATIONS</td>
<td>157</td>
</tr>
<tr>
<td>4.3 CONCLUSION</td>
<td>180</td>
</tr>
<tr>
<td>4.4 METHODS</td>
<td>181</td>
</tr>
<tr>
<td>HARDWARE DETAILS</td>
<td>181</td>
</tr>
<tr>
<td>SYSTEM PREPARATION</td>
<td>181</td>
</tr>
<tr>
<td>CLASSICAL MOLECULAR DYNAMICS SIMULATION</td>
<td>181</td>
</tr>
<tr>
<td>ACCELERATED MOLECULAR DYNAMICS SIMULATIONS</td>
<td>183</td>
</tr>
<tr>
<td>PRINCIPAL COMPONENT ANALYSIS</td>
<td>184</td>
</tr>
<tr>
<td>BINDING FREE ENERGY CALCULATIONS</td>
<td>185</td>
</tr>
<tr>
<td>4.5 BIBLIOGRAPHY</td>
<td>187</td>
</tr>
</tbody>
</table>

### SECTION 5: NS5 METHYLTRANSFERASE AS A DRUG TARGET

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 INTRODUCTION</td>
<td>191</td>
</tr>
<tr>
<td>A PROMISING DRUG TARGET</td>
<td>192</td>
</tr>
<tr>
<td>STRUCTURE</td>
<td>193</td>
</tr>
<tr>
<td>POSSIBLE MOLECULAR TARGET SITES</td>
<td>195</td>
</tr>
<tr>
<td>MECHANISM OF ACTION</td>
<td>198</td>
</tr>
<tr>
<td>A FRAGMENT-BASED DRUG DESIGN STUDY</td>
<td>199</td>
</tr>
<tr>
<td>AIMS AND OBJECTIVES</td>
<td>200</td>
</tr>
<tr>
<td>5.2 RESULTS AND DISCUSSION</td>
<td>201</td>
</tr>
<tr>
<td>5.2.1 FRAGMENT GROWING DRUG DESIGN</td>
<td>201</td>
</tr>
<tr>
<td>5.2.2 FRAGMENT LINKING DRUG DESIGN</td>
<td>214</td>
</tr>
<tr>
<td>5.2.3 FRAGMENT 2G3 MODE OF ACTION</td>
<td>229</td>
</tr>
<tr>
<td>5.3 CONCLUSIONS</td>
<td>234</td>
</tr>
<tr>
<td>5.4 METHODS</td>
<td>235</td>
</tr>
<tr>
<td>HARDWARE DETAILS</td>
<td>235</td>
</tr>
<tr>
<td>PROTEIN PREPARATION</td>
<td>235</td>
</tr>
<tr>
<td>DESIGN OF SMALL MOLECULE DATABASES</td>
<td>235</td>
</tr>
<tr>
<td>PHARMACOPHORE QUERIES</td>
<td>235</td>
</tr>
<tr>
<td>MOE MEDCHEM TRASFORMATIONS</td>
<td>236</td>
</tr>
<tr>
<td>MOE COMBINATORIAL BUILDER</td>
<td>236</td>
</tr>
<tr>
<td>LIGBUILDER</td>
<td>237</td>
</tr>
<tr>
<td>MOE LINK MULTIPLE FRAGMENTS</td>
<td>237</td>
</tr>
<tr>
<td>MOE DOCKING</td>
<td>238</td>
</tr>
<tr>
<td>MAESTRO GLIDE DOCKING</td>
<td>238</td>
</tr>
<tr>
<td>MOLECULAR DYNAMICS SIMULATIONS</td>
<td>239</td>
</tr>
<tr>
<td>5.5 BIBLIOGRAPHY</td>
<td>241</td>
</tr>
</tbody>
</table>
## SECTION 6: NS5 POLYMERASE AS A DRUG TARGET

### 6.1 INTRODUCTION
A PROMISING DRUG TARGET 245
STRUCTURE 245
POSSIBLE MOLECULAR TARGET SITES 247
MECHANISM OF ACTION 248
AIMS AND OBJECTIVES 249

### 6.2 RESULTS AND DISCUSSION
6.2.1 ALLOSTERIC INHIBITORS BINDING SITE D 250
6.2.2 ALLOSTERIC INHIBITORS BINDING SITE C 279

### 6.3 CONCLUSION
289

### 6.4 BIOCHEMICAL ASSAYS ON DENV NS5 POLYMERASE
291

### 6.5 COMPUTATIONAL METHODS
294
HARDWARE DETAILS 294
HOMOLOGY MODELLING 294
SITE D VIRTUAL SCREENING INPUT PREPARATION 295
MAESTRO GLIDE DOCKING SIMULATIONS 295
FLEXX AND PLANTS RESCORING 296
CONSENSUS SCORE 296
N-SULPHONYLANTHRANILIC ACID DERIVATIVES SAR STUDY 298
SITE C VIRTUAL SCREENING INPUT PREPARATION 298
SITE C VIRTUAL SCREENING SHAPE FILTER 299
LEADIT FLEXX DOCKING SIMULATIONS 299
MOE AND PLANTS RESCORING 299

### 6.6 EXPERIMENTAL SECTION
301
6.6.1 BIOCHEMICAL ASSAYS 301
6.6.2 CHEMICAL SYNTHESIS 304

### 6.7 BIBLIOGRAPHY
314

## SECTION 7: GENERAL CONCLUSIONS

## SECTION 8: APPENDIXES

APPENDIX 1: VACCINES AND DRUGS IN DEVELOPMENT 323
APPENDIX 2: INTRODUCTION TO USED MOLECULAR MODELLING PROGRAMS 337
APPENDIX 3: SHAPE-BASED VIRTUAL SCREENING SELECTED MOLECULES 358
APPENDIX 4: NS3HEL NTP SITE VIRTUAL SCREENING SELECTED MOLECULES 360
APPENDIX 5: NS3HEL POTENTIAL INHIBITORS TESTED DURING THE SECONDMENT PERIOD 362
APPENDIX 6: FRAGMENTS SOLVED IN COMPLEX WITH NS5 METHYLTRANSFERASE 365
APPENDIX 7: COMPOUNDS DESIGNED FROM COMPOUND 44 MODIFICATIONS 366
APPENDIX 8: SELECTED MOLECULES FROM RDRP SITE D VIRTUAL SCREENING STUDY 368
APPENDIX 9: NS5 RDRP POTENTIAL INHIBITORS TESTED DURING THE SECONDMENT PERIOD 372

APPENDIX 10: PUBLICATIONS 375
**ABBREVIATIONS**

### GENERAL ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>Monodimensional</td>
</tr>
<tr>
<td>1H-NMR</td>
<td>Proton (1H nucleus) NMR</td>
</tr>
<tr>
<td>13C-NMR</td>
<td>Carbon (13C nucleus) NMR</td>
</tr>
<tr>
<td>2D</td>
<td>Bidimensional</td>
</tr>
<tr>
<td>3'dATP</td>
<td>3'-Deoxy Adenosyl Triphosphate</td>
</tr>
<tr>
<td>3D</td>
<td>Tridimensional</td>
</tr>
<tr>
<td>6-FAM</td>
<td>6-Carboxyfluorescein</td>
</tr>
<tr>
<td>α</td>
<td>AMBER Force Field-Dependent Constant</td>
</tr>
<tr>
<td>α</td>
<td>Tuning Parameter (in aMD calculations)</td>
</tr>
<tr>
<td>α\textsubscript{dihed}</td>
<td>Dihedral Tuning Parameter</td>
</tr>
<tr>
<td>α\textsubscript{pot}</td>
<td>Potential Energy Tuning Parameter</td>
</tr>
<tr>
<td>β</td>
<td>MMFF94x Force Field-Dependent Constant</td>
</tr>
<tr>
<td>βOG</td>
<td>β-N-octyl-glucoside</td>
</tr>
<tr>
<td>δ</td>
<td>Chemical Shift</td>
</tr>
<tr>
<td>ΔE\textsubscript{coul}</td>
<td>Electrostatic Component of the Binding Free Energy</td>
</tr>
<tr>
<td>ΔE\textsubscript{sol}</td>
<td>Solvation Component of the Binding Free Energy</td>
</tr>
<tr>
<td>ΔE\textsubscript{vdw}</td>
<td>Van der Waals Interaction Component of the Binding Free Energy</td>
</tr>
<tr>
<td>ΔSA\textsubscript{weighted}</td>
<td>Penalization for the Solvent Exposed Surface Area of the Ligand</td>
</tr>
<tr>
<td>ΔG</td>
<td>Binding Free Energy</td>
</tr>
<tr>
<td>ΔG\textsubscript{0}</td>
<td>Empirical Constant for Binding Free Energy Calculations</td>
</tr>
<tr>
<td>ΔG\textsubscript{bind}</td>
<td>Binding Free Energy</td>
</tr>
<tr>
<td>ΔG\textsubscript{conf}</td>
<td>Energy Involved in Conformational Changes</td>
</tr>
<tr>
<td>ΔG\textsubscript{rot}</td>
<td>Rotational Free Energy</td>
</tr>
<tr>
<td>ΔG\textsubscript{rotor}</td>
<td>Loss Of Entropy Linked to the Rotatable Bonds Freezing</td>
</tr>
<tr>
<td>ΔG\textsubscript{solv}</td>
<td>Solvent Effect Energy</td>
</tr>
<tr>
<td>ΔG\textsubscript{vib}</td>
<td>Energy Involved in Vibrational Modes Changes</td>
</tr>
<tr>
<td>ΔT\textsubscript{m}</td>
<td>Melting Temperature Shift</td>
</tr>
<tr>
<td>ΔV(r)</td>
<td>Bias (or Boost) Potential</td>
</tr>
<tr>
<td>a</td>
<td>Acceleration (in F=m*a)</td>
</tr>
<tr>
<td>a</td>
<td>Weighting Factor for E\textsubscript{rep}</td>
</tr>
<tr>
<td>ACO</td>
<td>Ant Colony Optimisation</td>
</tr>
<tr>
<td>A.D.</td>
<td>Anno Domini</td>
</tr>
<tr>
<td>ADE</td>
<td>Antibody-Dependent Enhancement</td>
</tr>
<tr>
<td>ADME</td>
<td>Adsorption, Distribution, Metabolism, Excretion</td>
</tr>
<tr>
<td>AFMB</td>
<td>Architecture Et Function Des Macromolécules Biologiques</td>
</tr>
<tr>
<td>AMBER</td>
<td>Assisted Model Building with Energy Refinement</td>
</tr>
<tr>
<td>aMD</td>
<td>Accelerated Molecular Dynamics</td>
</tr>
<tr>
<td>AMPPNP</td>
<td>Adenosine 5'- (β,γ-imido) Triphosphate</td>
</tr>
<tr>
<td>anchC</td>
<td>Anchored Capsid Protein</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>B</td>
<td>Weighting Factor for $E_{\text{elec}}$</td>
</tr>
<tr>
<td>BDV</td>
<td>Border Disease Virus</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Load Alignment Search Tool</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>BVDV</td>
<td>Bovine Viral Diarrhoea Virus</td>
</tr>
<tr>
<td>c</td>
<td>Weighting Factor for $E_{\text{pair}}$</td>
</tr>
<tr>
<td>C</td>
<td>Capsid Protein</td>
</tr>
<tr>
<td>CC</td>
<td>Non-Infected Cells Control</td>
</tr>
<tr>
<td>CC$_{50}$</td>
<td>Half Maximum Cytotoxic Concentration</td>
</tr>
<tr>
<td>CDC</td>
<td>Center for Disease Control</td>
</tr>
<tr>
<td>CG</td>
<td>Conjugated Gradient</td>
</tr>
<tr>
<td>CH$_3$COOH</td>
<td>Acetic Acid</td>
</tr>
<tr>
<td>CHCl$_3$</td>
<td>Chloroform</td>
</tr>
<tr>
<td>cMD</td>
<td>Classical Molecular Dynamics</td>
</tr>
<tr>
<td>CPE</td>
<td>Cytopathic Effect</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>CS</td>
<td>Consensus Score</td>
</tr>
<tr>
<td>CSFV</td>
<td>Classical Swine Fever Virus</td>
</tr>
<tr>
<td>d</td>
<td>Doublet</td>
</tr>
<tr>
<td>D1</td>
<td>Domain 1</td>
</tr>
<tr>
<td>D2</td>
<td>Domain 2</td>
</tr>
<tr>
<td>D3</td>
<td>Domain 3</td>
</tr>
<tr>
<td>DALYs</td>
<td>Disability-Adjusted Life Years</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DENV</td>
<td>Dengue Virus</td>
</tr>
<tr>
<td>DENV1</td>
<td>Dengue Virus Serotype 1</td>
</tr>
<tr>
<td>DENV2</td>
<td>Dengue Virus Serotype 2</td>
</tr>
<tr>
<td>DENV3</td>
<td>Dengue Virus Serotype 3</td>
</tr>
<tr>
<td>DENV4</td>
<td>Dengue Virus Serotype 4</td>
</tr>
<tr>
<td>DENV5</td>
<td>Dengue Virus Serotype 5</td>
</tr>
<tr>
<td>DF</td>
<td>Dengue Fever</td>
</tr>
<tr>
<td>DHF</td>
<td>Dengue Haemorrhagic Fever</td>
</tr>
<tr>
<td>DHODH</td>
<td>Dihydroorotate Dehydrogenase</td>
</tr>
<tr>
<td>DIHED</td>
<td>Average Dihedral Energy</td>
</tr>
<tr>
<td>DPEA</td>
<td>N,N-diisopropylethilamine</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothrethiol</td>
</tr>
<tr>
<td>dsRNA</td>
<td>Double Stranded RNA</td>
</tr>
<tr>
<td>DSS</td>
<td>Dengue Shock Syndrome</td>
</tr>
<tr>
<td>E</td>
<td>Envelope Protein</td>
</tr>
<tr>
<td>$E_{\text{th}}$</td>
<td>Threshold Energy (in aMD calculations)</td>
</tr>
<tr>
<td>$E_{\text{attr}}$</td>
<td>Attractive Energy Contribution</td>
</tr>
<tr>
<td>$E_{\text{bend}}$</td>
<td>Energy Related to Angle Bending</td>
</tr>
<tr>
<td>$E_{\text{Coul}}$</td>
<td>Coulombic Energy</td>
</tr>
<tr>
<td>$E_{\text{des}}$</td>
<td>Desolvation Energy for Each Ligand's Atom</td>
</tr>
<tr>
<td>$E_{\text{dihed}}$</td>
<td>Dihedral Threshold Energy</td>
</tr>
<tr>
<td>$E_{\text{elec}}$</td>
<td>Electrostatic Energy Contribution</td>
</tr>
<tr>
<td>$E_{\text{flex}}$</td>
<td>Energy Correlated to the Ligand's Loss in Flexibility</td>
</tr>
<tr>
<td>$E_{\text{H-bond}}$</td>
<td>Energy Associated With Hydrogen Bonds</td>
</tr>
<tr>
<td>$E_{\text{metal}}$</td>
<td>Energy Associated With Metal Interactions</td>
</tr>
<tr>
<td>$E_{\text{MM}}$</td>
<td>Molecular Mechanics Related Energy</td>
</tr>
<tr>
<td>$E_{\text{pair}}$</td>
<td>Desolvation Energy Contribution</td>
</tr>
<tr>
<td>$E_{\text{pot}}$</td>
<td>Potential Threshold Energy</td>
</tr>
<tr>
<td>$E_{\text{rep}}$</td>
<td>Repulsive Energy Contribution</td>
</tr>
<tr>
<td>$E_{\text{sol}}$</td>
<td>Solvation Energy</td>
</tr>
<tr>
<td>$E_{\text{str}}$</td>
<td>Energy Related to Bond Stretching</td>
</tr>
<tr>
<td>$E_{\text{tors}}$</td>
<td>Energy Related to Dihedral Torsions</td>
</tr>
<tr>
<td>$E_{\text{tot}}$</td>
<td>Total Energy</td>
</tr>
<tr>
<td>$E_{\text{vdW}}$</td>
<td>Van der Waals Energy</td>
</tr>
<tr>
<td>EC$_{50}$</td>
<td>Half maximum Effective Concentration</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic Acid</td>
</tr>
<tr>
<td>EJC</td>
<td>Exon Junction Complex</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>EM</td>
<td>Energy Minimization</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray Ionisation</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>MMPBSA</td>
<td>Molecular Mechanics Poisson-Boltzmann Surface Area</td>
</tr>
<tr>
<td>MODV</td>
<td>Modoc Virus</td>
</tr>
<tr>
<td>MOE</td>
<td>Molecular Operating Environment</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MTase</td>
<td>Methyltransferase</td>
</tr>
<tr>
<td>MVD</td>
<td>Mavalonate(diphospho) Decarboxylase</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NaOH</td>
<td>Sodium Hydroxide</td>
</tr>
<tr>
<td>NCS</td>
<td>Normalised Consensus Score</td>
</tr>
<tr>
<td>ND</td>
<td>Not Determined</td>
</tr>
<tr>
<td>NI</td>
<td>Nucleoside Inhibitor</td>
</tr>
<tr>
<td>NIAID</td>
<td>National Institute of Allergy and Infectious Diseases</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institute of Health</td>
</tr>
<tr>
<td>NITD</td>
<td>Novartis Institute for Tropical Disease</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear Localisation Sequence</td>
</tr>
<tr>
<td>NME</td>
<td>New Molecular Entity</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NMRC</td>
<td>Naval Medical Research Center</td>
</tr>
<tr>
<td>NNI</td>
<td>Non-nucleoside Inhibitor</td>
</tr>
<tr>
<td>NPT</td>
<td>Constant Number of Atoms, Pressure and Temperature Conditions</td>
</tr>
<tr>
<td>NS</td>
<td>Non Structural Protein</td>
</tr>
<tr>
<td>NS1</td>
<td>Non Structural Protein 1</td>
</tr>
<tr>
<td>NS2A</td>
<td>Non Structural Protein 2A</td>
</tr>
<tr>
<td>NS2B</td>
<td>Non Structural Protein 2B</td>
</tr>
<tr>
<td>NS3hel</td>
<td>Non structural Protein 3 Helicase</td>
</tr>
<tr>
<td>NS3pro</td>
<td>Non Structural Protein 3 Protease</td>
</tr>
<tr>
<td>NS4A</td>
<td>Non Structural Protein 4A</td>
</tr>
<tr>
<td>NS4B</td>
<td>Non Structural Protein 4B</td>
</tr>
<tr>
<td>NS5 MTase</td>
<td>Non Structural Protein 5 Methyltransferase</td>
</tr>
<tr>
<td>NSS RdRp</td>
<td>Non Structural Protein 5 RdRp</td>
</tr>
<tr>
<td>NSAIDS</td>
<td>Non-Steroidal Anti-Inflammatory Drugs</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleoside Triphosphate</td>
</tr>
<tr>
<td>NTPase</td>
<td>Nucleoside Triphosphatase</td>
</tr>
<tr>
<td>NVT</td>
<td>Constant Number of Atoms, Volume and Temperature Conditions</td>
</tr>
<tr>
<td>o</td>
<td>Offset Constant</td>
</tr>
<tr>
<td>O$_{18}$</td>
<td>Orientation-Dependent Overlap</td>
</tr>
<tr>
<td>OPLS</td>
<td>Optimized Potential for Liquid Simulations</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>p.</td>
<td>Peptide</td>
</tr>
<tr>
<td>P-loop</td>
<td>Priming Loop</td>
</tr>
<tr>
<td>PB</td>
<td>Poisson-Boltzmann</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Conditions</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PC$_{i}$</td>
<td>Phosphorous Pentachloride</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PDB ID</td>
<td>PDB Identification</td>
</tr>
<tr>
<td>PIV</td>
<td>Purified Inactivated Virus</td>
</tr>
<tr>
<td>Pk</td>
<td>Pharmacokinetics</td>
</tr>
<tr>
<td>PLANTS</td>
<td>Protein-Ligand ANT System</td>
</tr>
<tr>
<td>PME</td>
<td>Particle Mesh Ewald</td>
</tr>
<tr>
<td>POCI$_{3}$</td>
<td>Phosphorous Oxychloride</td>
</tr>
<tr>
<td>polyU</td>
<td>Poly-Uridine</td>
</tr>
<tr>
<td>POT</td>
<td>Average Potential Energy</td>
</tr>
<tr>
<td>ppA-RNA</td>
<td>RNA With 5’ Diphosphate Adenine</td>
</tr>
<tr>
<td>PPI</td>
<td>Protein-Protein Interaction</td>
</tr>
<tr>
<td>pppA-RNA</td>
<td>RNA With 5’ Triphosphate Adenine</td>
</tr>
<tr>
<td>PR-MD</td>
<td>Position Restrain Molecular Dynamics</td>
</tr>
<tr>
<td>prM</td>
<td>M Protein Precursor</td>
</tr>
<tr>
<td>PV</td>
<td>Powassan Virus</td>
</tr>
<tr>
<td>Py</td>
<td>Pyridine</td>
</tr>
<tr>
<td>q</td>
<td>Quartet</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Q₁</td>
<td>First Quartile</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum Mechanics</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative Structure-Activity Relationships</td>
</tr>
<tr>
<td>r.t.</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research And Development</td>
</tr>
<tr>
<td>RC</td>
<td>Replication Complex</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA Dependent RNA Polymerase</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviation</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root Mean Square Fluctuation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RNase</td>
<td>RNA Triphosphatase</td>
</tr>
<tr>
<td>(-)-RNA</td>
<td>Minus-Stranded RNA</td>
</tr>
<tr>
<td>S</td>
<td>Entropy</td>
</tr>
<tr>
<td>s</td>
<td>Gain/Loss of the Ligand Rotational And Translational Entropy</td>
</tr>
<tr>
<td>s</td>
<td>Singlet (in NMR spectra)</td>
</tr>
<tr>
<td>s.m.</td>
<td>Small Molecule</td>
</tr>
<tr>
<td>Sₜ,g</td>
<td>Shape Similarity Between Two Molecules (f and g)</td>
</tr>
<tr>
<td>SAH</td>
<td>S-Adenosyl-Homocysteine</td>
</tr>
<tr>
<td>SAM</td>
<td>S-Adenosyl-L-Methionine</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure-Activity Relationship</td>
</tr>
<tr>
<td>SASA</td>
<td>Solvent Accessible Surface Area</td>
</tr>
<tr>
<td>SD</td>
<td>Steepest Descend</td>
</tr>
<tr>
<td>SF</td>
<td>Superfamily</td>
</tr>
<tr>
<td>SF2</td>
<td>Superfamily 2</td>
</tr>
<tr>
<td>SI</td>
<td>Selectivity Index</td>
</tr>
<tr>
<td>SO</td>
<td>SYPRO ORANGE</td>
</tr>
<tr>
<td>SO₃H</td>
<td>Sulphonate</td>
</tr>
<tr>
<td>SP</td>
<td>Standard Precision</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Single Stranded RNA</td>
</tr>
<tr>
<td>(+)ssRNA</td>
<td>Positive Sense Single Stranded RNA</td>
</tr>
<tr>
<td>r.t.</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>T</td>
<td>Temperature</td>
</tr>
<tr>
<td>t</td>
<td>Triplet</td>
</tr>
<tr>
<td>Tₜ,g</td>
<td>Tanimoto Index for Molecules f and g</td>
</tr>
<tr>
<td>Tₚ</td>
<td>Melting Temperature</td>
</tr>
<tr>
<td>TBEV</td>
<td>Tick-Borne Encephalitis Virus</td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylethylenediamine</td>
</tr>
<tr>
<td>TGN</td>
<td>Trans-Golgi Network</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin Layer Chromatography</td>
</tr>
<tr>
<td>TPSA</td>
<td>Total Polar Surface Area</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminoethane</td>
</tr>
<tr>
<td>TSA</td>
<td>Thermal Shift Assay</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V(r)</td>
<td>Potential</td>
</tr>
<tr>
<td>V*</td>
<td>Modified Potential</td>
</tr>
<tr>
<td>VC</td>
<td>Infected Cells Control</td>
</tr>
<tr>
<td>VLDL</td>
<td>Very Low Density Lipoproteins</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
<tr>
<td>VS</td>
<td>Virtual Screening</td>
</tr>
<tr>
<td>VSV</td>
<td>Vesicle Stomatitis Virus</td>
</tr>
<tr>
<td>WEEV</td>
<td>Western Equine Encephalitis Virus</td>
</tr>
<tr>
<td>WT</td>
<td>Wild Type</td>
</tr>
<tr>
<td>XMP</td>
<td>Xantosine Monophosphate</td>
</tr>
<tr>
<td>XP</td>
<td>Extra Precision</td>
</tr>
<tr>
<td>YFV</td>
<td>Yellow fever Virus</td>
</tr>
</tbody>
</table>
**AMINO ACID ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Amo Acid</th>
<th>Abbreviation</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>(A)</td>
<td>Alanine</td>
</tr>
<tr>
<td>Asn</td>
<td>(N)</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Asp</td>
<td>(D)</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Arg</td>
<td>(R)</td>
<td>Arginine</td>
</tr>
<tr>
<td>Cys</td>
<td>(C)</td>
<td>Cysteine</td>
</tr>
<tr>
<td>Glu</td>
<td>(E)</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Gln</td>
<td>(Q)</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Gly</td>
<td>(G)</td>
<td>Glycine</td>
</tr>
<tr>
<td>His</td>
<td>(H)</td>
<td>Histidine</td>
</tr>
<tr>
<td>HID</td>
<td></td>
<td>His with H on the N&lt;sup&gt;δ&lt;/sup&gt;</td>
</tr>
<tr>
<td>HIE</td>
<td></td>
<td>His with H on the N&lt;sup&gt;ε&lt;/sup&gt;</td>
</tr>
<tr>
<td>HIP</td>
<td></td>
<td>Positively charged His</td>
</tr>
<tr>
<td>Ile</td>
<td>(I)</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>Leu</td>
<td>(L)</td>
<td>Leucine</td>
</tr>
<tr>
<td>Lys</td>
<td>(K)</td>
<td>Lysine</td>
</tr>
<tr>
<td>Met</td>
<td>(M)</td>
<td>Methionine</td>
</tr>
<tr>
<td>Phe</td>
<td>(F)</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>Pro</td>
<td>(P)</td>
<td>Proline</td>
</tr>
<tr>
<td>Ser</td>
<td>(S)</td>
<td>Serine</td>
</tr>
<tr>
<td>Thr</td>
<td>(T)</td>
<td>Threonine</td>
</tr>
<tr>
<td>Trp</td>
<td>(W)</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>(Y)</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Val</td>
<td>(V)</td>
<td>Valine</td>
</tr>
</tbody>
</table>

**NUCLEOTIDE ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Abbreviation</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>Adenine</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>Guanine</td>
</tr>
<tr>
<td>U</td>
<td></td>
<td>Uracil</td>
</tr>
</tbody>
</table>
Section 1:
INTRODUCTION
1.1 DENGUE VIRUS

EPIDEMIOLOGY

In the last 50 years, the increasing spread of dengue virus (DENV) has captured the attention of the Scientific World to this neglected, emerging and life-threatening pathogen. The exact period when this agent first affected humans is not certain, mainly because often the symptoms of the DENV related diseases are misinterpreted and records are not accurate, making evolutionary epidemiology studies challenging.  

The first report of DENV-related diseases in human derives from a Chinese medical encyclopaedia written between 265-420 A.D. In the text, the disease was defined as “water poison”, from the association with the mosquito-vector, which is a water-related insect. The subsequent reports belong only to the 17th Century and documentation asserts that by the 18th century, DENV was causing epidemics, in particular in two continents, Asia and America. In the following two centuries, sailing ships and the slave trade caused the virus to spread widely, mainly in tropical and subtropical regions, reaching pandemic proportions. Four DENV serotypes (DENV1, DENV2, DENV3, DENV4) are currently infecting humans and they are now all present in endemic regions. More details on DENV serotypes are in the TAXONOMY chapter of this introduction.

In recent decades the World Health Organisation (WHO) has registered a worrying increase in the number of new cases of dengue infection. For example, in the period between the years 2000-2004, the accrued cases (925,896 infections) were almost double of the ones between 1990-1999 (479,848 infections). The WHO 2014 report shows a 30-fold increase of the incidence in the last 50 years, reaching a value of around 50-100 million new cases per year. The disease is endemic in over 100 tropical and sub-tropical countries worldwide (Africa, the Americas, the Eastern Mediterranean, South-East Asia and the Western Pacific) and the most seriously affected are in the American, South-East Asia and the Western Pacific continents, where between 2008 and 2010 the number of cases has almost doubled (from 1.2 to 2.3 million cases).

Besides imported cases, DENV can also be responsible for epidemic outbreaks that have been registered also outside the endemic regions. In the last five years, outbreaks have been recorded worldwide, e.g. Yunnan, Honduras, Costa Rica, Mexico, Laos, Cook Islands, Malaysia, Fiji, Vanuatu and in Singapore. The latter country had been able to eradicate the mosquito vector and the virus, but now cases are currently reappearing. Similarly, the disease has re-emerged in Japan after 70 years, with novel cases of “domestic-acquired” infections. The
most severe outbreak in Europe was recorded in 2012 in the Madeira Islands (Portugal), where more than 2000 cases were diagnosed. Local DENV transmission was recognised in France and Croatia in 2010.\cite{6,9} DENV infections were also recorded in the USA, in Texas and in particular in Florida, where DENV is now autochthonous.\cite{9} The described global distribution and outbreaks of signalled DENV infection cases are represented in Figure 1.1.

![Figure 1.1. Geographic distribution of DENV.](image)

The countries where DENV is endemic are coloured in red, while the countries where DENV is absent are in green. Intermediate situations are coloured in yellow and in orange. The approximate location of the tropics is shown with blue lines. The outbreaks cited in the text are highlighted with black stars.

All this puts approximately 40% of the World’s population (2.5-3.5 billion people) at risk of infection. The risk applies to not only people that live in tropical and sub-tropical areas, but also travellers that come back from these places: for this class of people, dengue has become the most common cause of illness, beating even malaria and gastro-intestinal diseases. Statistics have shown that around 500,000 people (children in particular) need hospitalization for severe dengue every year and that 2.5% of the severely affected (about 22,000 people) die every year.\cite{2,6,7,10}

These are the official numbers, but the calculation of incidence cannot be considered accurate for more than one reason. First of all, primary dengue infections are sometimes asymptomatic or the symptoms are misinterpreted because they can be very similar to other illnesses, like influenza (see the CLINICAL FEATURES chapter of this introduction for a more accurate description of the clinical aspects of the infection). This leads to a miscounting of the real cases of dengue infection, in particular in non-endemic regions where there is lower familiarity with the disease. Another issue is that some Nations, particularly in Africa, do not have an adequate surveillance system for the detection of the disease cases among the population, decreasing the information about the exact extent of the disease.\cite{9,11}
A recent systematic search of published literature and online resources, reported an evidence-based map of dengue risk and estimated worldwide dengue-related infections on the basis of the population of 2010.\cite{12} In accordance with the WHO figures, the new infection estimates located the virus in the tropical and subtropical areas. The actual number of infections was, however, projected to be about three times higher than the one reported by the WHO. This disparity between actual (or modelled) and reported infection numbers has been explained, as mentioned above, by the fact that there is a low proportion of infected people that seek care from formal health facilities and that often misdiagnosis leads to cases miscalculation.

Similarly to other tropical diseases, including tuberculosis, the viral infection produces approximately 1,300 disability-adjusted life years (DALYs) per million people in the endemic countries.\cite{5} Adding this to the official number of cases and the costs for hospitalisation and therapies, it has been estimated that the global annual economic burden of DENV is around 1.7 billion dollars.\cite{13} From these numbers, it is clear that even only considering the WHO case reports, DENV infection has a high health and economic cost that needs to be tackled.

\section*{TRANSMISSION}

All four DENV serotypes exist in two possible transmission cycles that are ecologically and evolutionarily distinct: sylvatic and human. In both cases the transmission is aided by mosquito vectors belonging to the \textit{Aedes} family, but from different species according to the reservoir hosts.\cite{3} Sylvatic transmission occurs mainly in the arboreal areas of Malaysia and Senegal. Human-to-human DENV transmission has a more worldwide spread and is generally predominant, apart from West Africa where the sylvatic cycle is more diffused.\cite{14} In the sylvatic cycle, the amplification and reservoir hosts are non-human primates and the mosquito vectors are \textit{Aedes} species that live in arboreal environments. In the human transmission cycle, the only amplification and reservoir hosts are humans and the mosquito vectors, \textit{Aedes aegypti} and \textit{Aedes albopictus}, have adapted to live in domestic and peridomestic environments. Occasionally, arboreal \textit{Aedes} mosquitoes feed on humans and it has been hypothesised that this behaviour might be the cause of occasional transfers of DENV from the sylvatic to peridomestic environments, as well as the origin of the onset of human infections.\cite{1,3}

\textit{Aedes aegypti} and \textit{Aedes albopictus} have a very precise phenotype: they are small, dark with white markings and with banded legs. Because of these peculiar features, they are commonly called “tiger mosquitoes”. The most usual vector for DENV is the Africa native \textit{Aedes aegypti}, a highly adaptable species that lives closely with humans and within human habitations. The other type of mosquito, \textit{Aedes albopictus}, is originally from Southeast Asia. It is less common,
but is not a less efficient vector as its habitat is related to the peridomestic environment. Both arthropod larvae have been found not only in natural environments like tree holes, but also in artificial and occasional containers of water, such as water jars and buckets, or even discarded tyres.\textsuperscript{[2,14-16]}

The viral transmission occurs through the bite of a female mosquito belonging to one of these two species. The mosquito comes in contact with the pathogenic agent in consequence of its feeding on an infected person during the person’s period of viremia and after a brief phase of incubation (7-10 days) the insect probably remains infectious for the rest of its life (approximately 2 weeks).\textsuperscript{[11]} As human beings are the major host for DENV, the transmission is usually between people via the insect vector. Other methods of transmission, like transfusions, are rare.\textsuperscript{[2,4]}

There is a clear association between the regions with a high incidence of DENV infection and the large presence of \textit{Aedes} infestations, as the climate, in terms of temperature and precipitations, is favourable for the mosquitoes’ survival. Figure 1.2 shows how mosquitoes and DENV infection distributions overlap and that the two species often coexist in the same regions. These territories are frequently characterized by high grades of poverty that translates into a lack of suitable treatment and prevention of the disease, contributing to the endemic behaviour of the infection.\textsuperscript{[2]}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mosquito-distribution.png}
\caption{Geographic distribution of \textit{A. albopictus} and \textit{A. aegypti}.}
\end{figure}

The countries where DENV is endemic are coloured with grey shades, where the darker colour corresponds to higher presence of DENV. The presence of \textit{Aedes} mosquitoes is indicated with coloured circles: red for \textit{A. albopictus} and green for \textit{A. aegypti}.

The increase of the incidence of dengue-correlated infections is associated with the increase of the global population, in particular with the uncontrolled growth of the population in South East Asia and Africa, and it is also linked to the spread of the virus to new, uninfected
countries. The increasing movement of people and goods between different countries, in combination with the drought-resistance of the mosquito eggs, has allowed egg transportation and consequent settlement of the Aedes mosquito in countries other than their native ones in the tropical and subtropical areas of Africa (A. aegypti) and Southeast Asia (A. albopictus).\[2,11,14,15\]

The adaptation of the mosquitoes in new territories depends on both climatic factors and landscape features. Climatic factors refer mainly to winter temperatures and precipitation, as low values of both parameters are not compatible with the mosquito survival. As both A. aegypti and A. albopictus are weak flyers, the proximity with humans is also fundamental and makes land utilization another important parameter in vector distribution.\[14-17\] Not all areas where either mosquito is now established are also endemic for DENV infection. However, the presence of either mosquito increases the probability of the establishment of DENV as an autochthonous disease. In a first instance, autochthonous mosquitoes could then become infected from infected travellers; alternatively, infected mosquitoes could be transported through goods merchandising.\[9\] For example, in some European countries (e.g. Italy, Greece, France), newly established populations of Aedes albopictus have been observed, and most of Europe can be considered a favourable habitat for this arthropod. For the moment, the cases of dengue fever in Europe have been diagnosed mostly in returning travellers. However, in particular after the European outbreaks, it is expected that in the future there will be seen a diffusion of the infection also in this continent.\[2,11,15,18\] Similarly, it is estimated that the A. albopictus distribution in North-eastern USA will put at risk of DENV infection approximately 30 million people that inhabit those areas.\[17\]

**TAXONOMY**

Virology is a very young branch of science as the understanding of viral agents began only in the late 19th Century, when viruses were “infectious living fluids” that could pass the bacteria-retaining filters (with 0.3μm pores). Currently, viruses are defined as pathogenic entities that pass bacteria-retaining filters and possess a DNA or RNA genome that is used for their multiplication. As they are not able to grow and undergo binary fusion, to produce energy and do not have a complete translation machinery, they use the host structures and therefore they are obligatory parasites. The definition of virus is in continuous evolution. For example, the recent discovery of so called “giant viruses” (like Mimi-, Mega- and Pandora-) is introducing the possibility of the existence of viruses that can be bigger than some small bacteria, undermining the importance of particle dimension in the definition.\[19,20\]
Furthermore, in these last 100 years an even more basic question than the definition of a virus has not yet been resolved: are viruses a form of life or not?\textsuperscript{[20]}

Although virology is a young science, and major issues are still under discussion, a huge number of viruses have been discovered, leading to the need to categorize them. However, the lack of clarity to the basic question above renders difficult the application of classical taxonomy methods. For this reason, the International Committee on Taxonomy of Viruses (ICTV)\textsuperscript{[21]} has established a specific taxonomy classification for viruses, composed of: order (not for all the viruses), family, genus and species.\textsuperscript{[20,21]}

Dengue virus is classified within the \textit{Flaviviridae} family. This family includes several different positive sense single stranded RNA viruses that can be divided into 3 genera, all of which have common characteristics: they are all small (approximately 40-50nm) enveloped viruses with the same general replication cycle.\textsuperscript{[22]} One genus is \textit{Hepacivirus}, named after the Greek word “hepatos” (liver) because the most famous member of the genus is the hepatitis C virus (HCV) that targets mainly liver cells. Another genus is \textit{Pestivirus}, named after the Latin name “pestis” (plague) as the viruses implicated (bovine viral diarrhoea virus, BVDV; classical swine fever virus, CSFV; border disease virus, BDV) cause economically important diseases in animals. The viruses belonging to these two genera do not use arthropod vectors for their transmission, differing from DENV. For example, HCV is transmitted though direct blood contact with the virus, possible in certain circumstances, like blood transfusion or the use of infected medical equipment. The third genus belonging to the \textit{Flaviviridae} family is \textit{Flavivirus}, named after the Latin word “flavus” (yellow) as the first discovered virus belonging to this genus is the yellow fever virus (YFV) that causes very high fevers and in some cases liver damage that induces a yellow colouration to the patient (skin and eyes). DENV, as also West Nile virus (WNV), Japanese encephalitis virus (JEV) and many others (at least 73 distinct viruses have been determined to date), belong to this last genus and several of them are human pathogens.\textsuperscript{[22-24]}

In human infections, the known antigenically distinct serotypes of DENV, sharing about 65-70% of sequence homology, are four (DENV 1-4) and they coexist in the same geographic regions.\textsuperscript{[4,7,10]} A fifth DENV serotype (DENV5) has been discovered recently in Malaysia. However, to date it has only been observed in the sylvatic cycle and not in humans.\textsuperscript{[25]}

The causes of the serotypes differentiation are still not clear, but it has been hypothesised that each serotype has evolved from a different sylvatic progenitor that was transmitted from the sylvatic cycle to humans.\textsuperscript{[1,3]} From a clinical point of view, association of the serotypes with the transmission efficiency and with disease severity are still not clear.\textsuperscript{[2]} However, a severity increase of a secondary DENV infection has been observed when the sequential infections are DENV1 and then DENV2 or DENV3, or DENV3 and then DENV2.\textsuperscript{[5]}
CLINICAL FEATURES

Independently of the serotype, DENV is correlated to a broad spectrum of diseases, with different grades of severity, from asymptomatic to fatal. Infection with DENV starts after a variable period of incubation (from 2 days to 2 weeks) following viral inoculation from a vector bite.\cite{2,7,26} If it is a primary infection, it can be asymptomatic or cause a self-limiting flu-like disease called dengue fever (DF) that can give many different clinical responses. Children usually are affected by a flu-like fever with cough, vomiting and abdominal pain. Adults usually present a 5 to 6 days long fever with chills, headache (especially retro-orbital or frontal), lethargy, musculoskeletal pains, exanthema, leukopenia, transaminase elevation and thrombocytopenia.\cite{2,26,27} It has commonly been observed at skin and mucosal surfaces, minor bleeding and rash and in patients with peptic ulcer DF worsens the patient condition. The highest viremia titers (between $10^5$ and $10^6$ units/ml) can be found during the febrile period, after which the values drop. Sometimes, even if the infection is not too severe, the recovery can be complicated by fatigue and depression.\cite{2,7,11,27}

The primary infection, caused by any of the four serotypes and with any grade of severity, produces a lifelong immunity to that serotype, but only an approximately 6 months long immunity to the other serotypes. After this period, a secondary infection with a different serotype to the primary one, leads to more serious diseases: dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). Both of them are particularly severe in young children. Additionally to fever and many of the DF symptoms, DHF causes an increased vascular permeability with leakage of plasma into the interstitial space, thrombocytopenia, leukopenia, altered homeostasis and liver damage. The viremia titers are usually 10 to 100 fold higher that in DF. DSS is a complication of DHF caused by a very high loss of liquid in the interstitial space that produces hypovolemic shock, with severe and continuous abdominal pain, fainting, perspiration, persistent vomiting and high risk of circulatory failure and death.\cite{2,5,26,27}

The fact that a secondary infection is more severe than a primary infection is attributed to a mechanism called antibody-dependent enhancement (ADE). In this situation, antibody particles of a different serotype are able to bind viral particles, forming a complex, without the ability of neutralising the pathogen. This complex is then taken up into cells that express the Fc receptor (e.g. phagocytes). This produces an enhancement of the non-neutralized viral particle entry into the Fc expressing cells. Consequently, a higher number of viral particles successfully infect cells, leading to a more severe disease, like DHF and DSS.\cite{7,11,26-28}

In addition to the virulence of the virus and to the infection history (ADE), the severity of DENV infection is also linked to age, gender, host genotype and pre-existing comorbidities.
Probably due to intrinsic higher endothelium permeability, children have been found to have a greater prevalence of DHF and DSS with respect to adults. Interestingly, girls generally have higher DSS and death cases than males. This is probably due to physiologically immunological differences between genders, but unfortunately the cultural factor of difference in health-care seeking behaviour in endemic countries cannot be excluded. Furthermore, DENV interacts with many human receptors and factors (see BIOLOGY chapter), therefore mutations in the host genotype involving these structures can reduce viral susceptibility. Finally, DENV can be life-threatening (with 70% of fatalities in some cases) in individuals with pre-existing comorbidities. Examples of these comorbidities are asthma and diabetes.\(^{[5,24]}\)

**BIOLOGY**

![Figure 1.3. Images of a DENV mature virion.](image)

**A)** Virion surface (PDB 1K4R). Each envelope protein (E) surface is represented here. The three E monomers are differently coloured (red, light blue and green) for a better visualisation of the homotrimeric structure. **B)** Schematic representation of the viral structure. The envelope bilayer membrane that encloses the virion is represented in dark and light pink. E proteins are represented in green and membrane (M) proteins are in white. The disordered nucleocapsid is coloured in blue and the viral genome is indicated with the orange line binding to the capsid protein (C).

DENV is a small spherical virus with a diameter of approximately 50nm presenting a nucleocapsid (electron-dense core of 30nm) surrounded by an envelope. As shown in panel A of Figure 1.3, the overall shape of the mature viral particle is smooth and it resembles a golf ball.\(^{[22]}\) The external envelope has an outer shell constituted of 180 copies of two glycoproteins: envelope (E) and membrane (M). The internal portion of the envelope is constituted by a lipid bilayer that is derived from host cells. The nucleocapsid is a disordered construct, constituted by the complex of one (+)ssRNA molecule that interacts with multiple copies of the capsid protein (C).\(^{[29,30]}\) Panel B in Figure 1.3 shows a schematic representation of the structural elements composing a DENV virion particle.
The initial interaction between the virus and a target cell occurs between E and a cellular receptor. The exact cell targets for DENV are poorly known, as the diseases caused by this pathogen are systemic. Nevertheless, it is established that during the incubation period, DENV replicates in proximity to the site of inoculation, in dendritic cells, in macrophages and in lymphocytes. Through this route the virus eventually reaches the blood stream and at this point it is thought that the main cells targeted by DENV are serotype-specific depending on the serotype-specific E glycoprotein.\(^7,23\) Consequently, a variety of cell targets have been identified, like hepatocytes, lymphocytes, endothelial cells, neuronal cells and muscle satellite cells.\(^31\) The specific cell receptors that trigger the virus clathrin-mediated internalization are not known to date, but it has been shown that carbohydrate molecules on host cell surfaces can aid viral attachment through interaction with E with the function of concentrating the pathogen and facilitating the interaction with the specific receptor.\(^30\) It has been demonstrated that highly sulphated heparan sulphate (HS) is one of these binding factors and putative binding motifs were located in the E protein. Interestingly, the high heterogeneity of the structure of HS on cells from different tissues could drive the tropism of DENV in target cell binding.\(^32\) Viruses are “obligatory intracellular parasites”, meaning that they replicate exploiting the cellular structures as they are missing part of the molecular machinery.\(^20\) The general replication stages change between groups of viruses, but all viruses belonging to the Flaviviridae family follow the phases schematised in Figure 1.4. The first step of the viral replication cycle (the adsorption of the virus to the cell surface) is followed by the interaction with the specific receptor that triggers viral entry into the host cell through a clathrin-mediated endocytosis.\(^22-24\) Once the clathrin-coated vesicles have entered the cytoplasm, they fuse with endosomes. The consequent endosomatic pH decrease triggers the rearrangement of the E glycoprotein from a dimer into a monomer state and then into a “fusion-active trimeric state”. These changes in conformation of E allow the viral envelope to fuse with the endosomal membrane, releasing the viral RNA into the cytoplasm.\(^22-24\) The viral nucleic acid has three important roles in the viral replication cycle: it acts as a messenger RNA (mRNA) for the synthesis of the viral encoded proteins; it is a template for the synthesis of novel viral RNA; it is enclosed in novel viral particles as their genetic material. An untranslated region (UTR) at the 5’ end of the RNA (see later in this chapter) directs the viral genome to the cellular ribosome for the translation of the genome.\(^22-24\) The first step of this process is the synthesis of a polyprotein precursor that is then inserted in the endoplasmic reticulum (ER) membrane and co- and post-translationally cleaved and processed into the mature viral protein via the activity of a viral enzyme (NS3 protease, NS3pro) and host proteases.\(^22-24\) The viral RNA replication is operated by the replication complex (RC). The RC contains viral encoded proteins, RNA and probably host factors and is located in a
membranous system. These structures seem to be major rearrangements in host subcellular organelles, most probably rough and smooth ER, that host the viral replication steps.\textsuperscript{[23,24,33]} RNA replication undergoes two main steps, both of which are accomplished by the viral NS5 RNA-dependent RNA polymerase (RdRp). The first step employs the synthesis of a complementary (minus-strand) copy of the RNA. This is then used as a template for the synthesis of the plus-strand nucleic acid in the second step. The double stranded RNA (dsRNA) that is produced in this way is unwound by DENV NS3 helicase (NS3hel) and the positive sense RNA 5’ end is capped by the viral NS5 methyltransferase (MTase). The final (+)ssRNA is then used as mRNA, template for RNA transcription or encapsulated in novel virions (or viral particles).\textsuperscript{[22-24]}

The assembly of the virion starts with the association of the nucleocapsid. At this stage, the precursor of the nucleocapsid protein (C) anchors to the membrane, making possible its interaction with the envelope and membrane precursor proteins (E and prM). Multiple copies of C protein, featured by positively charged residues, bind to a single negatively charged RNA and form the nucleocapsid.\textsuperscript{[22-24]} Usually, enveloped viruses gather their lipid bilayer when they are released from the host cell into the extracellular environment. This is not the case for
DENV: the formation of the envelope comes as a consequence of the budding from the modified membranous system.\textsuperscript{23,24,33} Viral maturation occurs in the trans-Golgi network (TGN) where the low pH induces high rearrangement of the structure of the immature virus.\textsuperscript{30} At the end of the DENV replication cycle, the release of the virus occurs via exocytosis through fusion of vesicles containing the enveloped mature virus.\textsuperscript{22-24,33}

From the description of the replication cycle, it is clear that RNA plays a central role in the whole process. DENV genome RNA is composed of approximately 10,800 base pairs. At the 5' end there is a type I \textsuperscript{7Me}GpppA\textsubscript{2'Ome} cap structure that ensures stability to the nucleic acid and therefore it is essential for viral replication. The cap is followed by a short UTR (67-132 bases) responsible for genome targeting to the ribosomes and essential for the synthesis of the viral RNA. At the 3' end, a 114-585 nucleotide long segment constitutes another UTR. Between these sequences there is a single open reading frame (ORF) that encodes for approximately 3,400 amino acids that form the polyprotein precursor (schematized in Figure 1.5).\textsuperscript{22,24,27,30}

\textbf{Figure 1.5. DENV polyprotein processing and cleavage products.}

The structural proteins are indicated in cyan (C=nucleocapsid protein; prM=precursor of membrane glycoprotein M; E=envelope glycoprotein); the non structural proteins are in orange. Cleavage sites are shown by symbols: ◊ represents cellular peptidases cleavage sites, the ↓ arrow represents the NS3-2B serine protease sites. In multifunctional proteins, the activity of each portion is shown: "PROT"=protease, "HEL"=helicase, "MTase"=methyltransferase, "RdRp"=RNA-dependent RNA polymerase.

This polyprotein is co- and post-transnationally cleaved by viral and host proteases in three structural and seven non structural (NS) proteins (also schematized in Figure 1.5). The host cell enzymes involved are endoplasmic reticulum signalases (represented by ◊ in the figure) and they control the cleavage between nucleocapsid protein (C) and the precursor of the envelope glycoprotein M (prM); prM and envelope glycoprotein E; E and non-structural protein NS1; NS1 and NS2A; NS4A and NS5. The viral protein involved in the polyprotein processing is NS3 serine protease (NS3pro) in the presence of the NS2B cofactor. It operates the cleavage at the level of C/prM, NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B and NS4B/NS5 junctions.\textsuperscript{22,24,34}

The polyprotein precursor has transmembrane motifs that allow its insertion in the ER membrane. The mature viral proteins are also associated with the same membrane according to the topology reported in Figure 1.6.
In this model, the mature viral proteins are either in the ER lumen side of the membrane (prM, E, NS1), transmembrane (NS2A, NS2B, NS4A, NS4B), or on the cytoplasmic side of the ER membrane (C, NS3 and NS5). Most of them are in the ER lumen side of the membrane.\cite{22,24,34}

The mature viral proteins have different roles and on this basis they can be divided into two main groups: structural and non-structural proteins. Structural proteins are the ones involved in the structure of the virion itself.

Capsid protein C is a small protein of 120 residues of which a high percentage (25%) is constituted by basic residues (15 Lys and 9 Arg). This enables C to bind the RNA via electrostatic interactions with its negative charges during the encapsidation, enclosing the genome into the virion. C is inserted in the ER membrane through a small C-terminal sequence of 20 residues called signal sequence. In this form the protein is actually the C protein precursor (anchC or “anchored C”) that becomes C after the cleavage of the signal sequence.\cite{22,24,34}

prM is the glycosylated precursor protein of the membrane protein (M). It is a small (75 residues) protein with a C-terminal transmembrane anchor that forms ion channels that allow the passage of ions as calcium, sodium, chloride and potassium. On the lumen side of the ER membrane, prM forms a heterodimer with the envelope protein (E), protecting E from the acidic environment of Golgi vesicles during virion maturation. This low pH is similar to the one in the endosomes and in absence of prM these conditions could lead to a premature conformational change of E and to a consequent fusion of the envelope with the membrane, with a similar mechanism to the one described earlier. The proteolytic processing of prM occurs in a late stage of the virion maturation, when a cellular protease cleaves it into M, which remains intercalated in the envelope membrane, close to E, and “pr”, which is released into the extracellular environment.\cite{22,24,34}

The envelope glycoprotein (E) is associated to the viral envelope and it mediates host receptor binding and consequent viral internalization during the initial phases of the replication cycle. E forms homodimers on the envelope surface that undergo irreversible conformational changes.
(into monomers and then homotrimers) in the low pH environment of the endosome, allowing the fusion of the envelope with the endosome and the consequent viral RNA release into the host cell.\textsuperscript{22,24,34}

Non-structural (NS) proteins are not directly involved in viral morphology, but they are essential factors of the replication cycle. Non-structural protein 1 (\textbf{NS1}) is a 45 kDa water-soluble protein that is glycosylated for correct folding and released in the ER lumen. In infected cells it is present in a dimer form when associated to membranes (e.g. on cell membrane of infected cells) and as a hexamer in solution, also in the extracellular environment.\textsuperscript{35} Outside the infected cell, NS1 is able to aid the virus to elude the host immune system as it produces a protective humoral immune response. At the same time it has been observed that NS1 antibodies are associated with cross-reaction with platelets and endothelial cells. Therefore, it has been suggested that the inflammatory situation in DF is correlated to this non-structural protein and the fact that anti-NS1 antibodies protect against DENV infection is consistent with this hypothesis. NS1 also plays a role within the viral replication cycle, but its interactions with other viral proteins and the molecular mechanism are still not well understood.\textsuperscript{22,24,34}

Non-structural protein 2 is divided in two parts: NS2A and NS2B. The first one (\textbf{NS2A}) is a 22kDa protein that contains transmembrane domains. It has been localized in the replication site and it is essential for viral replication. It is able to bind NS3, NS5 and the 3’UTR of the genome. For these reasons it has been associated to RNA replication, but its exact role is not fully known to date. \textbf{NS2B} is associated with the membrane as it has several membrane spanning domains and only a 40 residues long fragment of hydrophilic amino acids. It is the cofactor for NS3 serine protease as it stabilises this structure in solution, allowing the correct function of this enzyme.\textsuperscript{22,24,30,34}

Non-structural protein 3 (\textbf{NS3}) is a multifunctional enzyme. The N-terminal one-third (170 residues) portion of NS3 is a serine protease (with the His51, Asp75 and Ser135 catalytic triad) that is functional only if bound to the cofactor NS2B. As listed previously in this chapter, NS3 protease is responsible of the cleavage of the junction between C/prM, NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B and NS4B/NS5. The C-terminal portion of NS3 has three enzymatic activities: helicase, nucleotide triphosphatase (NTPase) and RNA triphosphatase (RNase). The function of this portion of polyprotein (NS3hel) is to unwind in an NTP-dependent manner double strands of genomic RNA (dsRNA) and to remove the γ-phosphate at the 5’ end of the viral nucleic acid during the capping process.\textsuperscript{22,24,34}

Two peptides belong to non-structural protein 4: \textbf{NS4A} and \textbf{NS4B}. The function of these two proteins is not clear to date. NS4A has been associated with ER membrane alteration, while NS4B appears to be involved in the modulation of viral replication through the interaction
with NS3. It has been shown to cause the displacement of the NS3 hel from ssRNA, enhancing dsRNA unwinding.\cite{22,24,30,34}

The largest and most conserved non-structural protein is $\text{NS5}$, another multifunctional enzyme. Its C-terminal portion (residues from 270 to 900) is a RNA-dependent RNA polymerase (RdRp). The main function of this portion is the synthesis of viral RNA through the synthesis of an intermediate minus-stranded RNA template. The N-terminal part of NS5 has a S-adenosyl-L-methionine (SAM) dependent methyltransferase (MTase) activity. The reaction catalysed in this case is the transfer of two methyl groups from two molecules of SAM to the last two 5′ bases (a guanine, G and an adenine, A) of the viral RNA in positions N-7 of G and 2′-O of A. These are two of the four steps of RNA capping, essential for RNA stability.\cite{22,24,30,36}

**DIAGNOSIS AND CURRENT TREATMENT**

DENV infection diagnosis is usually clinical, but the broad spectrum of symptoms of DENV infections, makes laboratory diagnosis often crucial for DF or DHF/DSS recognition, in particular in Nations were the disease is not common.\cite{5} The main available laboratory methods are virus isolation, serological tests and nucleic acid amplification tests. Virus isolation in cell cultures is the technique that provides the most specific result. However, due to its high cost it is not always available. Easier to perform and much less expensive are serological tests. For instance, the detection in blood samples of the viral non-structural protein 1 (NS1) can confirm infection. Furthermore, the enzyme-linked immunosorbent assay (ELISA) can be applied for the recognition of two DENV-specific immunoglobulin proteins that are produced 5-7 days after infection: IgM and IgG. The first antibody becomes undetectable 2-3 months after the infection and can be used as an indicator for on-going infections. IgG remains detectable for over 60 years after infection and is a useful indicator of passed infections. Interestingly, the IgM/IgG ratio can be used to discriminate between primary and secondary infections, however cut-offs have not been fully calibrated to date. The nucleic acid amplification tests can be a very sensitive and rapid tool for the identification and quantification of the virus during the high viremia phase. Importantly, they can also distinguish between serotypes, but to date the tests are available only in the USA.\cite{5,10}

Once the infection has been diagnosed, the therapy is currently only symptomatic. WHO guidelines for the treatment of patients were published in 2009 and completed in 2012, as the “Handbook for clinical management”.\cite{37,38} In case of high fever, the administration of paracetamol is usually preferred to other non-steroidal anti-inflammatory agents (NSAIDS), such as acetylsalicylic acid, as they can worsen gastric ulcers and consequent bleeding.\cite{37,38}

Two of the common antiviral agents, Ribavirin and interferon, have shown no efficacy against
DENV and cannot be used for therapy.\cite{10,23} Most cases of DF are self-limiting and paracetamol support is sufficient.\cite{10,38} Conversely, for cases of DHF and DSS, hospitalization is required most of the time and the hypovolemia is treated with hydration or with aggressive fluid management in order to counteract the loss of plasma. This can be performed via body fluid replacement or with colloid solutions (e.g. with lactate, dextran, starch or crystalloid solutions). In the worst of cases with severe bleeding, blood transfusion might also be necessary.\cite{11,23,28,38}

**PREVENTION**

Other guidelines presented by WHO address the prevention of the disease that can be operationalized in two ways: vector control and vaccine production. Several strategies for the decrease of *Aedes* mosquito populations are reported in the “DENGUE Guidelines for diagnosis, treatment, prevention and control” (e.g. the use of insecticides or larvicidal fishes, proper water supply to households, periodic cleaning of water containers and waste management).\cite{37} However, to date they have mostly not been effective in both the short or the long term for several reasons. First of all, after an initial success the procedures are generally abandoned, allowing the residual populations to grow again. This is due to an over reliance upon the chemical agents used and a lack of monitoring procedures that assure that the vector has effectively been removed. Secondly, not all of the countries are applying measures for vector control. This, added to increased international travel, allows the vector to be easily reintroduced even after complete eradication. These issues can be related to inadequate financial and human resources or excessive investment requirements for involved Countries that often are underdeveloped or in development.\cite{11,14,18,39}

Vaccine development is the second strategy for DENV infection prevention, but it has proven to be challenging for several reasons. First, the pathogenesis of the virus is still not fully understood. It has been found that human antibodies mostly target the membrane precursor (prM) and envelope (E) structural proteins and non-structural protein NS1 (described in the BIOLOGY chapter of this introduction). However, studies to find the best epitope are currently continuing. The lack of complete comprehension of the antibody-enhanced (ADE) mechanism is also an issue for vaccine development. A monovalent vaccine could cause severe cross-reaction in case of DENV infection, in the same way as a primarily infection does. Therefore, current knowledge suggests it is necessary to develop a tetravalent vaccine.\cite{40,41} A second obstacle has been the lack of suitable animal models for research and development.\cite{41,42} A third complication in vaccine development has been to obtain antibody titers sufficient for long-lasting (life-long) protection. ADE will always be a worry even with tetravalent vaccines
Section 1: INTRODUCTION

1.1 DENGUE VIRUS

and therefore vaccine clinical trials must address this issue with long-term follow-up phases.\(^{[43]}\) Last, there are other criteria that a vaccine must fulfil, especially for developing and third world countries: immunisation must occur with no significant disease symptom manifestation; a full protection must be obtained with a maximum of two injections; there must be a reduced transmissibility and the cost of the vaccine must be low.\(^{[7,11,42,44]}\)

To date there are no vaccines on the market against DENV, but some candidates are in clinical trials, as reported in Table A1.1 of APPENDIX 1. The most common strategies applied to date for vaccine development have been: in vitro passaged life-attenuated virus (LAV) vaccines, recombinant life-attenuated virus vaccines, chimeric life-attenuated virus vaccines, purified inactivated viruses (PIV), DNA vaccines and recombinant subunits.\(^{[7,11,41,43]}\) The most advanced vaccine is the Sanofi Pasteur live chimeric vaccine (CYD-TDV), in which the licensed YFV 17D vaccine is used as a backbone for the expression of all four DENV serotypes prM and E proteins. It reached clinical trial phase III and it was expected to be launched in the market in 2015.\(^{[45]}\) However, phase IIb clinical trial results showed only a 30% efficacy and immunisation only for serotypes 1, 3 and 4, creating disappointment in the vaccine strategy.\(^{[41,43]}\) Aiming to gain stronger statistical analysis and to conclusively assess the CYD-TDV efficacy on larger populations, clinical III trials have been begun in Asia and in Latin America. In July 2014, the results of the Asian study were disclosed. The study was performed administering three doses (at 0, at 6 and at 12 months) of the vaccine or of a placebo to approximately 10,000 children, with a follow-up of two years. The vaccine showed a 56% reduction of the disease with a limited protection against DENV2. Inexplicably, it worked best against severe dengue disease in people who had already been infected in the past (88.5% protection).\(^{[46,47]}\) The results of the second study on approximately 20,000 children in Latin America were published in September 2014 and have been found to be in line with the previous trial.\(^{[48]}\) The positive finding of cross-trial consistency is however weakened by the low efficacy in dengue-naïve people that seems to preclude its use in non-endemic areas and for travellers.\(^{[47]}\)

**FUTURE POTENTIAL TREATEMENTS**

**THERAPEUTIC STRATEGIES**

The previous chapter has described how DENV prevention has proven to be challenging and the approaches tried so far have given poor results. In summary, vector control has shown to be unrealistic, while vaccine development to date has not given the required efficacy. A simulation of the effect of the introduction of a vaccine in the market was recently conducted for the analysis of the potential drug market for this virus.\(^{[13]}\) According to this work, even if
the Sanofi Pasteur vaccine was to reach the market in 2015, the clinical cases of dengue fever will not decrease below the values of 2006 before 2033 and even then, it will be very likely that the vaccinated population will only be around 16%.\[13\] This is due to the vaccine production time and to the fact that it might not be possible to vaccinate people that have been already infected once with DENV. The consequence of this analysis is that, even should a vaccine become available therapeutic drugs will have a fundamental role for dengue virus management. Furthermore, anti-DENV agents would not only increase the early search for medical treatment, reducing dramatically the incidence of DHF and DSS, but they could also be used prophylactically in geographical areas of high infection, preventing a further spread of the pathogen.\[49\] Despite the pressure to find an antiviral therapy, there are currently no approved specific treatments and there is therefore a great need for research in this field. Very similarly to the ideal vaccine, an ideal drug should meet some basic pharmacological criteria: it should have a good bio-availability through oral administration; have a low frequency of administration (maximum 3 times per day); cause a low genetic resistance; be safe also in paediatric patients.\[50\] As well as these features, it is important to remember that the countries with the highest prevalence of DENV are in tropical and subtropical regions and that they are low-income countries. Consequent, the ideal drug should also: have an easy and “low-cost” synthesis and preparation; have a good thermal stability and good hygroscopic characteristics.\[31,50\] A significant concern on DENV antiviral therapy is the rapidity of the viremia decline during the infection, making the rapidity of action a key factor in drug development, in conjunction with the optimisation and spread of reliable diagnostic tools.\[49\] Even though the pathogenesis of DENV is not fully understood, the knowledge acquired on DENV suggests three strategies for the development of novel treatments: to reduce the viremia during the early stages of infection, to block viral replication by inhibiting a viral encoded target, or to control the host response to the disease (important, in particular for DHF and DSS).\[31,51\] A relevant issue in the development of drugs against DENV has been the absence of suitable animal models in which the disease phenotype is similar to humans and that can therefore be used for pre-clinical tests of compounds. Non-human primates have been investigated, but the absence of clinical signs after DENV infection has made these models not suitable for antiviral research. The availability of genetic tools has made possible the manipulation of several cheaper and smaller animals, like mice, making them susceptible to DENV infection. However, a model that simultaneously resembles all disease features has not been created to date and the choice of the model is therefore linked to the study itself.\[31,52\] Great utility for antiviral research has been demonstrated by the AG129 mouse model, lacking both IFN-α/β
and IFN-γ, as it can reproduce mild and severe disease models as well as ADE according to the experimental setup.\textsuperscript{[52,53]}

Given the difficulties in the development of an \textit{in vivo} model, several \textit{in vitro} approaches were developed for the assessment of compound activity. Cell-based assays currently available consist of live-virus, viral replicon and virus-like particle assays. The most common used test is the cytopathic effect inhibition assay (CPE) that assesses the ability of potential drugs to inhibit viral-induced cell death using staining procedures that respond only to alive cells. It requires live infected cells, with biosafety containment issues, but it has been proven to be reliable. Another live-cell assay is an ELISA test that detects and quantifies the production of E protein in DENV infected cells. The viral replicon assay consists of a luciferase-based subgenomic replicon system in which a luciferase reporter replaces DENV structural proteins. No biosafety containments are necessary, but this test can only be performed with viral replication inhibitors and false positives can occur as compounds might interact with the luciferase reporter directly. The co-transfection of a viral replicon with a vector expressing DENV structural proteins gives a virus-like particle. As for viral replicon assays, assays that use virus-like particles do not need biosafety containment and have a high rate of false positive results. However, they can be used to test both replication and entry inhibitors.\textsuperscript{[51]}

As a compound might interact with several cellular pathways, the use of a cell-based assay is not always useful for structure-activity relationship studies. For this reason and as for pre-clinical development it is important to know the drug target, several protein-specific assays have also been developed for compound assessment. In particular, commonly used DENV proteins are: the protease, the helicase, the methyltransferase and the polymerase.\textsuperscript{[49,51]}
VIRAL TARGET DRUG DESIGN

In a viral-based drug design approach, the main aim is to block viral functions that are essential at any stage of its replication cycle. The rationale behind the direct antiviral approach relies on the awareness that virus titers are directly correlated to the severity of the disease and, consequently, the reduction of viral levels should prevent or reduce the progression to DHF and DSS.\[49\] For this reason and based on available knowledge, the main viral targets studied to date to find an anti-DENV molecule are the structural proteins (envelope, E and capsid, C) and the non structural (NS) protein NS3-2B serine protease, the NS3 helicase, NS4B, the NS5 methyltransferase (MTase) and NS5 RNA-dependent RNA polymerase (RdRp) (see BIOLOGY chapter for viral protein description).\[31,34,51\] A detailed list of potential inhibitors is reported in Table A1.2 of APPENDIX 1, while only the general aspects of drug design approaches against viral targets is discussed here.

The inhibition of viral entry into the cell is an antiviral strategy that has proven to be successful for other viruses, such as HIV.\[49\] As explained in the BIOLOGY chapter, dengue virus entry in the host cell is mediated by the envelope protein (E) that undergoes major conformational changes upon change of environmental pH. A range of techniques has been applied for the discovery of novel compounds that interact with E and therefore disrupt the E-host receptor interaction. Molecules have been selected through in silico approaches such as virtual screening of compound libraries or docking. In both cases, compounds were aimed to bind to a hydrophobic pocket of E, identified by β-N-octyl-glucoside (βOG) that was co-crystallised with the structural protein. The hypothesis behind this strategy is that inhibitors targeting this cleft would disrupt pH-dependent conformational changes in E. Unfortunately, none of the compounds were promising enough for further development.\[49,54\] Peptides derived from a portion of the E protein inhibited DENV in cells and some of them were computationally optimised. However, their intrinsic short shelf-life and the need of intravenous administration has reduced the interest in this class of molecules.\[49,54\] High throughput screenings (HTS) of compound libraries with DENV infection assays have identified two classes of antibiotic-derivatives that inhibit DENV entry, but their mechanism of action has not yet been defined. Preclinical studies are currently on-going for the doxorubicin analogues.\[49,54\] One of them is exemplified in panel A of Figure 1.7.\[55,56\] As described in the BIOLOGY chapter of this introduction, DENV E protein binds the cells polysaccharide glycosaminoglycan heparan sulphate. For this reason, this scaffold, shown in panel B of Figure 1.7, has also been used for the design of antivirals that interact with E, but no follow up studies have been reported.\[49,54\]
A further target for DENV is the **capsid (C)** structural protein. Its main function is to complex the viral RNA through electrostatic interactions and to form the nucleocapsid. Through a HTS strategy that analysed compounds belonging to a library and their ability to inhibit the infection in cells, a compound, ST-148 (Figure 1.8), was identified that is active against all four DENV serotypes. Mutations in resistant viral strains showed that the molecule binds to the C protein. In vivo studies confirmed the compound’s activity and further developments are currently in progress.\[^{49,57}\] The mode of action of this compound was recently published and it will be further discussed in this thesis (in the **CAPSID AS A DRUG TARGET** section).\[^{58}\]

**Figure 1.8. ST-148.**

DENV **NS3 serine protease** (NS3pro) belongs to the trypsin family and, as all viral proteases, it contributes to viral polyprotein processing. As this is an essential step in the viral replication cycle it is considered a valid antiviral target. It needs the cofactor **NS2B** that actively participates in the formation of the active site which results in being flat and composed of a
high percentage of negative charged residues for an easier binding to basic residues (Arg and Lys). To date, three approaches have been used to find inhibitors of this protein: designed peptidomimetics, library HTS and in silico drug design. The selection of peptides as inhibitors aimed to maximise the interaction with the protease active site and therefore basic residues are used as a scaffold. However, inserting positive charges on the molecule creates cell permeability issues. Furthermore, this approach requires complex (and therefore costly) chemical synthesis, representing an issue for a disease of the developing world. Several HTS were performed, using mainly a construct of NS3-2B protease and several compound series were identified. Both natural products and small molecules have been recognised to be active on the enzyme. They often present positively charged groups in their structure that retain the cell permeability issues (e.g. guanidines, as in compound 166347 reported in panel A of Figure 1.9). Ring rich compounds were found in silico to fit the flat and shallow binding site. From these, anthracene-based derivatives, obtained from compound ARDP0006 (panel B in Figure 1.9) modifications were used for structure-activity relationships (SAR) evaluations, but no further studies showing the activity of these compounds in cells have been reported to date.

![Figure 1.9. Protease inhibitors.](image)

**A)** Compound 166347. This is an example of guanidine rich inhibitor selected through HTS. **B)** Compound ARDO0006. This anthracene-based molecule is an example of inhibitors identified in silico.

**NS3 helicase/NTPase** (NS3hel) participates in the DENV replication cycle by using energy to unwind double stranded nucleic acid, making the single strand available to protein synthesis machinery and for virion formation. Like helicases from other (+)ssRNA viruses (e.g. HCV) also NS3hel is a promising antiviral drug target. As well as dsRNA unwinding, NS3hel has nucleoside triphosphatase (NTPase) and RNA triphosphatase (RNAse) functions. The two binding pockets (one for RNA and one for NTP) have proven to be not suitable for effective drug design: the first one is very shallow and the second one is not specific for any
nucleobase, making selective inhibitors a challenge.\textsuperscript{[51]} Attempts to overcome these issues were performed within this project and are described in the \textit{NS3 HELICASE AS A DRUG TARGET} section. Other research groups have employed other strategies and two compounds were discovered through \textit{in silico} docking and an enzymatic assay (ivermectin) and through HTS (ST-610).\textsuperscript{[49]} The first compound, reported in panel A of Figure 1.10, is a broad spectrum flavivirus helicase and protease inhibitor, but in cells it is more active against WNV and YFV than DENV.\textsuperscript{[63-65]} ST-610, in panel B of Figure 1.10, is a promising panserotypic DENV NS3hel inhibitor, active both in cell-based and biochemical assays and \textit{in vivo} (in the AG-129 mouse model). As it does not inhibit the ATPase activity and it induces A263T resistance mutation it probably binds to the RNA binding site.\textsuperscript{[66]}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{helicase_inhibitors.png}
\caption{Helicase inhibitors.
\textbf{A)} Ivermectin. This is an example of inhibitor identified \textit{in silico}. \textbf{B)} ST-610. This is an example of inhibitor identified with HTS.}
\end{figure}

\textbf{NS4B} is a hydrophobic transmembrane protein that associates to NS3hel. It probably participates in the association of the replication complex (RC) to membranes. Even if very little is known about its precise function within the DENV replication cycle, a few compounds have been shown to inhibit DENV infection through interaction with NS4B. These small molecules have been identified through HTS campaigns, screening libraries against DENV replicon.
Compound interaction with NS4B was then confirmed by mutations found in this protein in resistant viruses.\cite{49}

The N-terminal portion of NS5 is a methyltransferase (\textbf{NS5 MTase}), an essential enzyme for RNA capping. It is responsible for two (N-7 and 2'-O) methylations in a S-adenosyl-L-methionine (SAM)-dependent manner. Host methyltransferases are localised in the nucleus and cannot mediate these reactions, making NS5 MTase essential for DENV replication and therefore a promising drug target.\cite{31,34,51} Structural information about the enzyme in several complexes is available. Three identified binding sites have been used for structure-based drug design studies: one for SAM, one for the cap G base (also called “GTP binding site” because it was first seen when the protein was crystallised with GTP) and a shallow highly positively charged area that connects the two pockets and is supposed to interact with the RNA strand.\cite{36} Unfortunately, none of the designed compound were found to be neither highly active nor specific.\cite{49} S-adenosyl-homocysteine (SAH) was rationally modified into an analogue (panel A in figure 1.11) that was confirmed by crystallography to inhibit the NS5 MTase though binding to the SAM cleft. However, this molecule was not selective for DENV.\cite{67} Structural data confirmed also that ribavirin (panel B of figure 1.11) binds to the GTP binding site, but it is not active in patients.\cite{51} For the RNA binding site, a compound (ATA, in panel C of figure 1.11) was identified through virtual screening, but no follow up studies have been reported.\cite{68}

\begin{center}
\includegraphics[width=\textwidth]{methyltransferase_inhibitors.png}
\end{center}

\textbf{Figure 1.11. Methyltransferase inhibitors.}

\begin{itemize}
\item \textbf{A)} SAH analogue. Example of inhibitor rationally designed to bind SAM cleft.
\item \textbf{B)} Ribavirin. Example of inhibitor binding the FTP cleft.
\item \textbf{C)} ATA. Example of inhibitor identified \textit{in silico} to bind the RNA site.
\end{itemize}

Furthermore, a fragment based drug discovery (FBDD) approach was applied within this current project and details are reported in the \textit{NSS MTASE AS A DRUG TARGET} section of this
thesis. Novartis Institute for Tropical Diseases (NITD) also performed HTS with biochemical assays, but failed to identify putative NS5 MTase inhibitors.\[^{49}\]

The C-terminal portion of NS5 is the RNA-dependent RNA polymerase (**NS5 RdRp**). It is the most conserved DENV protein and it is responsible for RNA synthesis, an essential step within the DENV replication cycle. For these reasons it represents one of the most popular drug targets in antiviral research.\[^{54,69}\]

RdRp inhibitors are commonly classified in two categories: nucleoside/nucleotide analogue inhibitors (NI) and non-nucleoside inhibitors (NNI). The first class represents probably the largest group of drugs in antiviral therapy.\[^{49}\]

An example NI is compound NITD203 (panel A of Figure 1.12), that has shown good *in vitro* and *in vivo* activity. It is a prodrug, as the original compound (NITD449, in black in the figure) had cell permeability issues.\[^{70}\]

The popularity of NI as a drug strategy resides in the fact that molecules target the NS5 RdRp active site, which has a low resistance emergence rate as it is the most conserved portion of the enzyme. Furthermore, this region is conserved also among different serotypes, genotypes and viruses, allowing the possibility of development of broad-spectrum compounds. However, NIs present also important challenges, in addition to cell permeation. First of all, as several host cell enzymes are involved in NI activity, structure-activity relationship determination is almost impossible and therefore it is difficult to apply the classical medicinal chemistry methodology to this class of compounds. Secondly, as they are nucleoside/nucleotide analogues, they can interfere with cell processes, giving off-target toxicity problems.\[^{13,54}\]

NITD203 development was indeed interrupted for toxicity issues.\[^{70}\]

This challenge is less important for NNI compounds as they are usually highly specific as they do not target directly the active site. NNIs are non-competitive molecules that exert their function by binding to specific clefts, blocking changes in NS5 RdRp conformations that are necessary for the enzyme in order to accomplish its function. The disadvantages of this strategy are mainly related to the choice of the binding pocket: it should be well conserved for cross-serotype activity and for the reduction of resistance onset.\[^{31,34,49,51}\]

Crystal structures of NS5 RdRp closed conformation have been solved, but often allosteric binding sites are not defined in this state. For this reason, HTS has been a more successful tool for the identification of NNIs with respect to structure based drug design.\[^{13}\]

N-sulphonylanthranilic acid derivatives (like NITD2 in panel B of Figure 1.12) were identified with HTS, but their development was interrupted for poor CPE activity.\[^{71-73}\]

Even if more challenging, some putative NS5 RdRp NNIs have been identified *in silico* through virtual screening and are described in the **NS5 RDRP AS DRUG TARGET** section of this thesis.
HOST TARGET DRUG DESIGN

As for all viruses, also DENV is an obligate parasite that relies on host factors for key steps in its replication cycle. For this reason, targeting the required host cell factors is another therapeutic strategy that can be pursued for antiviral drug development. This approach has the advantages to be less serotype-dependent and to decrease resistance-related issues, as more viral mutations would be required in order to overcome the loss of a host cofactor. Additionally, it is easier to develop broad-spectrum compounds, as several viruses share the same host cofactor. The disadvantages of targeting host factors are mainly linked to the higher rate of toxic effects that might develop. However, as DENV is an acute disease, side effects might be less problematic than in the case of chronic diseases. Consequently, antiviral research has also moved in this direction.[31,49,51,74]

Host cofactors that are necessary in the DENV replication cycle and that could be used as potential targets are several (e.g. proteases, dihydroorotate dehydrogenase, glucosidases, kinases and factors of the cholesterol-biosynthesis pathway).[49,51,74] Some inhibitors that target these host factors are reported in Table A1.3 of APPENDIX 1. Only a general overview of host target drug design is presented here.

**Furin** and **signal proteases** are responsible for the cleavage of the viral polyprotein precursor together with NS3-2B serine protease and hence, their inhibition would have the same effect as the inhibition of NS3-2B protease itself in infected cells.[31,51] Unfortunately, small-molecule inhibitors of Furin have proven to be challenging; while targeting the signal protease would give too many side effects as it is involved in several other cellular functions.[31,51] Compounds that inhibit selectively flavivirus RNA **translation** have, however, been discovered through
HTS. These molecules inhibit host factors, but the identity of the target protein has yet to be determined.\[^{49}\]

For viral RNA synthesis, the availability of ribonucleic acid bases is fundamental. Consequently, the synthesis of purine and pyrimidine bases plays an important role in the viral replication cycle.\[^{49}\] This was confirmed by the fact that screening campaigns have identified compounds that have antiviral effect by inhibiting either the mitochondrial protein dihydroorotate dehydrogenase (DHODH) or the inosine monophosphate dehydrogenase (IMPDH). The first enzyme is responsible for one of the steps required for de novo pyrimidine synthesis (e.g. uridine, U), while the second one is responsible of the conversion of inosine monophosphate (IMP) to xanthosine monophosphate (XMP), precursor of GMP. However, to show in vivo efficacy of these compounds is a challenge as pyrimidine bases (as U) and GMP levels can be restored through the diet.\[^{49,75,76}\]

Host kinases are responsible for nucleosides phosphorylation, an essential step for RNA synthesis.\[^{31,51}\] Thus, these proteins constitute another potential host target. However, the biological activity profile of these enzymes is often not fully understood, making drug discovery challenging.\[^{31,51}\] For example, known kinase inhibitors (AZD0530 and dasatinib in panel A and B of Figure 1.13 respectively) were also found to be host Fyn kinase inhibitors. Further in vitro studies have shown that this enzyme is critical in viral RNA replication through interaction with NS4B rather than nucleoside phosphorylation. This was demonstrated by the fact that drug resistant viruses presented mutations in the viral protein NS4B.\[^{77}\]

Three of the viral proteins (prM, E and NS1) have to be glycosylated for correct folding. Misfolding produces protein malfunction and the arrest of the viral replication cycle.\[^{49,51,74}\]

Consequently, the inhibition of glucosidases, in particular \(\alpha\)-glucosidases, is a promising therapeutic strategy as these enzymes are responsible for the mentioned viral proteins glycosylation.\[^{49,51,74}\] They seem particularly successful targets as one drug candidate that has recently finished clinical Phase Ib trial for DENV is Celgosivir (Figure 1.14), an oral pro-drug of castanospermine, a natural product from Castanospermum australae. It is a \(\alpha\)-glucosidase I and II inhibitor whose effect reflects in particular on NS1 unfolding. It is active both in vitro and in vivo in the AG129 mouse model.\[^{78,79,80}\] Unfortunately the efficacy data on DENV
patients was not as good as expected. Future clinical trials will be improved, adjusting doses
treatment frequency and duration.\textsuperscript{[81]}

![Celgosivir](image)

\textit{Figure 1.14. Celgosivir.}

Both endogenous and exogenous cholesterol have been found necessary for DENV infection.\textsuperscript{[49]} The exact role of cholesterol has yet to be fully understood, but it has been hypothesised that plasma membrane cholesterol relocalization, in particular near viral proteins, could affect membrane flexibility and/or stabilise the DENV replication complex. The inhibition of factors involved in cholesterol biosynthesis and transport have indeed been shown to have an antiviral effect and could be therefore interesting for anti-DENV therapeutics development.\textsuperscript{[49,51,82]} In particular, enzymes involved in the cholesterol biosynthesis pathway whose inhibition prompts an antiviral response are mevalonate (diphospho) decarboxylase (MVD), HMG-CoA synthase and squalene synthase. These enzymes could be used as drug targets.\textsuperscript{[82]}

The most severe DENV-related clinical features, DHF and DSS, have been associated with an antibody-dependent enhancement (ADE).\textsuperscript{[49]} For this reason, modulation of the immune system could help to overcome the illness. Patients with severe disease have high levels of \textit{cytokines}, \textit{chemokines} and IFN-\textit{γ} and consequently these factors have been correlated with disease severity. For this reason these factors have been considered as potential host drug targets. However, ADE and its molecular pathways has not been fully elucidated and consequently this approach has not been yet explored.\textsuperscript{[31,49,51]}
1.2 MOLECULAR MODELLING

MOLECULAR MODELING AND DRUG DISCOVERY

Drugs discovered by pharmaceutical research have made a great contribution increasing life expectancy by an average of 2 months each year and improving quality of life by transforming many life threatening diseases to almost negligible problems.\[^{83-85}\] The process that leads to a new molecular entity (NME) becoming available on the market for an efficacious therapy is long, costly and most of the times unpredictable. The discovery of each novel drug has it’s own peculiar “story”, but the overall process can be generalised. It commonly takes between 8 to 14 years and costs an average of 1.3 billion dollars to develop a NME.\[^{83,86}\] Everything starts with the choice of the disease and, if known, of the molecular target. Many scientific factors, in particular the understanding of the biological processes involved, play a role in this decision. However, also economic aspects are of great importance at this stage, in particular in pharmaceutical industries. It is fundamental for them to estimate the potential return that the drug could give when it reaches the market and evaluate if this is enough to cover the costs and gain money for future investments in research and development (R&D). After having identified the target, it is necessary to identify a lead molecule that is able to produce a selective therapeutic effect. The discovery of these compounds can be achieved in several ways (e.g. from natural products, from HTS campaigns or from rational drug design), but usually the first identified molecule is not suitable for progressing through drug development and eventually for human administration. It has to be optimised in order to improve both its pharmacodynamics and pharmacokinetics properties. Many analogue molecules are generally synthesised and evaluated at this stage. The most attractive compounds are then tested with in vitro and in vivo models. This constitutes the preclinical phase, where absorption, distribution, metabolism and excretion (ADME) properties are also evaluated, as well as the toxicology profile. The aim at this stage is to understand whether the compound(s) are safe enough to be tested in humans and to have an estimate of the initial dose that should be used. The reliability of the models used is therefore crucial at this stage in order to obtain the correct toxicology profile. Molecules that prove to be safe are then tested in human through clinical trials that can be divided into three phases. Phase I represents the first time that the compound is tested in humans. It is carried out on healthy volunteers with ascending dose regimens with the aim of evaluating the safety, tolerability pharmacokinetics and bioavailability properties of the potential drug. Efficacy can be evaluated in a second Phase Ib trial if human volunteers are a suitable population for the putative medicine. For example,
potential hypnotics are often first evaluated for efficacy at this stage. However, usually the candidate molecule is first evaluated in a phase II clinical trial involving between 100 and 300 patients depending on the complexity of the trial design. The tested compound efficacy is usually compared to placebo and, when possible, to the gold standard marketed therapy. The last clinical studies (phase III clinical trial) before the registration of the drug are on a larger scale (exposing thousands of patients) with the aim of confirming what has been observed in phase II, but with a stronger statistics. These trials also look for unwanted effects that are only apparent when large patient populations are exposed to the drug. Drug evaluation does not stop, though, with the approval to commercialisation. As the compound has never been tested on a large scale as the World population, it is important to continue to monitor its effects also after reaching the market through pharmacosurveillance.\[83\] Figure 1.15 schematises the drug discovery pipeline described in the text.

![Figure 1.15. Drug discovery pipeline](image)

Schematic representation of the drug discovery process. See text for description of the steps. Abbreviations: ADME= Adsorption, Distribution, Metabolism and Excretion.

From 1950, 1365 NMEs, most of which are small molecules, were approved for commercialisation.\[85,87\] Apart from a temporary peak between 1995 and 1999 and the unusually high number in 2012 (39 NMEs), the trend for the numbers of NMEs reaching the market each year is flat, with an average of 24 NMEs/year. Conversely, the required overall expenditure to get these NMEs to the market has been exponentially growing at a rate of 12.3% per year, whilst the sales growth is flattening. This “pharmaceutical crisis” scenario could be related to the fact that the “easy” targets have already been tackled, to the failure rate of 80%, to the continuous increase of regulation, to the requirement of significant expected improvement in respect to available therapy, to competition and to difficulties in innovation.\[85\]
In order to improve the situation and since “time is money”, industry has been applying techniques to speed up the drug R&D process. For instance combinatorial chemistry for analogue synthesis during the lead optimisation step has been greatly applied as it is a rapid method for parallel synthesis of a large collection of compounds.\textsuperscript{[83,88]} Another way to speed up the process is the development of automated HTS assays to screen huge compound libraries for a faster lead discovery. This technique, an essential facilitator of combinatorial chemistry, has been shown to be able to identify leads with a success rate between 50 and 60%, which is lower than the initial forecasts. This is mainly because it relies greatly on the quality of the screened libraries and on the strength of the screening assay. Additionally, the equipment is very costly and the screening of big compound libraries sometimes takes a long time, differently from a more focused approach. Attempts on HTS improvement, in particular on the quality and chemical diversity of the library, are being carried out, but the cost of this technique is still a big issue.\textsuperscript{[89]}

As it is based on using information technology, an important aid in greatly reducing time and costs of R&D can be given by molecular modelling, whose popularity is indeed increasing. Molecular modelling is a scientific field that exploits computational techniques (this is why it is often described as synonymous to computational chemistry) in order to explain complex chemical systems in terms of models. In general terms, a model is heuristic, a simplified description of a system that has the goal to enable understanding and predict macroscopic properties, starting from knowledge on an atomic scale. This can be achieved either exploiting experimental data or using a theoretical model or combining the two strategies.\textsuperscript{[90,91]} Drug activity is strictly correlated to molecular recognition and consequent binding to biological targets, like enzymes, nucleic acid, glycoproteins or receptors. This recognition relies on the formation of specific attractive interactions between molecules, like hydrophobic interactions or hydrogen bonds. Therefore, the comprehension of the characteristics that an ideal drug molecule should have in order to achieve these interactions is fundamental for efficient drug design. In a live organism a molecule also undergoes a series of processes (adsorption, distribution, metabolism and excretion) and consequently the compound structure must be optimized to enhance the opportunity for the molecule to interact with the molecular target. Molecular modelling can be used to help to consider all of these aspects through virtual experiments that are intrinsically cheaper, faster and safer than real experiments. \textit{In silico} results can help scientists to select compounds, removing those that do not have the correct profile, improving the probability of success at lower costs. The continued development and improvement of the computers’ processing capacity and of the molecular modelling programs themselves are contributing notably to the reliability of the results obtained with computational techniques and for all these reasons, the drug discovery process has increasingly involved computational chemistry in recent years.\textsuperscript{[84,92]}
Computer aided drug design is generally divided into two major categories: ligand-based drug design and structure-based drug design. Structure-based drug design aims to identify and optimize drug target interactions and is based on the lock and key metaphor. However, the availability or the possibility to construct the three-dimensional structures of these molecules is an absolute requirement. If structural data is not present or cannot be produced, ligand-based drug design represents a valid alternative. This approach is based on the analysis of various descriptors of a set of ligands of a biological target for which experimental data is available, in order to predict and optimize biological properties of compounds.

**QUANTUM MECHANICS**

Quantum mechanics (QM) is mainly based on the molecular orbital theory and Schrödinger’s equation. In a molecular system, its properties derive from the description of the particles (e.g. electrons and protons) that construct each atom that is obtained through the evaluation of the wave function associated with these particles. Schrödinger’s equation is the most accurate way to obtain these wave functions, using mathematical variables (quantum numbers). The partial differential equation is very complex and can be analytically solved only in very simple systems. Theoretically, this is the most correct approach that could be used in molecular modelling studies, but the complexity of the calculations described above, limits its application only to small molecular systems. Therefore, before using QM methodology to solve problems of interest, it is important to assess whether this grade of accuracy is necessary. Usually, problems that exploit QM calculations regard properties that depend on electronic distribution, like electron density studies and chemical reactions in which bonds are broken or formed.
MOLECULAR MECHANICS

Molecular mechanics (MM) methods apply Newton’s classical mechanics to molecular systems. Bigger approximations than in QM are applied and the electronic motions are ignored. In fact, energies in the system are calculated as a function of the atom’s nuclear position and electrons are considered to be always in their ground state. This reduces greatly the amount of calculation needed and allows calculations on systems with a large number of atoms.\textsuperscript{[92,96]}

The approximations made are several. The most important ones are that: molecules are treated as balls-and-spring systems, atoms are handled as solid spheres with fixed Van der Waals radius and charge, atoms differ among themselves by atom-type, bonds are considered unbreakable springs and electrons and nuclei are not studied as separated entities in the calculations. Atom-types do not only contain information about the atomic number of the nuclei, but they also specify the hybridization of the atom and therefore the geometry (normal bond lengths, angles and dihedral values) of the system concerning the atom of interest. In this way, a molecule is translated into a series of atom types and bonding lists, considering each atom as a separate entity. A series of mathematical equations are then employed to compute the descriptors of the molecular system. The collection of the mathematical equations and the atom parameters is defined as force fields.\textsuperscript{[92,96,97]}

FORCE FIELDS

Force fields are sets of mathematical equations and parameters that allow the description, as accurately as possible, of the intra- and inter- molecular interactions within a molecular system. The mathematical equations give the possibility to generate the potential energy of a system, within molecular mechanics approximations, and their derivatives, the forces. The equation forms vary between different force fields, but a general procedure for the calculation of the potential energy of a system is the sum of the different components of bonded and non-bonded interaction for each atom, as reported in Equation 1.1.

\[
E_{\text{tot}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{Coul}} + E_{\text{vdW}} + \text{Additional term}
\]

\textit{Equation 1.1. General equation used by force fields.}\textsuperscript{[96]}

The potential energy is calculated as the sum of different energy components. Some of them calculate the energy correlated to bonded interactions: bond stretching ($E_{\text{str}}$), angle bending ($E_{\text{bend}}$), dihedral torsion ($E_{\text{tors}}$). Others calculate the energy correlated to non-bonded interactions: electrostatic ($E_{\text{Coul}}$) and van der Waals ($E_{\text{vdW}}$) interactions. Some force fields use also an additional term.

Within the bonded interactions (bond stretching $E_{\text{str}}$, angle bending $E_{\text{bend}}$ and dihedral torsions $E_{\text{tors}}$, as indicated in the equation), each energy component is calculated in respect to a reference value that represents the “unrestrained” or “equilibrium” value of bond length,
angle and dihedral. These values represent the value that the bond/angle/dihedral adopts when all the other terms in the force field equation are set to zero. The deviation from them is translated in terms of energy penalties, calculated using an empirically determined constant, as it will be more specifically illustrated in the description of Equation 1.2. All the equilibrium values and the constants constitute the parameters of a force field. Non-bonded interactions are usually described in terms of electrostatic and van der Waals interactions (Lennard-Jones interaction). More sophisticated force fields have also additional terms in order to give more accuracy to the energy calculation.

Looking at the energy calculation in more detail, one common functional form of a force field is reported in Equation 1.2. In order to simplify calculations for calculation rapidity enhancement, bond stretching and angle bending are approximated to harmonic systems that follow Hooke’s law. The energy contribution derives consequently from the cubic difference of the bond length/angle value to the reference value, multiplied by an empirical constant. This approximation is more accurate if the system is near its equilibrium. The equation is more complex for the calculation of the dihedral contribution to the total energy. Different force fields adopt different functions, but the dihedral contribution often follows a cosine trend. As often some specific torsion angles are preferred to others, force field parameters also consider an additional empirical term in the dihedral energy contribution to describe correctly the geometry of the molecules. Van der Waals forces are taken into account with an empirical expression called the Lennard-Jones function. This function relates both attractive and repulsive coefficients between pairs of atoms to their distance. For the calculation of the electrostatic properties of the molecule, the atom partial charges are placed in the atom’s nuclei and the electrostatic energies and forces are then calculated with Coulomb’s law. Both of the non-bonded interactions are considered pair-additive, which means that energy contributions are calculated for one pair of interacting atoms at the time. In reality these energy interactions are not pair-additive and this approximation, that simplifies calculations, might introduce some errors.

\[
E_{\text{tot}} = \sum_{\text{str}} k_b \left( l - l_{\text{eq}} \right)^2 + \sum_{\text{bend}} k_\theta \left( \theta - \theta_{\text{eq}} \right)^2 \\
+ \sum_{\text{tors}} k_\varphi \left[ 1 + \cos(n\varphi - \varphi_{\text{eq}}) \right] + \sum_{\text{vdW}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{\text{Coul}} \frac{q_i q_j}{r_{ij}} + \text{Additional term}
\]

*Equation 1.2. Common functional form of a force field.*

More detailed representation of a force field mathematical equation, with \( k_b, k_\theta, k_\varphi = \) force constants; \( l=\) bond length; \( \theta=\) angle; \( \varphi=\) dihedral angle; \( n=\) periodicity; \( A=\) repulsive term coefficient; \( B=\) attractive term coefficient; \( r=\) distance between atom i and atom j; \( q=\) atom partial charge)

Force fields are often empirical or semi empirical in order to fit the equations to the experimental or calculated data. For this reason, in the last few years several different force
fields have been developed and they differ one to the other for the function form or the parameters, or both. Each one of them has been developed in order to give a reliable description of a certain type of molecular system (e.g. small ligands, proteins or nucleic acids) and this means that to date a unique force field that fits every kind of situation has not yet been developed. Consequently, it is impossible to establish unequivocally which is the best force field, because the “best” is clearly dependent on what properties one wishes to calculate and what kind of system is the object of study.\cite{98}
STRUCTURE-BASED DRUG DESIGN

In general, the process of drug design aims to enhance specific drug-molecular target interactions in order to improve the drug’s activity and/or to reduce the side effects. The accessibility to structural information on the molecular target makes this process easier and therefore preferable. Methodologies that exploit available target structures are defined as structure-based drug design methods. The rapid increase in the number of determined three-dimensional structures of potential molecular targets (proteins, nucleic acids, etc.) has been decisive for the development of this kind of molecular modelling approach. The structure-based drug design process can be an iterative procedure in which at each iteration the aim is to improve possible drug-target interactions. Initially, a ligand with low activity might be found and then this can be improved through structure optimization. The first stage of a structure-based approach is the identification of the optimal molecular target for drug design and, if there is more than one available structure, the choice of the best one in terms of structure quality and biological relevance. Structural information can derive from four sources. The first and most common one is X-ray crystallography. The main advantages of the use of crystal structures are that ordered solvent molecules can be visible and that the quality of the structure itself is easily determinable through the examination of a few parameters (resolution, R value, R\text{free}, temperature factors, values of $\phi$ and $\psi$ backbone angles). The second source of structure data is nuclear magnetic resonance (NMR). In this case there are no similar parameters that could indicate the quality of the structure apart from the root mean square deviation value of the coordinates. Even if structural data availability is exponentially increasing, there is still a huge gap between known structures and sequenced genome portions. In the case of a lack of experimental structural data for a protein, Molecular Modelling can be a third source of structures. For instance, homology modelling is a computational technique that enables the “building” of macromolecular structures exploiting the existence of structural data of similar molecules. Proteins can be considered similar enough for the reliability of this approach if they are evolutionary related and they share at least 30% or the primary sequence. Following this rule, this technique could be used to model a human protein when the same protein, but non-human, has been solved. Alternatively, as proteins are generally at an energy-minimum state, de novo methods that “fold” primary sequences of proteins are being currently developed. Unfortunately, due to extensive conformation exploration, they are very time consuming methods. Even if models produced by Molecular Modelling have performed well, it is preferable to use experimental structural data if possible. A further source of structural data is cryo-electron microscopy (cryo-EM). This source is not commonly used because, even if always
Section 1: INTRODUCTION

1.2 MOLECULAR MODELLING

improving, the resolution of this technique is not at the same level of X-ray crystallography. However, these data can be used to aid the building of a homology model of the protein.\textsuperscript{[103]}

Once the structure has been chosen/produced, several computer-aided approaches can be used for structure-based drug design: an initial compound present in the structure can be modified; a database of compounds can be screened; a \textit{de novo} generation can be developed by virtual positioning fragments within the target site and linking them. All of these are based on docking and scoring methodologies.\textsuperscript{[100]}

DOCKING AND SCORING

A docking process allows the simulation of the binding mode that a molecule would adopt in its interaction with its molecular target. The ligand conformation and orientation within the binding pocket is called pose (exemplified in Figure 1.16) and the aim of this kind of methodology is to have a prediction of binding affinity (that often translates into activity) of the docked molecules. Docking and scoring can be described in a multi-step procedure that can be generalized in three moments: poses generation, poses scoring and final pose selection.\textsuperscript{[93,101]}

![Figure 1.15. Example of pose of a molecule in the RNA binding site of DENV NS3 helicase.](image)
The docked molecule is shown as a space-filled representation and the protein’s protein surface is coloured in green.

Ligand poses are generated by search algorithms. In the simplest model, the ligand can translate and rotate within the binding site. However, in reality both ligand and protein flexibility are important factors, greatly increasing the degrees of freedom involved in ligand-molecular target binding.\textsuperscript{[93,100,101]} As a consequence, the rigorous search for all the possible binding poses and their evaluation would be extremely expensive from a computational point of view. The first docking programs simplified this issue by computing a so-called rigid
docking. This approach does not take into account either the ligand or the target molecule flexibility, but only rotational and translational degrees of freedom. With the advances in computational performance and program improvements, it is nowadays possible to compute semi-flexible and flexible docking, where only ligand’s or both ligand’s and the target molecule’s flexibility are considered respectively. This has produced a great enhancement in the accuracy of results.\textsuperscript{\textit{93,104}} Even if computational capability is continuously increasing, a rigorous conformational search is still not feasible. In fact, only a small amount of the conformational space is sampled by docking programs and the sampling method can differ according to the searching method. On this basis, search algorithms can be divided into three groups. The first group, called systematic methods, achieves the construction of the possible conformations through the incremental (e.g. every 30°) rotation of torsional angles. Random or stochastic methods belong to the second group of search algorithms. They apply random changes to the conformation of the ligands and then they evaluate the poses with a pre-defined probability function. The third group of searching methods is by simulation, such as energy minimisation or molecular dynamics. By these methods, the attempt is to find the pose at the energy minima, generally considering both the ligand and the protein as flexible.\textsuperscript{\textit{101,104}} The evaluation of the generated poses is accomplished by the use of scoring functions that estimate binding free energy and consequently the molecule’s activity. As in the case of the pose searching procedure, the exhaustive calculation of the ligand-protein interactions is impractical. For this reason, some approximations have been developed and the scoring functions can be classified in force field-, empirical- or knowledge-based.\textsuperscript{\textit{93,101,104}} As the name suggests, the first category is based on the MM force fields described previously and on Equations 1.1 and 1.2. In general terms, the interaction energy is calculated through the sum of the receptor-ligand interactions and the internal energy of the ligand. In the case of flexible docking software, in which also protein flexibility is taken into account, there is an additional term that represents the internal energy of the protein. Empirical-based methods use scoring functions in which parameters are calculated in order to resemble the experimental data on binding energies or affinities. Knowledge-based scoring functions are mathematical equations that are built in order to reproduce experimental structures, with the assumption that crystallographic (or NMR) data resembles energetically favoured conformations.\textsuperscript{\textit{101}} Docking and scoring experience has shown that different molecular systems are well evaluated using different types of scoring functions and that each scoring function carries some imperfections, as they are approximate. If the ideal scoring method is not known for a molecular system, a common procedure is to repeat the scoring with more than one method and to use a consensus score that combines the obtained results, improving the probability to identify favourable ligands.\textsuperscript{\textit{95,101}} Despite the fact that a consensus score increases the procedure’s performance, it is still forecast that the scoring values might contain some
deficiencies. Consequently, the user’s contribution in selecting the final poses through visual inspection of the actual ligand-protein interactions is fundamental for the selection of the best molecules.\textsuperscript{100,101,105}

DE NOVO DRUG DESIGN

Unlike in virtual screening approaches in which more or less large databases of known molecules are docked, scored and rescored in the desired protein site, a \textit{de novo} drug design approach aims to build new chemical entities that are not present in the literature. This can be achieved through two main methodologies.\textsuperscript{101,106}

One method requires the binding of small fragments within the binding site. These molecules usually follow the “rule of three”, meaning that they have a molecular weight less than 300 g/mol, a logP less than 3, less than 3 hydrogen bond donors and less than 3 hydrogen bond acceptor groups.\textsuperscript{107} Fragments can be either co-crystallised with the protein or placed in the binding site with the docking and scoring methodologies explained above. The fragments that bound the best are then assembled into larger molecules. These new molecules can then be energy minimised and their best pose can be searched with the same procedure of the one described above.\textsuperscript{101,106,108}

The second method is principally used in lead optimization procedures. The methodology requires a “seed” molecule to be placed in the binding site. New fragments are then linked to this structure in specified “growing sites” in order to obtain an increase of the binding affinity. Also in this case, the new molecules obtained can then be energy minimised and the best pose can be searched with the same procedure of the one described above.\textsuperscript{83,84}

Despite the fact that with these techniques there is the possibility to explore novel chemical space, an opportunity that is not normally accomplished through common virtual screening approaches, \textit{de novo} drug design methods are not widely adopted because the novel designed molecules often are very complex and difficult to synthesise.\textsuperscript{101,106,108} Groups are trying at the present time to overcome this major problem, improving the available programs, like LigBuilder or the fragments based tools in the Molecular Operating Environment (MOE).\textsuperscript{106,108}

These programs are now able to rate the new described molecules on the base of their ease of synthesis and to use this information to improve the output.
LIGAND-BASED DRUG DESIGN

As already mentioned, if possible, structure-based drug design is preferable. In some cases, such as membrane bound proteins, the three-dimensional structure of the target is difficult to achieve with the present techniques or in another cases the structure is not yet available. If it is not even possible to build a homology model, or in the case the target molecule is not known, the only strategy that can be followed is to apply ligand-based drug design methods. The common feature of these techniques is the fact that the drug design process is based on drug-like molecules, chemical structures and on their activity data. Three techniques can be used: pharmacophore search, similarity search and quantitative structure-activity relationships (QSAR). Even if these methods are classified as ligand-based methods, it is not unusual to find that they are used in combination with structure-based approaches.[93]

PHARMACOPHORE SEARCH

A pharmacophore is a set of chemical features that are related to molecule activity. Common features usually are: hydrogen bond donor (e.g. –OH), hydrogen bond acceptor (e.g. O, N), hydrophobic group (e.g. –CH₃), aromatic group, anion (e.g. –OPO₃²⁻) or cation (e.g. –NH₃⁺).[93,108]

The interaction between a small molecule and a biological target does not only depend on the presence or absence of particular chemical features in the compound, but also on the space disposition of these groups. If the spatial relationship between the features is specified through coordinates and distances between them, then it is a three-dimensional (3D) pharmacophore.[93] 3D pharmacophores are commonly used also in structure-based drug design approaches because they can aid molecule conformation selection prior to a docking procedure in order to select only those molecules that can potentially establish the desired interactions.[101] More sophisticated 3D pharmacophores can also include the projection of features. In this case, the interacting group of the biological target is indicated and ligand features that are able to give that interaction and that are at the correct distance are accepted.[108] An example of a 3D pharmacophore is reported in Figure 1.17.

A pharmacophore search usually consists of comparing ligand’s chemical features with the features indicated in the pharmacophore. In the case of a 3D pharmacophore, it is a conformation-dependent procedure, so the search should be conducted on all the significant and acceptable conformations of the ligands.[93,101,108]
SIMILARITY SEARCH

If a lead compound is available it is possible to search molecule databases in order to find similar molecule that could display drug activity. Molecules can be similar across different characteristics, but in order to make a comparison and calculate a similarity index, it is important to be able to extrapolate these features as molecular descriptors. Molecular descriptors are numbers that can be calculated from the chemical structure of a molecule and not necessarily experimentally determined also if they could have an experimental counterpart. According to the information required for the calculation, molecular descriptors are usually categorised as one-dimension (1D) descriptors if they require only the molecular formula (e.g. molecular weight); two-dimension (2D) descriptors if they require the 2D chemical structure (e.g. partition coefficient, count of chemical features); three-dimensional (3D) descriptors if they require 3D chemical structure (e.g. shape, molecular volume, 3D pharmacophore features).[93]

The calculation of the similarity index from these descriptors can be achieved through different methods, but most of them give a numerical output that ranges from 0 to 1, where 0 signifies absence of similarity and 1 suggests the maximum grade of similarity. A similarity index of 1 does not necessarily translate to molecular identity. As an example, 2-chlorophenol and 3-chlorophenol (Figure 1.18) have the same molecular weight (MW) so their similarity index regarding this descriptor is 1, but they are not the same molecule.[93]
Figure 1.17. Example of molecules that have similarity index of 1 without being the same molecule.
The two described molecules have the same molecular weight and therefore the similarity index calculated on this
descriptor is 1. However using other properties they are not equal. A) 2-chlorophenol. B) 3-chlorophenol.

QSAR

Quantitative structure-activity relationships (QSAR) are mathematical models that correlate
molecular descriptors to activity. Their aim is to be able to predict how the molecular
descriptor could be modified in order to improve the molecules’ activity. In order to compute
these, it is necessary to have a set of molecules with known activity parameters (e.g.
inhibitory concentration IC\textsubscript{50}) and molecular structure. The QSAR model correlates the activity
data to properties inherent to the structure of the molecules of the set, with the aim to give
an indication as to which are the important features for new molecules with enhanced
activity.\textsuperscript{93,108} It is seen immediately that the choice of the adequate molecular descriptor is
crucial for a useful QSAR study that could be predictive. For this reason, more than one
descriptor is commonly taken into consideration in one drug design process.\textsuperscript{108}

MOLECULAR DYNAMICS SIMULATION

Together with energy minimization, molecular dynamics (MD) is a type of simulation that can
be performed with computational chemistry. It is considered to be an important investigative
technique that consists in the analysis of the temporal behaviour of a molecular system over
time. This means that it is possible to gain information about the motion of the atoms and
about the stability of the system itself. MD simulations apply classical mechanics and
molecular mechanics functions to the N interacting atoms of the system and Newton’s
equation of motion, reported in Equation 1.3, is applied to each one of them.\textsuperscript{90,93} According
to Newton’s second law of motion, if a force F is applied to an object with mass m an
acceleration a that is proportional to the original force will be produced. This expression can
be translated into another form in which the force is in relation to the distance (r) and to the
time (t) according to the derivative expression indicated in Equation 1.3. This equation is
coupled to the calculation of energies, forces and potentials on the atoms that is performed
through the force field functions described previously in the FORCE FIELDS paragraph.\textsuperscript{90,93}
In some aspects a MD simulation can be compared to a silent cinematographic movie. In a simplistic view, a movie consists of a series of snapshots taken at a very short time period one after another and shown at a high speed. In the same way, in a MD simulation the whole time period (usually from tens of ps to ns or even ms with powerful computers) is divided into small time steps (usually of the order of fs) and a snapshot (i.e. the state of the molecular system) is taken at each time step. Snapshots include the coordinates of each atom that is part of the system, but also other information like each atom’s energy, velocity, force, as well as system temperature and pressure. The change in the position and parameters of the atom are calculated from one step to the other by solving, simultaneously, the motion equations. In this way the atom’s motion is approximated to the sum of little linear movements dependent to the initial velocity and force applied to each atom at each time step. Usually the motion equations are solved making sure that the total temperature and pressure of the system are constant during the whole duration of the simulation.\cite{90,93}

As already discussed, the fact that MM is used rather than QM introduces into the simulation a series of approximations. Therefore, standard MD simulations cannot be used for the simulation of particular properties that depend on electronic distribution, like chemical reactions, bond or angle vibrations, and electronically excited state. These are more accurately described by QM methods.\cite{90}
1.3 INTRODUCTION TO THE BIOLOGICAL ASSAYS

The potential antiviral compounds identified with several Molecular Modelling approaches for DENV NS3 helicase, NS5 RNA-dependent RNA polymerase and NS5 Methyltransferase were tested in vitro using different types of assays. For a better understanding of the results, here is reported a brief introduction to the tests performed. Unless otherwise stated, the biochemical assays were performed by the author.

**CYTOPATHIC EFFECT (CPE) INHIBITION ASSAY**

The cytopathic effect (CPE) inhibition assay was performed by our collaborators at the Rega Institute for Medical Research, K. U. Leuven (Belgium) and, differently from the other tests, it is a cell based *in vitro* assay that aims to evaluate the general antiviral activity of compounds. The particular cell lines used are susceptible to infection when exposed to specific pathogens. For DENV infection, Vero cells were found to be suitable. DENV infection of untreated cells leads to cell death. Therefore, the ability of the compounds to inhibit viral infection can be measured through cell viability. Specific reagents are able to produce coloured derivatives when they react with molecules that indicate that the cell is alive (for instance ATP). Alternatively, substances (e.g. methylene blue) that stain only living cells can be used for the same purpose.

The assay protocol used consists of exposing infected cells to different compound concentrations and then observing cell viability through quantification of the obtained coloration through ultraviolet light (UV) absorption. By subsequently plotting compound concentration against cell viability it is possible to calculate the half maximal efficacious concentration ($EC_{50}$) for active compounds. Figure 1.19 shows a schematic representation of the protocol and of the analysis of the results. 8000 Vero cells per well were exposed to the compounds before being incubated for eight days with DENV2. After this, following fixing and staining with methylene blue the inhibition of CPE was evaluated by UV absorption against two control systems: infected cells and non-infected cells. Similar experiments conducted on uninfected cells were used to monitor compound toxicity.
**1.3 INTRODUCTION TO THE BIOLOGICAL ASSAYS**

![Figure 1.18. Schematic representation of the CPE assay.](image)

In each well the infected cells are incubated with the compounds at different concentrations for eight days. Then they are fixed and stained with methylene blue. In this example the test well plate, the cell control (non-infected cells) and a viral control (untreated infected cells) are indicated. An example of a slightly active compound is highlighted in green and an example of a cytotoxic compound is highlighted in purple. Data is then collected and plotted in concentration-activity charts for EC₅₀ calculation.

**THERMAL SHIFT ASSAY (TSA)**

The thermal shift assay (TSA) is usually performed to evaluate the ability of molecules to bind to proteins. It is based on the principle that the complex of a protein with a ligand is more stable than the protein alone. Consequently, when a protein is heated and it is induced to unfold, it will have a higher melting point (Tₘ) if it is bound to a ligand. The method used for the determination of the melting point is a fluorescent probe (e.g. sypro orange, SO) that is quenched in solution, while it is fluorescent upon binding to the hydrophobic regions of proteins. As a protein unfolds, it exposes more hydrophobic regions and therefore the fluorescence intensity grows proportionally to this phenomenon, until the protein is completely melted. The intensity of fluorescence is measured and plotted against temperature for the interpolation of the melting point, as shown in Figure 1.20. The melting point (Tₘ) is represented by the temperature in which 50% of the protein is unfolded and this value can be calculated from the curve. Ligands can induce some changes in the stability of
the protein upon binding (either stabilising or destabilising the system) and this is reflected in deviations in $T_m$. Therefore, the binding ability of compounds can be assessed through measurements of $T_m$ shifts, with respect to the protein alone. At higher temperatures, other phenomena take place, such as protein aggregation and/or dye dissociation, and lead to a loss of fluorescence. Thus, only the range of temperatures around the melting point is considered in data analysis.\(^{[109]}\)

*Figure 1.20. Fluorescence profile according to increase in temperature in a thermal shift assay.*

**FLUORESCENCE POLARISATION (FP) ASSAY**

A fluorescence polarisation (FP) assay is used to assess the formation of a complex between a protein and a ligand that contains a fluorophore (also called probe). The assay is based on the fact that when a fluorophore is excited with linearly polarised light, it emits light with a degree of polarisation, which is inversely proportional to its Brownian motion in solution. In this type of motion, the velocity of a particle is inversely proportional to its mass. Therefore, when the probe is alone in solution, the mass of the system is low and also the fluorescence polarisation results to be low. However, when the probe is bound to the protein, the mass of the system is significantly higher, as well as the FP value. Figure 1.21 shows a schematic representation of the basic principle of the fluorescence polarisation.
**Figure 1.19. Schematic representation of the FP assay.**

The monochromatic light, at the ideal absorption wavelength for the probe, is produced by a light source and then it is polarised through a polarisation filter (a). The polarised light interacts with the molecules in solution. Here two possible scenarios are shown: the probe in solution, that is able to move more than the probe in complex with the protein. The fluorescence produced is then analysed with a detector after the passage through a polarisation filter that selects only parallel waves with respect to the source polarisation (b), and one that selects only perpendicular waves with respect to the source polarisation (c).

Using the excitation light plane as a reference, the parallel and the perpendicular emission light intensity are detected and the FP is calculated according to Equation 1.4.\(^{[110]}\)

\[
FP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}
\]

**Equation 1.4. Fluorescence polarisation.**\(^{[110]}\)

\(I_{\parallel}\) is the parallel emission light intensity and \(I_{\perp}\) is the perpendicular emission light intensity.

According to this formula, the probe concentration and the fluorescence absolute intensity do not affect FP values, therefore they have to be chosen according to the instrument sensitivity.\(^{[110]}\) However, this implies that if dissociation constants have to be computed, the fluorescent ligand concentration must be kept constant and the test must be done on different protein concentrations.

The aim of the assay was to assess the putative competition activity of the compounds with the natural substrate for the active site. Consequently, the natural ligand with a fluorophore probe was used and after the calculation of the natural ligand binding dissociation constant, a fixed concentration of protein and probe were chosen for compound screening at multiple concentrations. A reduction of FP is correlated with the probe displacement from the protein's site and therefore it is also related to compound inhibition of the natural ligand-protein complex formation.
The probes used in this study were two, one for each DENV NS3 helicase binding site. For the RNA tunnel, a 10 nucleotide RNA strand corresponding to the 5’ end of the DENV genome was used. The small nucleic acid filament was coupled to the fluorophore (6-carboxyfluorescein, or 6-FAM) in the 5’ position. No problems related to RNA secondary structure formation were anticipated, as the length of the RNA strand is too small to give energetically feasible folding. The oligonucleotide is represented in Figure 1.22. The position of the 6-FAM probe at the 5’ end of the oligonucleotide was considered to improve the test performance as the protein binds the oligonucleotide in this region. Consequently, the difference in movement between the bound and the unbound probe is higher and therefore the variance in FP values as well.

For the helicase NTP binding site, a GTP molecule modified with a fluorophore (bodipy) was used. As DENV helicase has a NTPase activity, the assay was carefully performed in the presence of EDTA (ethylenediaminetetraacetic acid), a chelating agent for ions. The aim was to remove the catalytic ion from the binding site (Mg$^{2+}$ or Mn$^{2+}$), avoiding the probe’s hydrolysis. The structure of the GTP probe is reported in Figure 1.23.
The main interactions between DENV helicase and the NTP occur at the phosphate level, therefore it is important that the fluorophore is linked to the nucleotide at the level of the base or of the sugar moiety (as in this case) and not to the third phosphate group in order to assure binding. The probe used in the assay contained a mixture of the two possible substitutions in the sugar hydroxyl groups: either 2’ or 3’.

**HELICASE ATPASE ACTIVITY INHIBITION ASSAY**

The helicase ATPase activity inhibition assay was performed with the objective of studying the inhibition of the NTPase activity of DENV helicase exhibited by designed compounds. The assay principle concerns the liberation of a phosphate group for each starting molecule during ATP hydrolysis. If the enzyme is capable of this reaction, there is liberation of the anionic group that can be detected with a colorimetric method, described schematically in Figure 1.24.

![Figure 1.22. Schematic representation of the ATPase assay.](image)

The helicase hydrolyses ATP to ADP and phosphate. The phosphate is complexed by ammonium molybdate, forming an α-Keggin structure. After the reduction of a Mo atom by ascorbic acid, the complex converts to a β-Keggin structure, which is blue and adsorbs light at 595nm. The presence of an inhibitor impedes the formation of phosphate and therefore the development of the blue colour.

The ammonium molybdate coordinates the phosphate group, creating a complex 3D arrangement called an α-Keggin structure. The reduction of a Mo(VI) atom to Mo(II), with the transfer of 4 electrons from the oxidation of ascorbic acid is aided by the acid environment of sodium citrate in acetic acid and triggers a rearrangement of the α-Keggin cage to another complex 3D structure called β-Keggin that absorbs light at 595nm, giving a blue colour. Like the formation of the latter structure, also 595nm light absorption is proportional to the concentration of the phosphate ion in solution, resulting in a reliable method for the
assessment of the ATPase activity of the protein through the assessment of phosphate concentration.

For every experiment, a calibration curve is built from known concentrations of a phosphate salt, in order to reduce random errors correlated with slightly different reactant volumes used. The inhibition of the ATPase activity is assessed by comparing the concentration of phosphate ions produced by the helicase in presence of the compound and the one produced by the enzyme alone, taking into account the spontaneous ATP hydrolysis in solution.

**HELICASE UNWINDING ACTIVITY INHIBITION ASSAY**

The helicase unwinding activity inhibition assay was performed by our collaborators at AFMB in Marseille University (France) with the aim of studying the activity of designed compounds and the correspondence with the RNA binding activity measured with fluorescence polarisation. The assay is based on the ATP-dependent unwinding reaction performed by the helicase, as shown in Figure 1.25. A double stranded RNA (dsRNA) is prepared by annealing a 5′ $^{32}$P labelled RNA strand with a longer non-labelled RNA strand. The reaction is then started by adding ATP. If the helicase is active, the dsRNA is divided into two ssRNA filaments. As the labelled ssRNA and the labelled dsRNA have different molecular weights, they can be resolved in polyacrylamide gel and quantified through autoradiography. The relative concentration of
these two species indicates the unwinding activity of the helicase. If the enzyme has a good activity, ssRNA will be in greater quantity, while the action of an inhibitor will impede nucleic acid unwinding, giving higher concentration of the dsRNA species. By testing different compound concentrations and comparing the ssRNA and dsRNA amounts to the ones produced by the enzyme alone, IC\textsubscript{50} values can also be determined.

**POLYMERASE ACTIVITY INHIBITION ASSAY USING RADIOACTIVE PROBES**

The polymerase activity inhibition assay was performed by J.-C. Guillemot group at AFMB in Marseille University (France) with the aim of assessing the ability of compounds to inhibit the activity of the polymerase enzyme. The polymerase activity was observed by applying the test schematised in Figure 1.26.

![Figure 1.24. Schematic representation of the polymerase activity assay.](image)

See the text for further explanation of the steps.

Briefly, given a RNA template and the necessary nucleotides, the synthesis of a complementary RNA strand can be detected with a radiometric method. The RNA template (a polyU strand in the scheme, as an example) is usually a strand formed by “cold” nucleotides. The “bricks” necessary for the RNA synthesis are given with a percentage of radiolabelled (in the case of this particular experiment with tritium, \textsuperscript{3}H) complementary nucleotides (in green in the figure) that can be incorporated in the double strand RNA (dsRNA) during the enzymatic activity. After stopping the polymerase reaction, by sequestering the catalytic ions with EDTA, the buffer solution contains the dsRNA as well as the exceeding nucleotides. Therefore, the dsRNA is fixed on filter paper that allows the passage of the nucleotides alone. In this way
errors due to the presence of radiolabeled nucleotides that were not incorporated are eliminated. The radioactivity is then evaluated. In the presence of an inhibitor, the synthesis of dsRNA is reduced (or completely blocked) and consequently, less radiolabeled dsRNA is collected in the filter and the detected radioactivity is lower. The comparison of the radioactivity values obtained for protein-compound complexes at various concentrations compared to the ones obtained from the protein alone can therefore be used for testing the inhibitory activity of the molecules and IC$_{50}$ determination.

**POLYMERASE ACTIVITY INHIBITION ASSAY USING PICOGREEN**

A second assay that did not rely on radioactive labelling was also used by J.-C. Guillemot group at AFMB in Marseille University (France) to measure inhibition of polymerase using a fluorescent dye PicoGreen®. Also in this case the formation of dsRNA is monitored as an indication of the enzyme activity. The fluorescent dye is able to bind to both dsDNA and dsRNA superstructures that form spontaneously. The molecule structure of PicoGreen® is reported in Figure 1.27.

![Figure 1.25. PicoGreen.](image)

The quinolinium group that intercalates between base pairs is highlighted in green; while the benzo-thiazolinium moiety that interacts with the negatively charged phosphates is highlighted in green. The two groups are coupled. The aliphatic arms are able to span along 4 base pairs.

PicoGreen® has excitation and emission wavelengths of 485nm and 530nm respectively, both when it is free in solution and when it is bound to the nucleic acid. Figure 1.26 also shows the binding mode of the molecule: its quinolinium group intercalates between base pairs, while the benzo-thiazol moiety is able to create electrostatic interactions with the negatively charged phosphates of the RNA, stabilising the whole aromatic coupled system in a precise conformational state. This rigidification of the molecule upon dsRNA binding produces an up to 1000-fold increase of fluorescence intensity of PicoGreen®. In the assay, this increase is proportional to dsRNA formation itself and therefore it can be used to quantify the production
of double stranded nucleic acid.[111] High fluorescence values are expected for the protein control, while a decrease of fluorescence is expected in the presence of an inhibitor that blocks the synthesis of RNA and therefore the formation of dsRNA. Testing different compound concentrations in comparison to the results of protein alone, IC$_{50}$ values of the molecules can also be obtained.

**METHYLTRANSFERASE 2’-O AND N-7 METHYLATION ACTIVITY INHIBITION ASSAY**

The methyltransferase activity inhibition assay was performed by K. Barral group at AFMB in Marseille University (France). It is the same as the polymerase activity inhibition assay that uses radiolabeled probes. The only difference is that the radioactive probe used is S-adenosyl methionine (SAM), with a tritium ($^3$H) atom on the methyl group linked to the sulphur of the molecule. This group is transferred to the cap by the NS5 MTase and therefore incorporated in the capped RNA. Differently capped RNAs can allow the assessment of the two methylation processes ($^{7}$MeGpppA-RNA for 2’-O methylation GpppA$_{2’OMe}$-RNA for N-7 methylation). Similarly to the polymerase assay, also in this case the radioactivity values are proportional to the methylation reaction catalysed by the protein and they can be used to assess the ability of small molecules to inhibit the enzyme.
1.4 AIMS AND OBJECTIVES

The four serotypes of the mosquito-borne dengue virus (DENV) are a health burden responsible for around 50-100 million new cases and 22,000 deaths per year. This pathogen is worryingly spreading out of the endemic regions, putting at risk 2.5-3.5 billion people. Despite the efforts of the past 20 years only symptomatic therapy is available to date, making antiviral research on DENV a global health unmet need. As vaccine development has turned out to be more challenging than expected, the development of anti-DENV drugs is important for infection management.

Following this, the overarching aim of this PhD thesis was to identify novel targets to combat DENV infection and subsequently to discover potential original chemical leads exploiting those targets. This main aim was subdivided as follows.

I. Identify suitable drug targets. Anti-DENV drugs could target either host or viral proteins. Even if more sensible to serotype-specificity and to emergence of resistance, viral targets are currently better understood and usually different from host factors, allowing a reduction in off-target effects probability. The objective was therefore to find novel potential ways to target viral proteins essential in DENV replication that are preferably not present in humans.

II. To generate knowledge in order to understand better the mechanisms and roles of the chosen targets in DENV replication. The four chosen targets have been extensively studied during the last few years. However, knowledge generated about their mechanisms and roles is still incomplete and this could be partially responsible for the challenges that have been met in finding potent anti-DENV therapeutic agents. For this reason, basic research was included during this PhD to increase understanding of the mode of action of a promising capsid inhibitor (ST-148) and of the means of translocation of NS3hel along the RNA.

III. Find novel chemical leads to develop effective antiviral agents against DENV infection through applying classical Medicinal Chemistry approaches, chemical synthesis and in vitro assays. In particular, by carrying out different in silico approaches for the screening of publicly available libraries or for fragment-based drug discovery; starting either from known protein inhibitors or from modelling the interactions of the known structures of the four chosen targets with small molecules.
1.5 BIBLIOGRAPHY


Section 1: INTRODUCTION

1.5 BIBLIOGRAPHY


[84]: David C. Young, COMPUTATIONAL DRUG DESIGN - A Guide for Computational and Medicinal Chemists, (2009), John Wiley & Sons, Inc. Publication

[85]: B. Munos, Lessons from 60 years of pharmaceutical innovation, Nature Reviews Drug Discovery 8 (2009), p. 959-968, DOI: 10.1038/nrd296


[103]: M. Topf, A. Sali, Combining electron microscopy and comparative protein structure modelling,
Section 1: INTRODUCTION


[105]: P. J. Gane, P. M. Dean, Recent advances in structure-based rational drug design, Current Opinion in Structural Biology 10 (2000), p. 401-404, DOI: 10.1016/S0959-440X(00)00105-6


[108]: Molecular Operating Environment (MOE), 2011.10; online user tutorials, http://www.chemcomp.com/


Section 2: CAPSID AS A DRUG TARGET
2.1 INTRODUCTION

A PROMISING DRUG TARGET

Three structural proteins are encoded in the viral (+)ssRNA: Envelope (E), the membrane protein precursor (prM) and the capsid protein (C). All of these proteins are required for a correct and functional assembly of the virion. As described in the main INTRODUCTION section, 180 copies of the first two proteins are inserted in a host cell derived bilayer with which they constitute the viral envelope. This structure incorporates multiple copies of C that complex with one strand of viral RNA, forming the nucleocapsid. The exact mechanism of the construction of the nucleocapsid is still not fully understood, but it is supposed to involve prM, E and probably some host factors. Conversely, its critical importance in the formation of infective DENV particles has been demonstrated. In fact, the absence of C induces the formation of subviral particles that do not contain the viral genome and consequently are not able to replicate in new host cells. The interference with the formation or with the disassembly of the nucleocapsid should result in the inability of DENV virus to deliver its genome to the host cell, impeding the replication cycle. Despite the fact that C is the least conserved among the Flavivirus family (with homology between 15-90%), its critical function makes C an attractive target for anti-DENV antiviral therapy development. This hypothesis is supported by the evidence that the molecule ST-148, active against all four DENV serotypes both in vitro and in vivo, indeed binds and inhibits the capsid protein C.

STRUCTURE

The immature capsid protein contains a hydrophobic signal C-terminal portion (anchC) that anchors the protein to the ER modified membrane. This portion is cleaved by the viral protease NS3pro in complex with the cofactor NS2B during maturation. The mature form of C is a highly α-helical 12kDa protein that forms symmetric homodimers in solution that were also observed in the NMR structure that was obtained for residues 21-100 (PDB ID: 1R6R). Each monomer is constituted of four α-helices (from α1 to α4). The first three α-helices (α1-α3) form a 3-helix core, while the fourth (α4) is longer and extends from this bundle. Panel A of Figure 2.1 shows the surface of C dimer, where monomers are differentiated by colour.
panel B of the same figure the ribbon structure of the C dimer is represented, where α1 is coloured in green, α2 in red, α3 in blue and α4 in silver, with different shades in order to highlight the monomers.

\[\text{Figure 2.1. DENV capsid protein. (PDB: 1R6R)}\]

\textbf{A) Surface of the C dimer.} The surface corresponding to one monomer is coloured in red, while the one of the other monomer is in light blue. \textbf{B) C dimer shown as ribbon.} In both monomers the α-helices are represented with the same colour but lighter in one monomer and darker in the other monomer, in order to distinguish them. α1 is shown in light/dark green, α2 is in orange/red, α3 is in light/dark blue and α4 is represented in light/dark grey.

As can be observed in the figure, the majority of monomer contacts in the dimer involve α2-α2’ and α4-α4’ and the bigger α4 helix was indeed found to be essential for dimerization. Furthermore, dimer-dimer contacts are mainly hydrophobic, with very few electrostatic interactions.\[^6,8\]

A high percentage (25%) of the protein is constituted by basic residues (15 Lys and 9 Arg). The peculiarity of this structure is that the total charge is distributed in an asymmetric manner, placing the highest density of positive charges in the solvent exposed portion of the α4-α4’ region. On the opposite side, in the cleft formed by α1-α1’ and α2-α2’, residues are highly apolar, forming a concave hydrophobic region.\[^8\]

Nuclear localisation sequences (NLS), involving random coils (before α1, between α3 and α4 and after α4) and a solvent exposed site of α4 (Arg85-Lys86) have also been identified. The involvement of these regions (in particular the one between α3 and α4 and the one within α4) in capsid localisation within the host cell nucleus has been confirmed with mutational studies. As these sequences resemble NLS in other proteins (e.g. NS5 RdRp), it has been suggested that nuclear translocation is mediated by host proteins of the importin family, but neither the reason for C nuclear localisation nor the mechanism with which it occurs have been fully understood to date.\[^10,11\]
**MECHANISM OF ACTION**

It has been shown that the capsid protein is fundamental for DENV replication as its deletion produces “empty” sub-viral particles that do not carry the viral genome and therefore are not infectious.\textsuperscript{[6]} It appears that the active form of C is the homodimeric one observed in solution, but the exact mechanism of C recruitment and nucleocapsid formation is not yet fully understood. However, it has been suggested that the different charge and hydrophobicity distribution on the C homodimer play an important role for nucleic acid incorporation in the virion structure. According to this hypothesis (Figure 2.2) the highly conserved hydrophobic portion of C interacts with the inner part of the hydrophobic membrane of the envelope, whilst the positive charged residues binds the viral RNA, allowing the new virion particles to carry the viral genome and therefore to be infective. The interactions with the RNA are non-specific, resembling the histone/chromatin interaction.\textsuperscript{[8,12,13]}

![Figure 2.2. Model of capsid interactions with the envelope membrane and viral RNA.](image)

In this model of DENV virion structure, C interacts with the envelope membrane through its hydrophobic cleft, coloured in green in the figure. On the other side, the highly positively charged surface under the α4-α4’ helices interacts electrostatically with the viral RNA, rich in negative charges as indicated by the magenta “.” symbols. The membrane is represented by the light and dark red rectangle and the RNA is shown as a purple line.

It has been shown that DENV infection causes lipid droplet (LD) redistribution within the host cell.\textsuperscript{[14]} Even if the exact mechanism is not fully understood, this recruitment is fundamental for viral replication. It has been observed that C co-localises with these LDs within the cells during infection.\textsuperscript{[14]} Further experiments have demonstrated that C is able to bind LDs through the same highly conserved hydrophobic region on the α2-α2’ helices of the dimer.\textsuperscript{[10,14]} On the basis of the fact that C-LD association occurs early in DENV infection, it has been suggested that LD might aid nucleocapsid formation by providing a support for C-RNA complex construction. However, as it cannot be excluded that C binds the RNA after its transfer to the ER for particle morphogenesis, the exact role of LD and the C-LD complex remains unclear.\textsuperscript{[14]}

66
C has other unclear behaviours that have yet to be elucidated. Similarly to LD association, C has shown to be able to bind plasma low-density and very-low density lipoproteins (LDLs and VLDLs respectively) through the conserved highly hydrophobic α2-α2' helices of the dimer. However, the importance of this phenomenon is unclear as these complexes have not yet been observed in vivo, probably because of the transient nature of the complex itself.\textsuperscript{[15]} Moreover, as already mentioned, the capsid translocates to the nucleus during infection. The reason and the exact mechanism of this delocalisation are unknown, but the NLS sequences identified have proven to be essential for this phenomenon.\textsuperscript{[10,11]}

**POSSIBLE MOLECULAR TARGET SITES**

C does not interact with natural small molecule ligands, differently from other viral proteins, like NS3hel that binds NTP.\textsuperscript{[6]} Therefore, the protein does not present putative druggable binding sites. However, C displays regions with different properties that allow the protein in the homodimeric form to interact with different partners.\textsuperscript{[6,8]} Even if C has many unclear functions in the DENV replication cycle (see above), the main established purpose of C remains nucleocapsid formation for the inclusion of the viral genome in the novel virion particles. As explained in the model schematised in Figure 2.2, C interacts with both RNA and the envelope membrane and C homodimer constitutes the minimum functional unit.\textsuperscript{[8,12,13]} Different studies, including the cryo-electron microscopy (cryo-EM) structure of DENV virion determination, have shown that the homodimers interact with each other in a disordered manner, creating bigger complexes to form the nucleocapsid.\textsuperscript{[6,13,16,17]} Consequently, as the interaction between homodimers appears to be important for the stability of the nucleocapsid, the interference with this protein-protein interaction is a promising strategy for drug design.
ST-148

Figure 2.3. ST-148 molecular structure.
Chemical name: 3-amino-N-(5-phenyl-1,3,4-thiadiazol-2-yl)-6,7,8,9-tetrahydro-5H-cyclohepta[b]thieno[3,2-e]pyridine-2-carboxamide

ST-148, in Figure 2.3, is a promising anti-DENV compound that was identified with a CPE-based HTS campaign, screening approximately 200,000 compounds. It is active in vitro against all four DENV serotypes (with IC₅₀ between 2.8 and 0.016 μM) and it is selective for Flaviviruses. It was also proven not to be mutagenic.⁴ Drug-resistant viruses were isolated in vitro and showed a single point mutation (S34L) on the C protein, suggesting that the antiviral activity of the compound involves this viral component. Furthermore, the mutation does not appear to interfere with neither the viral fitness nor C properties. Ser34 is a highly conserved residue among DENV serotypes positioned near the end of α₁ helix, in a likewise conserved region, explaining the compound’s efficacy against all serotypes.⁴ Furthermore, ST-148 binds directly with C, corroborating the importance of this viral protein in the antiviral activity of the molecule. In vivo studies in the AG-129 model have highlighted that ST-148 is well tolerated, but has low oral bioavailability compared to intraperitoneal administration. More and importantly, the compound has demonstrated its efficacy after administration by either route through decreasing viral titers both in tissues and in the blood, as well as the levels of inflammatory cytokines.⁴

Further in vitro assessments not only confirmed existing data, but also brought new insights about the compounds’ activity. Through time-of-addition studies, ST-148 was shown to inhibit DENV infection, but neither through a virucidal effect nor RNA synthesis interference.⁴,⁵ Instead it inhibits the production of infectious virions through both viral entry and assembly steps in the replication cycle. Specifically, ST-148 interferes with the assembly and the disassembly of the nucleocapsid, through the stabilisation of C self-interaction, in a dose dependent manner. Nevertheless, the compound does not appear to impede the interaction of C with either LDs or RNA, or the transport of the target protein to the nucleus of the infected cell, confirming that these phenomena are not related to the antiviral activity of the molecule.⁵
ST-148 has often shown to be a promising inhibitor of DENV infection through interaction with the C protein. However, the mechanism of this interaction is not understood. With the aim of building a model of ST-148 antiviral activity for future development and improvement on this molecule, its mechanism of action was investigated through Molecular Modelling techniques. Work was performed in collaboration with Heidelberg University and SIGA Technologies Inc. that provided the biological data used as a starting point for the in silico studies.
2.2 RESULTS AND DISCUSSION

MOLECULAR MODELLING STUDIES

The biological data summarised in the introduction is suggestive of the mode of action of ST-148 but is a poor indicator of how it functions at the molecular level. With the objective of understanding how ST-148 could stabilise C self-aggregation, several molecular modelling approaches, including protein-protein docking, molecular dynamics (MD) simulations and small molecule docking were used. As the functional form of C in solution is through homodimers, this structure was considered as a fundamental unit and C self-aggregation was studied through the simulation of the formation of tetramers both with wild type (WT) and S34L mutated C proteins. The effect of ST-148 in this process was then analysed. The details of all the procedures are explained in the METHODS chapter.

The structural starting point for this study was examining the 5 most different models out of the 20 belonging to the available NMR structure of DENV2 C (PDB ID: 1R6R) together with the S34L mutants obtained from the same models. Tetramers were then assembled through rigid protein-protein docking and the selection of the complexes to study was based on the following: biological data showed no reduction in nuclear localisation of C in presence of the compound and that NLS sequences have been shown to be essential for this process. Therefore, it was first assumed that the interaction between C homodimers did not involve these particular amino acids as otherwise they would be masked and no nuclear localisation would occur. Secondly, it was hypothesised that tetramer formation could also not involve the hydrophobic regions responsible for C-LD complex formation, as no effects on C-LD co-localisation after ST-148 exposure were observed. Thirdly, as ST-148 resistant viruses present a S34L mutation, it was inferred that C dimer-dimer interaction could involve the area where this residue is located on the protein surface and consequently, models of the tetramer in which this occurred were prioritised. Finally, as nucleocapsid structure has been observed to be disordered, the symmetry of the tetramer was not considered an important element.

The best model that fitted the above criteria was used for 30 ns MD simulations for both the WT and the S34L mutant structure, after the in silico insertion of the mutation. In line with biological data, the two molecular systems behaved in a similar way. In both cases, the tetramer changed during the first 2 ns into a stable conformation that was maintained for the rest of the simulation, as the systems had reached equilibrium.
The WT and S34L tetramer structures obtained after the energy minimisation step of the MD protocol were used for ST-148 docking. A 30 Å grid placed at the dimer-dimer interface of the tetramer complex and centred on either S34 or L34 was used for each molecular system. The best binding pose for the compound in the two molecular systems was then used for MD investigation, as previously carried out for the apo complexes. Both for WT-ST-148 and S34L-ST-148 complexes, during the first 2 ns of the simulation a similar change in conformation to the apo tetramers was observed. This new arrangement was then maintained for the rest of the total 30 ns as equilibrium was reached. As an example, Figure 2.4 shows the equilibrated structures for the WT tetramers in absence (on the left) and in presence (on the right) of the compound.

![Figure 2.4](image)

**Figure 2.4. Equilibrated structures of the WT tetramers in absence and presence of ST-148.**
The MD resulting structures for the WT tetramers are shown. On the left hand side of the arrow, the WT tetramer is not in complex with the compound, while the latter is present in the complex on the right. In all cases, one homodimer (indicated with HD1) is coloured in orange and red while the other one (indicated with HD2) is represented in cyan and blue. In all dimers, different shades of the same colour are used to highlight monomers. Ser34 at the interface of the dimers is highlighted with a space-filling model and ST-148 is shown in green.

In the WT tetramer, ST-148 binds at the interface of the two dimers, in a cleft formed by α1 and α3 helices of one unit and α1 helix of the other one. This pose stabilises after 20 ns and it is retained until the end of the simulation. Interestingly, the α2 and α3 helices of one dimer shift upon pose stabilisation in order to accommodate better the ligand in the cleft, as highlighted in Figure 2.5. The MD simulation of the S34L mutant in complex with ST-148

![Figure 2.5](image)

**Figure 2.5. Movements of α2 and α3 for ST-148 accommodation in the WT structure.**
The observed movement of α2 and α3 helices of one dimer for better ST-148 binding in the cleft is presented here. On the left the whole tetramer complex is shown, while the figure on the right is an enlargement of the binding cleft. The red ribbons represent the tetramer in absence of the ligand, while the blue ones represent the stable complex of the tetramer with the compound. ST-148 is shown in green using a space-filling model and the movements of the α-helices are indicated with arrows.
showed a different behaviour. In this case, the small molecule did not stabilise in the same cleft but explored one dimer’s solvent exposed surface through the transient interaction with more than one site.

With the aim of confirming the different binding of the compound to the WT and S34L mutant, ST-148 was docked in the stable WT tetramer-ST-148 complex structure. As previously done, the 20 Å grid was centred on either the Ser or the Leu34 at the interface of the two dimers. The docking results shown in panel A of Figure 2.6 were able to reproduce the ST-148 binding mode observed during the MD simulation of the WT system. In detail, the binding site is primarily composed by Val26, Leu29, Arg41 and Arg68 of one dimer and Ser34 of the other one. The binding of the compound arises mainly through hydrophobic interactions with the many hydrophobic residues of the pocket due to the high lipophilicity of the molecule. Ligand binding is further enhanced by the π-π stacking between Phe33 and the ST-148 phenyl ring together with the interaction of the electron-dense thieno[2,3-b]pyridine portion of the small molecule, tethered between Ser34 and Arg68, and the positive charge of the basic residue. This binding mode is consistent with the panserotypic antiviral activity of ST-148 as the main interacting residues (Phe33, Arg68 and Ser34) are highly conserved among DENV1-4 viruses and the other pocket forming amino acids are either conserved or maintain the same chemical-physical properties.

Panel B of Figure 2.6 shows clearly that the S34L mutation creates a steric hindrance that does not allow the binding of the compound, which occurs in such close contact with Ser34 in the WT system. This is consistent with the MD simulations that showed that ST-148 is not able to stabilise in the binding cleft and explores one dimer’s surface. All together, these results support that the S34L resistant mutation does not allow ST-148 to bind to the tetramer in the same stable mode of the WT.
Summarising, ST-148 stabilisation through dimer bridging is supported by the Molecular Modelling results and a model of the compound’s mode of action can be proposed on the base of in silico and biological data.

**MODEL OF ST-148 MODE OF ACTION**

As suggested from the available biological data, it appears that the mechanism of action of early and late stage inhibition by ST-148 is given by stabilisation of higher-order oligomeric C structures. According to the Molecular Modelling results, this is accomplished by hydrophobic and electrostatic interactions of this small molecule with the protein in a binding site between the capsid dimers and thus stabilises the tetramer complex itself. Accordingly, ST-148 could be classified as a direct protein-protein interaction (PPI) stabiliser. Moreover, the resistance inducing S34L mutation can also be explained with this model as ST-148 has been shown to bind to the mutant, but in a less stable manner, due to the steric hindrance caused by the larger Leu side chain that does not allow compound accommodation.
2.3 CONCLUSIONS

The Molecular Modelling study presented in this section was designed to uncover insights into the mechanism of action of ST-148, a potent panserotypic DENV inhibitor that interacts with the C protein, in accordance with the available biochemical, virological and imaging data. Several computational chemistry techniques were employed for the development of a methodology that, to the best of knowledge, was never applied to this molecular target and the results all corroborated the experimental data. Thus, a model of the mechanism of action of this compound has been developed and suggests that ST-148 acts as a protein-protein interaction stabiliser. Protein-protein interactions (PPIs) are involved in several important cellular processes and therefore are of huge interest for drug development. However, the inhibition or the stabilisation of PPIs have been historically challenging drug design strategies, in particular for small-molecules development.\textsuperscript{[18-20]} The interfaces between proteins tend to be larger and shallower than pocket binding sites and require a good understanding about interactions hotspots that contribute the most in binding. Fortunately, the advances in knowledge and in scientific tools in recent years, these strategies are becoming more popular and they are proving in a number of cases to be a winning strategy.\textsuperscript{[18-20]} In this study, a promising anti-DENV compound was shown to be a PPI stabiliser, confirming that this drug design strategy is achievable and a valid tool for drug discovery. The further development of ST-148-like molecules, aided by the information obtained with this collaborative study, could succeed in the discovery of a medicine that can successfully reach the market for DENV infection therapy.
2.4 METHODS

The computational methods for this study are presented here. A description of the molecular modelling programs used can be found in APPENDIX 2.

HARDWARE DETAILS

All docking simulations were performed on a 8 core computer with Inter Xeon 2.80 GHz E5462 CPUs, while MD simulations were performed on a 32 core computer with Inter Xeon 2.20 GHz E5-4620 CPUs.

PROTEIN-PROTEIN DOCKING

The NMR structure of C (PDB ID: 1R6R[8]) was downloaded from the Protein Data Bank website.[21] MOE 2010.10[22] Protein Consensus tool was used to compare the 20 dimer models present in the file for the selection of the 5 most different structures for rigid protein-protein docking. MOE 2010.10[22] Mutate tool was used for the construction of the S34L mutation structures starting from those of the wild type (WT), as both of them were used for each modelling approach undertaken. Prior to protein-protein docking simulation, each protein structure was prepared using Schrödinger Maestro 9.5[23] Protein Preparation Wizard. Schrödinger BioLuminate 1.2[24] Piper was then used for protein-protein docking. This simulation was performed by using as receptor and ligand the same C dimer. Complexes were generated by standard mode rigid docking, probing 70,000 ligand rotations. A 0.21 bonus was given to the tetramers presenting Ser34 in the dimer-dimer interaction surface. Finally, the best model was chosen through visual inspection.

MOLECULAR DYNAMICS SIMULATIONS

GROMACS 4.5.3[25] was used with AMBER99 force field for all the molecular dynamics (MD) simulations. Both WT and the S34L mutant were used for MD simulations. In the mutated molecular system, Ser34 was substituted by Leu34 in silico with MOE 2010.10[22] Mutate Tool, starting from WT. For MD simulation, a minimum 0.9 nm distance between the molecular
system and the cubic box employed and periodic boundary conditions (PBC) were applied. Water was added and described as an explicit solvent and the molecules were treated with the TIP3P model. The total charges of the systems were neutralised by adding 72 Cl\(^-\) ions. For the ST-148-tetramer complex simulations, the small molecule missing parameters were computed using AMBER12\(^{[26]}\) Antechamber software, using AM1-BCC charge method and GAFF force field for atom types. The obtained files were then converted with acype\(^{[27]}\) into a GROMACSm-compatible format. Two consecutive energy minimisations were initially performed, employing the steepest descendent (SD) method first and then the conjugate gradient (CG) method after, with the aim of a faster process and more accurate result. A force tolerance of 100 kJ mol\(^{-1}\) nm\(^{-1}\) was set for SD, while a 10 kJ mol\(^{-1}\) nm\(^{-1}\) force tolerance was applied with CG. However, in both cases the maximum number of iterations was set to 3,000 steps. Subsequently, a position restrain force of 1,000 kJ mol\(^{-1}\) nm\(^{-2}\) was applied to protein (and ligand) atoms for water molecules relaxation. Aiming a smoother equilibration, two consecutive 50 ps (50,000 steps, step size of 1 fs) position restrained MDs were consequently performed, saving coordinates, velocity and energy values every 500 steps. NVT conditions (constant number of atoms N, volume V and temperature T) were used for the first position restrained MD and a \(v\)rescale temperature coupling and 0.1 ps time constant was used to heat the system to 300 K. NPT conditions (constant number of atoms N, pressure P and temperature T) were used for the second position restrained MD, with the use of both temperature (\(v\)rescale, temperature 300 K and time constant of 0.1 ps) and pressure (Berendsen algorithm, 1 bar pressure and time constant of 0.5 ps) coupling. The production simulations, for a total 30 ns (15,000,000 steps with 2 fs step size) simulation time each, were run in the already described NPT conditions, saving coordinates, velocities and energy values every 1,000 steps. In all simulation stages, long-range electrostatic interactions were calculated with the Particle-Mesh-Ewald (PME) method with a 0.9 nm short range cut-off and short-range non-bonded interactions were computed only within a cut-off of 1.4 nm. Results were visually inspected with VMD\(^{[28]}\) and analysed with GROMACS 4.5.3\(^{[25]}\) tools and Grace\(^{[29]}\).

**SMALL MOLECULE DOCKING**

The small molecule docking simulations were performed with Schrödinger Maestro 9.5\(^{[23]}\). First, the protein complexes were prepared with the Protein Preparation Wizard tool and ST-148 was prepared with LigPrep, generating all possible tautomers and Epik ionisation states at pH 7.0 ± 2.0. Secondly, Glide was used for the definition of the grid that was centred on the Ser34 (or Leu34) at the interface of the two dimers in the tetramer complex, for all docking simulations. The grid size was set to 30Å for the first docking simulations with WT and
mutated systems; while it was set to 20Å for the second case. The inner grid box size was set to the default 10Å for both grids. The same software was then used for docking simulation, with the following settings: standard precision (SP) mode, nitrogen inversion and ring conformation sampling, penalisation of non-planar conformation for amides and no energy minimisation of binding poses. In each docking simulation, 25 poses were generated and refined with SP mode before visual inspection.
2.5 BIBLIOGRAPHY


[6]: S. Mukhopadhyay, R. J. Kuhn, M. G. Rossmann, A structural perspective of the flavivirus life cycle, Nature Reviews Microbiology 3 (2005), p. 13–22, DOI: 10.1038/nrmicro1067


[18]: A. Mullard, Protein-protein interaction inhibitors get into the groove, *Nature Reviews Drug Discovery* 11 (2012), p: 173-175, DOI: 10.1038/nrd3680


[21]: Protein Data Bank, [www.pdb.org](http://www.pdb.org) (Accessed, October 2014)


Section 3: NS3 HELICASE AS A DRUG TARGET
3.1 INTRODUCTION

A PROMISING DRUG TARGET

RNA replication is an essential step in the viral replication cycle. Using a positive sense single stranded RNA template, (+)ssRNA, the replication machinery generates a minus-stranded RNA, (-)RNA, that acts as a template for the synthesis of viral (+)ssRNA. During the synthesis of a new nucleic acid molecule, NS5 RNA-dependent RNA polymerase (NS5 RdRp) produces a double stranded genome (dsRNA) that has necessarily to be separated. The novel RNA is then modified with a cap structure at the 5’ end that protects the nucleic acid from degradation and allows its correct transcription. Due to the lack of appropriate host proteins, most of these steps are carried out by viral-encoded enzymes, mainly belonging to NS3 and NS5 proteins.\textsuperscript{[1,2]} Three enzymatic properties have been attributed to one of the most characterised viral proteins, the C-terminal helicase portion of the NS3 multifunctional enzyme (NS3hel): nucleotide triphosphatase (NTPase), helicase and 5’ RNA triphosphatase (RNAse). The first two activities work in concert, allowing NS3hel to unwind the dsRNA produced by NS5 RdRp in a NTP dependent manner. The third one, is enhanced by the interaction with NS5 and makes NS3hel responsible for the first of the four capping reactions further explained in the NS5 METHYLTRANSFERASE AS A DRUG TARGET section.\textsuperscript{[3-5]} All of this clearly indicates that NS3hel is highly involved in essential steps of RNA replication. Consequently, the inhibition of this viral protein would block the process in more than one step.\textsuperscript{[3-6]} Thus, NS3hel is a promising drug target for the design of novel potent anti-DENV drugs, as confirmed by the fact that some potential enzyme inhibitors start to appear in literature.\textsuperscript{[7-9]}

STRUCTURE

The overall geometry of the whole NS3 enzyme is reported in Figure 3.1. The N-terminal 170 residues (in orange in the figure) have a serine protease function (NS3pro), while the remaining 70% of the protein is the NS3hel. NS3pro is responsible for the cleavage of the polyprotein at the junctions between C/prM, NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B and NS4B/NS5. For its activity, the presence of the cofactor NS2B (in yellow in Figure 3.1) is necessary as it stabilises the protease in the active form. The interaction between these two
proteins occurs between the central hydrophilic region (residues 49-95) of NS2B and its binding site in the NS3pro structure. The protease active site is flat, shallow and negatively charged, explaining the binding preference of basic residues (Lys and Arg). As it is a serine protease, the residues that catalyse the hydrolytic brake of the amidic bond are a His (His51), an Asp (Asp75) and a Ser (Ser135).[^4][^6][^10]

![Figure 3.1. Complete NS3 enzyme (PDB 2WZQ).](image)

The protease portion is coloured in orange, while the helicase/NTPase portion is shown in blue. Carbon atoms of a small portion of the NS2B cofactor are represented in yellow.

NS3hel occupies the C-terminal portion of NS3 and it is constituted by 437 residues. It is divided into 3 equally sized domains, as reported in Figure 3.2. Domain 1 (residues 180-326), in blue in the figure, and domain 2 (residues 327-481), in red, are similar. They are characterized by a large six parallel β-sheets with an evident twist and four α-helices around them (as visible in the ribbon structure reported).[^6][^11][^12] The main difference between domains 1 and 2 is that the second one has a long β-hairpin that is in contact with domain 3, in the back of the RNA binding site in the region where the nucleic acid enters the cleft. Domain 3 (residues 482-618), in green in the figure, is mainly composed of α-helices, having 4 approximately parallel α-helices and other three short α-helices. It is the domain that differs the most to the helicases of other members of the Flaviviridae family, as for example HCV. DENV NS3hel domain 3 has indeed a unique fold as no significantly homologous structures could be found in the Protein Data Bank.[^13] Furthermore, domain 3 is not always present in helicases. For example, human helicases do not have a third domain.

The interfaces of the domains enclose two binding sites. The one located between domain 1 and 2 can accommodate the triphosphate moiety of either a nucleotide (NTP) or of the
triphosphate 5’ end of the RNA and it is where the phosphatase activity (both NTPase and RNase) takes place. A second, larger, binding site is between the three domains and it is responsible for the binding of 5 RNA nucleotides for the helicase function.\cite{6,11,12}

Figure 3.2. Helicase portion of NS3 (PDB:2JLV).

A) Protein surface. The three domain surfaces are coloured differently: domain 1 is blue, domain 2 is red and domain 3 is green. ATP, in pink, is in the NTP binding site, between domains 1 and 2. A RNA strand, in yellow, is in the RNA binding site, at the interface of the three domains. B) Protein ribbon. In this figure the protein is represented as ribbons. The same colour scheme of A) was applied here.

On the basis of conserved motifs, RNA and DNA helicases are divided in three superfamilies (SF). DENV NS3hel belongs to SF2 because it shares with these enzymes seven motifs, located in domains 1 and 2. Motif I (Gly196-Thr200), also called Walker A motif or “P-loop”, and motif II (Asp284-His287), also called Walker B motif or DEAH box, are situated in domain 1 and are involved in NTP phosphate and a divalent ion (either Mg\(^{2+}\) or Mn\(^{2+}\), catalytically necessary for NTP hydrolysis) binding respectively. In domain 2, two arginines (Arg460 and Arg463), also called “Arg fingers”, that belong to motif VI (Gln456-Arg463) are also involved in NTP binding and hydrolysis. The Walker A motif is the most conserved motif among all helicases. In fact, its GxGKT sequence is conserved among the whole 3 superfamilies. On the other hand, on the basis of the Walker B motif residues, SF2 is divided into 3 distinct subfamilies: DEAD (e.g. eIF4AIII, Vasa), DEAH (e.g. DENV NS3hel) and DEXH (e.g. HCV NS3hel) box proteins. Other motifs are involved in the binding and translocation of the nucleic acid: motif Ia (Pro223-Ala228), Ib (His262-Thr265), IV (Trp 360-Ala367) and V (Thr408-Ala416). Finally, motif III (Ile312-Pro324) couples the two enzymatic activities (NTPase and helicase).\cite{6,14,15}

The N- and C-terminal portions of NS3 have very different functions and influence each others activity.\cite{6,12} Crystal structures of the full-length NS3 protein have demonstrated that NS3pro is placed above the NTP binding site of NS3hel and that the two portions are connected through a flexible linker (residues 169-179) that confers high plasticity to the structure.\cite{12,16} Two
crystal structures in particular, showed that the full-length NS3 protein can adopt two distinguishable conformations (I and II). Conformation II is elongated, and NS3pro seems to be rotated by approximately 161° with respect to “conformation I”, causing the loss of some contacts between the helicase and the protease. Experimental data has shown that the two conformations have different helicase activities, with conformation I being less efficient in dsRNA unwinding. One hypothesis is that the lack of unwinding efficacy in this case is caused by the NTP binding inaccessibility, blocked by NS3pro. In addition to this, the fact that the protease is linked to the membrane-inserted protein NS2B, prompts the possibility that the rotation of the protease induces a different orientation of the whole NS3 with respect to the endoplasmic reticulum (ER) membrane. A model of this mechanism has been proposed and it is schematised in Figure 3.3. According to the model, the inhibition of the helicase activity also derives from the fact that conformation I (panel A) induces the RNA binding site of the helicase to face the membrane, obstructing the binding of the natural substrate.

Figure 3.3. Model of the two conformations of NS3 in respect of the membrane.
A) Conformation I. In this conformation NS3 is not able to bind either ATP or RNA. The NS3 structure for this conformation was taken from PDB ID 2WZQ. NS3pro (indicated as “Pro”) surface and ribbon are coloured in orange and for NS3hel (indicated as “Hel”) are in blue. The membrane is represented in pale green and the NS2B cofactor is coloured in red. The RNA molecule is represented as a purple circle, while ATP is represented by a dark green one. B) Conformation II. In this conformation NS3 is able to bind both ATP and RNA. The NS3 structure for this conformation was taken from PDB ID 2WHX. The same colouring scheme as A) was adopted here.
POSSIBLE TARGET SITES

NS3 helicase is responsible for double strand RNA unwinding, nucleoside triphosphate hydrolysis and 5’ RNA triphosphatase activity.\cite{6} In order to accomplish these functions, this enzyme presents two different binding pockets that could be used for rational drug design purposes: one for NTP and one for single strand nucleic acid molecules.\cite{19}

In the NTP binding site, the enzyme is able to bind either a nucleotide (NTP) or the triphosphate 5’ end of the RNA. In both cases, the enzyme hydrolyses the γ-phosphate in a divalent cation-dependent manner (either with Mn$^{2+}$ or Mg$^{2+}$).\cite{17,18} The interactions between the protein and the ligand are mainly at the level of the phosphates and some at the level of the ribose, but no particular interactions involving the base of the nucleotide have been observed. This explains the lack of specificity in regard to the base, which means that it is almost completely solvent-exposed. The main residues directly involved in ligand binding are: Thr200, Lys201, Arg418, Asn329, Asn416, Arg463, Gly198, Lys199 and Arg460 (reported in the schematic ligand interaction view of Figure 3.4).\cite{11,12,16,19}

![Figure 3.4. Schematic representation of ATP binding.](image)

The interaction rendering follows the legend on the right.

When the enzyme is bound to RNA, the NTP phosphates, three molecules of water and the OH moiety of Thr200 mediate the octahedral coordination of the divalent cation (Mn$^{2+}$ in the figure). In this conformation, the water molecule that takes part in the NTP γ-phosphate hydrolysis is positioned in the most favourable position for a nucleophilic attack. Furthermore, its interaction with Glu285 and Gln456 increases its nucleophilicity, facilitating the reaction. When RNA is not bound to the enzyme in the cleft at the interface of the three domains, Glu285 takes the place of one water molecule in the divalent cation coordination, leading to a
loss in NTPase activity. Consequently, RNA binding increases the NTPase efficiency of the enzyme.\textsuperscript{[11,12,16,19]}

The nucleic acid binding site is a shallow pocket at the interface of the three domains. From crystallographic data, five nucleotides with the bases stacked one to the other can take place in this site. In a similar manner to the NTP binding site, also in the case of the RNA binding site, there is no selectivity to the different nucleoside types as very few interactions are between the protein and the bases, while most of them occur at the level of the ribose-phosphate backbone (Figure 3.5).\textsuperscript{[19]}

![Figure 3.5. Schematic view of the interactions between the nucleic acid and NS3 helicase.](image)

The five co-crystallised RNA nucleotides are shown and protein residues are indicated. Colouring and interactions are shown as indicated in the legend.

From activity assays, DENV helicase has good unwinding activity for both DNA and RNA. However, it preferentially binds to RNA, explaining a slight specificity of DENV NS3hel for this nucleic acid.\textsuperscript{[11,20]} It has been hypothesised that this might be due to some interactions that involve the OH moieties in the 2’ position of the nucleotides. As the enzyme has to accommodate the negatively charged phosphate groups, it is not surprising that in this pocket there is a high presence of basic residues (Arg and Lys) that at physiological pH are positively charged.\textsuperscript{[14,19]}

Studies have shown that the mutation of some residues induces the inhibition of the enzyme’s activity. Surprisingly, some of these residues, like Arg376, Lys377, Lys380, Asp334, Glu335, Glu336 and Lys396 are located on the protein’s surface on domain 2, opening the possibility that the interaction of this domain with other proteins or even with the second chain of the unwinding dsRNA might be important. In the area of the NTP binding site, Gly198 and Lys199 mutation to Ala inhibited both NTPase and helicase activities, while four arginines (Arg460 and Arg463; Arg457; Arg458) seem to be important for the NTPase activity, but not very influential.
on helicase function. Only one interesting mutation concerned the nucleic acid binding site: Ile365Ala that inhibited only helicase activity.\textsuperscript{15,21}

As described in this chapter, the lack of base-specificity in NTP binding and the shape of the RNA binding site make both clefts challenging for drug design, explaining the limited advances in the discovery of NS3hel inhibitors despite the good availability of structural data.\textsuperscript{22}

**MECHANISM OF ACTION**

As for other SF2 helicases, like HCV NS3hel, RNA translocation occurs in a 3’-5’ direction. Crystal structures of different stages of the supposed mechanism of action have been solved, but they did not give valid insights into particular protein changes in conformation during the NTP binding-hydrolysis-release cycle, maybe due to particular constraints during crystallization.\textsuperscript{19} An approximately 40ns molecular dynamics (MD) study showed a scissors-like movement of the enzyme alone that was not shown when the enzyme was only bound to ATP. This suggests a mechanism of action similar to HCV NS3hel, in which a ratchet-like movement of the protein domains causes the translocation of the enzyme along a single stranded RNA (ssRNA) molecule.\textsuperscript{23-25} However, key aspects of the mechanism of action of DENV NS3hel remain elusive.\textsuperscript{19} Further knowledge about the translocation mechanism and on the 3’-5’ directionality in particular were obtained with molecular dynamics and binding energy studies computed in our laboratories. The new insights are presented in the **NS3 HELICASE MECHANISM OF TRANSLOCATION ALONG RNA** section.

The RNA unwinding mechanism is also not clear to date. There are two possibilities: a passive or an active unwinding. In the first case the helicase could bind to single stranded RNA taking advantage of the spontaneous opening of the duplex at the fork. The translocation of the helicase on the RNA would then exert the unwinding function. Even though this mechanism is serendipity-based, it should have high efficiency as the kinetics of local base opening are fast. Efficiency would, however, increase in the case of an active mechanism in which the helicase would operate an active disruption of base pairs hydrogen bonds. In any case, it has been proposed that the β-hairpin has a crucial role in disrupting the stacking interactions between bases and stabilizing the unwound state.\textsuperscript{14,19}
AIMS AND OBJECTIVES

As it is highly involved in the RNA replication cycle, NS3hel inhibition would lead to the arrest of DENV infection.\textsuperscript{[5]} Given its three enzymatic activities, NS3 could be targeted with several drug design strategies. For instance, competitive inhibitors could prevent natural substrate binding, or allosteric inhibitors could induce structural modifications that would stop the protein activity and molecules could also block RNA translocation. Furthermore, as NS3 interacts with other proteins in the replication complex, the disruption of these complexes with protein-protein inhibitors could lead to an antiviral effect.\textsuperscript{[5,26]}

With the aim of discovering novel, potent inhibitors of DENV NS3hel, three Molecular Modelling based approaches were adopted. One of them, discussed in chapter 3.2.1, was based on three commercially available compounds active against DENV helicase.\textsuperscript{[7]} As no structural data linked to these results was available, a ligand-based drug design approach was applied in order to find novel potential antiviral compounds on the basis of this information. Given the abundance of structural data on NS3hel, the other two studies were structure-based and aimed to design competitive inhibitors that could bind the RNA (see chapter 3.2.2) or the NTP (see chapter 3.2.3) binding sites of the protein.
3.2 RESULTS AND DISCUSSION

3.2.1 OUABAIN SHAPE-BASED DRUG DESIGN

Despite the available knowledge on DENV NS3hel, the lack of knowledge on the enzyme’s mechanism of action makes drug design of helicase inhibitors challenging. Nevertheless, at the 2011 International Conference on Antiviral Research (24th ICAR) organised by the International Society for Antiviral Research (ISAR), three compounds were reported as active against this viral enzyme for the first time. These molecules (Ivermectin, Paromomycin and Ouabain) are approved drugs for diseases not related to viral infection, for instance congestive heart failure and angina pectoris. No data specifies the exact site of interaction of these inhibitors with the enzyme, apart from a molecular modelling study that suggested that the binding site could be the entry of the nucleic acid binding site.

Consequently, even though several structures of NS3 helicase have been solved for DENV, it was decided to use the above reported promising results for a ligand-based drug design approach for the design of novel antiviral drugs. Briefly, the drug design approach applied was based on the selection of molecules belonging to a database on the basis of their shape similarity with one of the three active compounds. The selected compounds were then tested in vitro and further developed.

OUABAIN AS A REFERENCE MOLECULE

The optimal condition for shape screening requires the presence of data regarding the template’s active conformation. At the time of the study, knowledge regarding the interaction with DENV helicase was not available for the three molecules (Ivermectin, Paromomycin and Ouabain). For this reason and since the active conformation does not always correspond to the one at the absolute energy minimum, all the conformations corresponding to local energy minima of all three of the reference molecules were explored with MOE 2010.10, hypothesising that one of them corresponds to the active conformation. The screen was then performed by comparing the shape of molecules belonging to the database with shapes matching the query conformations. Molecules corresponding to the highest number of query shapes are more likely to adopt the unknown template active conformation. Therefore, the main hypothesis is that the selection of those molecules would lead to better screening results. An important consequence of this hypothesis is the relevance of the total number of
low energy conformations that the query molecule can adopt: the fewer the conformations the higher the probability that screened molecules match the active conformation.

Of the three reference molecules, Ouabain demonstrated the least number of conformations and was chosen as the reference molecule. From the structures reported in Table 3.1, it is possible to rationalise this with the fact that Ouabain is the most rigid molecule, characterised by a steroidal core; while the other molecules have a larger number of freely rotatable bonds.

Table 3.1. Chemical structures of the three template molecules.

<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivermectin</td>
<td><img src="image" alt="Ivermectin structure" /></td>
</tr>
<tr>
<td>Paromomycin</td>
<td><img src="image" alt="Paromomycin structure" /></td>
</tr>
<tr>
<td>Ouabain</td>
<td><img src="image" alt="Ouabain structure" /></td>
</tr>
</tbody>
</table>

The number of Ouabain conformations produced was 33. Since this number is too high to work with for shape comparisons, it was reduced through visual inspection. Duplicate conformations were eliminated and similar shapes were clustered. The final query selection
was of the three most different Ouabain templates, called Ouabain_1 (in green), Ouabain_10 (in blue) and Ouabain_22 (in red) as reported in Figure 3.6.

![Shape of the three Ouabain query conformations.](image)

*Figure 3.6. Shape of the three Ouabain query conformations.*
The surface of Ouabain_1 is in green, Ouabain_10 is in blue and Ouabain_22 is coloured in red.

**SHAPE SCREENING**

The three Ouabain shape queries were used as templates for the screening of molecules belonging to the SPECS database, a collection of purchasable drug-like small molecules, whose structures are available online. The version of the database that was used was the one updated in February 2012, containing approximately 209,950 drug-like molecules.

The shape screening was computed with the OpenEye ROCS program that operates a rigid superposition of molecules to the given queries. As the shape is a conformation-dependent descriptor, in order to compute an appropriate shape comparison it was necessary to apply a conformational search to the SPECS database, obtaining approximately 7,130,000 conformations.

The virtual screening was conducted comparing each small molecule conformation with all three of the Ouabain queries and scoring the match with the Tanimoto index. As no knowledge about which functional groups of Ouabain are important for its activity was available at the time of the study, only shape comparison was used for conformation scoring.

Conformations were ordered according to the score and only the best results were analysed, searching for conformations that could match more than one shape query. The screening was repeated five times, saving 1000, 2000, 3000, 4000 and 5000 best hits respectively. The final screening outputs were then collected and compared. There were 45, 29, 17, 8 and 3 conformations (corresponding to 14, 11, 8, 4 and 3 different molecules respectively) matching 2 queries when 5000, 4000, 3000, 2000 and 1000 hits were kept respectively. No conformation matched all three queries. Interestingly, the conformations that had good overlap with two queries always matched query Ouabain_1 and either Ouabain_10 or Ouabain_22. There were no matches that involved both Ouabain_10 and Ouabain_22, the most different shapes amongst the three.
According to the study hypothesis, a higher query matching corresponds to a higher probability in matching the active conformation and consequently only the conformations that matched two queries in each screen were considered for further investigation. Good Tanimoto scores belonged to the elements of this group of conformations and they ranged from a maximum of 0.7235 to a minimum of 0.6925. For each entry, the average score was computed for the two query matches. The chemical diversity of the molecules was also assessed by clustering them in 22 and 25 groups for structures with 75% and 85% similarity respectively. These two parameters and the visual inspection of the shape superposition were used for the selection of 11 compounds (reported in APPENDIX 3) that were purchased and tested in vitro.

IN VITRO EVALUATION OF THE SELECTED COMPOUNDS

All the purchased molecules were tested in vitro at the Rega Institute for Medicinal Research, K. U. Leuven (Belgium) by our collaborators. The biological assay was conducted on 8000 DENV infected Vero cells per well, with the aim of detecting inhibition of virus-induced cytopathic effect (CPE). The cells were exposed to different concentration of the compounds (50, 16.7, 5.56, 1.85, 0.62, 0.21, 0.069 and 0.023 µg/ml) and then incubated with DENV2. The result appraisal was done after fixing and staining the cells. The staining procedure was carried out with methylene blue that stains only live cells. The inhibition of CPE was visually evaluated against two control systems: infected cells and non-infected cells. As well as the eleven purchased compounds, also two reference molecules were tested in the same assay: Ribavirin, that shows to be cytostatic at high concentrations and dextran sulphate that shows inhibitory activity at high concentrations. Unfortunately, none of the 11 compounds selected with this procedure were found to inhibit DENV cell infection at any of the tested concentrations.

Nevertheless, five of the 11 selected compounds were also tested on DENV NS3 helicase with the binding and enzymatic assays discussed in chapter 3.4 of this section. All these assays were performed by the author in AFMB, Marseille University (France). In the first case, the aim was to assess the binding ability of the compounds to NS3 helicase (TSA) or to inhibit the formation of the protein-ligand complex (FP assay, using either a GTP or a RNA probe). The enzymatic assay evaluated the inhibitory effect of the compounds on the ATPase activity of NS3hel. In all the tests, the mother solutions were prepared by solubilising the compounds in 100% DMSO, for a final concentration of 20mM. Compound 1 (see APPENDIX 5) showed solubility problems already in the mother solutions and only a concentration of 10mM could be achieved. As reported in chapter 3.4, solubility problems occurred in the TSA and ATPase assays and unfortunately it regarded all five compounds of this series.
Despite solubility problems, encouraging results could be achieved from the FP assay, using RNA-6FAM as fluorescent probe: four of the five selected molecules were able to inhibit the binding of RNA to NS3hel. Two of them (compounds 2 and 3) were found to be active with low μM IC_{50} values, while the other two (compounds 1 and 4) demonstrated a weaker inhibition of RNA binding. The four molecules, their structure and their activity data (when it was possible to calculate values) are reported in Table 3.2. As already mentioned, none of these compounds inhibited DENV infection in cells as they were either cytotoxic (e.g. compounds 2 and 3) or they did not have antiviral effects. However, the inhibition of protein-RNA binding could be an appealing mechanism of action for an antiviral drug as the lack of interaction of the protein with the natural substrate would lead to a loss of unwinding activity. For this reason, compounds 2 and 3 were selected for further improvement with the aim of improving activity and reducing toxicity.

**Table 3.2. Structure of compounds that inhibit helicase-RNA binding.**

<table>
<thead>
<tr>
<th>Compound</th>
<th>NS3hel IC_{50} (μM)^a</th>
<th>CPE (μM)^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Structure</td>
<td>DV3</td>
</tr>
<tr>
<td>1</td>
<td><img src="image1" alt="Structure1" /></td>
<td>37.90</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="Structure2" /></td>
<td>7.86</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Structure3" /></td>
<td>15.77</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4" alt="Structure4" /></td>
<td>32.17</td>
</tr>
</tbody>
</table>

Abbreviations: NS3hel = Non-structural Protein 3 Helicase; IC_{50} = Half Maximal Inhibitory Concentration; DV3 = Dengue Virus Serotype 3; DV4 = Dengue Virus Serotype 4; CPE = Cytopathic Effect Assay; CC_{50} = Half Maximal Cytostatic Concentration; EC_{50} = Half Maximal Efficacious Concentration

^a = tests performed by the autor at AFMB, Marseille University (France)

^b = tests performed by the Rega Institute for Medicinal Research, K. U. Leuven (Belgium)
PRELIMINARY STRUCTURE-ACTIVITY RELATIONSHIP DETERMINATION

With the aim of confirming compound activity and of gaining preliminary structure-activity relationships, a series of analogues of compounds 2 and 3 were purchased and tested in vitro both on infected cells (CPE assay) and on the protein (FP assay, using the RNA-6FAM probe). Our collaborators (Rega Institute of Medical Research, KU Leuven and Marseille University) performed these assays. The CPE assay was performed as described before; while the evaluation of the inhibition of RNA binding to the helicase was assessed only on DENV4 NS3hel, using a different concentration of the RNA probe (100nM, see the EXPERIMENTAL SECTION for further details about the conditions of the biochemical assays). An enzymatic assay was also performed on compounds 2, 3 and on the compound 3 analogues that gave positive FP results (8) in order to establish if the inhibition of protein-RNA complex formation corresponded to the inhibition of the helicase activity of NS3hel. The assay was designed and performed by our collaborators in Marseille University, and monitored the unwinding activity of NS3hel. A 16 nucleotide strand of $^{32}$P labelled RNA was annealed with a 30 nucleotide RNA strand. The helicase unwinding reaction was then performed in presence of ATP. The quantity of ssRNA and dsRNA were then compared after a run on non-denaturating polyacrylamide gel. A brief description of this assay can be found in the INTRODUCTION section. Table 3.3 reports the structures of the analogue compounds and their activity data obtained from the RNA binding (FP), the enzymatic and the CPE assays. Compounds 2 and 3 were retested with the biochemical assays and the new data are reported in the table. Due to solubility issues, IC$_{50}$ of compounds 5, 6 and 7 could not be determined as it appeared to be higher that their maximum soluble concentration (250, 125 and 250µM respectively) in the assay conditions. Unfortunately, the FP results in Table 3.2 for compounds 2 and 3 could not be repeated. This was probably due to slightly different experimental conditions as the concentration of the RNA probe was different from the initial test. However, useful activity data were obtained and preliminary structure-activity relationships could be determined. An important observation is the coherence of the RNA binding inhibition and of the enzymatic activity inhibition results. This supports the hypothesis that compounds that inhibit the binding of RNA to NS3hel cause a loss in the enzyme’s function.

Consistently, as well as compound 2, all its analogues with the indole central scaffold (compounds 5, 6 and 7) were inactive in both biochemical and cell-based in vitro systems. Conversely, anti-DENV activity (either from biochemical or cell-based assays) could be observed for the other three molecules. These results suggest that a phenyl ring in the central scaffold, as in compounds 3, 8 and 9, enhances the activity of the compounds.
### Table 3.3. Structure and activity of compounds 2 and 3 analogues.

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>FP - NS3hel IC₅₀ (µM)ᵃ</th>
<th>Enzymatic - NS3hel IC₅₀ (µM)ᵇ</th>
<th>CPE (µM)ᵇ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><img src="image2" alt="Structure 2" /></td>
<td>&gt;1000</td>
<td>&gt;1000</td>
<td>11.10 12.00</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Structure 3" /></td>
<td>37.99</td>
<td>18.33</td>
<td>8.01 -</td>
</tr>
<tr>
<td>5</td>
<td><img src="image5" alt="Structure 5" /></td>
<td>&gt;250</td>
<td>-</td>
<td>51.11 10.13</td>
</tr>
<tr>
<td>6</td>
<td><img src="image6" alt="Structure 6" /></td>
<td>&gt;125</td>
<td>-</td>
<td>&gt;22.54 40.13</td>
</tr>
<tr>
<td>7</td>
<td><img src="image7" alt="Structure 7" /></td>
<td>&gt;250</td>
<td>-</td>
<td>58.98 10.51</td>
</tr>
<tr>
<td>8</td>
<td><img src="image8" alt="Structure 8" /></td>
<td>245.6</td>
<td>32.12</td>
<td>&gt;116 2.85</td>
</tr>
<tr>
<td>9</td>
<td><img src="image9" alt="Structure 9" /></td>
<td>&gt;1000</td>
<td>-</td>
<td>&gt;142 8.57</td>
</tr>
</tbody>
</table>

**Abbreviations:** FP= Fluorescence Polarisation Assay; NS3hel= Non-structural Protein 3 Helicase; IC₅₀= Half Maximal Inhibitory Concentration; DV4= Dengue Virus Serotype 4; Enzymatic= Enzymatic assay; CPE= Cytopathic Effect Assay; CC₅₀= Half Maximal Cytostatic Concentration; EC₅₀= Half Maximal Efficacious Concentration; SI= Selectivity Index

ᵃ tests performed by AFMB, Marseille University (France)
ᵇ tests performed by the Rega Institute for Medicinal Research, K. U. Leuven (Belgium)
Analysing more in detail compound 3 and its analogues (compounds 8 and 9), it appears that the adamantane group could be involved in compound 3 cytotoxicity and lack of antiviral effect in the cell, as less cytotoxicity and antiviral activity are seen in the analogues without this moiety. Unfortunately, the same pattern is not reproduced in the biochemical evaluation, where the most active compound of the three was indeed compound 3. While in the CPE assay a comparable but slightly better activity was registered for compound 8 respect to 9, in the biochemical assay while 8 retained some activity and compound 9 was unable to inhibit either RNA binding or NS3hel unwinding activity.

Since at least one compound (8) was able to inhibit both NS3hel and DENV infection in cells, it was hypothesised that the toxicity related to compound 3 did not allow a proper assessment of its antiviral activity in infected cells. Furthermore, as there is a small chemical difference (a Bromine substituent on the central phenyl ring) between compounds 8 and 9, such a difference on NS3hel inhibition was unexpected, but suggested that there could be cellular pathways, not investigated in this study, that modify the inhibitors, affecting the inhibition of DENV infection.

CONCLUDING REMARKS

The shape screening of the SPECS database allowed the selection of 11 chemically different commercially available compounds that were purchased and tested in vitro to evaluate the cell-based interference with DENV infection as well as the inhibition of the RNA binding to the NS3hel protein. Two compounds (2 and 3) showed inhibition of RNA-NS3hel complex formation and another 5 analogues of these two molecules were purchased and tested. A further evaluation was performed for these molecules, as also their enzymatic inhibition activity was explored. Furthermore, the original two compounds were retested in the biochemical assays. Interestingly, only compounds 3 and its analogues retained activity in either cell- or biochemical-based assays or in both, suggesting that a phenyl ring is a better central aromatic group than the indole moiety. However, it is more difficult to define structure-activity relationships within this family of three compounds as it is hypothesised that there might be other cellular mechanisms that have not been explored could be involved in the activity of these compounds. Nevertheless, the fact that compound 8 inhibited both NS3hel and viral infection in Vero cells, gives an important support to one of the founding hypothesis of this thesis: DENV NS3hel is a promising drug target as its inhibition stops DENV infection.

To the best of knowledge these molecules have never been discovered as anti-DENV compounds and represent an interesting novel class of potential antiviral drugs. As predicted,
the inhibition of RNA binding was also very well correlated with the inhibition of NS3hel unwinding activity, suggesting a direct or indirect interference of the compounds to the protein-RNA interaction. This is particularly interesting as it has been hypothesised that Ouabain could interact at the level of the RNA site. As explained more in details in APPENDIX 2, shape complementarity between two molecules, as well as their surface properties, is fundamental for the strength of the interaction between them. Consequently, as the compounds in Table 3.2 were selected on the base of the shape similarity with Ouabain and as the analogues purchased are similar to the parent compounds, they should have a similar behaviour in the interaction with NS3hel. The activity data presented here confirm that identified compounds have a similar activity behaviour to the drug design starting molecule, even if they present a different chemical structure.

To the best of knowledge the designed shape-based virtual screening method used for the identification of compounds 2 and 3 followed a novel workflow of *in silico* experiments based on the assumption that one of the selected queries represented a Ouabain-active conformation and that active shape-analogue inhibitors could be identified in those compounds matching the highest number of queries. Both the basic assumption and the method were confirmed by the activity data that validated the designed molecular modelling approach as successful in identifying novel potential antiviral agents with a chemical scaffold that has never been proposed for DENV NS3hel inhibition and much simpler than the original natural product Ouabain.

In conclusion, this study has determined the starting point for future drug development projects and in particular, the more detailed investigation of the effects of these types of compounds in the cell could be useful for directing the improvement of these molecules.
3.2.2 RNA BINDING SITE COMPETITIVE INHIBITORS

DRUG DESIGN

PREVIOUS WORK

Previous to this PhD project, a drug design study on DENV NS3hel was performed.\cite{29} The aim was to identify putative panserotypic antiviral agents that could bind the enzyme in the RNA cleft through a virtual screening approach. As several crystal structures of the helicase were available, but none of them had co-crystallised inhibitors, two structures from the same crystallographic series were chosen for the purpose: NS3hel alone (PDB ID: 2JLQ\cite{19}) and NS3hel in complex with both RNA and ATP (PDB ID: 2JLV\cite{19}). Given the large size of the cleft, an equivalent sub-site of the RNA cleft was selected for both molecular systems. As mutational studies have highlighted the importance of Ile365, the chosen site was in the nucleic acid entry site, between domains 2 and 3.\cite{15,21} The virtual screening technique was applied on two databases of approximately 200,000 drug-like molecules each (SPECS database updated to May 2011 and SPECS-EX database) and multiple steps were used, as schematised in Figure 3.7. Methods and parameters can be found in reference \cite{29}.

![Figure 3.7. Virtual screening workflow.](image)

This is a schematic representation of the workflow followed for the virtual screening study. The main steps and programs used are indicated and the numbers are referred to the number of molecules at every step. The plate represents the in vitro evaluation of the compounds with CPE assay.
With the aim of reducing the input number of compounds for the docking and scoring procedures, the conformations corresponding to the large set of around 400,000 molecules was first searched with two pharmacophores built for each one of the two protein structures. The pharmacophore for 2JLV (panel A of Figure 3.8) was obtained first, through the PLIF statistical analysis of the first two RNA nucleotides interactions with the protein in the six available helicase-RNA complex structures (PDB IDs: 2JLU, 2JLV, 2JLW, 2JLX, 2JLY and 2JLZ). This pharmacophore was subsequently transposed to the 2JLQ structure where excluding volumes and features were manually moved in order to adapt the features to the new residues positions (panel B of Figure 3.8). Both pharmacophores had nine features centred on the RNA molecule for the essential RNA-protein interactions reported in Table 3.4. Of interest, almost all of the residues involved in these interactions are well conserved among DENV serotypes. Only Leu429 is replaced by Met429 in DENV serotype 2, conserving the residue’s hydrophobic properties.

![Figure 3.8. Pharmacophores.](image)

**A) Pharmacophore built for 2JLV.** 2JLV structure and the pharmacophore features are superposed here. Some residues are indicated and shown in lines, while the first two nucleotides are represented with thicker lines. The features are represented with the following colour scheme: “hydrogen bond donor” features (F1, F2) are in purple; “hydrophobic” centroid (F3) is in green; “hydrogen bond acceptors” (F4, F6) and “anion of hydrogen bond acceptor” (F5) features are in cyan; “hydrogen bond donor or acceptor” feature (F7) is in dark red; “aromatic or hydrophobic” features are in orange. **B) Pharmacophore built for 2JLQ.** 2JLQ structure and the pharmacophore features are superposed here and it is possible to observe the amino acids movements and feature repositioning, with respect to A). Same colour scheme of A) was applied.

The two pharmacophore filters reduced the structures to a total of 9,459 molecules that were docked in the selected binding site using Maestro Glide with SP mode. A rescoring was performed with FlexX and PLANTS programs and a consensus score function allowed further reduction of the number of molecules to 471 for the 2JLQ structure and 504 for 2JLV.

Obtained poses were visually inspected and clustered according to their chemical structure and to the binding mode within the targeted site. In pose evaluation, molecules that bound to conserved residues among the four DENV serotypes were preferred to the others. A total
number of 50 compounds were selected and, of these, 37 were purchased and one was synthesised.\textsuperscript{[29]}

Table 3.4. Features of 2JLV and 2JLQ pharmacophores.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of interaction</th>
<th>Groups involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Hydrogen bond donor</td>
<td>A1 2’OH – Pro363 carbonyl O</td>
</tr>
<tr>
<td>F2</td>
<td>Hydrogen bond donor</td>
<td>G2 base NH\textsubscript{2} – Asp290 COO</td>
</tr>
<tr>
<td>F3</td>
<td>Hydrophobic</td>
<td>A1 5’-C – Leu443, Leu429, Asp603</td>
</tr>
<tr>
<td>F4</td>
<td>Hydrogen bond acceptor</td>
<td>A3 phosphate O – Arg387 α-NH, Thr408 OH</td>
</tr>
<tr>
<td>F5</td>
<td>Anion or hydrogen bond acceptor</td>
<td>G2 phosphate O – Arg387 guanidine</td>
</tr>
<tr>
<td>F6</td>
<td>Hydrogen bond acceptor</td>
<td>G2 phosphate O – Ile365 α-NH</td>
</tr>
<tr>
<td>F7</td>
<td>Hydrogen bond donor or acceptor</td>
<td>Region near Asp409</td>
</tr>
<tr>
<td>F8</td>
<td>Aromatic or hydrophobic</td>
<td>A1 purine – Pro431 and Leu443</td>
</tr>
<tr>
<td>F9</td>
<td>Aromatic or hydrophobic</td>
<td></td>
</tr>
</tbody>
</table>

CELL-BASED DENV INHIBITION OF THE COMPOUNDS

During this PhD study, the available 38 molecules were tested \textit{in vitro} at the Rega Institute for Medicinal Research, K. U. Leuven (Belgium) with the same assay described in the Ouabain-like shape screening study. The assay conditions were the same, but in this case the concentrations of compounds tested were 100, 50, 25, 12.5, 6.25, 3.13, 1.56 and 0.78 µg/ml. One of the 38 tested molecules displayed some activity against DENV infection: compound 10, reported in Figure 3.9.

![Figure 3.9. Compound 10.](image)

This molecule demonstrated 70% inhibition of virus-induced cytopathic effect (CPE) at 50µg/ml and 40% inhibition at 25µg/ml (50.25µM), but at a concentration of 100µg/ml
(200.99µM) it appeared to be cytostatic (data not shown). From the titration curves in Figure 3.10 the antiviral and antimetabolic effects of the compound, were evaluated at Leuven and an EC\textsubscript{50} equal to 201µM and a CC\textsubscript{50} higher than 201µM was found.

Figure 3.10. Concentration-dependent effect of compound 10 in Vero cells.

The red curve shows the antiviral activity of the compound, while the green one shows the cells viability at the same compound concentration. EC\textsubscript{50} and CC\textsubscript{50} values are reported on the right.

The previous study docking results, reported in Figure 3.11, suggested that compound 10 interacted with the NS3hel at the level of the entry of the nucleic acid binding site, preferably when the enzyme assumes the conformation that allows RNA binding. As shown in panel A of the figure, the molecule co-localises with the RNA backbone and it interacts with the protein similarly to the natural ligand.

Figure 3.11. Docking pose and interactions of compound 10 in the helicase entry site.

A) Docking pose of compound 6. Here the compound pose is superposed to 2JLV structure. The protein ribbons are shown in white and the mentioned residues are indicated and have green carbon atoms. The RNA strand is represented as purple lines and compound 10 is coloured by element. B) Compound 10 ligand interactions. The schematic representation of compound 10 interactions with the NS3hel follows the legend reported in the bottom of panel B.

In detail, the carboxylic moiety of the compound hydrogen bonds with the Arg387 backbone amine and the side chain guanidine groups and with the Thr408 hydroxyl moiety. Additionally, a hydrophobic contact is established between the thiadiazole ring and Thr224. The molecule does not interact directly with Ile365, but is positioned in its near vicinity. A schematic
representation of the interactions of the compound with the protein is reported panel B of Figure 3.11.

As already discussed in the INTRODUCTION chapter, DENV NS3 helicase does not bind specifically to the RNA bases, but only to its backbone. The docking simulation suggested that compound 10 recreates the same type of interactions, implying that the molecule could compete with the RNA, inhibiting helicase’s activity.

With the aim of improving both the antiviral activity and of reducing the cytostaticity of the compound, a preliminary structure-activity study was performed. For the investigation of the relationship between the cephalosporanic core and activity of compounds, four commercially available analogue compounds were purchased and tested for cell-based DENV infection inhibition. The scaffold structure and the substituents of the compounds, including compound 10, are reported in Table 3.5.

Table 3.5. Compound 10 and analogue structures.

<table>
<thead>
<tr>
<th>Compound</th>
<th>R₁</th>
<th>R₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td><img src="image" alt="Structure 10" /></td>
<td><img src="image" alt="Structure R₁ 10" /></td>
</tr>
<tr>
<td>11</td>
<td><img src="image" alt="Structure 11" /></td>
<td><img src="image" alt="Structure R₁ 11" /></td>
</tr>
<tr>
<td>12</td>
<td><img src="image" alt="Structure 12" /></td>
<td><img src="image" alt="Structure R₁ 12" /></td>
</tr>
<tr>
<td>13</td>
<td><img src="image" alt="Structure 13" /></td>
<td><img src="image" alt="Structure R₁ 13" /></td>
</tr>
<tr>
<td>14</td>
<td><img src="image" alt="Structure 14" /></td>
<td><img src="image" alt="Structure R₁ 14" /></td>
</tr>
</tbody>
</table>
These tabulated molecules were tested in the same CPE assay previously described. In this case, the tested compound concentrations were 50, 16.7, 5.56, 1.85, 0.62, 0.21, 0.069 and 0.023 μg/ml. The results of the biological test for the 4 analogue compounds are reported in Figure 3.12.

As highlighted by the green circles in the figure, three of the four molecules retained inhibitory activity against DENV infections, whilst one of them, compound 13, did not. Of the three active molecules, compound 11 shows higher activity as it displays partial activity at the concentration of 16.7μg/ml (34.76μM) as well as at 50μg/ml (104.06μM). However, no complete infection inhibition cold be achieved for any compound at any of the tested concentrations.

Titration curves were computed in Leuven for two of the four compound 10 analogues. Whilst compound 14 demonstrated toxicity (CC_{50}=40μM), compound 11 showed a similar activity pattern to the original molecule (EC_{50}=208μM; CC_{50}>208μM). Even if these concentrations are high, the chart reported in Figure 3.13 shows that the antiviral activity starts at a slightly lower
concentration, suggesting that improvements on the activity and on the toxicity of this compound series can be made.

From these data, preliminary structure-activity relationships could be determined. Firstly, the fact that three compounds retained activity in cells confirmed that the cephalosporanic core is associated with an antiviral effect. Compound 13 was the only compound 10 analogue not substituted in R₂ and as it was also the only one with no antiviral activity, it can be inferred that the presence of a R₂ group in necessary for inhibitory activity. Unfortunately, the same type of assessment could not be done for R₁ as no purchasable compounds without this substituent could be found. Furthermore, of the new molecule series, the compound with higher activity (compound 11) is also the one with a tetrazole R₂ moiety instead of a thiadiazole, emphasising the importance of R₂ and suggesting that the activity is linked to the chemical nature of this moiety. The tested modifications at R₁ did not produce a similar change in DENV infection inhibition.

**NS3HEL-BASED DENV INHIBITION OF THE COMPOUNDS**

Five of the total 42 compounds (38 from the virtual screening selection plus 4 compound 10 analogues) were tested on DENV NS3hel through most of the biochemical assays applied in the Ouabain shape-like virtual screening study: TSA, ATPase, FP with GTP bodipy and FP with RNA-6FAM. These assays were performed by the author at AFMB, Marseille University (France). Since the selection of compounds was done on the basis of chemical diversity, only one molecule of the cephalosporanic compound series was chosen. Being one of the most active molecules with a cephalosporanic core, compound 11 was amongst the selected molecules. The selected compounds’ structure and solubility are reported in APPENDIX 5. In all the tests, mother solutions were prepared by solubilising the compounds in 100% DMSO, for a final concentration of 20mM. Compound 15 (see APPENDIX 5) showed solubility problems already in the mother solutions and only a concentration of 10mM could be achieved. The assays were performed according to the protocol in the EXPERIMENTAL SECTION chapter.

Three of the five compounds precipitated in the TSA and in the ATPase assay and consequently could not be assessed for NS3hel binding and for the inhibition of the ATPase activity of the target protein. Unfortunately neither of the other two compounds bound to NS3hel in the thermal shift assay. One compound decreased the helicase ATPase activity by 31.6% and by 76.3% in two screens at 2mM concentration. However, these results were not trusted for the high variability of the inhibition percentages and for the lack of reliability of the assay itself (see chapter 3.4).
Unfortunately, no compounds inhibit the binding of either GTP bodipy or RNA-6FAM. As discussed in chapter 3.4, some molecules gave problems with the FP tests as they showed to interfere with the assay by increasing the fluorescence polarisation values in a concentration-dependent manner also in absence of the protein. As the results were not clear, it was chosen to not consider the assay results for these compounds. Of the total 15 tested compounds, all the problematic molecules in the FP assays (two and four with the GTP bodipy and RNA-6FAM probes respectively) were selected with this structure-based drug design study for the inhibition of NS3hel through RNA cleft binding. No explanation of this phenomenon could be found, but a direct interaction with the probe was hypothesised. Summarising, no NS3hel interaction or inhibition could be determined for the tested molecules. Compound 11 displayed neither solubility problems nor interactions with the FP assay probes and consequently it could be assessed with the biochemical assays. Unfortunately, none of the tests was positive for this molecule, suggesting that its antiviral activity does not involve NS3hel.

**CONCLUDING REMARKS**

A previous drug design study identified 38 molecules for the inhibition of NS3hel through their interaction with the RNA binding site, competing with the natural ligand. Cell-based *in vitro* assays showed that one of the molecules, compound 10, had antiviral activity with EC$_{50}$ of 201µM. Unfortunately, high concentration cytotoxic effects could be associated to the compound as well. From the assessment in the same assay of four purchased analogues, preliminary structure-activity relationships could be gathered, in particular on the role of the R$_2$ substituent. Unfortunately, even if activity improvement was obtained with compound 11, the EC$_{50}$ values were still very high and the selectivity index very low. Five of the total 42 compounds (38 from the original virtual screening and the 4 compound 10 analogues) were also assessed with biochemical assays, but no positive results could be obtained. Compound 11 was part of this group of molecules, but no direct NS3hel binding or inhibition of the protein activity or formation of complexes with the natural substrates could be observed. Consequently, it was hypothesised that the cell-base activity observed for this class of molecules is not due to interaction with NS3hel, but to other mechanisms that could involve either viral or host pathways. The exact mechanism of action of these molecules has not been determined to date. Unfortunately, the lack of correspondence between cell-based and biochemical assays results did not validate the binding mode identified for compound 10 with the molecular modelling study that was designed previously to this PhD project. The method itself could not be validated either. This could be due to several reasons linked to either one of the steps in the
screening workflow (from the selection of the RNA sub-site to the parameters used for the docking and rescoring procedures). The evaluation of each factor could be interesting and could improve the general in silico approach for the identification of novel active molecules.

In conclusion, the previous Molecular Modelling study and the small structure-activity relationship assessment performed in this PhD study identified a class of molecules with a cephalosporanic scaffold that inhibit DENV infection in cells. Unfortunately, the activity was displayed at high concentrations, and very close to the appearance of cytotoxicity, and it could not be related with NS3hel interaction or inhibition. Despite the possibility of improvement given by lower antiviral concentrations with respect to antimetabolic ones, based on the unclear mechanism of the low activity of this class of compounds and their difficult chemistry, it was decided not to further develop these molecules in the future.
3.2.3 NTP BINDING SITE COMPETITIVE INHIBITORS

DRUG DESIGN

DENV NS3 helicase has proven to be a difficult drug design target and this might be related to the already discussed disadvantages of the two binding sites within the protein.\(^6,^{22}\) Of the two pockets, the dimensions of the helicase NTP binding site appear to be suitable for small molecule drug design and it would not require the selection of a sub-site within it, as applied in the study discussed in the previous chapter. NTP binding site competitive inhibitors would, however, be potentially more toxic as they mimic NTPs that are ubiquitous and bind to many human proteins.\(^6,^{22}\)

Nevertheless, the work reported here aimed to develop a method that could allow the discovery of novel compounds that target the NTP site and inhibit selectively the NTPase activity of DENV NS3hel, without interacting with host helicases. This was chosen as a target site as most of the residues that form it are highly conserved among the four DENV serotypes, aiding the possibility of designing panserotypic inhibitors. Furthermore, mutational studies showed that some of these residues are essential for the ATPase activity of DENV NS3hel and consequently, molecules that bind to them could be good competitive inhibitors.\(^{15,21}\) As DENV NS3hel lacks a base preference upon NTP binding, the design of nucleoside inhibitors could not be applied for the scope of the study. Alternatively, a virtual screening workflow was adapted from a procedure reported in the literature.\(^{30}\) In this approach, docking and scoring procedures were performed not only on three DENV NS3 helicases, but also to three human helicases. The latter were used as anti-targets and the compound selection was based in finding molecules that showed selective binding to the viral enzyme with respect to the human ones.

TARGETS AND ANTI-TARGETS SELECTION

In order to find compounds that would selectively bind to the viral helicase, three DENV and three human helicase structures were chosen for the virtual screening approach and used as target and anti-target structures respectively. In both cases, the primary criteria for the structure selection was the presence of the ligand (generally ATP or an analogue) and of the catalytic ion (usually Mn\(^{2+}\) or Mg\(^{2+}\)) in the NTP cleft. Afterwards, the resolution and the biological relevance of the structure were taken into account. The presence of the divalent ion in the structures was an important aspect because it is essential for substrate binding and for the hydrolysis of NTP to NDP + P\(_i\).
Two of the three DENV NS3hel structures (PDB IDs: 2JLR and 2JLV)\textsuperscript{19} were co-crystallised with an ATP non-hydrolysable analogue (ANPPNP); while in the third one (PDB ID: 2WHX)\textsuperscript{12} an ADP molecule is bound to the enzyme. The latter structure was the only full length NS3 structure complexed with the natural ligand in the NTP site available. As it would have been interesting to observe if interactions with the protease domain could be used for selectivity improvement, it was selected as one of the targets despite the fact that the ligand was the hydrolysed form of ATP.

With the aim of designing highly selective compounds that could discriminate between highly similar enzymes, human helicases as comparable as possible to DENV NS3hel were selected. For this exercise, the 2JLV FASTA sequence was analysed with DELTA-BLAST (Basic Local Alignment Search Tool) program, searching structures within the Protein Data Bank database. Forty-eight PDB entries that were similar to the query sequence and belonged to the Homo Sapiens genus were obtained. At this stage, also protein entirety and function were considered, aiming the selection of three different types of anti-target helicases for better sampling of the possible interactions of the small molecules. Differently from the viral crystal structures, all of the chosen human structures had an ATP analogue (ANPPNP) bound to the protein.

One of the human proteins used for compound selection was Upf1 (PDB: 2GJK), which is a SF1 (superfamily 1) RNA-dependent ATPase involved in eukaryotic nonsense-mediated mRNA decay.\textsuperscript{21} The second human enzyme was the eukaryotic initiation factor 4A-III (eIF4AIII or DDX48, PDB: 2J0S), a DEAD-box RNA helicase that, in the exon junction complex (EJC), binds mRNA during the splicing and displaces the ribosome from mRNA, acting as a “memory” of the location of the non-coding region.\textsuperscript{22} The third anti-target chosen was the DEXD/H-box RNA helicase DDX19B (PDB: 3FHT) that aids the detachment of protein from mRNA during mRNA nuclear export.\textsuperscript{33}

Before starting the virtual screening procedure, all six structures were superposed, constraining the alignment on the NTP binding site, and prepared according to the procedure described in the METHODS chapter. The divalent cation present in all the structures was maintained for all the docking and rescoring simulations, as it is essential for NTP binding and catalysis. This was done because it was hypothesised that compounds that could interact with the ion could also bind and inhibit better to the enzyme.
IDENTIFICATION OF SELECTIVE NS3HEL INHIBITORS

As anticipated, this study was designed with the aim of selecting compounds that could bind to DENV NS3hel NTP site, but not to the corresponding pocket in human helicases. The small molecule database for this study was SPECS, a collection of purchasable drug-like small molecules, which structures are available. At the time of the study it contained approximately 209,000 compounds. The general procedure of the designed molecular modelling method is schematised in Figure 3.14.

The conformations and stereoisomers of the molecules belonging to the SPECS database were first searched with three pharmacophores, built for the DENV NS3 helicases as described more in detail later in this chapter. Of the positive results, two conformations for each molecule were kept for docking evaluation. ATP was added to these structures because a comparison with the natural ligand scores was useful later in the workflow for compound selection. A total of 1,444 molecules (2,580 conformations) were subsequently further evaluated.

The new database of molecules was then processed with the Maestro LigPrep program that elaborated all possible tautomer and ionisation states for each conformation, considering a pH of 7 (±2). The chemical evaluation of the ligands allowed a more precise scoring in the Maestro Glide docking application because poses are scored on the base of the energy (and likelihood) related to the chemical structure of the molecule itself as well as on the interactions with the receptor. As molecules had already been filtered with the pharmacophores, basic information about the preferred chirality and E/Z conformation of molecules were known and stereoisomers were maintained as in the input 3D structures.

The prepared ligands were then docked and refined in each target structure with Maestro Glide SP mode and rescored with LeadIT FlexX and PLANTS. Three poses for each molecule were analysed at this stage. The scores from these three programs were then evaluated with a consensus scoring function (CS1, Equation 3.1) and with a normalised consensus scoring function (NCS1, Equation 3.2), as reported in the METHODS chapter. These functions allowed the ranking of the molecules according to a unique score and to keep those that performed within the best 25% results in all the programs used. Only 348, 351 and 509 molecules for 2JLV, 2JLR and 2WHX structures respectively were consequently kept at this stage. These molecules were then compared and 2 conformations of the molecules that were kept in the docking and scoring procedure of at least two target structures were saved for further evaluation. The obtained database (412 molecules, 824 conformations) represented the compounds belonging to the SPECS database that displayed the predicted best binding at this stage to the explored viral target and it was used for the final selection based on the specificity of the molecules. This was performed by docking and scoring the conformations in both the target and the anti-target structures with Maestro Glide XP mode and rescoring.
them with LeadIT FlexX and PLANTS. Three poses were generated for each conformation entry.

The best pose for each molecule was saved and all the scores were then elaborated with a novel consensus scoring function (CS2, Equation 3.3 reported in the METHODS chapter), with the aim of ranking the molecules according to the preferentiality for viral helicases over human ones. With this function, a unique score that collects Glide XP, FlexX and PLANTS evaluations in one normalised value was computed for each target and anti-target structure used, obtaining a number that was directly proportional to the binding capacity of the molecule. The final score was then calculated by subtracting the anti-target-related scores to the target-related ones. Consensus scores varied from a minimum of -1.5011 to a maximum of 1.8671. From a qualitative point of view, positive CS2 values suggest that the molecule binds better to targets than to anti-targets, while negative ones indicate the opposite scenario. A score around 0 should mean that the molecule is not selective between the two protein sets.

For the purpose of a virtual screening approach, it was important to select a reference for which a compound could be considered selective or not. This was achieved comparing the CS2 values of the molecules with the natural ligand: ATP. This molecule is not selective between viral and human enzymes and consequently it was hypothesised that molecules scoring better than ATP bound preferably to the viral helicases. As expected, CS2 value for ATP was around 0 (0.0129), confirming that the designed methodology could be consistent with experimental data. The molecules that scored better than ATP were 293. In order to select highly selective molecules, the 84 molecules that scored 0.6 or above were kept for the last stage of the procedure.

Chemical structure clustering and visual inspection of the poses were applied for the final selection of 10 compounds. As the discovery of panserotypic agents would be preferred, molecules that bound to conserved residues among the four DENV serotypes were preferred to the others. As intended also in the target and anti-target structure preparation and in the pharmacophore query design, all the selected molecules (reported in APPENDIX 4) had at least one carboxylic group for metal chelation. Unfortunately, no molecules displayed interactions with the DENV NS3 protease domain in structure 2WHX.
Figure 3.14. Virtual screening workflow.
This is a schematic representation of the general procedure followed for the virtual screening study. The main steps are indicated and the numbers refer to the number of molecules (conformations in brackets) at every step. Targets (in green) and anti-targets (in red) are reported with the corresponding PDB ID. Molecular modelling programs and functions are abbreviated as follows: SP = Maestro Glide SP mode; FX = LeadIT FlexX; PL = PLANTS; CS = consensus score; NCS = normalised consensus score; XP+R = Maestro Glide XP mode docking and rescoring with LeadIT FlexX and PLANTS.
**Pharmacophores**

The DENV helicase pharmacophores were built with the PLIF tool that analysed the interactions between the protein and the co-crystallised ligand in the NTP binding site of the three target structures. All the pharmacophores were then manually modified in order to maintain the same seven features across the three structures. The pharmacophore for 2WHX was the most extensively modified because features regarding the missing γ-phosphate had to be transferred from the other crystal structures. Two features were also added in all the queries with the aim of involving NS3hel residues at the NTP base level, but in a region that is distant from that of human helicase responsible for adenosine-specificity in these proteins. Glu233 and Arg418 were suitable for this scope, as shown in Figure 3.15, and the projection of the features were positioned on each residue side chains’ functional group (features were hydrogen bond donor and acceptor respectively). This figure shows an example of DENV-human helicases superposition (PDBs: 2JLV and 3FHT) where base interactions of human proteins and the targeted residues in the pharmacophores are highlighted.

![Figure 3.15. Target and Anti-target superposition.](image)

Here, 2JLV (residues carbon atoms in yellow and AMPPNP ones in orange) is superposed to 3FHT (residues carbon atoms in cyan and AMPPNP ones in dark green). Intermolecular interactions are shown with green dashed lines. The two AMPPNP molecules are well overlapped in the phosphate region and both coordinate a divalent cation (cyan sphere). In the human helicase two aromatic residues (in this case Phe112 and Phe433) stack with the ATP base and adenine N7 and NH$_2$ groups are hydrogen bonded with protein residues (in this case Arg114 and Gln119), achieving base specificity.

The three pharmacophores are shown in Figure 3.16; while Table 3.6 describes their features, the type of interaction involved and in which structural pharmacophore they are present. For pharmacophore design, all the selected residues or at least their properties (for Arg418 and Lys201) were conserved among the four DENV serotypes, aiding the design panserotypic inhibitors. Furthermore, four of the amino acids were also shown to be essential for NS3hel ATPase activity as their mutation to Ala impaired this function. These residues were the
“arginine fingers” (Arg460 and Arg463) and the Gly198 and Lys199, belonging to the Walker A motif.\textsuperscript{[15,21]}

In the pharmacophore search, compounds that matched at least three essential features were kept. Two searches were computed for each structure: one with F3, F4 and F6 essential features and one with F3, F4 and F7 features. The results were then combined together with the aim of obtaining molecules that could mimic the \(\alpha\) and the \(\beta\) phosphates of NTP and interact either with Glu233 or Arg418.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{pharmacophore_queries.png}
\caption{Pharmacophore queries.}
\end{figure}

A) Pharmacophore for 2JLR structure. The 2JLR structure and its pharmacophore are superposed. Important residues are indicated, the divalent cation is shown as a green sphere and the ligand is shown with thicker lines. The features are represented with the following colour scheme: “hydrogen bond donor and acceptor” feature (F1) is in dark red; “hydrogen bond acceptor” features (F2, F5, F8) are in cyan; “anion and hydrogen bond acceptor and metal ligator” features (F3, F4) are in blue; “hydrogen bond donor projection” (F6) and “hydrogen bond acceptor projection” (F7) features are in yellow. B) Pharmacophore for 2JLV structure. The 2JLV structure and its pharmacophore are superposed. The same colour scheme as A) was applied. C) Pharmacophore for 2WHX structure. The 2WHX structure and its pharmacophore are superposed. The same colour scheme as A) was applied.
Table 3.6. Pharmacophores features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of interaction</th>
<th>Groups involved</th>
<th>Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Hydrogen bond donor and acceptor</td>
<td>NTP 3’ and 2’ OH and Arg418, Asn329, Arg463</td>
<td>All</td>
</tr>
<tr>
<td>F2</td>
<td>Hydrogen bond acceptor</td>
<td>N7 of Adenine and Lys201</td>
<td>All</td>
</tr>
<tr>
<td>F3</td>
<td>Anion and Hydrogen bond acceptor and metal ligator</td>
<td>NTP γ-phosphate and Mn²⁺, Arg460, Lys199</td>
<td>All</td>
</tr>
<tr>
<td>F4</td>
<td>Anion and Hydrogen bond acceptor and metal ligator</td>
<td>NTP β-phosphate and Mn²⁺, Thr200, Arg463, Gly198 and Lys199</td>
<td>All</td>
</tr>
<tr>
<td>F5</td>
<td>Hydrogen bond acceptor</td>
<td>NTP α-phosphate O and Lys201 α-NH</td>
<td>All</td>
</tr>
<tr>
<td>F6</td>
<td>Hydrogen bond donor projection</td>
<td>Glu233</td>
<td>All</td>
</tr>
<tr>
<td>F7</td>
<td>Hydrogen bond acceptor projection</td>
<td>Arg418</td>
<td>All</td>
</tr>
<tr>
<td>F8</td>
<td>Hydrogen bond acceptor</td>
<td>Other NTP α-phosphate O</td>
<td>Only 2WHX</td>
</tr>
</tbody>
</table>

**ACTIVITY OF THE SELECTED COMPOUNDS**

Five of the 10 molecules identified *in silico* were purchased molecules and tested *in vitro* at the Rega Institute for Medicinal Research, K. U. Leuven (Belgium) by our collaborators, with the same methodology reported for the Ouabain shape virtual screening study discussed in chapter 3.2.1. Unfortunately, none of the tested compounds were found to inhibit DENV cell infection at any of the tested concentrations (with CC₅₀ and EC₅₀ above 50 µg/ml).

Nevertheless, all 5 of them were tested on DENV NS3 helicase using biochemical assays, aiming to assess the binding ability of the compounds to NS3hel (TSA), to inhibit the formation of the protein-ligand complex (FP assay, using either a GTP or a RNA probe) and the inhibitory effect of the compounds on the ATPase activity of NS3hel. The choice of the molecules was mainly based on the chemical similarity of the structures, intending to explore as much chemical space as possible. In all the tests, the mother solutions were prepared by solubilising the compounds in 100% DMSO, for a final concentration of 20mM. No solubility problems occurred for these compounds neither in mother solution preparation nor in the biochemical assays. Details about biochemical assays protocols are reported in the *EXPERIMENTAL SECTION* chapter.
Unfortunately, no compounds inhibited helicase-ligand complex formation in either of the two FP assays. However, one of them (compound 16, in \textit{APPENDIX 5}) showed a negative shift in the melting temperature (Tm) with respect to the protein alone, with a Tm decrease at 1mM concentration of 3.31°C and 1.94°C for DENV3 and DENV4 NS3hel respectively. In the same assays AMPPNP, a known helicase binder, presented a Tm increase of 4.27°C and 3.75°C for DENV3 and DENV4 NS3hel respectively. These data suggest that compound 16 might interact with NS3hel, but destabilises it. As the same molecules did not exhibit positive or reliable results in the other biochemical assays, no titration curves with the thermal shift assay have been performed and the compound was not further developed.

Another compound (compound 17, in \textit{APPENDIX 5}) decreased the helicase ATPase activity by 62.7% in one of two screens at 2mM concentration, but did not affect the enzymatic activity in the other one. However, these results were not trusted because of high variability of the inhibition percentages and for the lack of reliability of the assay itself (see chapter 3.4). As no other promising data were obtained for compound 17, it was not taken further.

\textbf{CONCLUDING REMARKS}

A molecular modelling method was designed for the discovery of DENV NS3hel inhibitors that could target the NTP binding site. The main objective was to identify compounds that could selectively bind to the viral helicase and consequently reduce the toxicity risk related to the use of this pocket as a target. Particular attention was also drawn to the conservation of the targeted residues because the design of panserotypic inhibitors would be preferable. This virtual screening approach allowed the gradual selection of small drug-like molecules belonging to the SPECS database from around 209,000 to 10 final compounds. Using multiple docking and scoring programs and steps, exploiting structures for three DENV helicases and three human helicases, a specifically built consensus score was applied for compound selection. In comparison with ATP, identified molecules in comparison to ATP, exhibited a higher discrimination towards the viral protein with respect to the human one. To the best of knowledge, no similar approaches have been applied previously that target the DENV NS3hel NTP binding site for the discovery of novel, potential virus inhibitors.

Unfortunately, the selected compounds did not show significant activity in the \textit{in vitro} assays performed. As other compounds obtained from other approaches within this PhD study were more promising, none of the compounds identified and discussed above were further developed.

In conclusion, the experimental data suggest that this molecular modelling approach was not successful in identifying anti-DENV agents. However, the methodology seems to mimic reality in the lack of ATP preferentiality for viral helicase proteins. For this reason, it would be
worthwhile to apply again the methodology using different experimental conditions. For example, another virtual database of small molecules could be assessed at the NS3hel NTP binding site to compare to SPECS. In addition, it would also be interesting to change docking programs and/or their settings in order to explore the general virtual screening workflow.
3.3 CONCLUSION

Given its important role in the viral replication cycle, it was hypothesised that the inhibition of DENV NS3hel could produce an antiviral response. Thus, with the aim of designing novel anti-DENV agents, three drug design approaches were performed. All three of them were based on the virtual screening of drug-like small molecule databases and when active hit molecules were found, preliminary structure-activity relationships were searched with the assessment of analogue molecules. The designed molecular modelling methodologies were novel and never applied on DENV NS3hel before. One of them was a ligand shape-based virtual screening that used Ouabain as a reference molecule for the search of novel inhibitors, whilst the other two were structure-based and aimed to discover compounds that could bind the RNA and the NTP binding site of the protein, giving particular attention to target residues that were conserved in identity or in properties amongst DENV serotypes for the design of potential panserotypic inhibitors. Unfortunately, structure-based molecular modelling approaches showed that DENV NS3hel is a challenging drug target and that even if it is one of the better characterised DENV proteins, further knowledge on its mechanism of dsRNA unwinding is needed for drug design improvement. Conversely, the ligand-based approach was particularly successful in identifying a novel class of NS3hel inhibitors that could be further developed in the future and that supported the basic hypothesis of the in silico approach. Most importantly, a 4-benzylidene-1-phenylpyrazolidine-3,5-dione compound with antiviral activity applied to infected cells inhibited the unwinding activity of enzyme through the disruption of the protein-RNA complex. Thus, DENV NS3hel was validated as a promising anti-DENV target that can be used for either ligand or structure-based drug design approaches.
3.4 BIOCHEMICAL ASSAYS ON DENV NS3 HELICASE

As part of a secondment period in Marseille University, four different biochemical tests involving NS3hel were performed and are reported in Table 3.7 with the aim of assessing in vitro the activity of compounds designed with the three drug design approaches described in chapters within this section. Fifteen molecules (reported in APPENDIX 5) were tested, five from each virtual screening approach. The choice of the compounds among the small molecules found was mainly based on chemical diversity. Mother solutions were prepared for all tested compounds by dissolving them in 100% dimethylsulfoxide (DMSO). Two different helicase domains of the NS3 protein were used for the biochemical assays: DENV 3 and DENV 4 NS3 helicase domains. Some tests were performed on both serotypes, while others were performed only on one of them, according to the enzyme and time availability. A brief description of these assays can be found in the INTRODUCTION section, while general procedures, results and test calibrations are reported here. Further details of compound evaluation are discussed in the RESULTS AND DISCUSSION chapter of this section, in the corresponding paragraph for each drug design study. Other in vitro tests (cytopathic effect, CPE, inhibition and NS3hel unwinding assay) were performed by our collaborators and consequently are not discussed in this chapter.

Table 3.7. Summary of assays performed for DENV helicase.

<table>
<thead>
<tr>
<th>Name</th>
<th>What it measures</th>
<th>Enzyme(s) used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal Shift Assay (TSA)</td>
<td>Protein melting temperature (Tm) shift due to compound binding.</td>
<td>DENV3 NS3hel, DENV4 NS3hel</td>
</tr>
<tr>
<td>ATPase assay</td>
<td>Compound inhibition of ATP hydrolysis, through the measurement of the phosphate ions produced and detected with a colorimetric method.</td>
<td>DENV4 NS3hel</td>
</tr>
<tr>
<td>Fluorescence Polarisation (FP) using a RNA probe (RNA_BS9.2012: RNA-6FAM)</td>
<td>Fluorescence polarisation, different for probe bound or unbound with the protein. It was used to identify compound inhibition of RNA binding.</td>
<td>DENV3 NS3hel, DENV4 NS3hel</td>
</tr>
<tr>
<td>FP using GTP bodipy</td>
<td>Fluorescence polarisation, different for probe bound or unbound with the protein. It was used to identify compound inhibition of NTP binding.</td>
<td>DENV3 NS3hel, DENV4 NS3hel</td>
</tr>
</tbody>
</table>

**Abbreviations:** TSA= Thermal Shift Assay; Tm= Melting Temperature; DENV3= Dengue Serotype 3; NS3hel= Non-Structural Protein 3 Helicase; DENV4= Dengue Serotype 4; ATPase= Adenosine Triphosphatase; ATP= Adenosine Triphosphate; FP= Fluorescence Polarisation; RNA= Ribonucleic Acid; GTP= Guanosine Triphosphate; NTP= Nucleoside Triphosphate
THERMAL SHIFT ASSAY (TSA)

The thermal shift assay (TSA) was performed on both DENV helicase serotypes with the aim of determining the binding capacity of the compounds. As a reference, the known ATPase inhibitor Adenosine 5’-(β,γ-imido) triphosphate (AMPPNP), that binds to the enzyme, was used. Initially the assay was performed using Mg\(^{2+}\) in the buffer because the presence of the catalytic divalent ion is important also for NTP analogues (like AMPPNP) binding. Unexpectedly, in these conditions not even the reference molecule was showing positive results. Therefore Mg\(^{2+}\) was substituted with Mn\(^{2+}\) and in the new conditions it was possible to observe binding between the helicase and AMPPNP through a significant shift in the protein melting point (Tm).

All the compounds were screened at a concentration of 1mM (or 0.5mM for less soluble compounds), while AMPPNP was screened at 0.5mM. The final content of DMSO was equal to 5%, and the Tm were confronted with the Tm of the protein in buffer and in buffer with 5% DMSO. Eight of the total fifteen compounds precipitated during the assay and consequently their binding to NS3hel could not be assessed. Unfortunately, it has not been possible to test the precipitated compounds at lower concentration. The shift in the melting point (ΔTm) was calculated subtracting the melting temperature of the protein in buffer with 5% DMSO to the one of the assessed molecular system. Only the reference molecule (AMPPNP) gave a significant increase in melting temperature. In all the other cases a reduction of Tm values could be observed and this was meaningful only in the case of one compound designed to bind the NTP binding site (see chapter 3.2.3).

ATPASE ASSAY

The ATPase assay was executed in order to assess the capability of all the fifteen compounds to inhibit the hydrolysis of ATP performed by DENV helicase. Only DENV4 NS3hel was used for this assay. The ATPase enzymatic activity of the protein is assessed through the evaluation of the concentration of phosphate ion after a time period. The reaction time, ATP and protein concentrations were chosen following assay calibration. The compounds were screened at 2mM and the results were confronted with a positive (Protein alone in buffer and 10% DMSO) and a negative (Protein with AMPPNP, known ATPase inhibitor) control. All compounds and controls were screened in duplicates. Blanks with ATP and the compound (but without protein) were analysed as well for background noise detection. A titration curve of known concentrations of potassium phosphate was used for quantitative determination. As
happened in the previous assay, the same eight compounds precipitated and therefore could not be assessed with this test.

The results were dubious for AMPPNP as it was not possible to reproduce the same inhibitory effect in the duplication of the assay, as shown in Figure 3.17. Unfortunately, the causes of this could not be established. Of the tested compounds, three (two selected by targeting the NTP binding site and one by targeting the RNA cleft) showed some positive results. However, since these molecules also absorb at the wavelength used for the phosphate detection, there was not enough confidence on the reliability of this assay. Consequently, it was decided to assess NS3hel-NTP binding inhibition through fluorescence polarisation, using GTP bodipy as probe and the original assay was not further considered.

![Figure 3.17. ATPase assay results for AMPPNP.](image)

The phosphate concentrations values are reported for the Positive control and the system with AMPPNP in duplicate (AMPPNP-1 and AMPPNP-2). The phosphate concentration of the positive control is also shown with the red line.

**FLUORESCENCE POLARISATION (FP) ASSAY**

The fluorescence polarisation (FP) assay was used for the analysis of compounds binding to the two helicase binding sites, exploiting two fluorescent probes. The binding was assessed on both DENV helicase serotypes. For the NTP binding site the used probe was GTP bodipy, while for the RNA binding site a RNA oligonucleotide linked to a 6-carboxyfluorescein (6-FAM) molecule was chosen. As DENV helicase is not base specific in either binding sites, the probe selection was based on the best assay performance that could be obtained.
FLUORESCENCE POLARISATION (FP) ASSAY WITH GTP BODIPY PROBE

The assay calibration aimed to find the ideal concentrations of the probe and of the protein to use in order to maximise the test sensibility. It was also assessed the importance of the use of a ion chelating agent (EDTA) that, removing the catalytic divalent ion, impeded the hydrolysis of the probe from GTP bodipy to GDP bodipy that would decrease the probe binding and consequently alter the fluorescence polarisation (FP) values. The dissociation constant ($K_d$) of the helicase – probe complex was determined during the calibration by testing different protein concentrations with the minimum probe concentration possible (100nM). As the $K_d$ was found to be around 1µM for both helicase serotypes, a concentration higher than the $K_d$, but not in the plateaux range of the chart was chosen for compound screening: 3 µM. Figure 3.18 shows an example of calibration curve (DENV3 NS3hel). Here the dissociation constant and the chosen protein concentrations are indicated.

![Figure 3.18. FP-GTP calibration curve for DENV3 NS3hel.](image)

The variation of the fluorescence polarisation (FP) is plotted against the protein concentration. The dissociation constant of 1.10µM is indicated with $K_d$ and the chosen working protein concentration is shown as well.

All compounds were tested at eight concentrations with 2X dilution, starting from 1mM or from the maximum possible concentration of the compound (see APPENDIX 5) in order to avoid precipitation. Dilutions were done in DMSO in order to have the same DMSO content (5%) in all the wells. Similarly to the ATPase inhibition assay, AMPPNP was used as a negative control. Blanks with all compounds concentrations and the GTP probe in buffer were also prepared in order to exclude background noise. The six compounds that gave a reduction of the fluorescence polarisation (which means a reduction in GTP bodipy binding) with concentration increase and the ones that were found positive with the ATPase assay were selected and the test was repeated other two times with the aim of improving the statistical relevance of the results. Three compounds in particular (one from each virtual screening study) gave interesting results. The chemical structure of these compounds (3, 15 and 17) can
be found in *APPENDIX 5*. IC$_{50}$ values for these molecule were calculated and ranged approximately between 8 and 130 µM. Interestingly, of these three, the compound designed to bind the NTP site (17) was also found positive in the ATPase assay.

A further check had to be performed for two compounds designed to target the RNA site as the blank showed an increase in polarisation correlated with compound concentration. GTP bodipy and the compounds (at the same concentrations) were tested in combination and alone. Also in this case the results were not clear and it was not possible to determine if the FP increase was correlated only with compound fluorescence and interaction with the protein, or if it involved also interactions with the probe. Unfortunately, one of these two molecules was also amongst the three compounds with positive results. Furthermore, despite GTP bodipy binding inhibition could be observed for some compounds, in any tested compound no consistency could be reproduced in the triplicates curves, as exemplified in Figure 3.19. This decreased the confidence in the obtained results, in particular in the statistical relevance of the calculated IC$_{50}$ values. For this reason, they are not reported here. As the whole results were so unclear and no reason that could explain this behaviour could be found, it was decided to not consider them for future compound developments.

![Figure 3.19. Example of GTP-FP assay triplicate curves.](image)

This example reports the three curves of FP assay on DENV4 NS3hel probe for a compound that showed to inhibit GTP bodipy binding. The polarisation is reported for the logarithm of the compound concentration (in µM). The three repeats (indicated as Test 1, blue curve and symbols; Test 2, green curve and symbols; Test 3, red curve and symbols) are different. The curve in black (indicated as Blank) refers to the system in absence of the protein.

**FLUORESCENCE POLARISATION (FP) ASSAY WITH RNA-6FAM PROBE**

Similarly to the other FP assay, also for this test an assay calibration was performed, following the same protocol. In this case EDTA was not used because no catalytic ions are involved in the site of interest. The probe (RNA-6FAM) concentration used was of 50nM and several concentrations of the two helicases domains were assessed. The dissociation constants of the complex were found to be very similar: 17.64nM for DENV3 NS3hel and 17.44nM for DENV4 NS3hel. Also in this case, a protein concentration higher than this value, but not in the
plateaux area of the chart was chosen for compound screening: 40nM. Figure 3.20 shows an example of calibration curve (DENV4 NS3hel). Here the dissociation constant and the chosen protein concentrations are indicated.

All compounds were tested at eight concentrations with 2X dilution, starting from 1mM or the maximum possible concentration of compound (see APPENDIX 5) in order to avoid precipitation. Small molecule dilutions were done in DMSO in order to have the same DMSO content (5%) in all the wells. AMPPNP was tested as well as a negative control. Blanks with all compounds concentrations and the RNA probe in buffer were also prepared in order to exclude background noise.

![Figure 3.20. FP-RNA calibration curve for DENV4 NS3hel.](image)

The variation of the fluorescence polarisation (FP) is plotted against the protein concentration. The dissociation constant of 17.44nM is indicated with $K_d$ and the chosen working protein concentration is shown as well.

The first screen of the compounds showed that almost all of them gave a reduction in FP values, which corresponds to the inhibition of the RNA-helicase complex formation. As predictable, AMPPNP (the negative control) did not give a positive result, as its binding site does not correspond to the RNA pocket. Excluding the two molecules that did not show FP reduction, compounds were retested twice in order to increase statistical relevance of the results. Differently to what experienced for the NTP binding site, the obtained triplicate curves were much more consistent, as exemplified in Figure 3.21, and results were assessed to be more reliable. Only one set of experiments (the first data obtained for DENV3 NS3hel) had to be removed, as the middle part of the curve was not well described and led to large bias in the calculation of the IC$_{50}$.

Similar to the other FP assay, a further check had to be performed for 4 of the total 15 compounds, as the blank showed an increase in polarisation correlated with compound concentration. Interestingly, all of them had been selected with the virtual screening targeting the RNA binding site and this will be further discussed in chapter 3.2.2. RNA-6FAM and the compounds (at the same concentrations reported above) were tested in combination and
alone. Also in this case the results were not clear and it was chosen to not consider the assay results for these compounds.

$IC_{50}$s were calculated for compounds that did not present problems with the assay and values ranged approximately between 8 and 41 µM for the four most attractive compounds (see study chapters for further details). Other values were generally above 200 µM. Intriguingly, the higher binding activity was associated with compounds selected through a ligand-based approach that employed a shape screening using Ouabain as a template molecule and this aspect will be further discussed in chapter 3.2.1.

**Figure 3.21. Example of GTP-RNA assay triplicate curves.**

This example reports the three curves of FP assay on DENV4 NS3hel probe for a compound that has been shown to inhibit RNA-6FAM binding. The polarisation is reported for the logarithm of the compound concentration (in µM). The three repeats (indicated as Test 1, blue curve and symbols; Test 2, green curve and symbols; Test 3, red curve and symbols) are more consistent that in the FP-GTP assay. The curve in black (indicated as Blank) refers to the system in absence of the protein.
3.5 COMPUTATIONAL METHODS

The computational methods for all the studies discussed in chapters from 3.2.1 to 3.2.3 are presented here. A description of the molecular modelling programs used can be found in APPENDIX 2.

HARDWARE DETAILS

All of the calculations were performed on an 8 core computer with Inter Xeon 2.80 GHz E5462 CPUs.

OUABAIN SHAPE-BASED DRUG DESIGN

QUERY SELECTION AND PREPARATION

A conformational search was performed with MOE 2010.10 software for Ouabain, Ivermectin and Paromomycin. The conformation search was conducted with the Import Conformation tool exploring 250 conformations per molecule, with the following parameters: stochastic strain limit of 7, superposed RMSD test of 0.15, refinement conformation limit of 300, stochastic search failure limit of 30, stochastic search limit of 300, energy minimisation iteration limit of 200, energy minimisation gradient test of 0.5. Subsequently, hydrogen atoms were added and the conformations were energy minimized using MMFF94x force field, with a 0.001 gradient, adjusting hydrogens and lone pairs and calculating the force field partial charges.

As Ouabain was selected as a query molecule, the number of its conformations was consequently reduced to obtain the smallest number of queries. The conformations were superposed and the molecular surface was calculated for each one of them. Through visual inspection of both the surfaces and atoms superposition, the conformations were clustered according to shape similarity and the three most different conformations were chosen as shape screening queries.
3.5 COMPUTATIONAL METHODS

DATABASE PREPARATION

The approximately 209,950 screened molecules were downloaded from SPECS database (updated to February 2012).\textsuperscript{[27]} Prior to the shape screening, a conformational search on these structures was performed, using MOE 2010.10\textsuperscript{[34]} Import Conformation tool. For each molecule 100 conformations were explored, with the following parameters: stochastic strain limit fixed to 7, superposed RMSD test to 0.15, refinement conformation limit to 300, stochastic search failure limit to 30, stochastic search limit to 300, energy minimisation iteration limit to 200, energy minimisation gradient test to 0.5. The molecules were also filtered, omitting compounds with: transition metals, more than 8 rings, d-hybridized atoms, molecular weight over 600 Da, more than 12 donor/acceptor groups, more than 4 chiral centres, over 3 unconstrained chiral centres, more than 7 rotatable bonds, single bond chain length over 6 and logP under -4 or over 8. Subsequently, hydrogens were added and the conformations were energy minimized with MMFF94x force field, with a 0.001 gradient, adjusting hydrogens and lone pairs and calculating the force field partial charges. The final conformations were approximately 7,130,000.

SHAPE SCREENING

The five shape screenings of the database molecules were computed with the OpenEye ROCS 3.1.2\textsuperscript{[35]} software, using all three Ouabain queries and considering all the database conformations as separate molecules. Only shape comparison between the molecules and the queries was considered for conformation scoring and the ranking followed the Tanimoto index values. Three maximum hits were kept for each conformation and only the best 1,000; 2,000; 3,000; 4,000 and 5,000 hits, according to the screening protocol, were transcribed in the output files.

RESULTS ANALYSIS

The results for each screening were observed and compared. The most promising molecules were then observed in the MOE 2010.10\textsuperscript{[34]} environment. The molecular surface of these molecules was calculated with the MOE Surfaces and Maps tool and all the selected conformations were compared to the queries on the basis of the volume and atom superposition. In this way, the best molecules were chosen and purchased from SPECS.\textsuperscript{[27]} After \textit{in vitro} evaluation of the compounds, the analogues of compound 2 and 3 were purchased from SPECS\textsuperscript{[27]} as well.
NTP BINDING SITE COMPETITIVE INHIBITORS DRUG DESIGN

PROTEINS PREPARATION

All target (PDB IDs: 2JLV, 2JLR\textsuperscript{[19]} and 2WHX\textsuperscript{[12]}) and anti-target (PDB IDs: 2GJK\textsuperscript{[31]}, 2J0S\textsuperscript{[32]} and 3FHT\textsuperscript{[33]}) structures were downloaded from the Protein Data Bank\textsuperscript{[13]} and completed with MOE 2010.10\textsuperscript{[34]} software. The missing side chains of residues were added and energy minimised with AMBER99 force field and 0.001 gradient. This procedure was applied to the following amino acids: Arg184 and Arg480 in 2JLR; Lys33, Glu43, His60, Glu61, Arg62, Asn74, Glu92, Glu93, Lys107, Ile123, Lys143, Asp175, Lys213 in 2WHX; Glu21, Asp22, Arg105, Arg151, Lys195, Lys314, Glu314, Leu410 and Ile411 in 2J0S. In all of them, all water and co-crystallised molecules were removed, except for AMP\textsubscript{PNP} (or ADP) and the divalent cation. All structures were then protonated with MOE 2010.10\textsuperscript{[34]} Protonate 3D tool, considering a temperature of 300K and a pH of 7. All structures were aligned with MOE 2010.10\textsuperscript{[34]}, constraining alignments to have the best superposition possible of the NTP binding sites.

Additional preparation was performed for the docking simulations performed with Schrödinger Maestro 9.3\textsuperscript{[36]} Glide program, using the Protein Preparation tool.

LIGAND PREPARATION

The starting ligand database was originated from the SPECS database (updated to February 2012).\textsuperscript{[27]} Prior to pharmacophore filtering, the ligand preparation included a conformational search using MOE 2010.10\textsuperscript{[34]} Import Conformations tool with the following settings: 100 conformations explored per molecule, stochastic strain limit fixed to 7, superposed RMSD test to 0.15, refinement conformation limit to 300, stochastic search failure limit to 30, stochastic search limit to 300, energy minimisation iteration limit to 200, energy minimisation gradient test to 0.5. Furthermore, compounds with transition metals, more than 8 rings, d-hybridized atoms, molecular weight over 600Da, more than 12 donor/acceptor groups, more than 4 chiral centres, logP under -4 and over 8, over 3 unconstrained chiral centres, more than 7 rotatable bonds and single bond chain length over 6, were omitted. Ligands were subsequently also protonated.

The ligand databases used for docking and scoring simulations were prepared with Maestro 9.3\textsuperscript{[36]} LigPrep tool, using OPLS\_2005 force field. Tautomers for each molecule were designed considering all possible ionisation states at pH 7 ± 2 with Epik and maintaining the input...
chiralities. A filter was also applied to remove molecules with molecular weights higher than 600Da, more than 9 rings, more than 4 chiral centres and more than 12 hydrogen bond donor or acceptor groups.

**PHARMACOPHORE**

For each target protein (PDBs 2JLV, 2JLR and 2WHX) a pharmacophore was built with MOE 2010.10\cite{34} Protein Ligand Interaction Fingerprints (PLIF) tool. The annotation points were assigned with the Unified scheme, extra volumes were added for binding site size definition and the features were modified manually through visual inspection. 2JLR and 2JLV pharmacophores had 7 features and 2WHX had 8.

The prepared ligand conformations were searched with these pharmacophores and the best 2 conformations for each molecule that matched at least three essential features of the total were saved. Two pharmacophore searches were performed: one with F3, F4 and F6 essential features and the other one with F3, F4 and F7 ones (see chapter 3.2.3). The results of the two searches were merged and a maximum two conformations for each unique molecule were saved in a database.

**DOCKING**

An ATP molecule was added as a reference to the pharmacophore-filtered ligands prior Maestro 9.3\cite{36} LigPrep small molecule preparation (see above).

All docking simulations were performed with Schrödinger Maestro 9.3\cite{36} Glide program. Prior docking, the receptor grids were built selecting the co-crystallised AMPPNP (or ADP) as the centre and a 9Å grid length. In all cases, the catalytic ion present in the structures (Mn\(^{2+}\) or Mg\(^{2+}\)) was retained.

All the docking simulations were performed in a semi-flexible approach with the sample ring conformation option for an exhaustive conformational search for the ligands. Ligands with more than 300 atoms or more than 50 rotatable bonds were excluded. The non-bonded interactions were calculated with a 0.8nm van der Waals radii scaling factor and a 0.15nm partial charge cut-off. Only a maximum of 3 conformations were kept for each ligand and they were not energy minimised after the docking because a refinement procedure was applied on the docking output. The first docking calculations were applied only to target structures and were performed with standard precision (SP); while the ones implemented on both target and-antitarget receptors were executed with extra precision (XP). The resulting poses were, then, visually inspected with the MOE 2010.10\cite{34} software.
RESCORING

The docking results were rescored with other two programs: LeadIT FlexX 2.1.3\textsuperscript{[37]} and PLANTS.\textsuperscript{[38]} No molecule placement was performed in this case and poses were scored in their input conformation and position within the binding pocket. The prepared target proteins were used with both softwares.

The receptor binding site was prepared with LeadIT software, and the centres of the receptor sites were chosen in the same way as described for the Maestro Glide docking. In all cases the grid sphere had a radius equal to 20Å.

For PLANTS rescoring the AMPPNP or ADP molecules had to be removed prior to simulation.

The same grid centres as in the Glide docking procedure were used and in all systems the binding site radius was set to 9Å. The search algorithm used 20 ants and CHEMPLP scoring function was used.\textsuperscript{[39]}

CONSENSUS SCORE 1 (CS1)

All scores obtained with the first Glide docking and FlexX and PLANTS rescoring procedure were processed with a scoring function (consensus scoring function 1) that was previously developed in our laboratory. For each set of data, the value corresponding to the lowest 25% of the scores (first or lower quartile, $Q_1$) was calculated. Then, molecules that had a score lower than $Q_1$ (representing the best 25% of poses) were rewarded with a “+1” score, while the ones with score equal or higher than $Q_1$ were assessed with a “0” or “-1” score respectively. Then, the total consensus score (CS1) was obtained from the sum of the given scores for each set of data. Equation 3.1 shows the mathematical expression of the consensus score function.

\[
CS_1^i = \text{sign}(Q_{1,Glide} - x_{Glide,i}) + \text{sign}(Q_{1,FlexX} - x_{FlexX,i}) + \text{sign}(Q_{1,PLANTS} - x_{PLANTS,i})
\]


CS1$^i$ consensus score 1; $i =$ pose; $Q_{1,name} =$ first quartile of the “name” set of data (Glide, FlexX and PLANTS set of scores); $x_{name,i}$ score obtained by pose “$i$” with the “name” software (Glide, FlexX or PLANTS)

A normalised consensus score (NCS1) was also calculated, using Equation 3.2:

\[
NCS_1^i = \frac{x_{Glide,i}}{\text{Min}_{Glide}} + \frac{x_{FlexX,i}}{\text{Min}_{FlexX}} + \frac{x_{PLANTS,i}}{\text{Min}_{PLANTS}}
\]

Equation 3.2. Normalised consensus scoring function 1.

NCS1$^i$ normalised consensus score 1; $i =$ pose; Min$^{name}_{name}$ lowest value of the “name” set of data (Glide, FlexX and PLANTS set of scores); $x_{name,i}$ score obtained by pose “$i$” with the “name” software (Glide, FlexX or PLANTS)
Molecules were ranked according to the CS1 and the NCS1 in descending order and only molecules that had CS1 equal to 3 (and therefore performed in the best 25% with all the modelling programs) were kept, in a single conformation for further analysis. The saved molecules from all three docking and rescoring procedures were merged and two conformations for each compound that had CS1 = 3 in at least two of the three target molecular systems were kept for further proceedings.

**CONSENSUS SCORE 2 (CS2)**

A new consensus score function was designed on purpose for this study with the objective of finding molecules that bind viral helicase better than host ones, with a higher discrimination than the natural ligand (ATP). Therefore, for each molecule a normalised consensus score (NCS) was calculated for each molecular system, using Equation 3.2. Subsequently, the anti-target NCSs were subtracted from the target NCSs, obtaining the consensus score 2 (CS2). The mathematical form of the scoring function is reported in Equation 3.3.

\[
CS2_i = \frac{x_{Glide,i}}{Min_{Glide}} + \frac{x_{FlexX,i}}{Min_{FlexX}} + \frac{x_{PLANTS,i}}{Min_{PLANTS}}_{targets} - \frac{x_{Glide,i}}{Min_{Glide}} + \frac{x_{FlexX,i}}{Min_{FlexX}} + \frac{x_{PLANTS,i}}{Min_{PLANTS}}_{anti-targets}
\]

*Equation 3.3. Consensus scoring function 2.*

CS2 = consensus score - 2; i = pose; Min_name = lowest value of the “name” set of data (Glide, FlexX and PLANTS set of scores); X_name,i = score obtained by pose “i” with the “name” software (Glide, FlexX or PLANTS)

In general, higher NCS is associated with a better theoretical binding of the ligand to the protein. Consequently, a positive CS2 means that the docking procedure suggests that the molecule has preference to bind to the target enzyme.

Molecules with CS2 equal or higher than 0.6 were selected for visual inspection.
3.6 EXPERIMENTAL SECTION

**COMPOUND PREPARATION**

Of the 15 tested compounds, 14 were purchased from SPECS\(^ {27} \) and one (18) was chemically synthesised in house in a previous study.\(^ {29} \)

For each compound, a solution of 20mM in DMSO was prepared, starting from approximately 5mg of compound. Two compounds (1 and 15) did not dissolve at this concentration and therefore they were diluted to 10mM in DMSO. These solutions were then used as starting material for the biophysical and enzymatic assays.

**THERMAL SHIFT ASSAY (TSA)**

The thermal shift assay (TSA) was performed on two DENV NS3hel from different serotypes (EVA314, from serotype 3, 50.554 kDa and EVA186, from serotype 4, 50.912kDa). Both proteins were previously prepared and purified in AFMB, Marseille University laboratories.

The buffer used for the assay was constituted of Tris (pH 7.5) 50mM, NaCl 150mM, TEMED 200mM and MnCl\(_2\) 2mM in water. SYPRO ORANGE (SO, with excitation wavelengths at 300nm and 470nm and emission wavelength at 570nm) was used as fluorescent probe.

PCR plates with 96 wells were used and each well was prepared with 25\(\mu\)L of total solution (20\(\mu\)L of protein, 1.25\(\mu\)L of compound 20mM solution and 3.5\(\mu\)L of SO solution). The final concentration of the proteins were 0.81mg/mL for DENV3 and 0.72mg/mL for DENV4, while the final concentration of the compounds was of 1mM in each well, except for the poorly soluble compounds (1 and 15) that were at 0.5mM. Three controls were also prepared for each protein: one with 0.5mM of AMPPNP as it is known to bind the enzyme, one with 5% DMSO and one with the protein alone.

Before performing the assay, the wells were visually inspected in order to detect the presence of eventual precipitated compounds. The assay was performed in a modified PCR machine that reads the fluorescence as a function of temperature. It was set for 360 cycles of 12s each with a temperature increase of 0.2\(^\circ\)C, starting from a temperature of 20\(^\circ\)C.

The resulting charts of fluorescence on temperature were analysed with ORIGIN. For each protein, a temperature window was chosen in order to include the reported thermal shift and
the curves were fitted with a Boltzmann distribution. Melting points and fitting errors were, in this way, determined.

**ATPASE ASSAY**

The ATPase assay was performed on DENV4 NS3helm (EVA186, serotype 4, 50.912kDa) that was previously prepared and purified in AFMB, Marseille University the laboratories. Solutions (A, B, C and D) for the phosphate colouring reaction were prepared on the day of the assay. Solution A contained 12% ascorbic acid in HCl 1M; solution B contained 1% ammonium molybdate tetra hydrate in water; solution D contained 2% sodium citrate in 2% acetic acid. Solution C was prepared less than one hour before detection and it was obtained by mixing 2 parts of solution A with one part of solution B.

For each assay, two premix solutions were prepared in reaction buffer (Tris pH 7.5 50mM, MgCl$_2$ 2mM, BSA 0.1mg/mL): one containing ATP (premix I) and one containing the enzyme with or without inhibitor (premix II). At time 0, premix II was transferred in premix I and the reaction was incubated at 37°C.

At desired time points, 20µL of mixture were sampled. In the sampling volumes the reaction was stopped with 0.5µL EDTA 50mM in a 96 well plate with the flat bottom. A volume of 60µL of solution C was added to each well and the obtained mixture was incubated for 5 minutes at room temperature. Consequently, 60µL of solution D were added to the mixture and the wells analysed with Magellan spectrophotometer at a wavelength of 595nm. Eight concentrations (obtained with 2X serial dilutions from an initial concentration of 10mM) of KH$_2$PO$_4$ with EDTA were also always analysed with this colorimetric assay as a reference for the correlation between the absorbance and the phosphate concentration.

**ASSAY CALIBRATION**

The assay was first performed with four enzyme concentrations (0, 25, 50 and 100nM) and two DMSO percentages (0% or 10%) were tested for each one of them. In all the cases, an ATP concentration of 0.5mM was used. The reaction was sampled at the following times: 0, 3, 6, 9, 12 and 15 minutes from the beginning. Phosphate concentration curves were calculated with a spread sheet and the optimal protein concentration was chosen. Subsequently, the assay was performed with two ATP concentrations (0.5 and 0.25mM) and two DMSO percentages (0% or 10%) were tested for each one of them. In all the cases, a protein concentration of 25nM was used. The reaction was sampled at the following times: 0,
3, 6, 9, 12 and 15 minutes from the beginning. Phosphate concentration curves were calculated with a spread sheet and the optimal ATP concentration was chosen.

**COMPOUND SCREENING**

After calibration, it was chosen to use 0.25mM and 25nM ATP and enzyme concentration respectively for the assay. All 15 compounds designed for NS3hel inhibition were screened in duplicates at 2mM concentration, following the general procedure reported above. The poorly soluble compounds (1 and 15) were tested at 1mM concentration. In this way the system always contained 10% DMSO. For each compound a blank containing all substances except the enzyme was prepared for background effects exclusion. A known helicase inhibitor (AMPPNP) was also tested with 10% DMSO as a negative control. The protein without inhibitors and with 10% DMSO was used as positive control.

The reaction was stopped after 8 minutes and phosphate concentration was calculated as described above. At the beginning and at the end of the assay, the reaction solutions were visually inspected in order to detect the presence of eventual precipitated compounds.

**FLUORESCENCE POLARISATION (FP) GENERAL PROCEDURE**

This assay was carried out with two probes in order to assess the compounds’ competition with RNA or NTP binding. Helicases from two viral serotypes were used in appropriate buffer for the assay: DENV-3 (EVA314, from serotype 3, 50.554kDa) and DENV-4 (EVA186, from serotype 4, 50.912kDa). The enzymes were previously prepared and purified in AFMB, Marseille University laboratories.

The assay was carried out at room temperature in a PHERAtar multidetector plate reader, using the “480 520 520” optic module. Greiner black plates with flat bottom wells were used. The fluorescence polarisation values for each well were saved as a spread sheet and analysed with GraphPad PRISM 6.0 software.
FP WITH GTP PROBE

ASSAY CALIBRATION

For NTP binding competition analysis ATP-bodipy and GTP-bodipy probes were purchased. They are NTPs linked through the ribose to a fluorescent moiety (bodipy FL, with excitation wavelength of 488nm and emission wavelength of 512nm).

Titration curves were obtained varying the enzyme concentrations (12 concentrations from 40µM to 78nM, with 2X dilution factor for the enzyme and 0nM) and with a constant concentration of the NTP-bodipy probe of 100nM. The buffer used for the assay was constituted of Tris (pH 7.5) 25mM, DTT 2mM and EDTA 0.5mM in water.

Titration curves were obtained for both DENV NS3hel serotypes and both NTP probes. For assay validation, a titration curve was also performed with DENV3 NS3hel and GTP-bodipy in the presence of AMPPNP 1mM. Data analysis was performed and the NTP-bodipy binding constant (K_d) was calculated in all the tested conditions.

Greiner 96 half area wells plates were used and for each well a total volume of 100µL was prepared.

COMPOUND SCREENING

Compound screening was performed on both helicase serotypes in the same buffer as the assay calibration, in Greiner 384 wells and for each well a total volume of 20µL was prepared. Constant protein and GTP-bodipy concentrations (of 3µM and 100nM respectively) were used; while all compounds were tested at eight concentrations with 2X dilution, starting from 1mM or from the maximum concentration possible. Compound dilutions were carried out in DMSO in order to have the same DMSO content (5%) in all of the wells. AMPPNP was tested as well as a negative control. Blanks with all compound concentrations and the GTP probe in buffer were also prepared in order to exclude background noise. The test was repeated in duplicate for selected molecules.
FP WITH RNA PROBE

ASSAY CALIBRATION

For RNA binding competition analysis the RNA oligonucleotide (RNABS9.2012) was used as the fluorescent probe. It is a 10 nucleotides RNA oligomer bound to the fluorophore carboxyfluorescein (6-FAM, with excitation wavelength of 495nm and emission wavelength of 520nm) at the 5’end. All titration curves were obtained varying the enzyme’s concentrations (from 0 to 20μM) and with a constant concentration of the RNA probe, at 50nM. The buffer used for the assay was constituted of Tris (pH 7.5) 25mM, DTT 2mM and MgCl₂ 2mM in water. For assay validation, titration curves were also performed with the two enzymes and RNA probe in the presence of ATP 2mM or AMPPNP 2mM. In more detail, for DENV3 NS3hel, twelve concentrations of protein (from 10μM to 0nM, with 2X dilution factor for the enzyme) were tested with RNA and eight of them (from 10μM to 0nM, with 4X dilution factor for the enzyme) were tested also with the addition of ATP (2mM). For DENV4 NS3hel, fifteen concentrations of protein (from 20μM to 0nM, with 2X dilution factor for the enzyme) were tested with RNA, whilst eight of them (from 10μM to 0nM, with 4X dilution factor for the enzyme) were tested with the addition of ATP (2mM) or AMPPNP (2mM).

The possible interference of DMSO in the assay was also assessed for both proteins. In this case, the wells were prepared with a constant concentration of protein (10μM) and the RNA probe (50nM) and different DMSO concentrations: 0%, 1%, 2%, 3%, 4%, 5% and 10%.

Data analysis was performed and the RNA probe dissociation constant (Kₐ) was calculated in all the tested conditions.

The assay calibration was performed in Greiner 96 half area wells and for each well a total volume of 100μL was prepared.

COMPOUND SCREENING

Compound screening was performed on both helicase serotypes in the same buffer of the assay calibration in Greiner 384 wells and for each well a total volume of 20μL was prepared. Constant protein and RNA probe concentrations (of 40nM and 50nM respectively) were used; while all compounds were tested at eight concentrations with 2X dilution, starting from 1mM or from the maximum concentration possible. Compound dilutions were done in DMSO in order to have the same DMSO content (5%) in all the wells. AMPPNP was included in the
experiment as a negative control. Blanks with all compound concentrations and the RNA probe in buffer were also prepared in order to exclude background noise. The test was repeated in duplicate for selected molecules.
3.7 BIBLIOGRAPHY


[24]: M. Gu, C. M. Rice, Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism, PNAS 107 (2010), p. 521-528, DOI: 10.1073/PNAS.0913380107


[29]: I. M. L. Trist, In silico studies of Flaviviridae Helicases as potential new targets for antiviral drug design, Univeristà degli Studi di Padova (2011)


Section 4:
NS3 HELICASE MECHANISM OF TRANSLOCATION ALONG RNA
4.1 INTRODUCTION

NS3 HELICASE MECHANISM OF TRANSLOCATION

NS3hel involvement in essential steps of the viral RNA processing, including RNA unwinding, NTP binding and hydrolysis, RNA translocation and RNA triphosphatase activity, makes this protein an attractive antiviral drug target. Indeed, the knowledge on DENV NS3 helicase is increasing and some potential enzyme inhibitors that putatively target the RNA binding site have started to appear in literature. Deeper insights about this protein can be found in the NS3 HELICASE AS A DRUG TARGET section of this thesis.

The unwinding function of this enzyme occurs in a NTP-dependent manner, involving the protein translocation along the nucleic acid. The direction of the movement along the nucleic acid has been explored for several DNA and RNA helicases and a 3’-5’ directionality was found for helicases, belonging to superfamily 2 (SF2). The molecular mechanism behind the translocation activity of the enzyme has been described for the hepatitis C virus (HCV) NS3hel, which belongs to SF2 as well as for the DENV helicase. Two crystal structure series suggested that the helicase translocation is driven by the chemical energy of ATP hydrolysis and that it occurs mainly through the rotation of domain 2 (D2) that produces a “scissors-like” movement that prompts a “ratchet” translocation pattern. In the absence of the natural ligand, the NTP binding site is open and 5 nucleotides can be accommodated in the RNA cleft. Upon NTP binding, D2 rotates towards domain 1 (D1), closing the NTP cleft and reducing the RNA pocket size. In this stage, changes in D1 interactions with the nucleic acid strand causes the exit of the 3’ nucleotide from the binding pocket, with only 4 nucleotides remaining in the cleft. NTP hydrolysis products are then released through the rotation of D2 away from D1. During NTP cleft opening, D2 interactions with RNA shift and allow the entrance in the nucleic acid pocket of a novel nucleotide at the 5’ end of the RNA, bringing back the number of bound nucleotides to 5. Thus, according to this model, the hydrolysis of one molecule of NTP corresponds to the helicase movement of one nucleotide along the viral genome. Unfortunately, the same type of structural investigation could not give exhaustive insights for DENV NS3hel, probably due to the crystallographic conditions. A published molecular dynamics (MD) study suggested that there is an analogy between dengue virus and the HCV helicases as a similar dynamism could be observed for the helicase alone or in complex with ATP. As these molecular systems did not include the RNA strand, the simulations did not explain the DENV NS3hel translocation mechanism for this complex. Consequently, the
mechanism of translocation of DENV NS3hel along the RNA remain elusive to date, in particular the reasons why it occurs in a 3'-5' direction.

**AIMS AND OBJECTIVES**

Being highly involved in the RNA replication cycle, NS3hel inhibition would lead to the arrest of DENV infection.\textsuperscript{[17]} Although knowledge on NS3hel is increasing and inhibitors start to appear in the literature, the development of compounds that apply any of these strategies has been challenging, as the real mechanism of action of the helicase is not yet fully understood.\textsuperscript{[17,7-9]}

As this important knowledge could aid structure-based drug design efforts, this study aimed at obtaining a more detailed understanding of the protein activity, using molecular modelling techniques that is believed to have never been applied before to DENV NS3hel. As a dynamic process was investigated, the approaches used were based on MD. In addition to classical MD (cMD) simulations of systems involving DENV NS3hel, accelerated MD (aMD) simulations were applied as well in order to improve the exploration of the system energy surface.\textsuperscript{[18-20]} As molecular processes that involve large protein molecules usually involve collective changes in conformations, the use of powerful statistical methods such as Principal Component Analysis (PCA) was applied as it aids the exploration of collective low-frequency motions of biomolecules.\textsuperscript{[21,22]} MD trajectories were therefore used to extract linear correlated vibrational modes of group of atoms. The in silico thermodynamic study of the protein-nucleic acid recognition has been of considerable interest in recent years, as these types of interactions are very common.\textsuperscript{[23,24]} The evaluation of free energy, in particular, allowed not only to understand processes at the atomic level and how they proceed, but also to probe states that are not experimentally accessible.\textsuperscript{[25]} Several studies on molecular systems reported in the literature have demonstrated a good agreement between free energy calculations and experimental values, giving mechanistic insights of chemical and biochemical phenomena and sometimes being successfully predictive.\textsuperscript{[26,27]} As the aim of this study was to understand the 3’-5’ directionality also at a molecular level, the MD trajectories were also used for binding energy estimation with MMGBSA\textsuperscript{[28-35]} (molecular mechanics Generalised Born surface area) methodologies. A description of this technique is reported in *APPENDIX 2* in the *AMBER* chapter.

Classical and accelerated MD approaches were applied to two crystal structures representing the DENV NS3hel-RNA complex in presence and in absence of a bound nucleoside triphosphate (NTP). The trajectory analysis, including visual inspection, PCA and MMGBSA calculations, allowed the identification of dynamic features of DENV NS3hel that could give
interesting hints on the mechanism and rationale behind the translocation of the protein along RNA at the molecular level, in particular of the 3'-5' directionality. Here, the obtained results are presented and discussed, as well as a proposed hypothesis on the possible mechanism of DENV helicase movements along the nucleic acid.
4.2 RESULTS AND DISCUSSION

CLASSICAL MOLECULAR DYNAMICS SIMULATIONS

This study was applied to the two crystal structures of the DENV NS3hel-RNA complex in the presence and in the absence of a NTP bound to the protein. The PDB structures for these systems belonged to the same crystallographic series and their IDs were 2JLV and 2JLU respectively. [15] The RNA fragment used was of 5 nucleotides with the same sequence (5' AGACU-3') for both molecular systems. Following system preparation, three 40ns MD simulations were performed for each molecular system with the GROMACS program, as described in the METHODS chapter, with the aim of improving the conformational space exploration respect to a single 120ns simulation.

All the MD simulations were monitored with backbone RMSD and total energy fluctuation evaluation. Figure 4.1 shows an example of total energy (panel A) and backbone RMSD monitoring (panel B), taken from one of the MD simulations. From this analysis the MD duration appeared to be appropriate and the molecular systems showed to be equilibrated from early stages of the MD simulations. [26]

![Figure 4.1. Example of total energy and backbone RMSD monitoring.](image)

Examples of MD simulations RMSD and total energy analysis. **A)** Time evolution of the system total energy. This example is from a NS3hel-RNA complex without ATP. **B)** RMSD time variations of the protein backbone. This example is from a NS3hel-RNA-ATP complex.

Residue RMSD fluctuations (RMSF) were also computed on MD trajectories. All simulations had a similar residue RMSF pattern with no significant differences between the molecular systems with or without ATP, as shown in panel A of Figure 4.2. This is consistent with the fact that no significant differences between the B-factors of the original PDB structures (panel B of Figure 4.2) can be detected by comparison. The highest RMSF values calculated from the MD simulations were recorded in particular for residues 168-172, belonging to the helicase-
protease linker, which, due to the absence of the protease portion of NS3, was very flexible. Additional high RMSF values were mainly recorded for other solvent exposed residues. Comparing the charts in panel A of Figure 4.2 with the original PDB B-factors in panel B of the same figure, it is possible to see that the high RMSF residues correspond to protein regions with high B-factors, indicating that RMSF patterns and simulated protein vibration are compatible with experimental data.

**Figure 4.2. Residue RMSF and PDB B-factors.**

_A) Residue RMSF._ The RMSF values for each residue calculated from all the MD simulations are plotted here. The trajectories involving the NS3hel-RNA-ATP and the NS3hel-RNA complexes are indicated with their names in the legend and the MD simulations are specified with numbers from 1 to 3. _B) PDB B-factors._ The atom B-factors of the original PDB used for the two molecular systems (NS3hel-RNA-ATP and NS3hel-RNA complexes) are presented here. Data was taken from PDB ID 2JLV and 2JLU for the NS3hel-RNA-ATP and NS3hel-RNA systems respectively.
LARGE SCALE CONFORMATIONAL CHANGES

The MD trajectories were observed in order to assess the influence of ATP on the overall enzyme movements, but the visual inspection of the trajectories failed to highlight major protein movement differences between the two molecular systems nor the “ratchet-like” behaviour previously reported in literature.[16]

Low frequency collective movements of large molecules are often associated with large molecule function. However, they are difficult to detect through simple visual inspection. Consequently, for a better understanding of relative domain movements, Principal Component Analysis (PCA) was applied to all the trajectories.[21] PCA is a statistical tool that, if applied to a MD trajectory, can be used to separate the modes in the movement of the molecule and to identify the major collective fluctuations of a protein. This method is based on the construction of a covariance matrix of Cartesian coordinates, which diagonalisation is used to obtain eigenvectors and eigenvalues. The first are the vectorial description of the fluctuation, while the second are the representation of the contribution of the eigenvector to the total motion. Eigenvectors, or principal components (PC), are usually sorted according to their eigenvalues and it was observed that the modes that contribute most to the total movement are also representative of the slowest dynamical transitions, usually associated with protein function.[21,22]

PCA was computed on all RNA atoms, while only Cα atoms were considered for the protein, in order to highlight major protein movements and ignore artificial apparent correlations between slow side-chain fluctuations and backbone motions. The calculated eigenvectors were in descending order of contribution and for all the simulations the first 10 modes were found to represent between 56.8% and 71.7% of the total system motion, as shown in Table 4.1. Figure 4.3 shows an example of how the percentage associated to these 10 eigenvectors decreases (panel A) and how the cumulative percentage grows (panel B) according to eigenvector number.

Table 4.1. Percentage of total motion described by the first 10 eigenvectors for all MD simulations.

<table>
<thead>
<tr>
<th>First 10 eigenvectors of simulation</th>
<th>Percentage of total motion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NS3Hel:RNA complex</strong></td>
<td></td>
</tr>
<tr>
<td>MD1</td>
<td>58.7%</td>
</tr>
<tr>
<td>MD2</td>
<td>71.7%</td>
</tr>
<tr>
<td>MD3</td>
<td>67.7%</td>
</tr>
<tr>
<td><strong>NS3Hel:RNA-ATP complex</strong></td>
<td></td>
</tr>
<tr>
<td>MD1</td>
<td>68.5%</td>
</tr>
<tr>
<td>MD2</td>
<td>56.8%</td>
</tr>
<tr>
<td>MD3</td>
<td>63.7%</td>
</tr>
</tbody>
</table>

The molecular system is reported in the first column and the three MD simulations for each one of them is indicated by MD, followed by a number.
Eigenvectors were then used singularly and collectively for trajectory filtering and porcupine plots construction. In all of the simulations fluctuations were not involving major domains rearrangement, particularly in the NS3hel-RNA-ATP molecular system. Furthermore, an important contribution was given by the more flexible RNA 3’ and 5’ ends, particularly evident in those simulations that showed peculiar 3’ or 5’ patterns, described in the SMALL CONFORMATIONAL CHANGES chapter. Protein motions with high contribution were always associated to a fluctuation of RNA binding site dimensions, with a general trend of domains 1 (D1) and 2 (D2) to get closer to domain 3 (D3), representing between 17.6% and 37.4% of the total motion according to the simulation. Other motions regarding independent domain movements particularly for D2 which has been shown to be more flexible than the others, consistent with the translocation mechanism described for HCV helicase.\(^{12,13}\) Never in the first five eigenvectors and very few of the first modes were similar to the “scissors-like” pattern described for HCV helicase and they always represented vibrations with low contributions to the whole motion (never higher than 2.8%). These modes involved the more external portions of the domains and never with the D1 and D2 regions that interact with RNA. Figure 4.4 shows an example of a porcupine plot in which the main movements are associated with the protease-helicase linker region and vibration that causes the change in RNA binding site size.
Figure 4.4. Example of porcupine plot.
The porcupine plot of the second eigenvector (corresponding to 8.6% of motion) of a MD simulation of the NS3hel-RNA molecular system. The front and lateral view are reported on the left and on the right respectively. The protein is shown with a red ribbon and the RNA is represented as orange cartoon. The vectors associated to this mode are shown as blue needles.

The variation with time of the distance between key residues alpha-carbons (Cα) was also computed on all simulations. Table 4.2 reports the selected amino acid pairs and the relative domain movements that they were selected for, while Figure 4.5 shows the position of these residues in the protein. In both cases, domain 1 is indicated by D1, domain 2 by D2 and domain 3 by D3. The searched patterns in residue pair distances were the ones that would confirm a domain 2 lateral rotation consistent with previously published observations.\[16\]

Table 4.2. Residue pairs selected for relative domain movement monitoring.

<table>
<thead>
<tr>
<th>DOMAIN RELATIVE MOVEMENT</th>
<th>RESIDUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 – D2</td>
<td>Thr224 – Ile410</td>
</tr>
<tr>
<td></td>
<td>Glu285 – Met413</td>
</tr>
<tr>
<td></td>
<td>Lys201 – Asn416</td>
</tr>
<tr>
<td>D1 – D3</td>
<td>Arg225 – Asp541</td>
</tr>
<tr>
<td></td>
<td>Pro245 – Arg538</td>
</tr>
<tr>
<td></td>
<td>Thr267 – Arg538</td>
</tr>
<tr>
<td>D2 – D3</td>
<td>Ile365 – Asp603</td>
</tr>
<tr>
<td></td>
<td>Lys366 – Pro604</td>
</tr>
<tr>
<td></td>
<td>Arg387 – Arg599</td>
</tr>
<tr>
<td></td>
<td>Arg387 – Asp541</td>
</tr>
</tbody>
</table>

D1: domain 1; D2: domain 2; D3: domain 3.
In particular, a HCV-like NTP binding site closure would be associated with an increase of the distance between residues of D2 and D3, simultaneous to a decrease in distance between residues from D1 and D2. Opposite patterns would be expected in the case of NTP binding site opening. Generally, small distance variations could be observed for all residue pairs, indicating small overall changes in protein conformation. Within these small fluctuations, distance oscillations between 0.25nm and 0.70 could be observed for Ile365 – Asp603 and Lys366 – Pro604 pairs (representative of the D2-D3 relative motion) in both molecular systems. Although these distance fluctuations could look similar to what has been previously published, they were never accompanied by opposite direction distance variation of the D1-D2 amino acids pairs (e.g. Lys201 – Asn416), excluding the presence of a “scissors-like” movement of D2 in these MD simulations. This was in agreement with the PCA results that showed that this pattern was not an important component of the total motion. Furthermore, the RNA binding site dimension fluctuations and domain independent movements observed in the filtered trajectories, could be associated with some residue pairs fluctuations, as patterns D1-D3 and D2-D3 distance patterns were often similar. An example of these distance fluctuations is reported in Figure 4.6. Here, the small distance fluctuation patterns between the Cα of residues belonging to Ile365-Asp603 (for D2-D3 relative movements) and Arg225-Asp541 (for D1-D3 relative movements) pairs are similar, while the ones of Lys201-Asn416 (for D1-D2 relative movements) are not. This could be correlated with the fluctuation of RNA binding site size variation described above.
4.2 RESULTS AND DISCUSSION

NS3HEL COMPLEXES WITHOUT RNA

One hypothesis that could explain the lack of “scissors-like” movements is that the presence of the RNA in the system stabilises and constrains the protein movements. In order to verify this theory, new MD simulations were performed following the same procedure of the already described simulations. In this case triplicate MD simulations were performed for two molecular systems in which the helicase was not bound to the nucleic acid were considered for the new MD simulations: one with the apo form of the protein and one with the protein in complex an ATP molecule, with PDB IDs 2JLQ and 2JLR respectively. Already from visual inspection, it appeared that the protein was more flexible and free to explore the conformational space, in particular in the case of the apo protein, indicating that natural substrate binding (either RNA or ATP) stabilises the molecular system. Similarly to the previous molecular systems, also in the cases of NS3hel alone or in complex with RNA clear “scissors-like” patterns could not be directly observed, while the variation of the RNA binding site size was significant. This was confirmed by PCA, as movements associated to RNA cleft size variations were often described by the first eigenvector (representing between 18.8% and 31.9% of motion).

The pairwise residue Cα distance investigation confirmed a general domain movement, less important in the NS3hel-ATP system as the distances fluctuated generally less in this case. However, the D1-D2 and the D2-D3 fluctuations were neither contemporary nor inverted in direction. This suggests that in the presence of the ATP molecule, the helicase is more stable.
in the conformation that is more convenient for molecule binding, while in the absence of this molecule, the protein explores more the available conformational space.

The lack of consistency of the distances variations and of the observation of the RNA cleft opening and closing movement, did not confirm that the helicase, in the absence of ligands, gives a “scissors-like” conformational change that resembles the one observed in the case of the HCV helicase.\textsuperscript{[12,13]} Therefore, it was concluded that the lack of the “scissors-like” pattern in the MDs of the helicase complexes with RNA were not related to a dynamic constraint caused by this substrate.

**BINDING FREE ENERGY CALCULATIONS**

The average interaction energies and their components were calculated for each MD, following the procedure reported in the *METHODS* chapter. A microscopic system explores many states at a given temperature and these are reflected in thermodynamic properties. Consequently, free energy calculations need to sample as extensively as possible the configurations at that temperature. One of the most common approaches to do this is based on MD simulations from which snapshots are saved and used for system probing.\textsuperscript{[36]} As in all MD the system was well equilibrated from the early stages of the simulation, 1906 representative snapshots for the binding free energy calculations were taken from each trajectory and each one of them was used for free binding energy estimation, merging molecular mechanics with an implicit solvent model through MMGBSA methodology.\textsuperscript{[36,37]} In molecular complexes treated with implicit solvents, molecule entropy is defined through the translational and rotational freedom (vibrational entropy) and its possible low energy configurations (conformational entropy). It has been shown that upon ligand-protein binding, the overall loss of ligand entropy and the consequent increase of free energy of the system can make a big contribution, especially in high-affinity complex systems.\textsuperscript{[38]} Consequently, entropy can be an important contribution to binding energy. However, with MMGBSA methods, the largest standard deviations and errors have been related to the uncertainty calculation of this energy component.\textsuperscript{[22,39]} From these considerations and since in this study the aim was to compare the behaviour of different molecular systems but involving the same complex of interest counterparts, it was decided to estimate the free energy values ignoring the entropic term of Equation A2.13 in APPENDIX 2.

For all the MD simulations, the obtained binding free energy profile was overall constant for the duration of the simulation, as shown in the examples in Figure 4.7. Here the results obtained for two simulations of the NS3hel-RNA and NS3hel-RNA-ATP complexes are reported.
Figure 4.7. Time variation of binding free energy values.

One example of time variation of the NS3hel-RNA binding free energy is reported here for each molecular system. In both cases, the values obtained for the NS3hel-RNA and NS3hel-RNA-ATP complexes are indicated with “NS3hel-RNA” (light blue line) and “NS3hel-RNA-ATP” (light red line) respectively. Averages are also reported as “Average NS3hel-RNA” (dark blue line) and “Average NS3hel-RNA-ATP” (dark red line). In these examples, the averages values were -113.50 kcal/mol and -122.69 kcal/mol for NS3hel-RNA and NS3hel-RNA-ATP systems respectively.

Table 4.3. Binding free energy components of the DENV NS3hel-RNA interaction in absence of ATP.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>MD 1 (kcal/mol)</th>
<th>Std</th>
<th>MD 2 (kcal/mol)</th>
<th>Std</th>
<th>MD 3 (kcal/mol)</th>
<th>Std</th>
<th>Average (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔE_int</td>
<td>0.68</td>
<td>0.22</td>
<td>0.68</td>
<td>0.22</td>
<td>0.68</td>
<td>0.22</td>
<td>0.68</td>
</tr>
<tr>
<td>ΔE_vdW</td>
<td>-87.75</td>
<td>5.11</td>
<td>-106.07</td>
<td>8.66</td>
<td>-96.97</td>
<td>6.62</td>
<td>-96.93</td>
</tr>
<tr>
<td>ΔE_Coul</td>
<td>-776.06</td>
<td>42.90</td>
<td>-769.20</td>
<td>40.86</td>
<td>-766.28</td>
<td>36.31</td>
<td>-770.51</td>
</tr>
<tr>
<td>ΔE_gas</td>
<td>-863.12</td>
<td>43.48</td>
<td>-874.59</td>
<td>39.41</td>
<td>-862.56</td>
<td>37.48</td>
<td>-866.76</td>
</tr>
<tr>
<td>ΔG_sol</td>
<td>749.62</td>
<td>40.01</td>
<td>735.45</td>
<td>38.13</td>
<td>731.71</td>
<td>32.35</td>
<td>738.93</td>
</tr>
<tr>
<td>ΔG_tot</td>
<td>-113.50</td>
<td>6.78</td>
<td>-139.14</td>
<td>9.41</td>
<td>-130.85</td>
<td>8.30</td>
<td>-127.83</td>
</tr>
</tbody>
</table>

Values for all three MDs are reported, as well as the overall averages. The three simulations are indicated with MD1, MD2 and MD3. ΔE_int: internal energy; ΔE_vdW: van der Waals energy; ΔE_Coul: Coulomb energy; ΔE_gas = ΔE_int + ΔE_vdW + ΔE_Coul; G_solv: solvation energy based on Generalised Born method; ΔG_tot: total binding free energy calculated with MMGBSA; Std: standard deviation.

Table 4.4. Binding free energy components of the DENV NS3hel-RNA interaction in presence of ATP.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>MD 1 (kcal/mol)</th>
<th>Std</th>
<th>MD 2 (kcal/mol)</th>
<th>Std</th>
<th>MD 3 (kcal/mol)</th>
<th>Std</th>
<th>Average (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔE_int</td>
<td>0.68</td>
<td>0.23</td>
<td>0.68</td>
<td>0.23</td>
<td>0.68</td>
<td>0.22</td>
<td>0.68</td>
</tr>
<tr>
<td>ΔE_vdW</td>
<td>-102.21</td>
<td>6.40</td>
<td>-98.19</td>
<td>5.16</td>
<td>-95.44</td>
<td>5.62</td>
<td>-98.61</td>
</tr>
<tr>
<td>ΔE_Coul</td>
<td>-816.17</td>
<td>36.08</td>
<td>-878.09</td>
<td>55.05</td>
<td>-809.11</td>
<td>43.94</td>
<td>-834.46</td>
</tr>
<tr>
<td>ΔE_gas</td>
<td>-917.70</td>
<td>37.37</td>
<td>-975.59</td>
<td>56.40</td>
<td>-903.87</td>
<td>44.82</td>
<td>-932.39</td>
</tr>
<tr>
<td>ΔG_sol</td>
<td>782.56</td>
<td>32.67</td>
<td>845.17</td>
<td>51.75</td>
<td>781.19</td>
<td>40.17</td>
<td>802.97</td>
</tr>
<tr>
<td>ΔG_tot</td>
<td>-135.14</td>
<td>8.72</td>
<td>-130.42</td>
<td>8.46</td>
<td>-122.69</td>
<td>8.15</td>
<td>-129.41</td>
</tr>
</tbody>
</table>

Values for all three MDs are reported, as well as the overall averages. The three simulations are indicated with MD1, MD2 and MD3. ΔE_int: internal energy; ΔE_vdW: van der Waals energy; ΔE_Coul: Coulomb energy; ΔE_gas = ΔE_int + ΔE_vdW + ΔE_Coul; G_solv: solvation energy based on Generalised Born method; ΔG_tot: total binding free energy calculated with MMGBSA; Std: standard deviation.
In line with all the data discussed above, the interaction energy profile confirmed that in all the MDs the molecular systems were well equilibrated and stable. From the comparison of the results obtained for the two molecular systems, the presence or absence of ATP does not significantly affect the interaction energy values, in line with the RMSD and total energy profiles. This is confirmed by the average values of the total binding energy and of its components over whole trajectories (Tables 4.3 and 4.4).

The five nucleotides and residues in proximity of the RNA strand were selected for a more detailed binding free energy evaluation with the MMGBSA method. The selected amino acids were the ones within the ranges: Ala222-Val227; Tyr242-Ala246; Met260-Leu269; His287-Cys292; Val362-Ala367; Gln384-Thr389; Thr407-Ser411; Cys428-Val432; Ile442-Ala444; Asp484-Ala486; Met537-Pro543; Ala598-Leu607. The interaction energy was decomposed per residues and the contribution of each one of the computed amino acid was obtained from the same representative snapshots used for the total binding energy calculations, following the procedure described in the METHODS chapter.

Figure 4.8 shows the average values of the interaction energy contribution for all the analysed residues, as well as RNA nucleotides. The residues that give higher contribution to RNA binding are Arg225 and Arg387 in particular and also Thr224, Thr244, Ser364, Ile365, Ser386, Thr408, Ile410 and Arg599. Other residues with lower, but still significant, energy contributions are: Pro245, Cys261, Ala263, Thr264, Phe288, Asp290, Lys388, Leu429, Pro431 and Leu443 and Arg538. Not surprisingly, most of these residues correspond to the interacting residues reported in the crystallographic data (see Figure 3.5 in the NS3 HELICASE AS A DRUG TARGET section) and they belong to domains 1 and 2. This complies with the fact that most contacts that the enzyme has with the RNA are at the level of the sugar-phosphate backbone of the nucleic acid.\cite{15} In line with the RMSD, total energy and total binding free energy data, also the residue free energy contribution displays no relevant differences between the complexes in presence or absence of ATP. The only main dissimilarity is Arg538, but this can be connected to a different initial position of the side chain of the residue rather than to peculiar binding mechanism at this level.

Interestingly, residues involved in hydrogen bonds with the RNA 2’OH groups and considered to be responsible for DENV NS3 helicase specificity for RNA (Pro223, Thr264, Pro363 and Asp409), had a small contribution to the total interaction energy (see Figure 4.8).\cite{5,15} Counting the hydrogen bonds between these amino acids and the correspondent hydroxyl groups showed that these interactions were not stable in time, explaining the minor contribution to the total interaction energy of these amino acids. However, this does not exclude the importance of these contacts in RNA preferentiality, as the rapid formation and disruption of these interactions was observed throughout the simulations and could be important in nucleic acid recognition and translocation along the binding pocket.
4.2 RESULTS AND DISCUSSION

**Figure 4.8. Per residue binding free energy decomposition.**
The average per residue binding free energy values are reported here. Data was obtained by averaging the results of the triplicate MD results, for each molecular system. Results for NS3hel-RNA-ATP and NS3hel-RNA complexes are indicated as “AVERAGE NS3hel-RNA-ATP” and “AVERAGE NS3hel-RNA” respectively. The residues discussed in the text are indicated here.

**SMALL CONFORMATIONAL CHANGES**

In addition to the average values, the analysis of the time variation of the per residue binding free energy values gave useful insights into change in conformations of small residues, in particular near the RNA extremities. The two oligonucleotide ends (A1 and U5) themselves appeared to have a more variable interaction energy pattern, explaining why they have a lower binding free energy contribution respect to G2, A3 and C4 (see Figure 4.8). This was consistent with the high flexibility of 3’ and 5’ nucleic acid ends observed with visual inspection and principal component analysis. Both nucleotides were very flexible in all MD simulations, but in some of them particular movement patterns could be observed at the 5’ and 3’ end. However, no correlation with the presence of ATP was found.

In three trajectories (two from the complex with ATP and one from the one without) it was possible to observe a clear rotation of A1, prompting the G2 phosphate to move away from DENV NS3hel domain 2 and the increase of the nucleic acid strand curvature. Major transitions in the protein are accomplished by the α1’ helix and the loop before it (Motif IV), the loop between β2A’ sheet and α2’ and the α3’ helix and the loop before it (Motif V). While the latter two portions of DENV NS3 helicase followed the RNA trend, Motif IV, in particular Ile365, seemed to move in the opposite direction. This increase in space between the molecular partners triggers the break of the hydrogen bond between Ile365 backbone and G2 phosphate as the distance between the gap between the Ile365 α-nitrogen and G2 phosphate...
oxygen grew from 2.79Å to 5.67Å, negatively affecting the binding free energy contribution of the residue (panel B of Figure 4.9). The hydrogen bond break was accompanied by the rotation of the Ile365 side chain into a hydrophobic site given by Gln384, Thr389, Val406 and Val362 (panel A of Figure 4.9).

![Figure 4.9. Movement pattern around RNA 5' end.](image)

**A**. MD visualisation. The shift of G2 phosphate and Ile365 side chain are represented here by superposition of two frames of the trajectory of a MD simulation. The ribbons and the carbon atoms of the frame at the beginning of the trajectory are in orange, while the ones of the frame taken from around the middle of the simulation are in green. The shifts are indicated by black arrows. The hydrogen bond between Ile365 and G2 phosphate present at the beginning of the simulation is shown as a black line. Discussed residues are indicated.

**B**. Time variation of per-residue binding free energy. Data for Ile365 and Leu443 are described as an example. Here, the Ile365 free energy contribution increases upon hydrogen bond break in the first 2ns, while β-hairpin residues (exemplified by Leu443) remain constant.

The RNA shift exposed the A1 ribose and in particular its 5’ hydroxy moiety to the RNA binding site entrance, suggesting that this could favour the entrance of a novel nucleotide in the cleft and therefore be important in 3’-5’ translocation of the helicase along the nucleic acid. Importantly, in A1 rotation, its base did not move from its position, blocked by hydrophobic contacts mainly between the A1 base and residues Leu429, Pro431 and Leu443. These interactions appeared to be conserved during all MD simulations, explaining the unchanging interaction energy contribution pattern of these residues. On this base, it could be hypothesised that these hydrophobic residues play a key role in impeding NS3hel to translocate towards the RNA 3’ end by maintaining in place the base of the nucleotide in the 5’ end of the binding site.

Similarly to the other terminus, the 3’ end of the nucleic acid strand was flexible in all the MD simulations. However, in four trajectories (two from the complex with ATP and two from that
without the U5 nucleotide base moved out of the binding site and interacted with domain 1. Unlike the conformation variations described for the 5’ RNA end, U5 did not change conformation always in the same way and the protein residues were more static, suggesting that the free energy contribution of the amino acids varied as a consequence of U5 movement. Two examples of how different the U5 displacement from the RNA binding site was are presented in Figure 4.10 (panels A and B). The shift of U5 from the RNA cleft was reversible in two of the four simulations, confirming that the newly established interactions between the nucleotide and the protein were neither stable nor mechanistically relevant to explain the NS3hel translocation along the nucleic acid. Nevertheless, the fact that the 3’ end nucleotide, differently to the one in the 5’ end, was flexible and free to exit the RNA cleft, is relevant and supports the hypothesis that the β-hairpin residues (Leu429, Pro431 and Leu443) have an important role in the 3’-5’ directionality of NS3hel translocation mechanism.

**Figure 4.10. Movement pattern around RNA 3’ end.**
The two panels illustrate how different and unstable the 3’ nucleotide displacement from the RNA binding site could be. Each panel shows the superposition of three frames taken from the same MD trajectory. Different panels correspond to different MD simulations. In both panels, the protein surface is in grey and some residues are indicated with lines. The carbon atoms of the initial RNA conformation are in orange. An intermediate conformation is shown in magenta and a third conformation is shown in cyan.
Despite the novel insights on residue changes in conformations that prompt the 3'-5' directionality of the helicase along the RNA, no major protein conformational changes could be identified and the role of ATP could not be assessed.

Big molecules, such as proteins and nucleic acid, explore their conformational space by passing from one local energy minimum to another, crossing energy barriers, and this requires timescales of milliseconds to seconds or even longer, according to the molecular system.\[18,19\]

As MD simulations require considerable calculation, classical molecular dynamics are usually performed in a nanosecond timescale. If insufficient energy is supplied, the conformational space beyond high barriers cannot be easily explored, “trapping” the molecular system in a local minimum for periods of time that can be longer than the simulation itself. This means that nanosecond time frame classical MD allows the exploration of the free energy surface around the local minimum, usually close to the initial conformation, but fails in the study of long timescale large molecule conformation transitions that require a broader sampling of the free energy surface.\[18\] On this base, it has been hypothesised that the performed 40ns MD simulations failed to produce insights in broader protein changes in conformations because not enough simulation time was given to the systems for an appropriate exploration of the free energy surface. As the available computational power was not enough to perform millisecond time scale cMD, an alternative MD approach was used: accelerated molecular dynamics (aMD).

Accelerated molecular dynamics is a method that enhances sampling compared to cMD, without increasing the necessary simulation time. In the last decade this approach has been applied in several studies, thanks to its relatively simple application that does not alter the approximated shape of the free energy landscape of the system, making it ideal for the study of large scale global motions.\[18,19,40-44\] It is a powerful tool that has shown to be able to produce consistent results with millisecond time scale cMD simulations in conformational space exploration.\[20\]

The method reduces the amount of computational time spent in a local energy minimum by the molecular system by applying a bias potential $\Delta V(r)$ to the real potential $V(r)$ only when the system is below a chosen threshold energy $E$. Only in this case the simulation is performed on the modified potential $V^*$, while the true potential is applied in the case of an energy greater than $E$. The mathematical form of the potential alteration is reported in Equation 4.1.\[19,40\] In this way, the energy barriers near the local minima are reduced, facilitating the passage from a local minima to the other and reducing the time spent in the minima wells by
the molecular system. Figure 4.11 shows a schematic representation of the described potential modification.

\[
V^*(r) = \begin{cases} 
V(r), & V(r) \geq E \\
V(r) + \Delta V(r), & V(r) < E 
\end{cases}
\]

_Equation 4.1._ Calculation of the modified potential \(V^*(r)\).[40]

\(V^*(r)\): modified potential; \(V(r)\): real potential; \(\Delta V(r)\): bias or boost potential; \(E\): threshold energy

![Figure 4.11. Schematic representation of the potential modification.](image)

The real potential \(V(r)\) is modified by the application of the boost potential \(\Delta V(r)\) when the system is at an energy below the chosen threshold \(E\), as explained in the text.

If \(E\) or \(\Delta V(r)\) are too high, the minima wells will not be sampled sufficiently and in extreme cases might also disappear. In the opposite case, if they are too small the modified potential will be too similar to the real one and sampling will not be effective. Consequently, the choice of \(E\) and \(\Delta V(r)\) are important for a correct potential smoothening.[18] \(\Delta V(r)\) is calculated with Equation 4.2 and depends on two factors: \(E\) and a tuning parameter \(\alpha\). Thus, the needed parameters to set aMD simulations are \(E\) and \(\alpha\).[18,40]

\[
\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}
\]

_Equation 4.2._ Calculation of the boost potential \(\Delta V(r)\).[40]

\(\Delta V(r)\): bias or boost potential; \(E\): threshold energy; \(V(r)\): real potential; \(\alpha\): tuning parameter

These equations are applied to the dihedral potential of a molecular system because the major component that determines protein conformation is given by torsional rotations. However, it has been shown that if explicit solvent is used in a simulation, a better sampling is obtained if a boosting potential energy is applied also to solvent molecules as well in order to avoid the slowing down of large scale protein movements caused by diffusion of the solvent around the large molecule.[18,19,40] For this reason, also a potential energy bias was applied in this study, using a so called “dual-boost” a MD approach.

With the aim of revealing insights in DENV NS3hel mechanism of translocation along the RNA, the method was applied to the same molecular systems used previously: the NS3hel-RNA and the NS3hel-RNA-ATP complexes, taken from the same PDB IDs (2JLU and 2JLV respectively).[15]

From previous published work, it was seen that optimal \(E\) and \(\alpha\) parameters can be calculated
from residue and atom numbers and from the average potential and dihedral energies that can be obtained from cMD simulations.\textsuperscript{[18,19,40]} Thus, two 40ns cMDs were performed, one for each molecular system, following the procedure reported in the \textit{METHODS} chapter. The simulations were repeated because at the time of the study it was not possible to perform aMD with GROMACS, the software used for the cMD simulations presented above. Consequently, another MD program (AMBER) was used instead. The average dihedral and potential energies obtained from the 40ns cMD simulations were used in Equations 4.3, 4.4, 4.5 and 4.6 reported in the \textit{METHODS} chapter for the calculation of the $E$ and $\alpha$ parameters for the two boost potentials. Table 4.5 reports the average dihedral (DIHED) and potential (POT) energies obtained from the cMD simulations and the calculated $E$ and $\alpha$ parameters for the dihedral ($E_{\text{dihed}}$ and $\alpha_{\text{dihed}}$) and the potential ($E_{\text{pot}}$ and $\alpha_{\text{pot}}$) biases. Numbers of atoms in the system and of protein residues are also stated. All values are reported for the NS3hel-RNA and NS3hel-RNA-ATP complexes. ATP, where present, was not considered in the residue count as the changes in conformations of interest regarded the protein and the RNA strand only. The calculated parameters were then used for the setup of two 200ns aMD simulations, one for each molecular system, with the AMBER program. The obtained trajectories were then analysed in a similar way to that done for the cMD simulations computed with GROMACS.

\textit{Table 4.5. Parameters for aMD simulations.}

<table>
<thead>
<tr>
<th></th>
<th>NS3hel-RNA</th>
<th>NS3hel-RNA-ATP</th>
<th>Measure unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIHED</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein residues number</td>
<td>7992.46</td>
<td>8024.84</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>$E_{\text{dihed}}$</td>
<td>9816.46</td>
<td>9,848.84</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>$\alpha_{\text{dihed}}$</td>
<td>364.8</td>
<td>364.8</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>POT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System atoms number</td>
<td>-186,592.95</td>
<td>-205,418.33</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>$E_{\text{pot}}$</td>
<td>-174,222.39</td>
<td>-193,238.81</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>$\alpha_{\text{pot}}$</td>
<td>12,370.56</td>
<td>12,179.52</td>
<td>kcal/mol</td>
</tr>
</tbody>
</table>

During the aMD simulations an average of 102.21 kcal/mol and 101.93 kcal/mol total boost potential was applied for the NS3hel-RNA and NS3hel-RNA-ATP complexes respectively. As shown in Figure 4.12, the applied boost varied during the simulation as higher $\Delta V(r)$ was applied in correspondence to local minima. This variation suggested that different regions of the free energy surface were sampled, complying with the aim of the method. In the charts in the figure the time variation of $\Delta V(r)$ is shown in grey and the 5 period running average is shown in black for a better visualisation of the boost potential trend during the simulation. For the NS3hel-RNA complex (panel A of Figure 4.12) the applied $\Delta V(r)$ ranged between 169.36
kcal/mol and 47.16 kcal/mol; while for the NS3hel-RNA-ATP complex (panel B of Figure 4.12) it was between 168.02 kcal/mol and 53.66 kcal/mol.

![Diagram A and B](image)

**Figure 4.12.** Applied boost potential trend during aMD.
In both panels the boost potential is plotted against simulation time. The trend is shown with a grey line, while the 5 point moving average trend is represented with a black line. **A) Data for the NS3hel-RNA complex. B) Data for the NS3hel-RNA-ATP system.**

**LARGE SCALE CONFORMATIONAL CHANGES**

As shown in Figure 4.13, RMSD fluctuations were different for the two molecular systems and reach a maximum value of 4.3Å and 4.0Å for the NS3hel-RNA and NS3hel-RNA-ATP complexes respectively.

![Diagram C](image)

**Figure 4.13.** Protein backbone RMSD for aMD simulations.
The RMSD values of the protein backbone are plotted here against simulation time. Data for the NS3hel-RNA and the NS3hel-RNA-ATP complexes are shown in grey and in black respectively.

Comparing this figure with panel B of Figure 4.1 it appears that RMSD values were higher than the ones recorded for the cMD simulations, confirming a greater exploration of
conformational space of the protein with the aMD approach. Furthermore, higher fluctuations of the values are also consistent with a higher exploration of the free energy landscape.

The two panels of Figure 4.14 show the RNA RMSD time-dependent variations after fit on the nucleic acid strand (panel A) or on the protein backbone (panel B). A much higher RNA displacement was observed for the NS3hel-RNA system with both fitting methods. In particular, in this molecular system a shift in RNA position respect to the protein occurred approximately between 90ns and 150ns and higher changes in RNA conformations appeared after approximately 60ns for the rest of the simulation. In the NS3hel-RNA-ATP system, changes in RNA conformation are more relevant approximately in the last 20ns of simulation and no significant shift from the protein can be observed.

A) RNA RMSD with fit on RNA. B) RNA RMSD with fit on the protein backbone.

Per-residue RMSD fluctuations (RMSF) calculated for the two molecular systems are shown in Figure 4.15. In both cases, the most flexible residues of the protein were located in the protease-helicase linker sequence and mainly solvent exposed regions of the protein. Higher RMSF values were also generally associated with D2 and D3 (from residue 318 to residue 618), indicating a higher flexibility of these two domains. Importantly, in both types of simulation high RMSF values were associated with a sequence belonging to D1 and formed by the β2A sheet and a random coil (residues 241-257). Significantly higher values for the NS3hel-RNA complex respect to the NS3hel-RNA-ATP one were recorded for the protease-helicase linker region (residues 168-180) and the RNA strand (residues 619-623). The latter is consistent with the RMSD observations discussed above. Like for the cMD simulations, also for aMD the regions with high RMSF values corresponded to residues with high B-factors in the original PDB files (2JLU and 2JLV for the NS3hel-RNA and NS3hel-RNA-ATP complexes respectively),
confirming that the conformational changes occurred in the simulations are consistent with the experimental data.

![Residue RMSF from aMD simulations](image)

**Figure 4.15.** Residue RMSF from aMD simulations.

The RMSF values for each residue calculated from all the MD simulations are plotted here. The trajectories involving the NS3hel-RNA-ATP and the NS3hel-RNA complexes are indicated with these names in the legend and they are shown with a black and a grey line respectively.

Plotting the variation with time of residue RMSD values (Figure 4.16) indicated that even if regions with high RMSF are similar, in the two molecular systems the RMSD values changed in different moments in the simulation. Particularly relevant RMSD variations are highlighted in the figure. As already mentioned, the linker region between NS3pro and NS3hel (residues 168-180) was very flexible, in particular in the NS3hel-RNA system. The most relevant difference between the two simulations is the D1 region between residues 241-257, corresponding to the β2A sheet and a random coil. Even if the total RMSF values are similar (see Figure 4.15), in the aMD without ATP this sequence changes conformation after 20ns and maintains high RMSD values for the rest of the simulation, while in the other aMD RMSD increases only after approximately 170ns. Not surprisingly, differences between the two simulations also involved the ATP binding site as the presence of the natural ligand constrained the binding residues in their conformation. In particular two D2 sequences appeared to have higher RMSD values in the system without ATP: residues 317-330 (belonging to motif III, responsible for NTPase and helicase activities coupling) and residues 462-470 (partially belonging to motif VI, also involved in NTP binding). Importantly, less difference could be observed for the portions of these motifs in proximity of the ATP base. Consistently with DENV NS3hel lack of base selectivity for NTP binding, during the NS3hel-RNA-ATP simulation the ATP base was particularly flexible and interacted non-specifically with sequences of domains 1 and 2, while the phosphate portion of the ligand was binding stably in the cleft. This explained the higher RMSD values for the portions of motifs III and VI close to the ATP base in the simulation. Other
solvent-exposed regions with high RMSD values are not highlighted because they are not relevant for the discussion of the helicase translocation along the RNA.

With the aim of assessing global protein movements, also for the aMD simulations the variation upon time of the distance between key residues C$^\alpha$ atoms was computed on all simulations. The same residue pairs (reported in Table 4.2 and Figure 4.5) were used, but like in the case of the cMD simulation no relevant patterns that could explain domains relative movements were observed.

More informative was the principal component analysis that was computed for the two aMD simulations, using all heavy atoms belonging to the RNA strand and protein C$^\alpha$ atoms. Where present ATP was not selected because the aim for PCA was to identify protein and RNA conformational changes that could give an insight of the helicase mechanism of translocation. The calculated eigenvectors were in descending order of contribution and for all the simulations the first 10 modes represented 68.66% and 60.05% of the total system motion for the NS3hel-RNA and NS3hel-RNA-ATP complexes respectively. Table 4.6 shows the percentage (%) and the cumulative percentage (Cumulative %) of motion represented by the first 10 eigenvectors of the two aMD simulations. Values are also plotted in Figure 4.17.
Table 4.6. Percentage and cumulative percentage of motion represented by the first 10 eigenvectors of the two aMD simulations.

<table>
<thead>
<tr>
<th>Eigenvector</th>
<th>NS3hel-RNA %</th>
<th>NS3hel-RNA %</th>
<th>NS3hel-RNA-ATP %</th>
<th>NS3hel-RNA-ATP %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.05</td>
<td>20.05</td>
<td>24.55</td>
<td>24.55</td>
</tr>
<tr>
<td>2</td>
<td>16.50</td>
<td>36.55</td>
<td>7.39</td>
<td>31.94</td>
</tr>
<tr>
<td>3</td>
<td>7.91</td>
<td>44.46</td>
<td>5.89</td>
<td>37.83</td>
</tr>
<tr>
<td>4</td>
<td>6.12</td>
<td>50.58</td>
<td>5.79</td>
<td>43.63</td>
</tr>
<tr>
<td>5</td>
<td>4.40</td>
<td>54.99</td>
<td>3.87</td>
<td>47.49</td>
</tr>
<tr>
<td>6</td>
<td>3.13</td>
<td>58.12</td>
<td>3.38</td>
<td>50.88</td>
</tr>
<tr>
<td>7</td>
<td>2.80</td>
<td>60.92</td>
<td>2.95</td>
<td>53.82</td>
</tr>
<tr>
<td>8</td>
<td>2.20</td>
<td>63.12</td>
<td>2.24</td>
<td>56.07</td>
</tr>
<tr>
<td>9</td>
<td>2.01</td>
<td>65.13</td>
<td>2.09</td>
<td>58.15</td>
</tr>
<tr>
<td>10</td>
<td>1.85</td>
<td>66.98</td>
<td>1.89</td>
<td>60.05</td>
</tr>
</tbody>
</table>

A)

B)

Figure 4.17. First 10 eigenvectors contribution of motion.

Data for these charts was taken from Table 4.6. In both panels the NS3hel-RNA system is shown in grey and the NS3hel-RNA-ATP one is in black. A) Contribution percentage contribution of eigenvectors. B) Cumulative percentage contribution of eigenvectors.

The projection on the trajectories of the ten eigenvectors showed that in both molecular systems the most interesting time-dependent variations upon time of the eigenvalues belonged to the first five PC (Table 4.7). As they also represented approximately 50% of the total motion in both simulations, they were further analysed.

The investigation of the atomic RMSF for first five modes, showed a general increase of values with eigenvector number increment. This confirmed that largest amplitude principal components describe the slower dynamical transitions. The first 5 modes of the NS3hel-RNA system had general lower RMSF values than the other complex, indicating faster atomic vibration in the NS3hel-RNA-ATP simulation.
Table 4.7. Time variation of the first 5 eigenvalues for the NS3hel-RNA and NS3hel-RNA-ATP simulations.

<table>
<thead>
<tr>
<th></th>
<th>NS3hel-RNA</th>
<th>NS3hel-RNA-ATP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC1</strong></td>
<td><img src="image" alt="PC1 Graph" /></td>
<td><img src="image" alt="PC1 Graph" /></td>
</tr>
<tr>
<td><strong>PC2</strong></td>
<td><img src="image" alt="PC2 Graph" /></td>
<td><img src="image" alt="PC2 Graph" /></td>
</tr>
<tr>
<td><strong>PC3</strong></td>
<td><img src="image" alt="PC3 Graph" /></td>
<td><img src="image" alt="PC3 Graph" /></td>
</tr>
<tr>
<td><strong>PC4</strong></td>
<td><img src="image" alt="PC4 Graph" /></td>
<td><img src="image" alt="PC4 Graph" /></td>
</tr>
<tr>
<td><strong>PC5</strong></td>
<td><img src="image" alt="PC5 Graph" /></td>
<td><img src="image" alt="PC5 Graph" /></td>
</tr>
</tbody>
</table>
In Table 4.8, two examples of per-atom RMSF charts for each simulation are reported, showing how values varied according to mode number and between the two molecular systems. Here RMSF values are reported for each atom selected for PCA. Atoms 168-618 correspond to protein residues 168-618 because only C\(^\alpha\) atoms were selected for the protein molecule. Some protein and RNA regions were associated with high RMSF values in both simulations. In particular, these regions were mainly solvent exposed random coils, the protease-helicase linker (residues 168-180), the D1 flexible portion discussed above (residues 241-257), the two ATP binding motifs (residues 317-330 and residues 462-470), D2 \(\alpha1'\) and \(\alpha2'\) helices, the D2 loop that follows \(\alpha3'\) helix (residues 410-416). Within the RNA, vibration could be associated mainly with the two strand extremities, in particular to U5. In all the modes obtained for the two aMD simulations, the movements involved mainly D2 and D3, indicating that these domains were generally more flexible than D1. RNA movements involved mainly a general backbone bent and U5 flexibility that prompts its displacement from the RNA binding site.

Table 4.8. Atom RMSF according to motion mode.

<table>
<thead>
<tr>
<th>NS3hel-RNA</th>
<th>NS3hel-RNA-ATP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC1</strong></td>
<td><strong>PC1</strong></td>
</tr>
<tr>
<td>RMSF</td>
<td>RMSF</td>
</tr>
<tr>
<td>Atoms</td>
<td>Atoms</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Differently from the cMD simulations, a “scissors-like” pattern could be observed in both simulations. The “scissors-like” pattern was mainly due to D2 lateral rotation towards D1, affecting the size of the ATP cleft and of the RNA binding site entrance, but not the distance from D3. Figure 4.18 shows the porcupine plots for the corresponding principal components that describe the “scissors-like” movement for the NS3hel-RNA (panel A) and the NS3hel-RNA-ATP (panel B) molecular systems.
D2 rotation involved also regions in contact with the nucleic acid and in particular the interaction between motif IV (on the α1’ helix, residues 363-366) and G2 phosphate moiety. In the NS3hel-RNA system, this pattern is described already in PC2, representing 16.50% of the total motion. Conversely, in the other molecular system, it was only described by PC4, representing only 5.79% of motion, making this vibration mode much less relevant. This
suggested that the presence of the natural substrate constrains the molecular system in the
closed conformation, while when the protein was unbound to ATP it was more free to explore
the conformational space. This was consistent with the translocation model proposed for HCV
NS3hel in which the binding of a NTP molecule prompts the protein to pass from an open to a
closed conformation. Furthermore, in the NS3hel-RNA-ATP simulation D2 rotation was not
only lateral, but the oscillation was also affecting the distance from D3 (as shown in panel B of
Figure 4.18), making the movement pattern not completely consistent with the HCV “scissors-
like” oscillation.

NS3pro-NS3hel linker and D1 residues 241-257 movements were present in all eigenvectors
and represented different components of their total motion, clearly visible also in trajectory
visual inspection. In the NS3hel-RNA simulation, the latter protein region shifted from its
initial position in the first 40ns, almost contemporarily to U5 displacement from the RNA
binding site, which was described in all five eigenvectors, with different extent. Differently, in
the NS3hel-RNA-ATP simulation, the displacement of this protein region started only in the
last part of the simulation, consistently with PC1 trend (Table 4.7). Of interest, also in this case
the U5 displacement from the RNA binding site occurred at the end of the simulation, in the
last 15ns, suggesting that the two movements could be connected. U5 fluctuations were
described mainly by the third and fifth eigenvector and had consequently low contribution to
the total motion (5.89% and 3.85% of motion respectively). Conversely, the RNA 5’ end
movement was described also by the first mode.

As predictable, flexibility of loops involving part of the ATP binding site and including parts of
mainly motif III (residues 317-330) and motif VI (residues 462-470) was more important for
the NS3hel-RNA complex and it was recorded already in the first eigenvector (PC1, 20.05% of
total motion). In this simulation the α3’ helix and the loop that follows it (residues 410-416)
showed some flexibility already in PC1 and was a more important component in principal
modes 4 and 5 (6.12% and 4.40% of total motion respectively). Here, these residues shift away
from RNA and protrude towards the ATP binding site, affecting its size. This protein sequence
is part of D2 motif V, responsible of RNA binding, through interactions between Thr408 and
A3 phosphate group and Asp409 and G2 2’OH moiety (see Figure 3.5 in the NS3 HELICASE ANS
A DRUG TARGET section). In this aMD simulation, an additional interaction between Ile 410
and A3 phosphate was observed. However, the displacement of residues 410-416 caused the
disruption of this interaction, as well as a weakening of that of Thr408. This shift was observed
also in the visual inspection of the trajectory and occurred mainly between 103ns and 125ns
of simulation, compatibly with the PCS eigenvalue time pattern around 100ns (Table 4.7). In
the NS3hel-RNA-ATP molecular system, the ATP binding motifs vibrations (residues 317-330
and residues 462-470) had a lower component in the total motion. From visual inspection of
the aMD trajectory the amplitude of these movements was also smaller and the presence of
the ligand did not allow changes in the cleft size. Importantly, also the shift of residues 410-416 was impeded by the presence of ATP. Thus, not only the fluctuation was less important than in the NS3hel-RNA molecular system simulation (only 3.87% of total motion), but the Ile410-A3 phosphate hydrogen bond was stabilised by the presence of ATP.

BINDING FREE ENERGY CALCULATIONS

The average interaction energies and their components were calculated for each aMD from 2000 frames saved from the two trajectories, following the same procedure applied for the cMD simulations. As reported in Table 4.9, the estimated binding free energy was not significantly different between the two molecular systems, with average values of -107.72 kcal/mol and -114.10 kcal/mol for the NS3hel-RNA and the NS3hel-RNA-ATP systems respectively.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Value (kcal/mol)</th>
<th>Std</th>
<th>Value (kcal/mol)</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta E_{\text{int}} )</td>
<td>(-1.7 \times 10^{-6})</td>
<td>8.59 ( \times 10^{-5})</td>
<td>1.65 ( \times 10^{-6})</td>
<td>8.72 ( \times 10^{-5})</td>
</tr>
<tr>
<td>( \Delta E_{\text{vdW}} )</td>
<td>-104.14</td>
<td>8.54</td>
<td>-100.22</td>
<td>7.86</td>
</tr>
<tr>
<td>( \Delta E_{\text{Coul}} )</td>
<td>-696.96</td>
<td>70.58</td>
<td>-775.80</td>
<td>58.70</td>
</tr>
<tr>
<td>( \Delta E_{\text{gas}} )</td>
<td>-801.12</td>
<td>70.61</td>
<td>-876.01</td>
<td>60.58</td>
</tr>
<tr>
<td>( \Delta G_{\text{solv}} )</td>
<td>699.40</td>
<td>57.11</td>
<td>761.91</td>
<td>51.03</td>
</tr>
<tr>
<td>( \Delta G_{\text{tot}} )</td>
<td>-101.72</td>
<td>21.28</td>
<td>-114.10</td>
<td>17.80</td>
</tr>
</tbody>
</table>

Average values for the aMD are reported, as well standard deviations. The molecular systems are indicated with NS3hel-RNA and NS3hel-RNA-ATP. \( E_{\text{int}} \): internal energy; \( E_{\text{vdW}} \): van der Waals energy; \( E_{\text{Coul}} \): Coulomb energy; \( \Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdW}} + \Delta E_{\text{Coul}} \); \( G_{\text{solv}} \): solvation energy is based on the Generalised Born method; \( \Delta G_{\text{tot}} \): total binding free energy calculated with MMGBSA; Std: standard deviation.

These averages were calculated over the whole trajectories of the aMD simulations. However, as already explained, the aim of the accelerated molecular dynamics method is to avoid that the molecular system gets trapped in energy minima wells for too long time and applies a boost potential when the energy of the system is below a certain threshold. In this way, the conformational space of the molecular system is explored better. This was confirmed for the helicase complexes by the results presented above. Consequently, also the snapshots used for the MMGBSA calculations belonging to different conformations and the averages calculated on the whole trajectories are not indicative of the free energy status of the systems and the time variation of the values were much more informative. Figure 4.19 shows the binding free energy profile for the two molecular systems.
Figure 4.19. Time variation of binding free energy values for aMD simulations.

Time variation of the NS3hel-RNA binding free energy is reported here for each molecular system. The values obtained for the NS3hel-RNA and NS3hel-RNA-ATP complexes are indicated with “NS3hel-RNA” (light blue line) and “NS3hel-RNA-ATP” (light red line) respectively. Averages are also reported as “Average NS3hel-RNA” (dark blue line) and “Average NS3hel-RNA-ATP” (dark red line). The average values were -101.72 kcal/mol and -114.10 kcal/mol for NS3hel-RNA and NS3hel-RNA-ATP systems respectively. Values ranged between -46.14 kcal/mol to -164.88 kcal/mol and -61.01 kcal/mol to -163.92 kcal/mol for the NS3hel-RNA and NS3hel-RNA-ATP complexes respectively, showing a higher binding energy fluctuation for the simulation performed without ATP. The variation patterns for the two aMD simulations do not resemble any of the first five PCA eigenvector time fluctuations (in Table 4.7), indicating that the NS3hel-RNA binding energy change does not depend on a single component of the protein total motion. Furthermore, no particular similarities with backbone and RNA RMSD variation could be observed, suggesting that binding energy changes were not only related to global protein or RNA motions.

Aiming to obtain a better insight into the causes of binding energy variations, the five nucleotides and residues in proximity of the RNA strand were selected for a more detailed binding free energy evaluation using the MMGBSA method, with the same procedure adopted for the cMD simulations. As for the total average binding energy, also in this case average values could not be indicative of the free energy status of the system. However, as they could give an idea about the most important residues upon RNA binding, they were calculated and are shown in Figure 4.20. The residues that gave the highest contribution to RNA binding were Arg225 and Arg387 in particular, but also Thr224, Thr244, Ser364, Ser386, Lys388, Thr408, Ile410 and Arg538. Other residues with lower, but still significant, energy contributions were: Pro245, Cys261, Ala263, Thr264, Phe288, Thr289 Asp290, Ile365, Lys366, Leu429, Pro431, Leu443 and Arg599. Not surprisingly, most of these residues corresponded to the interacting residues reported in the crystallographic data (see Figure 3.5 of the **NS3 HELICASE AS A DRUG TARGET** section) and they belong to domains 1 and 2. This complied with the fact that most
contacts that the enzyme has with the RNA are at the level of the sugar-phosphate backbone of the nucleic acid.\cite{16}

![Figure 4.20. Per residue binding free energy decomposition.](image)

The per residue binding free energy average values are reported here. Data was obtained by averaging the results for each molecular system. Results for NS3hel-RNA-ATP and NS3hel-RNA complexes are indicated as “NS3hel-RNA-ATP” and “NS3hel-RNA” respectively. Residues discussed in the text are indicated here.

Generally, the same per-residue pattern could be observed for the two simulations. Also in aMD simulations, residues involved in hydrogen bonds with the RNA 2’OH groups and considered to be the responsible for DENV NS3 helicase specificity for RNA (Pro223, Thr264, Pro363 and Asp409), had a small or even positive contribution to the total interaction energy, even if the expected interactions could be observed in the trajectories inspection.\cite{5,15}

However, this does not exclude the importance of these contacts in RNA preferentiality, as the rapid formation and disruption of these interactions was observed throughout the simulations, explaining the average value.

A few key and relevant differences between the molecular systems involved mainly Thr244, Phe288, Thr289, Asp290, Lys366, Lys388 and Ile410 protein residues. A big dissimilarity in binding energy averages was recorded for RNA nucleotides C4 and U5 in particular. In the latter case, binding energy decomposition average values were positive in both aMD simulations, in particular in the one with ATP. As it will be further discussed in the next chapter, this could be linked with an unfavourable binding to the protein when U5 is inside the RNA binding site that justified the displacement of the RNA 3’ end form the cleft.

The per-residue decomposition variations against simulation time were also computed. The most interesting patterns are discussed in the next chapter with the conformational changes that could explain them.
SMALL CONFORMATIONAL CHANGES IN AMD SIMULATIONS

The per-residue decomposition average and time variations of the binding free energy, together with the visual inspection of the trajectories highlighted small conformational changes that could be relevant in explaining the helicase mechanism of translocation. Similarly to that observed in the cMD simulations, from trajectory visualisation the 3’ and 5’ end of the RNA strand were more flexible than the other nucleotides. The central part of the nucleic acid (A3 and C4 in particular) appeared to be bound to the protein in a more stable manner, explaining their lower per-residue binding free energy values. Consistently, a bent in the RNA strand could also be observed in both trajectories, but it was more evident in the simulation without ATP and occurred only at the end of the simulation in the other complex. At the 5’ end of the RNA strand, the nucleotide flexibility involved mainly A1 and G2. In both aMD simulations, the G2 phosphate moved away from D2, causing a rotation of A1 nucleotide and a shift in G2 base. The latter was particularly evident in the NS3hel-RNA complex, in relation with the more evident RNA bent. In this simulation, when the G2 base shifted, it interacted with Phe288, Thr289 and Asp290, explaining the binding free energy difference between the two molecular systems for these residues. In analogy with what observed in the cMD simulations, the A1 rotation generally did not involve a displacement of this part of nucleotide from the occupied area of the binding site, but caused the shift of the ribose portion, placing the 5’ OH group in a more exposed region of the RNA binding site entrance. Importantly, this change in conformation would facilitate the entrance of a novel nucleotide in the RNA cleft, as would happen in the 3’-5’ helicase translocation along the nucleic acid. The A1 base was constrained in its position by the D2 β-hairpin through hydrophobic contacts mainly between A1 base and residues Leu429, Pro431 and Leu443. These interactions were conserved in both aMD simulations and the interacting energy contribution pattern of these residues was constant during the simulation. This supported the hypothesis that these hydrophobic residues play a key role in impeding NS3hel to translocate towards RNA 3’ end by maintaining in place the base of the nucleotide in the 5’ end of the binding site. Interestingly, during a small time frame of a few nanoseconds in both simulations, A1 base shifted forward, towards the binding cleft entrance. However, the base was unable to exit the binding site, blocked by the β-hairpin, giving a further support in favour of the hypothesis. As well as Ile365 backbone amino group, which showed to bind G2 phosphate also in the crystallographic structures (see Figure 3.5 of the NS3 HELICASE AS A DRUG TARGET section), in the aMD also Ser364 and Lys366 side chains interacted with the group during the simulations. The G2 phosphate displacement was accompanied by a shift of the α1’ helix, already discussed in the PCA results, that moved the interacting residues in the opposite direction, causing a disruption on the phosphate-protein hydrogen bonds. For example, the distance
between Ile365 backbone nitrogen and the interacting phosphate oxygen oscillated between 2.70Å and 9.62Å in the NS3hel-RNA simulation and between 2.63Å and 11.06Å in the other one (Figure 4.21, panel A). A distortion of the loop before the α1' helix, composed of residues from Pro363 to Lys366, could also be observed. This was caused by the rotation of Ile365 side chain into a hydrophobic region of D2 given by the conserved residues Gln384, Thr389, Val406 and Val362. The Ile365 change in conformation had also been observed in the cMD simulations, but in the aMD ones it did not impede the interaction of this residue with RNA. The distortion of the Pro363-Lys366 loop was important as Ser364 interactions with the RNA occurred only when the Ile365 side chain was interacting with the hydrophobic groove. Furthermore, this also exposed the Ile365 backbone amino group more to the solvent, making it hypothetically available for the interaction with the phosphate moiety of an incoming nucleotide. An example of G2 phosphate shift and of Pro363-Lys366 loop distortion with Ile365 interaction with the hydrophobic area is shown in panel B of Figure 4.21, where two frames of the NS3hel-RNA simulation are superposed. As it occurs at the same time of the phosphate shift, in this example the conformational change disrupts the interactions between Ile365 and the RNA and no additional interactions are present. However, it can be observed here that the loop distortion allows Ser364 to get closer to G2 phosphate.

The positions of the residues and of the phosphate moiety oscillated between the bonded and the unbounded conformations during both aMD simulations, indicating that these interactions are not stable. The oscillation of the G2 phosphate position is probably also connected to the presence of Asp603 in the RNA binding site entrance (not displayed in Figure 4.21). When the G2 phosphate shifted away from D2, it got closer to the negatively charged Asp603, creating an electrostatic repulsion. This repulsion could explain the positive value of this residue binding free energy contribution as well as the oscillating pattern in the phosphate position. In the NS3hel-RNA-ATP complex, the bound conformations were more important than in the other molecular system, as shown by the distance pattern in panel A of Figure 4.21. Furthermore, the bound-unbound oscillations were less frequent and slower, allowing the observation of clear bound-unbound patterns in the per-residue binding energy decomposition of the involved residues, in particular Ser2364 and Ile365. The interactions of Lys366 side chain with G2 phosphate were more frequent in the NS3hel-RNA simulation, suggesting that the residue flexibility plays an important role in its interactions with RNA.
Figure 4.21. Movement pattern around RNA 5’ end.

A) Time variation of the distance between Ile365 N and the interacting G2 phosphate O atoms. Data for NS3hel-RNA and NS3hel-RNA-ATP simulations is presented in blue and in red respectively. B) Example of G2 phosphate shift and Ile365 side chain change in conformation. The shift of G2 phosphate and Ile365 side chain are represented here by superposition of two frames of the trajectory the NS3hel-RNA simulation. The ribbons and the carbon atoms of the frame at the beginning of the trajectory are in orange, while the ones of the frame taken from around the middle of the simulation are in green. The shifts are indicated by arrows. The hydrogen bond present at the beginning of the simulation is represented as a black dotted line. Discussed residues are indicated.

Similarly to the G2 phosphate, also for that of A3 more interactions were observed during the aMD simulations than in the crystal structure (see Figure 3.5 of the NS3 HELICASE AS A DRUG TARGET section). Consistently with the good binding free energy contribution averages (in Figure 4.20), as well as Arg387, also Ser386, Lys388 side chains and the backbone amino group of Ile410 interacted with this moiety through hydrogen bonds and electrostatic interactions (for the basic residues). During aMD visual inspection, relevant changes in Arg387 conformations could be observed, as shown in Figure 4.22, where snapshots from the NS3hel-RNA trajectory and the NS3hel-RNA-ATP one were superposed in panels A and B respectively. During the simulations, the residue’s side chain turned, breaking the interactions with A3 and G2 phosphates and forming new unstable hydrogen bonds with C4 and U5 phosphates and bases. The side chain rotation was not permanent in either of the simulations, but it was more stable in the simulation of the NS3hel-RNA system. The observed flexibility of Arg387 could be relevant for helicase translocation mechanism as it shows the ability of this residue to move along the RNA, facilitating the helicase shift along the nucleic acid.
Section 4: NS3HEL MECHANISM OF TRANSLOCATION ALONG RNA

4.2 RESULTS AND DISCUSSION

Figure 4.22. Arg387 change in conformation.
The ribbons and the carbon atoms of the frame at the beginning of the trajectory are in orange, while the ones of the frame taken later on in the simulation are in green. The shift is indicated by a black arrow. The hydrogen bonds formed between Arg387 and the RNA phosphates are represented as black dotted lines. RNA nucleotides are indicated. A) Shift in NS3hel-RNA simulation. The shift of Arg387 side chain is represented here by superposition of two frames of the NS3hel-RNA simulation. B) Shift in NS3hel-RNA-ATP simulation. The shift of Arg387 side chain is represented here by superposition of two frames of the NS3hel-RNA-ATP.

As already discussed in the PCA results, in NS3hel-RNA the shift of residues 410-416 is more important than in the other molecular system as the loop is allowed to occupy the ATP binding site by the absence of the ligand. Some fluctuations in the loop conformations occurred also in the presence of ATP, but they were smaller, as shown in panel A of Figure 4.23.

Figure 4.23. Ile410 change in conformation.
A) Time variation of the distance between Ile410 N and the interacting A3 phosphate O atoms. Data for NS3hel-RNA and NS3hel-RNA-ATP simulations is presented in blue and in red respectively. B) Example of G2 phosphate shift and Ile365 side chain change in conformation. The maximum shifts of Ile410 and Thr408 residues and of RNA backbone are represented here by superposition of two frames of the NS3hel-RNA simulation. The ribbons and the carbon atoms of the frame at the beginning of the trajectory are in orange, while the ones of the frame taken at around 100ns of the simulation are in green. The shifts are indicated by black arrows. The hydrogen bonds formed between A3 phosphate and Ile410 and Thr408 are represented as black dotted lines. Discussed residues and RNA nucleotides are indicated.
Here, the distance between Ile410 backbone nitrogen and the interacting A3 phosphate oxygen is plotted against time, with values ranging between 2.68Å and 11.08Å in the NS3hel-RNA simulation and between 2.68Å and 6.54Å in the NS3hel-RNA-ATP one. After approximately 80ns of the NS3hel-RNA simulation, residues 410-416 shifted away from RNA, which increased its curvature at the same time. This produced a permanent break of Ile410 interactions with the phosphate and a consequent increase of free energy contribution of this residue to nucleic acid binding, explaining the difference in Ile410 average contribution to the total binding energy between the two molecular systems. The effect on Thr408 contribution was less pronounced as the position of this residue changed less, leading to a weakening of the hydrogen bonds with the phosphate, but not to their complete disruption. The described residues’ changes in conformation are shown in panel B of Figure 4.23, where two frames of the NS3hel-RNA simulation are superposed.

Similarly to cMD simulations, the 3’ end of the nucleic acid strand was very flexible, prompting a permanent U5 displacement from the RNA binding site in all the aMD simulations. The fact that the 3’ end nucleotide, differently to the one in the 5’ end, was flexible and free to exit the RNA cleft, supports the hypothesis that the β-hairpin residues (Leu429, Pro431 and Leu443) have an important role in the 3’-5’ directionality of the NS3hel translocation mechanism. From visual inspection of both aMD trajectories, the U5 base displacement from the RNA binding site was almost simultaneous with the D1 residues 241-257 large scale change in conformation described previously. These two changes in conformations were not stable nor specific as they changed within and between the aMD simulations. As already mentioned, the U5 displacement and D1 residues 241-257 change in conformation started after approximately 40ns and 170ns in the NS3hel-RNA and NS3hel-RNA-ATP simulations respectively. Not surprisingly, the U5 displacement from the RNA cleft caused a general change in the nucleotide conformation. In particular, the shift of the U5 phosphate position caused an increase of the Thr244 binding free energy value in relation with the disruption of the residue’s interaction with this moiety. Nevertheless, when out of the RNA cleft, U5 formed new interactions with D1, mainly, and D3, which enhanced the binding energy component of this nucleotide. These interactions involved residues like Pro245, Thr267 and Arg538 and were neither stable nor specific, but prompted a better contribution of this nucleotide to the total binding free energy. Even residues that in the crystal structure are far from the RNA binding site like Tyr242 or Gln243, interacted with U5 phosphate during the nucleotide displacement, aided by the D1 241-257 region change in conformation. Even if these interactions were only transient and different between the simulations, they suggested that the 241-257 region could be involved in the 3’ end displacement from the RNA cleft though its transitory stabilisation. The U5 displacement occurred much later in the simulation for the NS3hel-RNA-ATP system respect to the NS3hel-RNA one, explaining the much more positive value of U5.
average binding energy and the more negative Thr244 one. An example of the U5 displacement from the RNA binding site is presented in Figure 4.24.

![Figure 4.24](image)

To exemplify the non-specific flexibility of U5 nucleotide and of D1 241-257 residues three frames from the NS3hel-RNA trajectory are superposed here. The ribbon and the atoms from the three snapshots are coloured in orange, green and blue. For an easier visualisation RNA nucleotides from A1 to C4 are in a lighter shade of the correspondent colour for the snapshot. U5 and D1 241-257 sequence are indicated with black arrows.

**PROPOSED MECHANISM OF TRANSLOCATION ALONG THE RNA**

The results obtained from the accelerated molecular dynamics approach applied to the helicase-RNA complex in the presence and in the absence of ATP, revealed important novel insights in the DENV NS3hel mechanism of translocation. As presented earlier in the chapter, HCV NS3hel crystal structures showed that the enzyme translocates along the nucleic acid with a ratchet-like mechanism that involve the NTP binding site opening and closing through D2 lateral rotation. In the absence of the natural ligand, the NTP binding site is open and 5 nucleotides can be accommodated in the RNA cleft. Upon NTP binding, D2 rotates towards domain 1 (D1), closing the NTP cleft and reducing the RNA pocket size. In this stage, changes in D1 interactions with the nucleic acid strand causes the exit of the 3’ nucleotide from the binding pocket, with only 4 nucleotides remaining in the cleft. NTP hydrolysis products are then released through the rotation of D2 away from D1. During NTP cleft opening, D2 interactions with RNA shift and allow the entrance in the nucleic acid pocket of a novel nucleotide at the 5’ end of the RNA, bringing back the number of bound nucleotides to 5. At the end of the cycle, HCV NS3hel has thus shifted by one nucleotide towards the 5’ end of the viral genome.\(^{[11-13]}\)

The starting system structures (PDB IDs 2JLU and 2JLV) superposition RMSD is only 0.38Å, suggesting that the starting protein conformations were the same for the two molecular
systems, in line with the fact that only observation of the crystal structures could not explain the DENV NS3hel mechanism of translocation.\textsuperscript{[15]} Since large and small scale changes in conformations were generally more evident in the molecular system without ATP, it was hypothesised that the crystal structures corresponded to DENV NS3hel closed conformation, with the natural ligand bound to the NTP pocket. In HCV NS3hel, in this state the protein only binds four nucleotides and not 5 as shown by DENV NS3hel crystal structures. According to the hypothesis that the crystal structures represent the closed conformation, as the protein opens and a novel nucleotide enters the RNA site, six nucleotides should bind to the helicase. However, as U5 displacement from the RNA binding site was observed in both amD simulations in different measure, no support for six binding nucleotides in the open conformation was obtained. Thus, a clear hypothesis on the number of nucleotides bound to the DENV protein in the open and in the closed conformations is difficult to formulate.

RMSD patterns and principal component analysis of the amD simulations were consistent with crystal structures B-factors, confirming the reliability of the observed changes in conformations. The large scale system vibration involved mainly domain 2, suggesting that the key role of the flexibility of this domain in the protein function. A D2 rotation, similar to what reported for HCV, could be described with PCA, in particular for the system without ATP, suggesting that also DENV NS3hel could translocate with a ratchet-like mechanism, driven by NTP binding thatfavours the protein closed conformation.

According to the ratchet-like model, the molecular system without ATP would pass to the open conformation through D2 rotation and changes in D2-RNA interactions. Most of the changes in contacts described in the previous chapter involved indeed D2-RNA interactions. Additionally, contacts variations were more important in the NS3hel-RNA complex, supporting the hypothesis that the crystal structures represent the closed conformation and that the molecular system without ATP tends to pass to the open conformation. In particular, a significantly bigger shift of Ile410 was only possible when the NTP binding site was not bound to the ligand. Furthermore, Ser364-Lys388 sequence interactions with the RNA were less stable in the closed conformation complex without ATP. All of these transient losses of contact would allow the shift of D2 over the RNA during enzyme opening. Moreover, the higher flexibility of Arg387 in the NS3hel-RNA system could be relevant for the helicase translocation mechanism as it shows the ability of this residue to move along the RNA, facilitating the helicase shift along the nucleic acid.

Some small scale residue changes in conformation observed in the amD simulations and confirmed in the free energy time fluctuations, also gave important insights about DENV NS3 helicase 3’-5’ translocation directionality. Both RNA ends were very flexible as they were less involved in stacking interactions than the other nucleotides. Not constrained by strong interactions with the protein, the base at the 3’ end was able to exit from the original site,
breaking the π-π stacking with the neighbour C4 and allowing more space in the RNA cleft. On the other end, A1 base showed that its hydrophobic contact with the conserved β-hairpin residues Leu429, Pro431 and Leu443 is constant and stable throughout the simulation and sufficiently strong to maintain the base in the same site, impeding the helicase to translocate in the 5'-3' direction along the RNA. It was therefore hypothesized that these hydrophobic residues are crucial for DENV NS3hel translocation directionality.

Some indications on the mechanism of insertion of a new nucleotide at the 5' end were given by the motion patterns observed in the RNA and in the protein at the level of the binding site entrance region. The fluctuations in interaction between the region around Ile365 and G2 phosphate, were linked to the opposite movement of the RNA and the α1' helix and to the Ile365 side chain turn towards a D2 hydrophobic cleft formed by the conserved residues GLn384, Thr389, Val406 and Val362. The distortion of the Pro363-Lys366 loop exposed more the Ile365 backbone amino group to the solvent, making it hypothetically available for the interaction with the phosphate moiety of an incoming nucleotide. The change of hydrogen bonds for Ile365 could be aided by the observed interaction between Ser364 and G2 phosphate as it only occurred when the Ile365 side chain is facing the hydrophobic cleft. The observed importance of Ile365 side chain hydrophobic interactions was also supported by mutational studies that showed that the replacement of this residue with an alanine impaired the unwinding activity of DENV NS3hel without affecting the ATPase activity.\textsuperscript{[45,46]} The movement of RNA away from D2 caused a bend in the RNA that placed the 5’ OH group of A1 in front of the RNA binding site entrance. This conformation would facilitate the entry of a novel nucleotide in the RNA cleft, supporting the importance of the described small changes in conformation in the helicase 3'-5' translocation along the RNA.

Summarising, it is here proposed that DENV NS3hel translocation along the RNA depends on both large and small scale changes in conformation. According to this hypothesis, D2 rotation is linked to NTP binding and produces a ratchet-like movement that resembles HCV mechanism of translocation aided by transient changes in residue contacts with RNA. The 3'-5' directionality of the protein movement along the nucleic acid is mainly depending on the presence of hydrophobic residues on domain 2 β-hairpin that impede the movement in the opposite direction, just like a pawl in a ratchet. Additionally, the instability of G2 contacts with D2 and the interactions of Ile365 side chain with a conserved hydrophobic site in domain 2 could be important factors in the mechanism of entry of a novel nucleotide at the nucleic acid 5’ end in the RNA binding site.
4.3 CONCLUSION

In this study two different molecular systems were analysed with molecular dynamics simulations with the aim of investigating the behaviour of DENV NS3hel during its translocation along the RNA strand, still elusive to date.\textsuperscript{15} Classical and accelerated MD simulations have been performed for the NS3hel-RNA complex with and without ATP and the trajectories were analysed and used for binding free energy estimation with the MMGBSA methods.

The aMD method was more successful in producing relevant results for the scope of the study, namely to understand the mechanism of NS3hel translocation along the RNA strand. In these simulations the complete transition from a closed to an open conformation of the protein has not been obtained, probably because a high energy barrier must be overcome and the applied boost potential to achieve this was not sufficient. Thus, it might be informative to repeat the \textit{in silico} experiment with a higher boost and compare the results with the ones discussed here. Furthermore, it would also be revealing to extend the RNA strand by adding nucleotides at both RNA ends and evaluate the differences in the conformational changes discussed here.

Nevertheless, the results obtained with this study represent a substantial increase in the understanding of this enzyme mechanism of action and it is believed that for the first time a hypothesis on DENV NS3hel translocation along the RNA has been formulated. Results indicated that both large scale domain movements and small residue conformation changes are important and that the protein translocates along the viral genome in a NTP-driven “ratchet-like” manner, similarly to HCV.\textsuperscript{12,13} Residues that are particularly important in the translocation mechanism and directionality were identified and could be used for future drug design approaches. For example, a conserved hydrophobic region of D2 formed by Val362, Gln384, Thr389 and Val406 could be used as a target site for a structure-based drug design approach, since it unexpectedly showed to be important and could be used for the design of novel inhibitors of DENV NS3hel.
4.4 METHODS

The computational methods for all this study are presented here. A description of the molecular modelling programs used can be found in APPENDIX 2.

HARDWARE DETAILS

The GROMACS 40ns cMD simulations were performed on a 32 core computer with Inter Xeon 2.20 GHz E5-4620 CPUs, while the AMBER14 MD simulations were carried out with a Tesla K20c GPU on a 32 core computer with Inter Xeon 2.20 GHz E5-4620 CPUs. Trajectory analysis calculations were performed on an 8 core computer with Inter Xeon 2.80 GHz E5462 CPUs.

SYSTEM PREPARATION

The selected DENV NS3hel structures (PDB IDs: 2JLU and 2JLV for NS3hel-RNA and NS3hel-RNA-ATP systems respectively) were downloaded from the Protein Data Bank site. The systems were prepared removing water molecules, non-catalytic ions and glycerol whenever these elements were present, using MOE 2010.10. In PDB 2JLV AMPPNP molecule co-crystallised with the enzyme was modified to ATP with MOE Builder. The same tool was also used for the deletion of incomplete RNA nucleotides in the original crystal structures. The missing side chain of Lys388 in 2JLU structure was built with MOE Mutate tool and then energy minimised with AMBER99 force field and 0.001 gradient, using the same modelling program.

CLASSICAL MOLECULAR DYNAMICS SIMULATION

For each of the molecular systems, the first set of classical MD (cMD) simulations were computed with the GROMACS 4.5.3 program, using a modified version of the implemented AMBER99 force field. The modification regarded the insertion of missing ATP topology parameters. Polyphosphate parameters to be used with AMBER force field have been calculated at the University of Michigan using molecular orbital calculations at the RHF/6-31+G* level. The same research group also reported that the application of these
parameters to ATP were suitable and gave good agreement with CHARMM. Consequently, the polyphosphate parameters were used for ATP topology building for the MD simulations reported in this work, according to the methodology used by R. Bryce group. In all the simulations the system was automatically protonated according to the selected force field during topology generation. The simulation was conducted in PBC conditions, using a cubic box with 0.9nm minimum distance between the molecular system and the box. Explicit water molecules described with the TIP3P model were added and the systems were neutralised with the addition of monovalent ions: 3 Na⁺ and 5 Na⁺ in the NS3hel-RNA and NS3hel-RNA-ATP system respectively.

Two consecutive energy minimisations were initially performed, employing the steepest descendent (SD) method first and then the conjugate gradient (CG) method after, with the aim of a faster process and more accurate result. A force tolerance of 100 kJ mol⁻¹ nm⁻¹ was set for SD, while a 10 kJ mol⁻¹ nm⁻¹ force tolerance was applied with CG. In both energy minimisations the maximum number of iterations was set to 3,000 steps. Subsequently, a position restrain force of 1,000 kJ mol⁻¹ nm⁻² was applied to protein and ligand atoms for water molecules relaxation. Aiming for a smoother equilibration, two consecutive 50 ps (50,000 steps, step size of 1 fs) position restrained MDs were performed, saving coordinates, velocity and energy values every 500 steps. NVT conditions (constant number of atoms N, volume V and temperature T) were used for the first position restrained MD and a v-rescale temperature coupling and 0.1 ps time constant was used to heat the system to 300 K. NPT conditions (constant number of atoms N, pressure P and temperature T) were used for the second position restrained MD, with the use of both temperature (v-rescale, temperature 300 K and time constant of 0.1 ps) and pressure (Berendsen algorithm, 1 bar pressure and time constant of 0.5 ps) coupling. A 40 ns (20000000 steps with 2 fs step size) production simulation with leap-frog algorithm was run in triplicates for each molecular system in the same NPT conditions described above. Coordinates, velocities and energy values were saved every 1500 steps. In all simulation stages, long-range electrostatic interactions were calculated with the Particle-Mesh-Ewald (PME) method with a 0.9 nm short range cut-off and short-range non-bonded interactions were computed only within a cut-off of 1.4 nm. Trajectories were visually inspected with VMD 1.9.1, analysed with GROMACS 4.5.3 tools and Grace 5.1.22 and used for principal component analysis (PCA) and binding free energy estimation.
ACCELERATED MOLECULAR DYNAMICS SIMULATIONS

For the application of the accelerated MD (aMD) method, calculations were performed with the AMBER 14\textsuperscript{[54]} program, using the AMBER12SB force field. The molecular systems were automatically protonated according to the selected force field during topology generation. The simulation was conducted in PBC conditions, using a cubic box with 1.0nm minimum distance between the molecular system and the box. Explicit water molecules described with the TIP4P-Ew model were added and the systems were neutralised with the addition of monovalent ions: 3 Na$^+$ and 5 Na$^+$ in the NS3hel-RNA and NS3hel-RNA-ATP system respectively.

Two consecutive energy minimisations were initially performed, one restraining protein and ligand atoms with a 10 kcal mol$^{-1}$ Å$^{-1}$ force and one with no restraints. Both of them were performed for 6000 steps, employing the steepest descendent (SD) method in the first 3000 steps first and the conjugate gradient (CG) method for the last 3000 steps, with the aim of obtaining a faster process and more accurate result. The systems were subsequently equilibrated in three steps. First, a 1ns (500,000 steps, step size of 2 fs) NVT (constant number of atoms N, volume V and temperature T) equilibration was computed, restraining the main chain atoms with a force of 10 kcal mol$^{-1}$ Å$^{-1}$. Here, temperature was linearly raised from 0 to 300K, using Langevin dynamics and collision frequency of 2 ps$^{-1}$. Secondly, a 2ns (1,000,000 steps, step size of 2 fs) NPT (constant number of atoms N, pressure P and temperature T) equilibration was computed, restraining the main chain atoms with a force of 10 kcal mol$^{-1}$ Å$^{-1}$. Here, temperature was maintained at 300K, using Langevin dynamics and collision frequency of 2 ps$^{-1}$ and pressure was set to 1atm by applying the Berendsen algorithm coupling with a relaxation time of 2 ps. Thirdly, a 3ns (1,500,000 steps, step size of 2 fs) NPT (constant number of atoms N, pressure P and temperature T) equilibration was computed, with no restraints and the same temperature and pressure coupling of the second equilibration. In all equilibration stages, long-range electrostatic interactions were calculated with the Particle-Mesh-Ewald (PME) method with a 1.0 nm cut-off and a SHAKE bond constraint was applied.

Prior to the aMD simulation a 40ns (20,000,000 steps with 2 fs step size) cMD with the leapfrog algorithm was computed for both the NS3hel-RNA and NS3hel-RNA-ATP molecular systems with the same NPT conditions described for the last two equilibration stages.

The aMD method was performed on the equilibrated systems and a dual boost (for dihedral and potential energies) was applied. The threshold energies ($E_{\text{dihed}}$ and $E_{\text{pot}}$, see Equation 4.2) and tuning parameters ($\alpha_{\text{dihed}}$ and $\alpha_{\text{pot}}$, see Equation 4.2) were computed using the cMD results in the following Equations:
\[ \alpha_{\text{dihed}} = \frac{[4 \text{ kcal mol}^{-1} \text{res}^{-1} \times (\text{RES})]}{5} \]

*Equation 4.3. Dihedral tuning parameter.*\(^{[40]}\)

\( \alpha_{\text{dihed}} \): tuning parameter for dihedral potential boost; RES: number of residues in the large molecule.

\[ E_{\text{dihed}} = \text{DIHED} + 5 \times \alpha_{\text{dihed}} \]

*Equation 4.4. Dihedral threshold energy.*\(^{[40]}\)

\( E_{\text{dihed}} \): threshold energy for dihedral potential boost; DIHED: average dihedral energy obtained from cMD simulations.

\[ \alpha_{\text{pot}} = 0.16 \text{ kcal mol}^{-1} \text{atoms}^{-1} \times (\text{ATOMS}) \]

*Equation 4.5. Potential energy tuning parameter.*\(^{[40]}\)

\( \alpha_{\text{pot}} \): tuning parameter for potential energy boost; ATOMS: number of atoms in the molecular system, including water molecules.

\[ E_{\text{pot}} = \text{POT} + \alpha_{\text{pot}} \]

*Equation 4.6. Dihedral threshold energy.*\(^{[40]}\)

\( E_{\text{pot}} \): threshold energy for potential energy boost; POT: average potential energy obtained from cMD simulations.

The values of the number of residues of the large molecules (RES), the number of atoms in the systems (ATOMS), the average dihedral energy (DIHED) and the average potential energy (POT) used for a MD set up are reported in Table 4.5.

A 200ns (100,000,000 steps with 2 fs step size) aMD with the leap-frog algorithm was consequently computed for both the NS3hel-RNA and NS3hel-RNA-ATP molecular systems with the same NPT conditions described for the last two equilibration stages.

In all AMBER cMD and aMD simulations, long-range electrostatic interactions were calculated with the Particle-Mesh-Ewald (PME) method with a 1.0 nm cut-off and a SHAKE bond constraint was applied.

aMD trajectories were visually inspected with VMD 1.9.1\(^{[52]}\), analysed with AMBER 14\(^{[54]}\) tools and Grace 5.1.22\(^{[53]}\) and used for principal component analysis (PCA) and binding free energy estimation.

**PRINCIPAL COMPONENT ANALYSIS**

Principal component analysis (PCA) was computed on trajectories obtained with GROMACS cMD or AMBER aMD. In both cases, the overall translational and rotational motions of the system were eliminated by centring the system and by a least square fit to the backbone of the protein. The covariance matrix \( C_{ij} \) of the analysed atomic coordinates was generated according to Equation 4.7. Here, \( i \) and \( j \) are two atoms and represent all the possible \( ij \) atom pairs of Cartesian coordinates \( x \). The <> brackets indicate average values. \( C_{ij} \) is a symmetric \( 3N \times 3N \) matrix, where \( N \) is the number of atoms considered. The matrix can be diagonalised with an orthonormal transformation matrix \( R \), whose columns correspond to the eigenvectors or principal or essential modes. Eigenvectors represent the vectorial description of each
component of the motion, while eigenvalues (the diagonalisation results) correspond to the contribution of that particular mode to the total motion of the system.

\[ C_{ij} = \langle (x_i - \langle x_i \rangle) \cdot (x_j - \langle x_j \rangle) \rangle \]

*Equation 4.7. Covariance matrix.*[55]

\( C_{ij} \): covariance matrix of all possible \( ij \) atom pairs; \( i,j=\) atoms; \( <>\): average values.

For the cMD simulations performed with GROMACS 4.5.3,[49] PCA was performed with the same program on \( C^\alpha \) atoms of the protein and all atoms of RNA. In order to investigate the main system fluctuations, the first 10 principal models were analysed for each cMD simulation and for each of these eigenvectors, the filtered trajectory was visually inspected and a porcupine plot was built.

For the aMD simulations performed with AMBER 14,[54] PCA was performed with the same program on \( C^\alpha \) atoms of the protein and RNA heavy, when present in the simulation. In order to investigate the main system fluctuations, the first 10 principal models were analysed for each aMD simulation. The first 5 eigenvectors were further investigated through trajectory projection, atom RMSF calculation, visual inspection of the filtered trajectory and porcupine plot design.

**BINDING FREE ENERGY CALCULATIONS**

The trajectories of cMD and aMD simulations were used for the calculations of the binding free energy of RNA to DENV NS3 helicase. The system coordinates, excluding water molecules and non-catalytic ions, were saved with VMD 1.9.1.[52] A total of 1906 and 2000 frames were saved for each cMD and each aMD trajectory respectively. The binding free energy (see Equations A2.12 and A2.13 of APPENDIX 2) of the nucleic acid to the helicase was calculated with the MMPBSA program of the AMBER suite through a Generalized Born Surface Area (MMGBSA) approach computed with the inbuilt MMPBSA program. AMBER 12[56] and AMBER 14[54] versions of the program were used for the cMD and aMD simulations respectively. For polar solvation contribution calculations a salt concentration of 0.100M was chosen and the dielectric constants for the solvent and for the solute were set to 80 and 1 respectively. For nonpolar solvation contributions the surface tension and the offset constant were set to 0.0072 kcal mol\(^{-1}\) Å\(^2\) and -1.008 kcal mol\(^{-1}\) and the SASA was computed using a water probe with 1.4 Å radius.

Binding and free energies were calculated for each trajectory frame saved and plotted in spread sheet charts, where averages and standard deviations were calculated. The entropic
term of the free energies was not calculated, as the objective of this work was to compare binding energies rather than calculating their absolute values. For a deeper analysis of the interactions between RNA and the DENV NS3 helicase residues of interest, further MMGBSA calculations with binding free energy decomposition to each residue were computed with the same variables values described above.
4.5 BIBLIOGRAPHY


[12]: M. Gu, C. M. Rice, Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism, *PNAS* 107 (2010), p. 521-528, DOI: 10.1073/PNAS.0913380107


[29]: S.-F. Chen, S. Cao, S. Han, J.-Z. Chen, Insight into the structural mechanism for PKBα allosteric inhibition by molecular dynamics simulations and free energy calculations, *Journal of Molecular
Section 4: NS3HEL MECHANISM OF TRANSLOCATION ALONG RNA

4.5 BIBLIOGRAPHY


[44]: C. A. F. de Oliveira, B. J. Grant, M. Zhou, J. A. McCammon, Large-Scale Conformational Changes of Trypanosoma cruzi Proline Racemase Predicted by Accelerated Molecular Dynamics Simulation, *PLoS


[47]: Protein Data Bank, [www.pdb.org](http://www.pdb.org) (Accessed, October 2014)


Section 5: NS5 METHYLTRANSFERASE AS A DRUG TARGET
5.1 INTRODUCTION

A PROMISING DRUG TARGET

Viral RNA plays a fundamental role in the replication cycle because it is necessary for the synthesis of the viral proteins and of a new genome for the nascent virions.\textsuperscript{1,2} In order to be able to be recognised by host cell ribosomes, the RNA must present a chemical modification at the 5' end called cap. The same structure is also fundamental for increasing the RNA stability itself as it impedes the action of several cellular 5' exonucleases that are responsible for nucleic acid degradation. The cap structure is given by a guanosine nucleoside that is linked to the 5' end RNA nucleoside (usually adenosine) through a 5'\textsuperscript{-}5' triphosphate bridge. This structure can then be methylated, giving different types of caps, depending on the number of methylations as shown in Figure 5.1. A cap 0 is methylated in the guanine N-7 position; a cap 1 is methylated in the guanine N-7 position and in the 5' end RNA nucleoside 2' hydroxyl group; while cap 2 is methylated as a cap 1 with an additional methyl group on the 2' hydroxyl group of the second nucleotide of the 5' end of the RNA.\textsuperscript{3}

DENV ribonucleic acid is modified with a cap 1 structure that is usually indicated in the following way: $\text{7MeGpppA}_{2\text{OMe}}\text{RNA}$.\textsuperscript{4,5} It has been shown that the methylation in the N-7 position is fundamental for an efficient RNA translation, making it an essential feature for viral

![Figure 5.1. General cap structure.](image)

The general structure of a RNA cap is reported here. The guanosine is linked to the 5' end nucleoside through a 5'\textsuperscript{-}5' triphosphate bridge. Different methylation states give different cap types. Cap 0 (green arrows) is methylated once, cap 1 (orange arrows) is methylated twice and cap 2 (light blue arrows) is methylated three times. Added methyl groups are highlighted with circles and the 5' and 3' end of the RNA are indicated.
replication. Conversely, the lack of the 2′-O methylation does not stop viral replication, but
the viruses formed are highly attenuated because this methylation is critical for avoiding
cellular recognition that the viral RNA as non-self, that would lead to immune system
activation.[6]
Also eukaryotic mRNA is capped co-transcriptionally in the cell nucleus, but as dengue virus
replication occurs in the cell cytoplasm, there are no possibilities for the virus to exploit the
host capping proteins and therefore it has evolved its own capping machinery.[3,4] In general,
capping occurs in four stages. Prior to capping, the nascent RNA strand has a triphosphate A at
the 5′ end (pppA-RNA). This nucleoside is first processed by the NS3 RNA triphosphatase
(RNase) that cleaves the γ-phosphate, resulting in an A diphosphate 5′ end (ppA-RNA).
Secondly, a not yet identified guanylyltransferase, transfers a GMP moiety from GTP to ppA-
RNA, forming the non-methylated form of the cap: GpppA-RNA. At this point, NS5
methyltransferase (MTase) is responsible for the third and fourth steps as it methylates first
the N-7 of G and then position 2′-O of A.[3,4,7]
From that which is stated above, it can be inferred that the cap is essential for viral RNA and
therefore it is essential for the viral replication cycle. NS5 MTase is responsible for two of the
four steps of cap formation and it has been shown that it’s activity is essential for effective
viral replication.[8,9] Consequently, NS5 MTase is a promising target for drugs that would be
potentially potent and selective anti-DENV agents.[4,10-12]

**STRUCTURE**

NS5 methyltransferase is the N-terminal portion of the 104 kDa multifunctional non structural
protein NS5. NS5 is not only the largest dengue virus protein, but also the most conserved one
within the Flavivirus genus and among DENV serotypes with 64% amino acid homology.[12-14]
Within NS5, three principal functional domains have been identified. At the N-terminal (amino
acids 1-263), NS5 has a S-adenosyl-L-methionine (SAM) dependent methyltransferase (MTase)
function that allows the completion of the RNA capping. The region between residues 320 and
368 is responsible for the binding of another viral encoded protein: NS3. In fact, these two
enzymes appear to regulate each other’s function as it has been reported that NS5 enhances
the NTPase and 5′ RNA triphosphatase activity of NS3. The C-terminal region of NS5 is a RNA-
dependent RNA polymerase (RdRp). The function of the RdRp is to synthesize a double strand
RNA from a RNA template. This allows the synthesis of a negative stranded RNA molecule that
serves as a template for the transcription of the viral genome.[4,14]
Unfortunately, the structure of the full-length protein has not yet been solved and therefore
the exact interaction between the RdRp and the MTase is not known. However, SAXS
experiments have shown that the two NS5 domains can adopt several relative orientations, due to the flexibility of the linker that connects them (His263 to Pro271). One of these orientations was proposed for West Nile virus (WNV) using an in silico model. By means of a protein-protein docking algorithm and mutational studies, it has been proposed that the Lys46, Arg47 and Glu49 loop of the MTase interacts with Leu512 of the RdRp (numeration refers to DENV3). As NS5 is highly conserved in the Flavivirus genus, this model could be transferred also to DENV. A second orientation has been proposed on the basis of two recent crystal structures: the full length Japanese Encephalitis Virus (JEV) NS5 protein and the DENV NS5 RdRp with a small linker portion (residues 268-271). The conformation of the linker atoms of the latter are consistent with the JEV NS5 structure and suggest that the MTase domain is in contact with the base of the RdRp fingertips, on the opposite side with respect to the other model.

The MTase is a monomeric protein that takes a globular form with dimensions of 55 x 45 x 40Å. It can be divided into three sub-domains. The enzyme’s catalytic domain is in the core sub-domain (in yellow in Figure 5.2), which is constituted by residues 55-222. The secondary structure of this region shows 7 β-sheets fringed by 4 α-helices and it resembles the catalytic domain of other SAM-dependent methyltransferases, making NS5-MTase a challenging molecular target. The only differences between NS5 MTase and the consensus methyltransferase scaffold are the complete or partial absence of two α-helices. At the N-terminal of the enzyme (residues 1-54) there is the N-terminal sub-domain (in red in Figure 5.2) that contains a helix-turn-helix motif followed by a β-sheet and an α-helix. Between these
two sub-domains there are two α-helices and two β-sheets that constitute the third sub-domain, at the C-terminal portion of NS5 MTase (residues 223-267; in blue in Figure 5.2).\(^{4,13,18}\)

**POSSIBLE MOLECULAR TARGET SITES**

![Figure 5.3. NS5 MTase bound to a short cap analogue and to a SAH molecule. (PDB: 2P3L)](image)

In this structure a cap analogue, on the left, is bound in the GTP binding site and a SAH molecule, on the right, is bound to the SAM binding site. Between these areas there is a large positively charged surface that might aid the binding of the RNA molecule. Protein surface is in blue and small molecules are shown with a space-filling model.

Available crystallographic structures show that NS5 MTase has three main binding sites.\(^{5,7,9,18-21}\) One of them allows the interaction with a molecule of SAM (on the right in Figure 5.3). This cofactor is essential for the methylation reaction, as it is the source for the methyl group, but it has been recently demonstrated that it is not essential for protein folding.\(^{21}\) A second binding site is named the “GTP binding site” because the first solved crystallographic structure of this enzyme in the presence of GTP showed that this molecule binds in this region (on the left in Figure 5.3). Between these two areas there is a large shallow surface rich in basic residues (Lys and Arg) that at physiological pH are positively charged. There is no structural evidence, but it is thought that such a large area of positive charge could assist the binding of the negatively charged phosphates of the RNA molecule.\(^ {4,18}\)

Most of the NS5 MTase crystal structures solved to date, have a molecule of S-adenosylhomocysteine (SAH), the demethylated form of SAM, bound to the protein and it is assumed that SAM binds in the same place during the methylation reaction. The most recent published structures either lack the methylation cofactor or have a SAM molecule bound to the described cleft, confirming that SAM indeed binds in the same mode of SAH.\(^{9,21}\) The ribose of the SAM molecule interacts with the protein through water-mediated hydrogen bonds with Gly106, Glu111 and Thr104 and the rest of the molecule binds to Ser56, Thr219, Glu86 and Trp87.\(^ {4,18,20}\) The adenine moiety of the methylation cofactor is positioned in a hydrophobic
pocket constituted by residues Thr104, Lys105, Val132 and Ile147 and the most important hydrogen bonds interactions occur between the base and Asp131, Lys105 and Val132 residues. Above this region, there is a unique to flaviviruses MTases additional hydrophobic pocket (Phe133, Ile147, Gly148, Glu149, Arg160, Arg163, Val164) that has been exploited for the design of SAM-analogue inhibitors (“SAH analogues” reported in the FUTURE POTENTIAL TREATMENTS chapter of the main INTRODUCTION section) that present hydrophobic groups that can interact in this area. The ligand interaction schemes for these two molecules are reported in Figure 5.4.

There is a highly conserved K-D-K-E tetrad among different methyltransferases and it is present also in the NS5 MTase with residues Lys61, Asp146, Lys182 and Glu218. Even if they are distant in the primary structure, they are located in near positions in the 3D conformation and they are thought to be the active site of the methylation reaction. The exact mechanism is not yet completely known, but mutational studies have shown that while the 2'-O methylation requires all four residues, the N7 methylation only requires Asp146, suggesting that the two methylations might be slightly different. The GTP binding site is at approximately 12.3 Å from the SAM binding site and therefore it cannot be the site where the nucleotide is positioned during the methylation. However, all crystallographic structures with cap analogues show that these molecules bind in this highly
Section 5: NS5 METHYLTRANSFERASE AS A DRUG TARGET

5.1 INTRODUCTION

Conserved region. Studies have also outlined that the mutation of residues in this area (Phe25, Lys29, Ser150, Asn18, Ser216) inhibits the enzyme’s function, in particular the 2’-O methylation.\cite{4,7,18} Biological assays have demonstrated that the enzyme binds only capped RNA molecules that have AG as the two RNA bases at the 5’ end. In fact, 5’-AG sequence is strictly conserved in the Flavivirus genus.\cite{4,5} The requirement for DENV MTase-RNA binding of a cap with these characteristics is supported by the specificity of Guanine binding given by the fundamental hydrogen bond between the aromatic NH$_2$ of G and the carbonyl oxygen atoms of Leu17 and Leu20. Furthermore, the base interacts with Phe25 through a π-π stacking, but no interactions occur at the N-7 level, leaving space for a methyl moiety in this position, as shown in some crystal structures (e.g. PDB: 2P3O). The ribose creates hydrogen bonds with Asn18 and Lys14, while the phosphates interactions were defined with less precision in the structures and involve residues like with Ser214, Lys14, Arg215 and Ser150.\cite{4,7} The main MTase-RNA interactions described here are schematised in Figure 5.5.

![Figure 5.5. Schematic representation of the GTP - MTase interactions.](image)

In order to show the interactions, the complex of the MTase with a methylated cap analogue (PDB 2P41) was used. The interaction rendering follows the legend on the right.

Structures with different small cap analogues have been solved. One particular series, showed that the guanine has always the same binding mode, while a second base can assume three different conformations.\cite{5} This information, together with the fact that the residues in this region are important for 2’-O but not for N7 methylation, suggest that this could be the region for the binding of the cap during the methylation at the A ribose. Two of the three conformations place the second base above the guanine, stacking with it (for instance shown in Figure 5.3). The third conformation is thought to be flexible and elongated towards the SAH molecule, but the density map was not sufficient to prove this. These structures do not give a
valid insight into the methylation process, but imply the hypothesis that the missing guanylyltransferase in the capping procedure could be the NS5 MTase itself, but further studies are needed to support this.\textsuperscript{[5]}

Like the SAM cleft, also the GTP binding site could be used for drug design. In particular, one study has demonstrated that a Ribavirin, G analogue, is able to interact with the protein in a similar manner and to have some inhibitor activity.\textsuperscript{[19]} Even if this compound has no efficacy in treating DENV-infected patients, this could be a starting point for the development of novel antiviral agents.\textsuperscript{[4,19]}

A shallow, highly positively charged area sits between the described binding sites. There is no structural proof, but it’s position and high prevalence of conserved basic residues, suggest that it could host the non-specific interactions with the RNA phosphates. The structure of the enzyme with a longer RNA molecule (8 nucleosides) has been solved.\textsuperscript{[17]} Unfortunately it was not able to give useful evidence as it shows a complex of two RNA molecules with four enzyme molecules that is not stable in solution. The determination of the exact binding mode would be useful for a better understanding of the enzyme’s mechanism of action and could aid the drug design of inhibitors that exploit this site.\textsuperscript{[4,17]}

**MECHANISM OF ACTION**

The exact mechanism of action of the enzyme is not fully understood to date, but two models have been proposed.\textsuperscript{[4,13]} The first model is also called “Single-MTase model” as it requires only one NS5 protein for each RNA strand. According to this model, after the first SAM-dependent methylation at the N7 position of the cap guanosine, the MTase slides along the RNA chain, positioning the cap guanosine in the GTP binding site. This places the A in the active site for the 2’-O methylation, after the replacement of the SAH molecule with a SAM molecule.\textsuperscript{[4,13]}

The second model, called “Double-MTase model”, requires the presence of two methyltransferase for each RNA molecule. According to this model, after the N-7 SAM-dependent methylation, the complex dissociates and the RNA binds to a second NS5 MTase, positioning the cap guanine in the GTP binding site and the adenine in the optimal position for the 2’-O methylation.\textsuperscript{[4,13]}
A FRAGMENT-BASED DRUG DESIGN STUDY

The previous chapters outlined the importance of DENV MTase and why it is a good drug target. Unfortunately, drug design studies reported in the literature based on virtual screening campaigns or on the modification of natural substrates have not produced promising inhibitors to date. An alternative drug design strategy could be to apply a fragment-based drug discovery (FBDD) methodology. This drug discovery tool has now been used for over 15 years and the Food and Drug Administration (FDA) approval of the first drug discovered with this methodology was Vemurafenib in 2011. This, together with many successful examples, has increased FBDD popularity within big and small pharmaceutical companies as well as in Academia. The basis of FBDD is the use of small molecules (fragments) with a molecular weight (MW) lower that 300 Da as a screening tool instead of the bigger compounds (with MW around 500Da) used in HTS campaigns. The idea that starting from smaller molecules leads to higher success rate is corroborated by the analysis of drugs on the market and the starting point (or “prototype”) from which they were developed. Commonly, less complex molecules as “prototype” are much more common than more complex ones. The general procedure for a FBDD approach consists in the design of a library of fragments that is then screened in vitro. The positive resulting fragments are then elaborated though in silico design and Medicinal Chemistry evaluations. The design of a good library is therefore essential for the outcome of the study. Fragments generally follow the rule of three (MW lower than 300Da, up to three hydrogen bond acceptors, up to three hydrogen bond donors and a maximum calculated logP of 3), avoiding unwanted scaffolds (e.g. toxic, unstable or reactive). The advantage of the screen of fragments over drug-like molecules is the fact that as the number of atoms involved is lower, the possible chemical space is much smaller (around $10^7$ possible molecules against $10^{60}$ minimum) and therefore it is possible to explore a larger portion of the space with reasonable-size libraries. However, being smaller molecules, fragments have also lower affinity for the target (around 0.1-10 mM). Consequently, in screening, fragments are evaluated with several biochemical techniques searching generally for consensus. A pivotal feature and limitation of FBDD is the requirement of the determination of the structure (either through X-ray crystallography or NMR) of the protein-fragment complexes for the elaboration of the hit fragments.

In collaboration with Marseille University, a FBDD study is being carried out for the discovery of novel inhibitors against DENV NSS MTase. As several structures of the MTase are available, it seemed a suitable target for a FBDD and it was indeed possible to crystallise the protein and to soak fragments for co-crystallisation. An initial pool of 500 fragments with molecular weight under 300Da was screened against the MTase with a thermal shift assay (TSA). The 30
compounds that were shown to bind to the protein by giving an increase in melting temperature of at least 0.5°C were then selected for X-ray crystallography and high concentration screening for inhibition of MTase enzymatic activity. The “bottle neck” of this procedure was the X-ray crystallography as only seven molecules were solved in complex with DENV3 NS5 MTase after crystal soaking. These compounds were then assessed for inhibition of MTase activity at several concentrations for IC_{50} determination. This is the starting point for fragment elaboration, as described later in this chapter. Figure 5.6 shows an image of the protein co-crystallised with the fragments and the IC_{50} values for the seven molecules were calculated and are reported in APPENDIX 6.

![Image](image.png)

**Figure 5.6. Image of the crystal structure of the NS5 MTase in complex with seven fragments and SAM.**
The protein is represented with a blue surface, SAM is coloured in orange and the fragments are represented with the ball and stick model. In order to show all fragments, two views with 90° rotation are displayed.

### AIMS AND OBJECTIVES

As documented in the introduction, DENV NS5 MTase is a promising drug target and FBDD is a successful approach for the development of novel drugs. With the aim of discovering novel potent inhibitors of NS5 MTase that could be developed as anti-DENV drugs, a study in collaboration with Marseille University has been carried out. Starting from the available structural data of NS5 MTase in complex with fragments, Molecular Modelling techniques have been applied on three of the seven fragments for an iterative process of fragment elaboration via molecule “growing” and “linking”. A fourth fragment was also explored with the aim of understanding its mode of action, as its binding pose did not appear to justify its inhibitory activity.
5.2 RESULTS AND DISCUSSION

5.2.1 FRAGMENT GROWING DRUG DESIGN

![Figure 5.7. Fragment 2A4.](image)

Here, fragment 2A4 is shown in complex with NS5 MTase. The position in respect of the whole protein of fragment 2A4 is reported on the left and a closer view of the two identical 2A4 fragments bound to the MTase is reported on the right. The protein complex on the left is the same as reported in Figure 5.6. The protein is shown with a blue surface and 2A4 is represented with the ball and stick model on the left and with lines in the magnified view on the right.

The first de novo drug design approach was applied to fragment 2A4 (thieno[2,3-b]pyrazin-7-amine). The structure of NS5 MTase in complex with SAM and 2A4 has recently been made available in the Protein Data Bank with PDB ID 4DTK. As shown in Figure 5.7, two fragment molecules were solved in the crystal structure and they both were located in the GTP binding cleft. These molecules show a π-π stacking interaction between them, but only the lower one is bound to the enzyme through a second π-π stacking with Phe25, resembling the same type of interaction of the GTP.

Superposing the protein-fragment structure with one available in the literature that contains guanidine (PDB ID 2P1D[18]), a good overlap of the lower 2A4 fragment with the GMP molecule could be observed, as shown in Figure 5.8. Additional to the π-π stacking with Phe25, the G base interacts with the enzyme also through hydrogen bonds between the amine group in the 2 position and the carbonyl oxygen atoms of Leu17 and Leu20. As already mentioned, this interaction is the reason for the specificity of the viral RNA cap binding that has a very conserved 7MeGpppA2'OMe structure.[4,7] Fragment 2A4 does not present a similar group, but interestingly, in the protein-fragment structure a highly ordinate molecule of water is...
approximately co-localised with the G amino group and interacts with Leu17 and Leu20, confirming the importance of these hydrogen bonds.

![Figure 5.8](image)

**Figure 5.8.** Overlap of 2A4 fragment with GMP molecule from a published crystal structure (2P1D). 2A4 fragment’s carbon atoms are coloured in cyan, while the GMP ones are in orange. Some protein residues are displayed as lines and a molecule of water (red sphere) from the fragments’ crystal structure is also reported.

The inhibition assay for the fragments showed that 2A4 is not one of the most active as it displays IC\textsubscript{50} values higher that 10mM for both the 2’-O and the N-7 methylations. However, as one of the two 2A4 fragments resembles the GTP binding mode and as GTP analogues (e.g. Ribavirin) can inhibit NS5 MTase, it was hypothesised that 2A4 could be used for the design of a novel competitive inhibitor of the NS5 MTase cap binding. Consequently, the lower 2A4 fragment was chosen for a “growing” de novo drug design strategy with the aim of adding functional groups that could enhance molecule binding and activity. The addition of three groups (from now on called R\textsubscript{1}, R\textsubscript{2} and R\textsubscript{3}) on the possible sites reported in Figure 5.9 was attempted through iterations for progressive molecule improvements with the aid of de novo molecular modelling tools, of a pharmacophore and of docking simulations.

![Figure 5.9](image)

**Figure 5.9.** Fragment 2A4 growing sites. 2A4 structure is shown here with atom numbering. The three growing groups are indicated with R\textsubscript{1}, R\textsubscript{2} and R\textsubscript{3}. 
GENERAL PROCEDURE

The general workflow of the molecular modelling study is reported in Figure 5.10. The first modification of 2A4 was based on the observations of the crystal structures and regarded group $R_1$. The resulting molecule was then used as an input for three de novo design computational tools (MOE MedChem Transformations, MOE Combinatorial Builder and LigBuilder) for growing procedures on groups $R_2$ and $R_3$. The results were then visually inspected and used for the design of a database of 2A4 derivatives that were docked and scored (with Maestro Glide) in the MTase structure. The binding poses were then visually inspected and used for the construction of a novel database of 2A4 derivatives. The docking-inspection-database design process was repeated in an iterative manner for molecule improvement. Due to the large number of obtained poses, the first docking results were filtered using a pharmacophore for a faster identification of good binding poses. All of the
procedure reported here was performed prior to the chemical synthesis and evaluation of designed compounds.

PHARMACOPHORE

The three computational tools used for molecule growing were: MOE MedChem Transformations, MOE Combinatorial Builder and LigBuilder. All three of them use a pharmacophore query in order to add to the molecule functional groups that can improve its binding to the protein.

For the two MOE-based tools, the pharmacophore query (shown in Figure 5.11) was built as described in the METHODS chapter of this section.

![Figure 5.11. Pharmacophore used for 2A4 growing approaches.](image)

Some MTase residues are indicated and shown as lines. A GMP molecule (from 2P1D) is reported with orange carbon atoms and the pharmacophore features are represented by the dotted spheres. F1 and F2, in orange, indicate aromatic groups; F3, in purple, indicates a hydrogen bond donor group; F4, in pink, indicates a hydrogen bond donor or acceptor group; F5, in pink, indicates a hydrogen bond acceptor group; F6 and F7, in yellow, indicate the projection of a hydrogen bond acceptor group; F8, in yellow, indicates the projection of a hydrogen bond donor group.

Features were built on the basis of the conserved GTP interactions with NS5 MTase displayed in the published crystallographic data. In particular, the \( \pi-\pi \) stacking between the G base and Phe25 was represented by “aromatic group” features (F1 and F2, in orange), while the hydrogen bonds with Leu17 and Leu20 were considered through a “hydrogen bond donor” feature (F3, in purple) placed on the highly ordered water molecule crystallised in the 2A4 structure. Furthermore, the interactions between the GTP 2’ hydroxyl group with both Asn18
main chain carbonyl oxygen and Lys14 side chain amine moiety were considered with a “hydrogen bond donor or acceptor” feature (F4, in pink). Additional interesting interactions for pharmacophore building were the ones involving Arg22, Lys29, Lys14 and Glu157 (features F5 to F8). As no indication was available for the preferred position in these cases, the projection (in yellow) of the feature was used in most of them. Of the total 8 pharmacophore features, only the three most important ones in GTP binding (F1, F2 and F3) were also considered essential in the MOE-based growing procedures.

All the residues considered in the design of the pharmacophore (Phe25, Leu17, Leu20, Asn18, Lys14, Arg22, Lys29 and Glu157) are conserved among all DENV serotypes and three of these residues (Phe25, Lys29 and Asn18) are essential for the methylation processes, in particular for the 2'-O methylation, as demonstrated by mutational studies.\(^4,7,18\) This opened the opportunity of designing potent panserotypic DENV inhibitors.

Differently from MOE, LigBuilder, the third computational tool used for 2A4 growing, builds a pharmacophore on the basis of the receptor-fragment complex that is given as input. Interestingly, the most important features of the pharmacophore built by LigBuilder and of the one built for MOE were in common, as shown by the superposition in Figure 5.12. Not surprisingly, these features were the ones derived from GTP binding mode.

Figure 5.12. Superposition of the pharmacophores built with LigBuilder and MOE.

The LigBuilder pharmacophore is represented by the grids, while the MOE one is represented by the solid spheres. It is possible to observe that the hydrogen bond donor features of LigBuilder (blue grids) correspond to the same features in MOE (F3, F4). In the same way also the hydrophobic LigBuilder feature (grey grid) corresponds to the aromatic MOE features (F1 and F2) and the hydrogen bond acceptor LigBuilder feature (red grid) corresponds to the same MOE features (F5 and F6).
**R₁ MODIFICATION**

Based on the observations from the protein-2A4 and GTP-protein complexes superposition, R₁ was the first growing position of 2A4 that was evaluated. The aim was to reproduce the hydrogen bonds with the carbonyl oxygens of Leu17 and Leu20, essential for G specific recognition. For this group, the MOE MedChem Transformations tool was used. Two possible groups were evaluated: a hydroxyl and an amino group. They were both placed in the same position of the ordered water molecule, but as only the amine could interact with both residues, this group was selected and compound 19 was developed further. Figure 5.13 shows compound 19 structure and binding pose.

![Compound 19](image)

*Figure 5.13. Compound 19.*

A) Chemical structure of compound 19. B) Compound 19 binding pose. The new amine group of compound 19 shows a good overlap with the water molecule and it interacts with both Leu17 and Leu20. Here, a few protein residues are indicated and shown as lines, while the original 2A4 fragment and compound 19 have purple and blue carbon atoms respectively.

Later in the molecule improvement process, it was observed that the original amino group, in position 7, did not display interactions with the protein as this area is designated to accommodate the N-7 methyl group of the capped RNA for 2'⁰-O methylation. Therefore, a small docking study was computed on three compounds with the amino group in position 6 (compound 20), in position 7 (compound 21) or in both positions (compound 22). Informatively, while the molecules with NH₂ in position 6 had very similar binding poses, compound 21 flipped, but they all maintained the hydrogen bond interactions with Leu17 and Leu20. This further confirmed the importance of these residues in ligand binding and suggested that of the two amino groups, only the one in position 6 was fundamental and that
the one in position 7 could be removed. The molecules designed for further improvement were indeed based on the scaffold reported in Figure 5.14.

![Molecular structures](image)

**Figure 5.14. Compounds 20, 21 and 22 and new molecular scaffold.**

A) Compound 20 binding pose. Protein residues are indicated and designed with lines, while ligands are shown with thicker lines. 2A4 carbon atoms are coloured in grey, while compound 20 ones are in cyan. B) Compound 21 binding pose. The same rendering as in A) was applied. C) Compound 22 binding pose. The same rendering as in A) was applied. D) New molecular scaffold for molecule improvement. Only the NH₂ in position 6 was retained. Atomic numbering is reported in the structure.

**R₂ AND R₃ “GROWTH”**

With the aim of growing the fragment with R₂ and R₃ groups, three computational tools were used: two MOE-based applications (MedChem Transformations and Combinatorial Builder) and LigBuilder software. As the binding pocket is shallow, LigBuilder software built huge molecules with side chains exiting the cleft as a result. As this fell outside the exploration objective, dummy atoms had to be added to “close” the binding site and improve the results for this computational approach.

The results from the three applications were visually inspected and used for the design of a database of molecules given by the combination of R₂ and R₃ groups. The total of 109 molecules (from the combination of 4 R₂ groups and 27 R₃ chains) were then docked with Maestro Glide. As already mentioned in the general procedure description, given the large
number of poses obtained, the same pharmacophore built for the MOE-based tools was used to detect the best binding molecules. For this reason, only molecules that matched in their absolute position at least four of the eight designed features (with F1 and F2 essential) were kept. The resulting 15 best molecules showed that $R_2$ interactions with the protein rarely involved Lys29, but always Arg22 through either carbonyl or carboxyl groups. Only two types of $R_3$ were interesting as they interacted with Lys14 in almost all the combinations with $R_2$. Of the two, the best interacting $R_3$ side chain appeared to contain a phenol group that also interacted with Ser213, as shown by compound 23 in Figure 5.15.

![Figure 5.15. Compound 23 binding pose.](image)

Example of result from the first docking procedure. In compound 23 $R_2$ interacts with Arg22 through the terminal carbonyl moiety, while $R_3$ presents the phenol portion that binds to Lys14 and Ser213. The protein residues are indicated and represented as lines, while small molecules are shown with thicker lines. The original 2A4 fragment carbon atoms are coloured in grey, while the compound ones are in orange.

**R$_2$ AND R$_3$ OPTIMISATION**

The results from the growing and docking procedure and $R_1$ optimisation were used for further development through the iterative procedure discussed above. Several combinations of $R_2$ and $R_3$ groups were assessed as well as the possibility to move their position and the most relevant outcomes are discussed here.

The movement of $R_3$ from position 3 to position 4 (with the substitution of the N atom in 4 to C) was examined with the aim of improving the interactions with Lys14 and Ser213 and removing the vinyl portion of the linker that bridges the phenol moiety and the molecule scaffold (as in compound 23). The position of the phenol hydroxyl group was also investigated, as shown in Figure 5.16. At the same time, $R_2$ improvement was attempted with the design of a series of $\alpha$-, $\beta$- and $\gamma$-ketoacids with different linker lengths, aiming for the simultaneous interaction with Arg22 and Lys 29.
Generally, in absence of α-, β- and γ-ketoacids in R₂, almost all the new R₃ moieties in 4 displaced the bicyclic scaffold, resulting in the loss of the Phe25 stacking and often of the Leu17 hydrogen bond. The ketoacid moiety in R₂ allowed the establishment of a second hydrogen bond between the molecules and Arg22, but no new interactions that involved Lys29 could be observed. Only two compounds gave partial desired MTase binding.

<table>
<thead>
<tr>
<th>“Original”:</th>
<th>“New”:</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Original" /></td>
<td><img src="image2.png" alt="New" /></td>
</tr>
</tbody>
</table>

**Figure 5.16. First R₂ and R₃ optimisation attempt.**

*A* Structure of the modification of the R₃ scaffold. The changes in position and structure of R₃ are represented here and atom numbering is reported. *B* Compound 20 and compound 24 overlapping binding poses. A comparison between the original R₃ moieties is presented through the overlap of the binding poses of compound 20 and 24. Protein residues are indicated and represented with lines, while small molecules are shown with thicker lines. Carbon atoms of Compound 20, those of compound 24 and of 2A4 are coloured in cyan, orange and grey respectively. *C* Compound 25 binding pose. The same rendering as in B) was applied and compound 25 has orange carbon atoms.

Compound 24 (in Figure 5.16) was able to bind to Lys14 and Ser213, but a molecule shift gave a less favourable position for the interactions involving the amine in position 6 as well as a weaker stacking with Phe25. Compound 25 (in Figure 5.16) interacted with both Arg22 and Lys29, but unfortunately with the loss of the hydrogen bond with Lys14. From these first optimisation attempts it could be inferred that the new R₃ moiety did not improve the molecule with these R₂ groups that did not represent a huge step in molecule optimisation.
As the pocket is solvent exposed, the flexibility of a α-, β- and γ-ketoacid R₂ group could give low stability to the complex and therefore further modifications were designed. The hydrophobicity of the substituent was increased in order to reduce the solvent-dependent ligand displacement and to substitute the electrostatic interaction involving Arg22 with a cation-π one. Docking simulations showed that in order to place a benzyl group correctly it was necessary to move R₂ from position 2 to 1. A hydroxyl group in the meta- or para- position was also inserted for the interaction with Asp26. These two OH positions appeared to be equivalent from docking simulations.

Interestingly, for this type of R₂ moiety, a better binding was obtained with a R₃ group in position 4, in particular with meta-hydroxyl benzyl substituents. Figure 5.17 shows two examples of the designed molecules with a benzyl R₂ group with the OH group (compound 26, in orange) or without (compound 27, in green).

Even if a cation-π interaction with Arg22 was achieved and hydrogen bonds with Leu17, Leu20, Lys14 and Ser213 were generally maintained, all the compounds with the R₂ benzyl group displayed a decrease in π-π stacking with Phe25. Therefore, the phenyl ring was reduced in size to a five-membered one (pyrrole, furan and thiazole). Of these, the best substitution resulted to be the pyrrole ring that could establish a hydrogen bond with Asp26. Compound 28 in Figure 5.18 is an example of this series.
A further iteration was undertaken with modifications aimed to replace the sulphur in the benzenthioephene core. The thiophene ring was substituted with a pyrrole, a furan or an indole moiety in compound 27. Unfortunately, the binding poses did not give insights on the best ring to be used as all the molecules superposed very well (data not shown).

For synthetic purposes, the methylene bridges were substituted with ether moieties in the most interesting compounds (e.g. compounds 24, 26, 27 and 28). The obtained binding poses were compared with those of the original compounds, but no significant differences could be observed, suggesting that this substitution could be applied for the synthesis of novel compounds.

**BIOLOGICAL EVALUATION OF SYNTHESISED COMPOUNDS**

The chemical synthesis of this chosen scaffold appeared to be challenging. Consequently, of the most promising molecules described above only one final compound and two intermediates (see Table 5.1) were synthesised and tested for MTase inhibitory activity by our collaborators at AFMB, Marseille University. Unfortunately, neither of them showed inhibition activity below the maximum tested concentration (5mM). Moreover, soaking experiments with these compounds was not successful and no co-crystallised structures could be obtained by our collaborators.
5.2 RESULTS AND DISCUSSION

Table 5.1. Synthesised compounds with activity data.

<table>
<thead>
<tr>
<th>Compound</th>
<th>R₁</th>
<th>R₂</th>
<th>N-7 methylation IC₅₀ (mM)</th>
<th>2’-O methylation IC₅₀ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>![Image]</td>
<td>![Image]</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>30</td>
<td>![Image]</td>
<td>![Image]</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>31</td>
<td>NH₂</td>
<td>![Image]</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
</tbody>
</table>

CONCLUDING REMARKS

Starting from the GTP site binding fragment 2A4 molecule growing procedures and docking simulations produced some molecules that could bind to the NS5 MTase GTP binding site and inhibit the NS5 methyltransferase enzyme via the competition with the natural ligand. In the attempt of directing the drug design towards serotype-nonspecific compounds, conserved amino acids were selected as target residues. Different growing locations were explored and the position of the amine present in the 2A4 fragment was evaluated. In summary, the amine is preferably located in position 6, complying with the mechanism of MTase G preferentiality nucleotides binding. The other side chains are preferentially situated in positions 2 (R₂) for the interaction with Arg22 and eventually Lys29, and 3 (R₃) for the interaction with Lys14 and Ser213. However, when substituting R₂ with an aromatic group the preferred positions become 1 and 4. This new scaffold (e.g. compounds 26 and 27) gave the most interesting binding results and was considered to be a good starting point for further optimisation. The reduction of the phenyl group in R₂ to a smaller five-membered ring aided a better accommodation of the molecules in the binding site and the most promising molecule of this series appeared to be 28. The sulphur in the benzothiophene core was also investigated and substituted, but no improvements could be observed from the Molecular Modelling results. Lastly, for synthetic purposes the methylene-ether substitution of the 1 and 4 side chains was assessed to be equivalent from an in silico binding point of view. The synthesis, biological
evaluation and soaking experiments were performed in Marseille University. The synthesis of this scaffold was found to be challenging and only three molecules were obtained, but unfortunately none of them showed activity against DENV 3 NS5 MTase and the soaking experiments failed to produce crystal structures of these molecules in complex with the same protein.

The negative experimental results might suggest that the Molecular Modelling approach used in this study failed to optimise the 2A4 fragment and to obtain a novel potent anti-DENV drug. However, it is important to underline that only one final compound analysed (31) corresponded to a compound 24 analogue. Other molecules different from this compound (e.g. 26, 27 and 28) showed a better binding than 24 but have not been assessed to date, nor have their ether analogues. Furthermore, the other two tested molecules were synthetic intermediates and consequently they were not expected to have a good inhibitory activity. In order to have a good evaluation of the Molecular Modelling and to validate 2A4 as a good starting fragment for drug design, it would be necessary to increase the amount of experimental data. Nevertheless, due the synthetic difficulties and to better results obtained from the optimisation of other fragments (see next chapter), this might not be done in the near future.
5.2.2 FRAGMENT LINKING DRUG DESIGN

Here, the fragment 2E11 and 2C3 are shown. The localisation of the two molecules within the protein-fragments complex is reported on the left and a closer view of 2E11 and 2C3 bound to the MTase is reported on the right. The protein complex on the left is the same reported in Figure 5.6. The protein is shown with a blue surface and the fragments are represented with the ball and stick model on the left and with lines on the right. 2E11 carbon atoms are coloured in orange and for 2C3 they are in yellow.

A second *de novo* drug design procedure was applied to two fragments: 2E11 (4-chloro-5-methylbenzene-1,2-diamine) and 2C3 (2-amino-4-methylbenzoate). The selected structures (2E11 in orange and 2C3 in yellow in Figure 5.19) bind to a MTase surface region that is well conserved among DENV serotypes. 2E11 and 2C3 were chosen because a linking strategy could be applied to them, as they appeared to be at an appropriate distance. Moreover, they showed a better inhibitory activity than 2A4, in particular on the 2'-O methylation. IC$_{50}$ values for 2E11 were 4.2mM and greater than 10mM for 2'-O and N-7 methylation respectively; while the ones for 2C3 were 4.4mM and 6.5mM respectively.

The original crystal structure reported two binding modes for 2E11: one as reported in Figure 5.19 and one on the other side of the protein. As only one is located at an appropriate distance from 2C3 fragment, only the first binding mode was considered for this drug design study. Furthermore, the selected 2E11 fragment is well superposed with another fragment (3C2) which, having poorer inhibitory activity (IC$_{50}$ higher than 10mM for both methylations), was therefore ignored.

The general approach adopted in this study was based on the connection of the two selected fragments through the addition of a linker chain. Four linker strategies were adopted and the best results were used for a docking procedure that allowed the selection of the molecules that best improved the binding of the initial fragments.
Experimental data of some of the designed molecules were then used for preliminary structure-activity relationship (SAR) determination and for further molecule growing and optimisation, using docking and molecular dynamics (MD) simulations.

**FRAGMENT LINKING**

*Four linking strategies*

*Table 5.2. Original fragments and linking strategies.*

<table>
<thead>
<tr>
<th>Name</th>
<th>2E11</th>
<th>2C3</th>
<th>Linking mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td><img src="image1" alt="Original-fragments" /></td>
<td><img src="image2" alt="Original-fragments" /></td>
<td>-</td>
</tr>
<tr>
<td>fragments</td>
<td><img src="image3" alt="Original-fragments" /></td>
<td><img src="image4" alt="Original-fragments" /></td>
<td></td>
</tr>
<tr>
<td>Link-1</td>
<td><img src="image5" alt="Link-fragments" /></td>
<td><img src="image6" alt="Link-fragments" /></td>
<td></td>
</tr>
<tr>
<td>Link-2</td>
<td><img src="image7" alt="Link-fragments" /></td>
<td><img src="image8" alt="Link-fragments" /></td>
<td></td>
</tr>
<tr>
<td>Link-3</td>
<td><img src="image9" alt="Link-fragments" /></td>
<td><img src="image10" alt="Link-fragments" /></td>
<td></td>
</tr>
<tr>
<td>Link-4</td>
<td><img src="image11" alt="Link-fragments" /></td>
<td><img src="image12" alt="Link-fragments" /></td>
<td></td>
</tr>
</tbody>
</table>
The two fragments were used as starting points for four linking strategies that reflected the possible combinations of those connection sites of each structure that showed suitability for a linking procedure. A total of four growing sites for 2C3 and two growing sites for 2E11 were chosen, as schematised in Table 5.2, where the red hydrogen atoms represent the connection sites for each structure and the sketch shows the pertinent linking mode. Other combinations from the ones shown here did not give positive results and will not be discussed. Fragment linking was performed with the MOE Link Multiple Fragments tool, as explained in the METHODS section.

The visual inspection of the results showed that all the most promising structures were belonging to the first linking mode (from now on called “Link-1”) and it was concluded that only in Link-1 the connectors were at the optimal distance for the linking of the fragments. The computational results for “Link-1” often connected the two fragments with complex structures, but they could be clustered and manually simplified. Interestingly, all of these patterns connected the two fragments through three atoms in the main linker chain.

Novel molecules with linkers derived from these patterns or from their modification, as well as from other three atom chain possibilities, were subsequently designed and evaluated through docking simulations using the MOE Dock tool.

**Pharmacophore**

![Figure 5.20. 3D pharmacophore for 2E11 and 2C3 fragments.](image)

Here, the 3D pharmacophore is superposed to the original fragments. 2E11. Carbon atoms are shown in orange, those for 2C3 are in yellow and a few protein residues are shown as lines. The pharmacophore features are represented as dotted spheres. Features F1 and F3 (in orange) indicate the presence of an aromatic ring, while F2 (in purple) indicates the presence of a hydrogen bond donor group and F4 (in green) shows a hydrogen bond acceptor.
With the purpose of assuring that the new molecules were placed with the same position and orientation of 2E11 and 2C3, a 3D pharmacophore was used for molecule placement in the MOE docking simulations. A pharmacophore-constrained placement was chosen because as the fragments are bound to the protein’s surface and not in an area that could be defined as a pocket, unconstrained docking procedures conducted with several programs were not able to place compounds correctly.

The designed features indicated four main groups that could be identified both in the original fragments and in all the novel molecules: two aromatic rings (F1 and F3, in orange), the amine moiety recognised as a hydrogen bond donor group (F2, in purple) and the carboxylic acid indicated as a hydrogen bond acceptor group (F4, in green). All four features were considered essential for a correct placement of the docked structures. Figure 5.20 shows the pharmacophore features superposed to the 2E11 and 2C3 fragments.

Optimisation iterations

Structures giving the best “Link-1” results, together with molecules containing different short linkers, were used to design a database of 25 molecules that were then docked into the 2E11 and 2C3 fragments binding sites.

The best results obtained had a three atom linker with an sp² carbon in the middle, linked to an out of chain heteroatom (O, N or S) through a double bond. An sp³ carbon in the middle position did not perform as well and a carbon-carbon double bond was not as effective as the carbon-heteroatom. Furthermore, the three most promising docking results (reported in Table 5.3), displayed the presence of two nitrogen atoms: one in each of the other two positions of the linker.

Table 5.3. Three best binding molecules.

<table>
<thead>
<tr>
<th>Scaffold</th>
<th>Compound</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Scaffold Image" /></td>
<td>32</td>
<td>O</td>
</tr>
<tr>
<td>33</td>
<td>NH</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>

The presence of these two NH groups at the 1 and 3 position of the linker improved the binding by adding a two hydrogen bonds with Glu112. This was also due to an overall slight movement in the new designed molecules of the ring from 2E11 that placed the hydrogen bond donor group at the right distance for the interaction with Glu112. An example of a binding pose (for compound 32) is reported in Figure 5.21.
Section 5: NS5 METHYLTRANSFERASE AS A DRUG TARGET

5.2 RESULTS AND DISCUSSION

Figure 5.21. Compound 32 binding pose.
The original fragments (2E11 on the left and 2C3 on the right) are reported with green carbon atoms while compound 32 carbons are in orange. A few protein residues are indicated and represented by lines. It is possible to see that while the two original fragments establish only one hydrogen bond with the protein (between 2C3 NH and Glu112), the new designed molecule demonstrates three hydrogen bond interactions with Glu112.

Figure 5.22. Compounds 35 and 36.
A) Structure of compound 35. B) Structure of compound 36. C) Compound 35 binding pose. Some protein residues are indicated and shown as lines. Carbon atoms belonging to the original fragment are coloured in green while those of the new compound are coloured in orange. D) Compound 36 binding pose. The same rendering of C) is used here. Both binding poses are discussed in the text.
Two other molecules (compound 35 and compound 36, in Figure 5.22) did not give an overlap with the original fragments as good as 32, 33 and 34, but they showed some of the interactions reported above. In both cases, the linker has only one of the two NH groups and therefore only two of three hydrogen bonds with Glu112. However, as 32, 33 and 34 are very rigid molecules, the conformation adopted in MTase binding could not be favoured due to conformational strains. Conversely, the compound 35 linker has both a rigid portion (the amide) and a flexible part (the CH₂) that allows the two aromatic rings to maintain the correct orientation.

EXPERIMENTAL DATA OF DESIGNED COMPOUNDS

Table 5.4. MTase inhibition data for synthesised compounds.

<table>
<thead>
<tr>
<th>Compound</th>
<th>X</th>
<th>R₁</th>
<th>R₂</th>
<th>2'-O methylation IC₅₀ (mM)</th>
<th>N-7 methylation IC₅₀ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>NH</td>
<td>2-NO₂</td>
<td>2-COOEt</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>38</td>
<td>NH</td>
<td>2-NO₂</td>
<td>2-COOH; 5-Me</td>
<td>&gt;2.5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>39</td>
<td>NH</td>
<td>2-NH₂</td>
<td>2-COOH; 5-Me</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>40</td>
<td>NH</td>
<td>2-NO₂</td>
<td>2-COOH</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>41</td>
<td>NH</td>
<td>2-NH₂</td>
<td>2-COOEt</td>
<td>&gt;2.5</td>
<td>&gt;2.5</td>
</tr>
<tr>
<td>42</td>
<td>NH</td>
<td>2-NO₂; 4-CF₃</td>
<td>2-COOH; 5-Me</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>43</td>
<td>NH</td>
<td>2-NO₂; 4-Cl</td>
<td>2-COOH; 5-Me</td>
<td>2.5</td>
<td>5</td>
</tr>
<tr>
<td>44</td>
<td>NH</td>
<td>2-NH₂; 4-CF₃</td>
<td>2-COOH; 5-Me</td>
<td>2.5</td>
<td>&gt;2.5</td>
</tr>
<tr>
<td>45</td>
<td>NH</td>
<td>2-NH₂; 4-Cl</td>
<td>2-COOH; 5-Me</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>46</td>
<td>NH</td>
<td>2-NHC00T-Bu; 5-Cl</td>
<td>2-COOEt</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>47</td>
<td>CH₂</td>
<td>2-NO₂</td>
<td>2-COOEt</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>48</td>
<td>CH₂</td>
<td>2-NO₂; 5-Cl</td>
<td>2-COOEt</td>
<td>1.2</td>
<td>2.5</td>
</tr>
<tr>
<td>49</td>
<td>CH₂</td>
<td>2-NH₂; 5-Cl</td>
<td>2-COOEt</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>50</td>
<td>CH₂</td>
<td>2-NH₂; 5-Cl</td>
<td>2-COOH</td>
<td>&gt;5</td>
<td>&gt;5</td>
</tr>
<tr>
<td>51</td>
<td>CH₂</td>
<td>2-NH₂; 5-Cl</td>
<td>2-C0O0Me</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>52</td>
<td>CH₂</td>
<td>2-NH₂</td>
<td>2-COOH</td>
<td>&gt;2.5</td>
<td>&gt;2.5</td>
</tr>
</tbody>
</table>
A collection of 16 molecules, representing either final compounds or synthetic intermediates were synthesised and evaluated by our collaborators in Marseille University. Structures and activity data for these compounds are reported in Table 5.4.

Of the 16 compounds, six showed an improved inhibition activity against either MTase 2'-O or N-7 methylation or both and preliminary structure-activity relationships (SAR) could be derived. They are discussed according to the scaffold presented in Table 5.4. In the case of the urea linker (compounds from 37 to 46), all active molecules (from 42 to 45) display either a chloro- or a trifluoromethyl- R₁ substituent on the left ring, in position 4, while inactive molecule do not, making an electron-withdrawing halogen group on this ring is essential for activity. Only one inactive compound (46) has a chlorine substituent on the left ring, but in position 5 instead of 4. However, this single molecule is not enough to state that the para-position with respect to the linker is the preferred one for activity. A carboxylic group is also present in all active molecules, making it another fundamental feature for MTase inhibition. It is not clear whether an amino substituent on the left ring is preferable to a nitro group. Similar SARs occur in the amide linker series (compounds from 47 to 52). However, in this series, the position of the electron-withdrawing halogen group is moved to position 5 and the ester seems to be preferred to the carboxylic acid in the right ring.

Soaking experiments were also performed in Marseille University and two active compounds (44 and 45) were co-crystallised with DENV NS5 MTase. As shown in panel A of Figure 5.23, a new binding site with respect to the original fragments and closer to the SAM binding cleft was found. As an example of the new contacts occurring between these ligands and the protein, panel B of Figure 5.23 reports a schematic view of compound 44 interactions.

---

**Figure 5.23. Compounds 44 and 45 new binding site.**

A) Binding mode. Here is reported the overlap of the binding of the co-crystallised compounds 44 and 45, of the 2E11 and 2C3 fragments and of the docked compound 32. The protein surface is coloured in grey and SAM is shown in purple with the space-filling model. Compound 44 carbon atoms are shown in dark red, those for compound 45 are in green, whilst those for 2E11 and 2C3 are in yellow and for compound 32 are in cyan. B) Ligand interactions scheme for compound 44. The rendering of the interactions follows the reported legend.
Correlations between structural and activity data could only be found by analysing the symmetric unit of the complex, suggesting that the mode of action of this series of molecules could be the stabilisation of an inactive protein homodimer. The necessary halogen containing substituent (either Cl or CF3) interacts in a hydrophobic pocket of the symmetric monomer, involving Leu237, Met233, Leu238, Val234, Arg206. Similar interactions engage the methyl substituent of the other ring and residues Met70, Ile220 and Val205 of the symmetric monomer. A third hydrophobic contact involves the Pro108 of the original monomer. Hydrogen bonds are established between the carboxylate and Arg84 and between the aromatic amine and His110 and Glu112. The carbonyl oxygen of the urea linker is very close to the Gly109 backbone amine, but distance and angle are not ideal for hydrogen bond formation. Not all of the residues in the symmetric monomer are identical among DENV serotypes, whereas the hydrophobicity of the amino acids is. Conversely, all interacting residues in the original monomer are well conserved.

**MOLECULE GROWING**

*Molecule optimisation*

The symmetric dimer was reconstructed and used for compound optimisation. A comparison of this structure with the ones available in the Protein DataBank\[28\] showed a different Glu111 side chain conformation in most of them, correlated to the presence of a sulphate ion in the position indicated in Figure 5.24.

*Figure 5.24. Sulphate ion localisation and Glu111 new conformation.*

Here is reported the modified reconstructed symmetric unit of the protein-compound 44 complex overlapped with the sulphate ion from available crystal structures. The monomer surface and ribbon are coloured in blue, while the symmetric monomer ribbon is in pink. A few protein residues are reported and shown as lines, apart from Glu111 that is shown with thicker lines. SAM carbon atoms are in orange and those for compound 44 are in white.
In this conformation the acidic residue is able to interact with the basic Arg84. The consensus of the presence and of the position of the sulphate in many crystal structures and the fact that in absence of this ion Glu111 locates its carboxylate group in the same area, suggested that a similar feature could be included in the small molecule for binding improvement. In order to do so, the Glu111 side chain conformation was modified and energy minimised in the reconstructed dimer. The new Glu111 conformation is shown in Figure 5.24.

As before, databases of new molecules were designed and docked in an iterative manner. The docking and scoring procedures were performed with Maestro Glide, using a 10 Å grid centred on compound 44 as reported in the METHODS chapter.

From visual inspection of the results, interactions of compound 44 were retained in molecules that overlapped well with compound 44, they also had better docking scores. For an easier explanation of the docking results, Figure 5.25 highlights the designed modifications to compound 44.

The carboxylic acid (blue circle) was removed to avoid electrostatic repulsion with Glu111 and it was substituted with OH (that could give hydrogen bonds) or removed completely. In general, the hydroxyl group was not fundamental, but improved the position of the aromatic ring. With the aim of mimicking the sulphate group, the molecule was extended from the methyl substituent (yellow circle) and an anionic moiety was inserted. Best docking molecules had a linker that connected the scaffold to a new ring with a carboxylate moiety in the para-position. The correct curvature in the shape of the molecule was achieved only with sulphonic esters or sulphonamides. As shown in Figure 5.26 (with the two best docking results), this construction allows the placement of the carboxylate in the same area of the sulphate ion and its interaction with the backbone amino group of Glu111. The position of this side chain in the phenyl ring indicated in grey in the scheme was also explored. Docking simulation showed that both the original substitution and the new position indicated by the yellow arrow were able to give good binding poses. In order to increase molecule flexibility, the left NH group of the urea moiety (green circle) was substituted with a methylene, obtaining an amide. However, the increase in flexibility did not produce binding improvements. The aromatic
amine (red circle) was substituted with a nitro group for a better understanding of the role of this group and docking simulations suggested that this substitution is not favourable as it removes two hydrogen bonds. The ring in grey and the third ring were substituted with a furan or a pyrazole ring and this replacement was sometimes successful in producing a good pose. Taken together, these results suggested that the iterative docking procedure was successful in improving the binding mode in the *in silico* model. The six best compounds, including these two (compounds 53 and 54), and their binding poses can be found in *APPENDIX 7*.

![Molecular dynamics simulations](image)

With the aim of understanding the role and the impact of the change of Glu111 side chain conformation in the docking simulations, 5 ns molecular dynamics (MD) simulations for four molecular systems were carried out with GROMACS. The four molecular systems were the symmetric unit with no ligands and in complex with either compound 44, 53 or 54. Interestingly, even if all the starting structures had the Glu111 in the crystal structure conformation, in all MD simulations Glu111 turned and interacted with Arg84, confirming that the modification done for the docking simulations is reasonable in physiological conditions.
More stable binding could be observed for compounds 53 and 54 than for 44. As shown in Figure 5.27, for compound 44 MD, the shift of the Glu111 side chain placed the residue negative charge near the small molecule’s carboxylate group, producing an electrostatic repulsion and a distortion in the urea moiety and a greater solvent exposure of the ligand.

![Figure 5.27. Compound 44 decreased binding observed with MD.](image)

The Glu111 movement and the consequent decrease in compound 44 binding are represented here. Two snapshots from the MD simulation (at 0 and 2ns) were superposed to show different conformations of Glu111 and compound 44. Shifts are highlighted with arrows. The original MTase monomer ribbon and carbon atoms are coloured in blue, while the ones of the symmetric monomer are in pink and SAM carbon atoms are in orange. In all cases, light and dark shades for 0ns and 2ns snapshots respectively are used. Compound 44 carbon atoms are shown in green and in dark red for the 0ns and 2ns snapshots respectively.

Conversely, the Glu111 flipped conformation of the side chain was particularly stable in the other two molecules simulations. Consequently, the MD results gave a mechanistic explanation as to why compound 44 could have a decreased binding compared to compounds 53 and 54, supporting the docking results that suggested that the new series of compounds could potentially have an improved MTase inhibitory.

**Experimental data for the last series of compounds**

From the last suggested series of compounds, five molecules without the OH substituent (compounds 57, 60, 63, 66 and 69) and their intermediates were synthesised, assessed in vitro and used for soaking experiments in Marseille University. Table 5.5 shows the synthesised compounds and their MTase inhibitory activity. Compound 69 was synthesised but its inhibitory activity results are not available to date.

As shown in the table, the compounds of this series are more active inhibitors of the 2’-O methylation than N-7 methylation. This pattern was found also with the most active
compounds (60 and 63) giving IC\textsubscript{50} values of around 100\(\mu\)M and above 700\(\mu\)M and for 2'-O and N-7 methylation respectively.

\textit{Table 5.5.} MTase inhibition data for synthesised compounds.

<table>
<thead>
<tr>
<th>Compound</th>
<th>X</th>
<th>SO\textsubscript{2}NH position</th>
<th>R\textsubscript{1}</th>
<th>R\textsubscript{2}</th>
<th>2'-O methylation IC\textsubscript{50} ((\mu)M)</th>
<th>N-7 methylation IC\textsubscript{50} ((\mu)M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>NH</td>
<td>3</td>
<td>4-COOMe</td>
<td>2-NO\textsubscript{2}; 4-Cl</td>
<td>&gt;500</td>
<td>&gt;500</td>
</tr>
<tr>
<td>56</td>
<td>NH</td>
<td>3</td>
<td>4-COOMe</td>
<td>2-NH\textsubscript{2}; 4-Cl</td>
<td>715</td>
<td>&gt;500</td>
</tr>
<tr>
<td>57</td>
<td>NH</td>
<td>3</td>
<td>4-COOH</td>
<td>2-NH\textsubscript{2}; 4-Cl</td>
<td>452</td>
<td>&gt;500</td>
</tr>
<tr>
<td>58</td>
<td>NH</td>
<td>3</td>
<td>4-COOMe</td>
<td>2-NO\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>&gt;500</td>
<td>&gt;500</td>
</tr>
<tr>
<td>59</td>
<td>NH</td>
<td>3</td>
<td>4-COOMe</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>398</td>
<td>&gt;500</td>
</tr>
<tr>
<td>60</td>
<td>NH</td>
<td>3</td>
<td>4-COOH</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>91</td>
<td>1,200</td>
</tr>
<tr>
<td>61</td>
<td>NH</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NO\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>676</td>
<td>&gt;500</td>
</tr>
<tr>
<td>62</td>
<td>NH</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>237</td>
<td>&gt;500</td>
</tr>
<tr>
<td>63</td>
<td>NH</td>
<td>4</td>
<td>4-COOH</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>110</td>
<td>742</td>
</tr>
<tr>
<td>64</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NO\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>559</td>
<td>&gt;500</td>
</tr>
<tr>
<td>65</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>476</td>
<td>&gt;500</td>
</tr>
<tr>
<td>66</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOH</td>
<td>2-NH\textsubscript{2}; 4-CF\textsubscript{3}</td>
<td>332</td>
<td>&gt;500</td>
</tr>
<tr>
<td>67</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NO\textsubscript{2}; 4-Cl</td>
<td>473</td>
<td>&gt;500</td>
</tr>
<tr>
<td>68</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOMe</td>
<td>2-NH\textsubscript{2}; 4-Cl</td>
<td>430</td>
<td>&gt;500</td>
</tr>
<tr>
<td>69</td>
<td>CH\textsubscript{2}</td>
<td>4</td>
<td>4-COOH</td>
<td>2-NH\textsubscript{2}; 4-Cl</td>
<td>In progress</td>
<td>In progress</td>
</tr>
</tbody>
</table>

Importantly, the activity data agreed with the \textit{in silico} results presented above. As suggested by docking results, the substitution of the urea linker with a more flexible methylene-amide decreases activity. Furthermore, the final molecules were always more active than the synthetic intermediates, confirming that the carboxylic acid and the aromatic amine are important for molecule binding. The CF\textsubscript{3} substituent appeared to increase activity with respect to the Cl substituent, while no particular activity pattern regarding the sulphanilamide position could be observed.

A soaking experiment produced a structure for the MTase in complex with compound 57. Comparison of this structure with the protein-compound 44 complex showed that most of the protein residues have retained their conformation. Conversely, the Glu111 side chain was
crystallised in two conformations with 0.5 occupancy. One of them is exactly the same as in the compound 44 complex. The other one is similar to the conformation that was built for the docking simulations, confirming that this modification was reasonable in physiological conditions. Furthermore, as a consequence of the latter Glu111 conformation, the Arg84 side chain is shifted as well, as shown in Figure 5.28.

The co-crystallised inhibitors (compounds 44 and 57 in dark red and in blue in Figure 5.28 respectively) overlap well in the same binding site. Due to a small shift of the molecule, the interactions between His110 carbonyl oxygen and urea and aromatic amine groups are no longer recognised by the modelling program. However, the other interactions with the MTase monomer are maintained and they involve hydrophobic contacts with Pro108 and hydrogen bonds with Gly109 NH (and the urea carbonyl oxygen) and Glu112 COO⁻ (and the aromatic amine). The latter hydrogen bond explains the improved activity conferred by an amino group with respect to the nitro moiety in the same position. By reconstructing the unit cell it was also possible to confirm that the halogen group (either CF₃ or Cl) sits nicely in a lipophilic pocket, formed mainly by Leu237, Val234 and Met233 of the symmetric image of the MTase.

![Figure 5.28. Structure of MTase in complex with compounds 44 and 57.](image)

A) Front view. Here the superposition of the crystallographic structures of the complexes with compounds 44 and 57 are presented. Compound 44 and 57 carbon atoms are shown in dark red and in blue respectively. Protein ribbons and residues carbon atoms follow a similar colour code: those of the compound 44 complex are orange and those of the compound 57 are in light blue. SAM carbon atoms follow the same colour code. The interacting residues are indicated and hydrogen bonds are shown with red dotted lines. Both side chains conformations for Glu111 are shown. B) Side view. As panel A. Additionally, also Arg84 side chain conformations are shown.

Unfortunately, the new compound side chain in 57 is solvent exposed and does not bind the MTase close to the SAM site differently from the drug design aim and docking results, as shown in panel B of Figure 5.28 and in Figure 5.29. This could be partly due to the interaction of one of the sulphonamide oxgens with Arg68 of the symmetric image. The only other contact of this part of the molecule is a hydrophobic interaction with the symmetric image.
Met70. Consequently, no explanation of the improved activity of the carboxylate substituent over the ester could be gathered from the structural results. A hypothesis on why the added portion of the molecule presents this binding conformation could be that a clash or close contact with SAM that does not allow the compound to bind to the protein as predicted. However, the docking simulations were performed in presence of the natural cofactor and this did not affect molecules poses. This is shown in Figure 5.29, where the same view of panel B of Figure 5.28 is superposed with the docking pose of compound 53 (in green in the figure), as an example. Here, the compound 53 portion extends towards the SAM binding site and the carboxylate moiety is correctly positioned according to drug design aims (see previous chapter) and clashes with the natural cofactor can be observed.

![Figure 5.29. Comparison of compounds 44 and 57 crystal structures with compound 53 docking pose.](image)

The superposition of the crystallographic structures of the complexes with compounds 44 and 57 and of the docking pose of compound 53 is presented here. The same colour code as Figure 5.28, panel B was applied. Compound 53 carbon atoms are reported in green.

**CONCLUDING REMARKS**

With the aim of discovering potent anti-DENV drugs, 2E11 and 2C3 fragments were linked in the initial drug design procedure seeking interactions with conserved residues among the four DENV serotypes. An iterative process of designing and docking allowed molecule optimisation. Of these, three (compounds 32, 33 and 34) demonstrated a good overlap with the original fragments and an improvement of the binding properties to the protein, and another two (compounds 35 and 36) were shown to be interesting as they could reduce the rigidity of compounds 32, 33 and 34, but conserving most of the interactions with the methyltransferase.
A set of 16 compounds (including both final designed molecules and synthetic intermediates) were synthesised and biologically evaluated in Marseille University. Two of the active molecules were also successfully co-crystallised with DENV3 NS5 MTase. The experimental data provided preliminary structure-activity relationships and to construct a hypothesis for which the mode of action of these molecules could be MTase inactivation through homodimer stabilisation. One of the co-crystallised molecules was then used for further compound optimisation.

The crystal symmetric dimer was reconstructed, the Glu111 side chain was moved according to available structural data and the model was then used for docking simulations. With the aim of reproducing the interactions of a sulphate ion present in several available crystal structures of DENV MTase, the co-crystallised compound (44) was grown. Docking and design iterations produced six molecules (e.g. compounds 53 and 54) that could potentially improve binding and consequently activity.

The change in the Glu111 side chain conformation was confirmed to be physiologically relevant through molecular dynamics simulations. This change also supported the fact that compounds like 53 and 54, which bind more stably NS5 MTase in the new Glu111 conformation respect to 44, should represent an improvement with respect to the previous molecule series. Moreover, the latest experimental data showed that an increase of over 100-fold activity has been achieved for inhibiting 2'-O methylation and of over 14-fold for inhibiting N-7 methylation compared to the starting fragments 2E11 and 2C3, confirming that the linking of two fragments has been a successful approach for the development DENV NS5 MTase inhibitors through the designed fragment-based drug design study discussed here.

Unfortunately, the latest crystallographic data did not confirm completely the binding mode of the last series of designed molecules, as they do not extend towards the SAM binding site. However, the synthesis and evaluation of other already designed molecules different in this portion of the structure (e.g. compound 54, which has a smaller aromatic ring) could give an improved knowledge of the mechanism of action of these compounds and improve future fragment-based drug design iterations.

Taken together, the results suggest that the general de novo drug design approach applied to 2E11 and 2C3 fragments was successful in finding inhibitors of DENV3 NS5 MTase with novel chemical structures and further development of these molecules will lead to the discovery of compounds that could one day be used in DENV-infections treatment.
5.2.3 FRAGMENT 2G3 MODE OF ACTION

Figure 5.30. Fragment 2G3.

Here, fragment 2G3 is shown in complex with NS5 MTase. The position in respect of the whole protein of fragment 2G3 is reported on the left and a closer view of the 2G3 fragment bound to the MTase is reported on the right. The protein complex on the left is the same as reported in Figure 5.6. The protein is shown with a blue surface and 2G3 is represented with the ball and stick model on the left and with lines on the right.

The most active fragment that was co-crystallised with DENV3 MTase was 2G3 (2-(2,5-dimethyl-1H-pyrrol-1-yl)benzoic acid), displaying IC_{50} values of 0.3 mM and 2.0 mM against 2’-O and N-7 methylation respectively. However, the crystal structure did not aid the understanding of the fragments’ activity as 2G3 is placed on the surface, with a big solvent exposed fragment portion, as shown in Figure 5.30.

It was theorised that 2G3 binds to another site of the MTase for its inhibition activity. As the structure was obtained by soaking the already formed crystal of the MTase-SAM complex in a concentrated solution of 2G3, it was also hypothesised that in these conditions 2G3 was not able to displace the SAM molecule from its site and consequently to bind correctly. Furthermore, SAM binding site residues are well conserved among DENV serotypes, so if 2G3 really bound to this cleft it could be an interesting starting point for the design of a novel potent serotype-nonspecific DENV inhibitor.

In order to validate this hypothesis, a binding mode for 2G3 in the SAM binding site was studied through docking and MD simulations and is discussed here.

2G3 BINDING MODE

Fragment 2G3 was docked in the SAM binding site using Maestro Glide software and a 9Å grid centred on the natural substrate. In the best binding pose obtained, 2G3 superposed well with SAM, as shown in Figure 5.31. In particular, overlap between the pyrrole and phenyl rings with
the SAM base and ribose respectively could be observed. As simplified by the ligand interaction scheme in panel B of this figure, in this binding pose the two rings of 2G3 give a hydrophobic interaction with Ile147, while the carboxylate moiety exhibits electrostatic and hydrogen bond interactions with the Lys105 backbone amino group and His110 side chain, retaining some of the SAM interactions with DENV MTase.

![Figure 5.31. 2G3 binding mode.](image)

**Figure 5.31. 2G3 binding mode.**

**A** 2G3 binding pose overlapped to SAM molecule. The binding pose of 2G3 is overlapped to the SAM molecule for comparison. Some protein residues are indicated and represented with grey carbon atoms and part of the protein surface is in yellow. 2G3 carbon atoms are in orange, while those for SAM are in green. **B** Ligand interaction scheme of 2G3 pose. The main interactions of the fragment with the protein are reported with the rendering indicated by the legend on the bottom of the panel.

### 2G3 BINDING MODE VALIDATION

For a validation of the 2G3 binding mode and to further investigate the interaction of the ligand carboxylate with His110, three 3ns MD simulations of the complex were performed with GROMACS: one for each possible protonation state of the amino acid. The neutral His110 with the hydrogen on the ε nitrogen position (HIE110) was the preferred protonation state according to the force field used for the MD simulations. Only in this case 2G3 maintained the docking binding pose, retaining the observed interactions. Conversely, HIE110 changed conformation for hydrogen bond optimisation, causing a shift of the protein loop. In the other two cases (with positive charged His110, called HIP110, and with the neutral His110 protonated on the δ nitrogen, called HID110), the amino acid retained the original conformation, while 2G3 rotated, loosing the SAM-ring co-localisation of the docking pose and reducing the hydrophobic interaction with Ile147. All the new fragment conformations were
stable for the length of the simulation. Figure 5.32 shows the conformations adopted by 2G3 and the protein main changes observed in the three MD simulations and described above.

![Figure 5.32. 2G3 different conformations adopted in the 3 MD simulations.](image)

The three 2G3 conformations described in the text are superposed here for comparison. The molecular system with HIE110 is coloured in blue, the one with HID110 is in green and the one with HIP110 is in orange. In all three, 2G3 is coloured with a darker shade and His110-carboxylate hydrogen bonds are shown as red dashed lines.

Summarising, these results confirmed that the ligand can assume a stable pose in the SAM binding site and the carboxylate-His110 interaction is one of the most important in fragment binding. However, a stronger validation of the docking-derived binding mode of 2G3 would come from a good correlation between MTase inhibitory activity and docking results of 2G3 analogues. For this reason, a database of 2G3 analogues was designed and docked with Maestro Glide, as done for the 2G3 fragment. The most interesting molecules from docking results and their activity prediction are reported in Table 5.6.

2G3 analogues from 70 to 76 retained 2G3 interactions and most of them (70, 72, 73 and 74) showed binding improvement. In detail, the addition of hydroxyl groups to the pyrrole substituents (as in compounds 70, 72, 73 and 74) enhanced binding by the addition of hydrogen bonds with the Thr104 hydroxyl moiety and Gly148 backbone amino group, while a methylamino substituent of the phenyl ring in meta- position with respect to the carboxylate group (as in compounds 70 and 74) could interact with the Gly81 carbonyl oxygen. Panel A of Figure 5.33 shows compound 70 binding pose as an example of these interactions. Furthermore, a hydroxyl substituent in the phenyl ring in the ortho- position with respect to the carboxylate group (compound 71) overlaps with the 2’OH group of SAM and could therefore be interesting. However, no additional interactions could be detected in this case, as shown in panel B of Figure 5.33, and consequently the activity profile of this residue cannot be predicted in silico.

Conversely, a loss of activity would be expected for compounds 75 and 76. These two 2G3 analogues were designed to confirm the importance of the carboxylate-His110 interaction.
observed both in docking and MD simulations and to investigate the role of the negative charge in small molecule binding. The methyl ester was chosen over more bulky substituents for compound 75 in order to minimise the steric effect in molecule binding.

Table 5.6. Structure and activity prediction for improved 2G3 analogues.

<table>
<thead>
<tr>
<th>Compound</th>
<th>R₁</th>
<th>R₂</th>
<th>R₃</th>
<th>R₄</th>
<th>R₅</th>
<th>Activity prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>OH</td>
<td>OH</td>
<td>COOH</td>
<td>H</td>
<td>CH₂NH₂</td>
<td>Improvement</td>
</tr>
<tr>
<td>71</td>
<td>H</td>
<td>H</td>
<td>COOH</td>
<td>OH</td>
<td>H</td>
<td>Not predictable</td>
</tr>
<tr>
<td>72</td>
<td>H</td>
<td>OH</td>
<td>COOH</td>
<td>H</td>
<td>H</td>
<td>Improvement</td>
</tr>
<tr>
<td>73</td>
<td>OH</td>
<td>OH</td>
<td>COOH</td>
<td>H</td>
<td>H</td>
<td>Improvement</td>
</tr>
<tr>
<td>74</td>
<td>H</td>
<td>OH</td>
<td>COOH</td>
<td>H</td>
<td>CH₂NH₂</td>
<td>Improvement</td>
</tr>
<tr>
<td>75</td>
<td>H</td>
<td>H</td>
<td>COOCH₃</td>
<td>H</td>
<td>H</td>
<td>Decrease</td>
</tr>
<tr>
<td>76</td>
<td>H</td>
<td>H</td>
<td>CH₃</td>
<td>H</td>
<td>H</td>
<td>Decrease</td>
</tr>
</tbody>
</table>

Figure 5.33. Binding poses of compounds 70 and 71.

A) Compound 70. Compound 70 pose is superposed with the 2G3 fragment. Some protein amino acids are reported as lines and interacting residues are highlighted with thicker lines. Compound 70 carbon atoms are in cyan and those for 2G3 are in orange. B) Compound 71. Compound 71 pose is superposed with the 2G3 fragment and SAM. 2G3 and some protein amino acids are reported as lines and interacting residues are highlighted with thicker lines. Compound 71 carbon atoms are in cyan, those for 2G3 are in orange and for SAM are in green.
CONCLUDING REMARKS

As structural data from Marseille University failed to explain why 2G3 was the most active fragment against DENV MTase activity, a Molecular Modelling study was performed on this fragment for the investigation of its inhibitory mode of action. It was hypothesised that the fragment binds to the SAM cleft and a binding mode was proposed and evaluated with MD simulations. In order to validate this premise, 2G3 analogues were docked in the same site and their biological activity was predicted. These compounds will be synthesised and tested for their MTase inhibition activity in Marseille University. A consensus between experimental and docking results would support the hypothesis of 2G3 binding to the SAM cleft. This would aid a successful optimisation of the fragment and as the protein site is highly conserved, and could lead to the design of a novel, potent and panserotypic drug against DENV.

Differently to all other structures available in the literature, a DENV MTase crystal structure has been recently solved in the absence of SAM. The published protocol consists of protein production, unfolding and refolding, after SAM (or SAH) removal. It would be illuminating to apply the same protocol as described in this reference and to use the crystals without the natural ligand for soaking experiments of the 2G3 fragment and its derivatives. This could be a reliable approach for the validation of the proposed binding mode of 2G3.
5.3 CONCLUSIONS

The presence of a cap structure at the 5’ end of the viral genome is required for DENV replication within host cells.\textsuperscript{[3-5]} As this is a cytosolic process, for cap maturation the virus cannot exploit the host methyltransferases, which are located in the nucleus. This makes the viral MTase an essential enzyme for viral replication and therefore a promising drug target.\textsuperscript{[3-5]}

In collaboration with Marseille University, a fragment based drug discovery (FBDD) study was started, aiming to discover a novel, potent anti-DENV drug. On the basis of structural and biochemical experimental data obtained in this institution, Molecular Modelling was used to aid the development of four of the seven active fragments that successfully inhibited the enzyme and produced soaked crystals of DENV3 MTase. On one of the active fragment (2A4) growing \emph{de novo} drug design strategies were applied for the improvement of the small molecule binding. With the same purpose, two other fragments (2E11 and 2C3) were linked and then grown with \emph{de novo} drug design approaches. Finally, a binding mode different from that observed in the crystal structure was suggested and evaluated for the most active compound, as the experimental data available could not explain such MTase inhibition. Even if DENV MTase is one of the most conserved DENV protein (with 60-70\% homology among the four serotypes), particular attention has been given to how well conserved interacting residues were, in order to direct the design exercise towards serotype-nonspecific agents.

Unfortunately, as both crystal structures and biochemical assays were based on DENV serotype 3, it cannot be assessed to date if this additional objective was accomplished.

FBDD is a multidisciplinary approach that has proven to successfully discover drug candidates (with one drug approval) starting from fragments with millimolar activity concentrations and for which structural data can be obtained. To the best of knowledge, this iterative process of small molecule optimisation and evaluation and the developed \emph{in silico} methodologies have never been applied before to this molecular target. Even if in the case of this project very few iterations have been performed to date, activity improvement (of even over 100-fold for 2’-O methylation) has already been achieved for one of the three studies presented here, indicating that the designed drug design method is moving in the right direction, confirming to be a suitable approach also for the development of MTase inhibitors. Future drug design and evaluation iterations will prompt further improvement of the molecules designed and discussed in this chapter, leading to potential novel, potent drugs against DENV infection.
5.4 METHODS

The computational methods for all the studies discussed in chapters from 5.2.1 to 5.2.3 are presented here. A description of the molecular modelling programs used can be found in APPENDIX 2.

HARDWARE DETAILS

All calculations were performed on a 8 core computer with Inter Xeon 2.80 GHz E5462 CPUs.

PROTEIN PREPARATION

In all the applications, the protein structure obtained from Marseille University was prepared by removing water molecules and adding missing hydrogen atoms, optimised with the MOE 2010.10\(^{[29]}\) Protonate 3D tool, considering a temperature of 300K, a pH of 7 and a salt concentration of 0.1M. Ligand molecules were retained or deleted according to necessity.

DESIGN OF SMALL MOLECULE DATABASES

All small molecule databases were prepared with MOE 2010.10\(^{[29]}\). Compounds were designed with the Builder tool, energy minimised with MMFF94x force field and steepest descendent method, using 500 iterations and gradient of 0.001 and preserving existing chiralities. Hydrogens and lone pairs were adjusted and partial charges were calculated. One conformation was saved for each molecule.

PHARMACOPHORE QUERIES

All pharmacophores were built with MOE 2010.10\(^{[29]}\) Query Editor, using the Unified scheme. Features characteristics and size were chosen according to occurrence. Essential features and partial matches were set according to necessity.
**MOE MEDCHEM TRANSFORMATIONS**

MOE 2011.10[30] MedChem Transformations was used for fragment 2A4 growth. The starting fragment was set as ligand and the atoms that had to be subjected to the medicinal chemistry transformations were selected according to the purpose of the calculation and the receptor was set as “Receptor atoms”. The first drug design attempt (R1 group modification for 2A4 fragment) used a MOE implemented transformation database ($MOE/sample/mol/GrowthRXN.mdb), while all the other design attempts used a database that contained the transformations implemented by the software ($MOE/sample/mol/GrowthRXN.mdb and $MOE/lib/medchem_rxn.mdb) and other defined transformations that were added. The new transformations were prepared singularly with ChemDraw Ultra 12.0 and imported in a mdb database through a svl command reported in the MOE online manual:[31]

```
run ['$MOE/lib/svl/run/medchemtrns.svl', [database, directory, opt], '_MedChemTrns_MakeMDB']
```

The number of iterations was decided on the basis of the results obtained and it was never higher than 5. Only molecules with molecular weight under 500 Da, SlogP between -4 and 8, total polar surface area (TPSA) between 40 and 140 and that matched at least 3 (for R1 group modification for 2A4 fragment) or 4 of the 8 features of a 3D pharmacophore were retained from the iterations. Three features of the pharmacophore query were set as essential. The generated structures were refined through a 500 iterations procedure with gradient of 0.001 and a pharmacophore restraint force constant. The resulting molecules were, then, visually inspected and pharmacophore searched, increasing the number of matching features, usually from 4 to 5. The most promising molecules or fraction of molecules were annotated for further investigation.

**MOE COMBINATORIAL BUILDER**

MOE 2011.10[30] Combinatorial Builder was used for fragment 2A4 growth. The starting fragment was set as ligand and the atoms that had to be subjected to molecule growing were selected according to the purpose of the calculation and the receptor was set as “Receptor atoms”. For molecule growing the default substituent’s library, with 69 R-groups, was used ($MOE/lib/buildfrag.mdb). Only molecules with molecular weight less that 500 Da, SlogP between -4 and 8, TPSA between 40 and 140 and that matched at least 4 of the 8 features of a
3D pharmacophore were retained from the iterations. Three features of the pharmacophore query were set as essential. The generated structures were refined through a 500 iteration procedure with gradient of 0.001 and a pharmacophore restraint force constant. The resulting molecules were, then, visually inspected and pharmacophore searched, increasing the number of matching features from 4 to 5. The most promising molecules or fraction of molecules were annotated for further investigation.

**LIGBUILDER**

The following LigBuilder 1.2\textsuperscript{[32]} tools were used for molecule growing: POCKET, GROW, PROCESS.

NS5 MTase structure, a modified 2A4 fragment (compound 19) and dummy atoms were used for binding site determination with POCKET. For pharmacophore determination construction the maximum number of features was set to 8 and the minimal feature distance was set to 3.5Å. The GROW process was performed with 20 generations and a maximum population of 3000 molecules for the genetic algorithm. The bioavailability rules were set as default: molecular weight between 300Da and 600 Da; logP between 3.00 and 6.00; hydrogen bond donor groups between 2 and 6; hydrogen bond acceptor groups between 2 and 6. Only molecules with a similarity index smaller than 0.90 were selected for the mating pool. Default chemistry rules and forbidden and toxic exclusions were used.

The results obtained with the LigBuilder program were then visually inspected and evaluated in MOE\textsuperscript{[29]} environment.

**MOE LINK MULTIPLE FRAGMENTS**

MOE 2011.10\textsuperscript{[30]} Link Multiple Fragments was used for 2E11 and 2C3 fragments linking. Three linking sites were chosen in the 2E11 fragment and three in the 2C3 fragment, adopting a total of four linking strategies employing two sites at the time. All the linking modes were based on modified structures of the 2E11 and 2C3 fragments. The atoms of the NS5 MTase were set as “Receptor atoms”. Linkers were taken from the default linker library ($MOE/lib/linker.mdb) that contained 22,603 linker conformations. Only molecules with molecular weight less that 500 Da, SlogP between -4 and 8 and TPSA between 40 and 140 were retained. The generated structures were refined through a 500 iteration procedure with gradient of 0.001. The resulting molecules were, then, visually inspected and the most promising ones were annotated for further investigation.
**MOE DOCKING**

The database of 2E11-2C3 derivatives designed on the base of the MOE\textsuperscript{[30]} Link Multiple Fragments tool was docked with MOE 2011.10\textsuperscript{[30]} Docking tool. After protein and small molecule database preparation, the binding site was identified using the original fragments. Ligand placement was restrained with the use of a four (essential) features 3D pharmacophore. The pose generation was conducted in semi-flexible conditions, with a conformational search on the ligands. The first rescoring was calculated with the default London dG scoring function, while the second one used was the GBVI/WSA dG scoring function. Before the second rescoring an energy minimisation was computed with AMBER99 force field, considering only those amino acids at a maximum 6Å distance from the ligands and maintaining the protein residues as fixed. This refinement was computed with 500 iterations and a 0.01 gradient and it was constrained by the use of the same pharmacophore as in the placement step. Duplicates were not removed after the refinement and a maximum of 30 poses were kept after the first refinement, while a maximum of 10 were kept in the second one. The output was then visually inspected.

**MAESTRO GLIDE DOCKING**

Maestro 9.3 Glide software\textsuperscript{[33]} was used for the docking of small molecule databases of 2A4 derivatives, compound 44 derivatives and of 2G3 analogues. In all the cases, prior to docking simulation the protein was processed with Maestro Protein Preparation tool using OPLS-2005 force field and ligand databases were prepared as previously described. The docking site was identified with Maestro Receptor Grid tool, setting the centre and the grid size according to necessity. All the docking simulations were performed with XP precision, using a semi-flexible approach with the sample ring conformation option for exhaustive conformational search of ligands. Ligands with more than 300 atoms or more than 50 rotatable bonds were automatically excluded. The non-bonded interactions were calculated with a 0.8nm van der Waals radii scaling factor and a 0.15nm partial charge cut-off. Only a maximum of 10 conformations were kept for each ligand and they were not energy minimised after the docking because a refinement procedure was applied on the docking output.

For the 2A4 derivatives, the structure of the MTase in complex with the fragment was used and the grid was defined, using the 2A4 fragment in the crystal structure as the centre and a length of 10Å.
For the compound 44 derivatives, a modified structure of the MTase in complex with this molecule was used. The symmetric unit of the crystal was reconstructed with MOE 2010.10[29] with the Superpose tool. The pdb structure of the reconstructed symmetric unit of the MTase-2C3 fragment complex was available and was used as a template as it superposed very well with both the compound 44 original monomer and its symmetric image. A copy of the original monomer from the compound 44 complex was superposed with the symmetric image of the 2C3 complex. The two protein chains representing the symmetric unit and compound 44 were saved. The side chain conformation of the Glu111 near the SAM binding site was then modified with MOE 2010.10[29] Rotamer Explorer and protonated with the Protonate 3D toll, considering a temperature of 300K, a pH of 7 and a salt concentration of 0.1M. The modified Glu111 and the nearby Arg84 were then energy minimised with AMBER99 force field and steepest descendent method, using 500 iterations and a gradient of 0.001. Hydrogens and lone pairs were adjusted and partial charges were calculated. For small molecule docking, the grid was defined using compound 44 as the centre and a length of 10Å.

For the 2G3 analogues, the structure of the MTase in complex with the fragment was used and the grid was defined, using the SAM molecule in the crystal structure as the centre and a length of 9Å.

**MOLECULAR DYNAMICS SIMULATIONS**

GROMACS 4.5.3[34] was used with AMBER99 force field for all the molecular dynamics (MD) simulations. A minimum 0.9 nm distance between the molecular system and the cubic box employed and periodic boundary conditions (PBC) were applied. Water was added and described as an explicit solvent and the molecules were treated with the TIP3P model. The total charges of the systems were neutralised. Small molecule missing topologies were computed with AMBER 12[35] Antechamber software, using the AM1-BCC charge method and GAFF force field for atom types. The obtained files were then converted with acppype[36] in a GROMACS-compatible format. Two consecutive energy minimisations were initially performed, employing the steepest descendent (SD) method first and then followed by the conjugate gradient (CG) method, with the aim of a faster process and a more accurate result. A force tolerance of 100 kJ mol⁻¹ nm⁻¹ was set for SD, while a 10 kJ mol⁻¹ nm⁻¹ force tolerance was applied with CG. In both cases the maximum number of iterations was set to 3,000 steps. Subsequently, a position restrain force of 1,000 kJ mol⁻¹ nm⁻² was applied to protein (and ligand) atoms for water molecules relaxation. Aiming to a smoother equilibration, two consecutive 50 ps (50,000 steps, step size of 1 fs) position restrained MDs were consequently performed, saving coordinates, velocity and energy values every 500 steps. NVT conditions
(constant number of atoms N, volume V and temperature T) were used for the first position restrained MD and a v-rescale temperature coupling and 0.1 ps time constant was used to heat the system to 300 K. NPT conditions (constant number of atoms N, pressure P and temperature T) were used for the second position restrained MD, with the use of both temperature (v-rescale, temperature 300 K and time constant of 0.1 ps) and pressure (Berendsen algorithm, 1 bar pressure and time constant of 0.5 ps) coupling. The production simulations were run in NPT conditions and settings (number of steps, step size, frequency of data saving) were different in the two studies (2E11-2C3 derivatives optimisation and 2G3 binding pose validation). In all simulation stages, long-range electrostatic interactions were calculated with the Particle-Mesh-Ewald (PME) method with a 0.9 nm short range cut-off and short-range non-bonded interactions were computed only within a cut-off of 1.4 nm. Results were visually inspected with VMD and analysed with GROMACS 4.5.3 tools and Grace. Four MD simulations were performed for the 2E11-2C3 derivatives optimisation study. In all four cases the protein used was the reconstructed symmetric unit (with the original Glu111 conformation) of the MTase-compound 44 complex obtained with MOE 2010.10 Superpose tool (see MAESTRO GLIDE DOCKING chapter of this METHODS section). The four molecular systems were the protein without ligands and in complex with either compound 44, 53 or 54. The production simulation was 5ns long (2500000 steps with step size of 2fs) and coordinates, velocities and energy values were saved every 1,000 steps.

Three MD simulations were performed for 2G3 binding mode validation. In all three of them the molecular system was the MTase-2G3 complex obtained from crystallographic data. His110 protonation was manually set and different in the three simulations (HIE110 in the first, HID110 in the second and HIP110 in the third). The production simulation was 3ns long (1500000 steps with step size of 2fs) and coordinates, velocities and energy values were saved every 1,500 steps.
5.5 BIBLIOGRAPHY


[28]: Protein Data Bank, [www.pdb.org](http://www.pdb.org) (Accessed, October 2014)


[31]: David C. Young, COMPUTATIONAL DRUG DESIGN - A Guide for Computational and Medicinal Chemists, (2009), John Wiley & Sons, Inc. Publication


Section 6:
NS5 POLYMERASE AS A DRUG TARGET
6.1 INTRODUCTION

A PROMISING DRUG TARGET

The replication of the viral genome is an essential step of the DENV replication cycle. The lack of viral RNA synthesis would indeed produce viral particles devoid of the nucleic acid. Even if these pseudovirions would be able to enter new cells, the synthesis of novel virions would not be possible due to the absence of viral genetic information. As a consequence, the infection could not proceed. DENV genome replication occurs in particular membrane structures in the host cell cytoplasm and it consists of RNA synthesis from a RNA template. Human cells do not have polymerases that are able to use RNA as a template and this explains the reason why (+)ssRNA viruses, like DENV, have evolved to encode for a polymerase protein.[1,2] DENV polymerase is the NS5 RNA-dependent RNA polymerase (NS5 RdRp). It is the most conserved viral protein in DENV and as it is one of the key enzymes in RNA replication, its inhibition would arrest the DENV replication cycle within the host cell. This, in addition to the fact that human cells do not have this enzyme, the importance of RNA replication itself and the drug design experience from other viruses make NS5 RdRp one of the most attractive targets for DENV infection inhibition.[3-5]

STRUCTURE

DENV polymerase is the C-terminal portion (residues 273-900) of the NS5 protein. General considerations on the full length protein are reported in the introduction to the NS5 METHYLTRANSFERASE AS A DRUG TARGET section.

The RdRp domain assumes the typical right-handed shape, as shown in Figure 6.1. It has overall dimensions of 65 x 60 x 40 Å and it displays a total of 27 α-helices and 7 β-strands. It comprises a palm (residues 497-542 and 601-705, in green), fingers (residues 273-315, 416-496 and 543-600, in red) and a thumb (residues 706-900, in blue) sub-domains. Differently from other polymerases, the fingers and the thumb are connected through the protrusion of one fingertip.[5,6] Furthermore, it presents two nuclear localization sequences (NLS, residues 316-368, in yellow) that are recognised by cellular factors that transport the polymerase into the cell nucleus. These sequences of approximately 20 amino acids each are strictly
conserved, suggesting that the nuclear localisation of RdRp could play an essential role in DENV replication, but the reason for this is still unknown.\textsuperscript{[6]}

DENV RdRp is a primer-independent polymerase (or de novo polymerase), which means that it is able to initiate RNA synthesis without the aid of a co-factor usually defined as a primer. The initiation platform is therefore given by a loop that projects towards the palm sub-domain from the thumb, called the priming loop or the P-loop (residues 782-809, in purple in the figure).\textsuperscript{[5,6]}

![RdRp structure](image)

Figure 6.1. RdRp structure.

The domains are indicated with different colours: the palm is in green, the fingers are in red, the thumb is in blue, the nuclear localisation sequences are in yellow and the priming loop is in purple. Two arrows show the entrance to the two tunnels described in the text.

The right-hand shape allows the definition of two perpendicular tunnels that intersect in the polymerase active site, where the RNA synthesis reaction occurs. The active site is in the palm sub-domain and comprehends the catalytic triad (Gly662, Asp663 and Asp664) and two Mg\textsuperscript{2+} catalytic ions. The first tunnel allows the template ssRNA entry. It is located between the fingers and the thumb (as shown in Figure 6.1) and reaches the palm sub-domain. Differently, the second tunnel spans the whole protein. From the side shown in the figure, the incoming NTP enters and reaches the active site, where it is incorporated in the nascent RNA strand that forms a dsRNA with the template. The dsRNA, then, exits through the same tunnel, but on the other side of the protein.\textsuperscript{[5,6]} Five crystal structures of DENV RdRp are available in the apo form or in a complex with an inhibitor and in all of them, the exit of the second tunnel is too small to allow the exit a dsRNA molecule. This is due to the fact that in all the structures, the protein is in the pre-initiation state, or closed conformation.\textsuperscript{[6-8]} During the elongation
6.1 INTRODUCTION

process NS5 RdRp changes conformation and the clash with the P-loop is removed, enlarging the exit cleft. However, no structural knowledge about this process has been acquired to date.\textsuperscript{[5,6]}

\textbf{POSSIBLE MOLECULAR TARGET SITES}

Polymerases are validated drug targets in antiviral research and two main approaches are possible for this enzyme: the nucleoside and the non-nucleoside one.\textsuperscript{[3,5,9]}

Nucleoside inhibitors (NI) are, from the name, nucleoside analogues that aim to interact with the polymerase at its active site, on the palm domain, following host kinase conversion into its triphosphate form (see the main \textit{INTRODUCTION} section of this thesis). The polymerase catalyses the reaction of insertion of the compound into the nascent RNA strand. The compound then inhibits RNA synthesis either through a “chain termination” mechanism (it does not allow the addition of further nucleosides), or it introduces errors in the genetic information. An example of NI compounds is NITD203, whose development was halted at the preclinical stage for toxicity problems.\textsuperscript{[3,4,9]}

Non-nucleoside inhibitors (NNI) are drug-like small molecules that do not interact with the polymerase active sites. Their mechanism of action is usually allosteric, which means that they produce conformational changes in the protein, inhibiting its activity. This strategy is less likely to produce high toxicity, but is more prone to drug resistance because it involves areas of the enzyme that were less subject to evolutionary pressure.\textsuperscript{[3,5,9]} In DENV polymerase four allosteric sites can be defined, as shown in Figure 6.2, where two orientations of NS5 RdRp are presented and the entry site of NTPs is indicated as well. From the analysis of the crystal structure, two cavities (cavity A and B, in light blue and in light purple in the figure respectively) were found in the thumb sub-domain. Unfortunately, neither of them has the ideal characteristics for drug design: cavity A does not display residues that could be identified as essential using mutagenesis studies; while cavity B is too small to accommodate drug-like molecules.\textsuperscript{[3,5]} Another allosteric site (site C, in green in the figure) was used for the evaluation of a class of compounds (e.g. NITD-1 and NITD-107) that bound to the enzyme in a pocket, near Thr413, where the RNA template is supposed to interact with the protein. However, low cell-based activity is slowing the development of these compounds.\textsuperscript{[7,10,11]} Another target site (site D, in red in the figure) was selected in one of the studies presented here and is more deeply discussed further in the report. This site is located on the priming loop and the compounds binding with Trp795 and Arg737 were aimed at blocking the enzyme in a non-active conformation.
**MECHANISM OF ACTION**

The polymerases catalyse RNA replication in two stages. The first one is *de novo* initiation of synthesis. In this process, the enzyme uses the P-loop, probably Trp795, to place the incoming NTP ("n") in the correct position for hydrogen bonding with the 3’ end nucleotide of the template. Then, when the second NTP ("n+1") enters the cleft it stacks with the first NTP and interacts with the second 3’ end nucleotide of the template. The correct positioning of the molecular system is aided by the presence of residues (Arg737, Arg729 and Ser710) that interact with the NTP phosphates. Once this configuration is formed, the catalytic aspartates and the catalytic Mg²⁺ catalyse the nucleophilic attack of the “n” 3’OH to the “n+1” α-phosphate for the formation of a new phosphate link between the two nucleotides, with the elimination of a pyrophosphate molecule.\[^{3,5,6}\] No protein structures of NS5 RdRp in complex with the template are available and therefore there is no accurate knowledge of the initiation process of the RNA synthesis. However, it was possible to construct models of this mechanism using other structural data, like the bacteriophage Φ6 RdRp in complex with the template and two NTPs. These models allowed the formulation of the initiation hypothesis presented above.\[^{3,6}\]

Following the initiation process, there is an elongation step that increases the nascent RNA length, on the bases of the template sequence. As already mentioned, in order to pass from the initiation to the elongation step, the protein must change conformation, allowing a bigger cleft for the exit of the resulting dsRNA. Also in this case, there is no structural information that could explain the exact molecular changes that occur. However, the high temperature
factors of the finger sub-domain in DENV crystal structures suggest that movement of this portion could be involved in the conformational change.\textsuperscript{[6]}

**AIMS AND OBJECTIVES**

Essential for RNA replication and a validated target, DENV NS5 RdRp is probably the most attractive molecular object for antiviral drug design. Two strategies have been historically applied to the design of inhibitors of this enzyme: nucleoside (NI) and non-nucleoside inhibitors (NNI). Pro and cons can be characterised for both class of drugs and have been discussed in this thesis. Given the lower toxicity rate and the higher suitability for molecular modelling drug design of NNI, this was the preferred strategy in this PhD. With the aim of designing novel, potent DENV inhibitors, two virtual screening campaigns were carried out on two putative allosteric sites. In the first case, the protein region close to the priming loop (site D) was chosen for the design of compounds that could inhibit RNA synthesis initiation. This study is discussed in chapter 6.2.1. In the second study, on the basis of the structural knowledge of known compounds, a second allosteric binding site (site C) was chosen and the results are discussed in chapter 6.2.2.
6.2 RESULTS AND DISCUSSION

6.2.1 ALLOSTERIC INHIBITORS BINDING SITE D

RdRp is a well known target that has been used for drug design studies for dengue and other viruses. For instance, both nucleosides and allosteric inhibitors were designed against HCV, a pathogen that has been taken as a model of drug discovery in other Flaviviridae viruses.\[^9\] As the aim of this work was to design novel NNIs against DENV, it was appealing to capitalise on the knowledge gained from HCV NNIs. However, overlapping the two RdRp structures the sites of HCV for the NNIs do not correspond to similar pockets in the DENV structure and consequently HCV inhibitors cannot be used as a starting point for DENV.\[^5,9\]

In this study, the choice of the allosteric binding site was based on the knowledge that one particular residue (Trp795, on the priming loop, P-loop) is thought to be essential for RNA synthesis initiation.\[^5,6\] Consequently, it was hypothesised that compounds binding in the region near this residue could inhibit this essential step of viral replication. A structure-based virtual screening was therefore designed for the selection of putative RdRp NNIs using as a start the SPECS\[^{12}\] database (updated to February 2012) comprising around 209,945 molecules.

PROTEIN MODELS

At the time of the study, only one series of two crystal structures (PDB IDs: 2J7W and 2I7U) were available in the protein Data Bank.\[^6,13\] They both presented the structure of DENV3 NS5 RdRp domain, but in one of the structures (PDB ID 2I7W) the protein was co-crystallised with a 3'dGTP molecule, a non-specific RNA polymerase inhibitor. The small molecule phosphate portion was resolved, differently from its sugar and the base moieties. On the basis of the position of this portion of the inhibitor within the crystal structure, it was suggested that in the initiation of RNA synthesis the conserved residue Trp795 mimics a primer for the stacking of the first nucleotide of the nascent RNA strand.\[^6\] This residue was consequently chosen as a target residue for the drug design study discussed here, with the hypothesis that RNA synthesis initiation could be blocked through the disruption of the first nucleotide-Trp795 stacking. By superposition of the two structures belonging to this crystallographic series, it appeared that one of the residues responsible for nucleotide binding (Arg737) has two different conformations depending on the interaction with the inhibitor, as presented in
Figure 6.3. Thus, it was hypothesised that inhibition of DENV NS5 RdRp could be enhanced through blocking the conserved residue Arg737 in the 3’dGTP unbound conformation. On the base of these assumptions the enzyme conformation of the 2J7U crystal structure and the protein region highlighted in Figure 6.3 and presented as site D in the introduction to NS5 RdRp was chosen as a target site for the discovery of novel RdRp NNIs.

![Figure 6.3. NS5 RdRp site for allosteric inhibitor design.](image)

On the left is shown the whole protein with NLS regions in yellow, palm domain in green, thumb domain in blue, fingers domain in red and the priming loop in purple. On the right is represented the overlap of the 2J7U and 2J7W structures in proximity of the priming loop. Here the interactions of the triphosphate moiety of 3’dGTP molecule (from 2J7W) are displayed with blue dotted lines. The 2J7W residues are coloured in cyan and the 2J7U residues are coloured in orange. The only residue with a different conformation is Arg737.

In both structures of the crystallographic series available at the time of the study, three protein fragments (residues Val310-Ala316; Ala406-Asp419; Tyr451-Ala472) are missing.\[^6\] One of them in particular, identified as “loop 3” (L3, Ala406-Asp419) is located in direct proximity of the selected target binding site. Consequently, a model of the complete protein had to be built through a homology modelling approach prior to the application of the drug design exercise. As sufficient homology with complete polymerase structures could not be found, the more complete RdRp structure from WNV (PDB ID: 2HFZ\[^14\]) was chosen as a template because of the high residue homology (70%) between the two proteins. A higher residue identity generally corresponds to a higher probability of obtaining a reliable model of the protein of interest. Statistical analyses showed that for big proteins like RdRp, a minimum of 20-30% of residue identity should be aimed for in order to have confidence in the constructed model.\[^15,16\] Therefore, 70% homology was considered a good value. Aiming to maintain the crystallised part of the protein in the original conformation, only the missing parts of DENV RdRp were modelled, according to the procedure described in the METHODS section. The amino acids at the boundaries of the missing fragments were modelled as well with the aim of obtaining the most reasonable structure possible. The obtained model (model A) superposed very well with the original structure, with a RMSD value of 0.60Å. Even if not
conserved in all four DENV serotypes, the potential importance of Thr413 in allosteric inhibitor binding has been reported in the literature.\textsuperscript{[7,10,11]} The loop containing this residue was modeled in two different conformations. A second model (model B) was therefore built for DENV RdRp, with Thr413 conformation being the only difference between the two homology models. Also model B superposed very well with the original structure, with a RMSD value of 0.60Å. Figure 6.4 shows the difference in Thr413 conformation in the two models.

The quality of the models was assessed through Ramachandran plot evaluation that showed that more residues were outliers compared to the original structure. The outliers belonged to protein sites far from the selected target site and most of them were however very close to the allowed region of the psi-phi plot. The quality of the two models was consequently considered suitable for the purpose of this study.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.4.png}
\caption{Superposition of models A and B.}
Protein backbone is represented in grey and Trp795 is reported for better understanding of Thr413 position. Thr413 carbon atoms from model A are in cyan, while those from model B are in orange.
\end{figure}

Two crystal structure series of DENV3 NS5 RdRp were published in 2013.\textsuperscript{[7,8]} In one of them (PDB IDs: 4HHJ and 3VWS), a new crystallisation technique was applied. Of the two structures (PDB ID: 3VWS) one had an allosteric inhibitor co-crystallised and resolved and the protein adopted a different conformation with respect to all the other structures.\textsuperscript{[7]} In the other publication, additional residues belonging to the MTase-RdRp linker was shown to stabilise the protein. Some of the linker residues were resolved in the crystal structure (PDB ID 4C11).\textsuperscript{[8]}

In the two new crystal structures of the NS5 RdRp with no ligands (PDB IDs: 4HHJ and 4C11), the Val310-Ala316 missing loop of 2J7U could be defined, whilst the other two loops were still missing. Superposing the homology models on them, RMSD values of 0.73Å and 0.86Å were obtained for 4HHJ and 4C11 respectively, showing a good overlap. Furthermore, the superposition of the loop Tyr308-Gly322 that was modelled in the present study and that was resolved in the new structures returned RMSD values of 1.10Å and 1.13Å for 4HHJ and 4C11 respectively. These low RMSD values supported the reliability of the model and confirmed that this region has a random coil secondary structure. In the new structures more residues were resolved also in proximity of the other two missing sequences of 2J7U, but they were not informative enough on the quality of models A and B.
IDENTIFICATION OF ALLOSTERIC INHIBITORS THROUGH VIRTUAL SCREENING

A virtual screening strategy was designed for the selection of compounds from the small molecule drug-like database SPECS containing around 209,000 molecules. Differently to that done in previous studies the 2D ligand structures were directly prepared with Maestro LigBuilder tool that generated all the possible ionisation states at pH 7 (±2), all the possible tautomers and all combinations of possible stereoisomers for each molecule and scored each one of them. Approximately 498,000 structures were generated. As this was considered to be already a high number of structures to dock in the selected site, it was decided not to further explore the molecules’ conformations. The prepared ligands were then used for the virtual screening approach schematised in Figure 6.5. The same approach was used for both homology models (model A and B) separately.

Briefly, the compounds were initially docked and scored with Maestro Glide in the two models’ grid centred on Trp795 with high throughput virtual screening (HTVS) precision mode. The docking was implemented in a semi-flexible way in order to explore the molecules’ conformational space. Compared to other precision models, HTVS has a lower accuracy of score values, but it was considered sufficient for a starting point to reduce the number of compounds. A quick visual inspection of the results showed that not all the compounds were binding the enzyme near Trp795. Consequently, the output databases were filtered with a one-feature pharmacophore in MOE environment that selected only those compounds that bound to the protein in the desired area. Only the best 5,000 unique molecules that had a negative docking score were then saved for further analysis.

These structures were then docked a second time in the models using a semi-flexible approach with Maestro Glide SP mode and rescored with LeadIT FlexX and PLANTS. For each one of the 5 compound conformations saved, the scores obtained with the three softwares were collected with a consensus score and a normalised consensus score function (Equations 6.1 and 6.2 in the METHODS chapter). In this way, the two best scoring conformations for each compound that resided within the best 25% of the molecules in all the used programs were saved.
Figure 6.5. Schematic representation of the virtual screening methodology.

The main steps are indicated and the numbers are referred to the number of molecules (conformations in brackets) at every step. Homology models are reported with the letters A for model A and B for model B. Molecular modelling programs and scoring functions are abbreviated as follows: HTVS= Maestro Glide HTVS mode; SP= Maestro Glide SP mode; FX= LeadIT FlexX; PL= PLANTS; CS= consensus score; NCS= normalised consensus score; XP= Maestro Glide XP mode.
The resulting conformations were then docked a third time in a semi-flexible manner in the two homology models using Maestro Glide with XP mode. The final results were collected and a comparison between the two models was made as well as the visual inspection of all the poses. The visual inspection of the compounds highlighted that there were different preferred binding modes, as displayed in Figure 6.6. Some molecules, like the one in yellow, bound in the cleft going across Trp795, from a smaller pocket near Arg737 (“P1”) to another small pocket on the other side (“P2”). Interactions primarily involved a \( \pi-\pi \) stacking interaction with Trp795 and often direct interactions with Arg737, as intended in the virtual screening approach design. Other poses showed that molecules, like the one in cyan, extended from P1 to another small pocket (“P3”) situated below Arg737, in the same region where the 3’dGTP was located in the 2J7W structure. In this case, molecules did not interact directly with Trp795, but their position could impede the movement from the “inactive” to the “active” conformation of Arg737, leading to a loss of nucleotide capability of the polymerase. Smaller molecules with respect to the ones described above (like the molecule in red and the one in dark green in the figure), bound the protein mainly in P1. This group of molecules gave a \( \pi-\pi \) stacking with Trp795 and most of the structures interacted directly with Arg737.

![Figure 6.6. Preferred compounds binding modes.](image)

Here, examples of molecules binding the RdRp site are reported as compounds coloured in red, yellow, cyan and dark green. The reference protein residues (Thr413, Arg737 and Trp795) are in light green and the priming loop backbone is shown as a purple ribbon. The protein surface is reported in grey. P1 = pocket 1; P2 = pocket 2; P3 = pocket 3. See text for pockets descriptions.

Molecules were ranked according to the final docking scores, classified on the basis of the pose (as described above) and their chemical structures were clustered with the aim of
selecting compounds as different as possible. In pose evaluation, molecules that bound to conserved residues among the four DENV serotypes were preferred to the others in order to direct drug design towards panserotypic inhibitors of DENV infection. As the only difference between the models was the Thr413 side chain conformation and since this residue was not involved in all the binding modes (Figure 6.6), several compounds matched both models selection. These molecules were generally preferred. Thus, 29 molecules were chosen and purchased from SPECS.\[12\] Their chemical structures are reported in APPENDIX 8.

### IN VITRO ACTIVITY OF THE SELECTED COMPOUNDS

All 29 purchased compounds were tested by our collaborators in the Rega Institute for Medical Research, K. U. Leuven (Belgium) with the cell-based in vitro assay described in the NS3 HELICASE AS A DRUG TARGET section. Two further compounds (77 and 78 in APPENDIX 9) were synthesised by a student in our laboratory as part of a structure-activity study of compound 79 (see APPENDIX 8) and tested in vitro.\[17\]

The results showed that one compound (compound 80, in panel A of Figure 6.7) is able to inhibit DENV2 infection with an EC\textsubscript{50} of 2.63μM and a CC\textsubscript{50} greater than 236μM. The assay was repeated twice more for results validation and the EC\textsubscript{50} values were confirmed to be very similar (2.71μM and 3.80μM). In the same assays, whilst the colorimetric assay displayed a CC\textsubscript{50} value greater than 117.95μM, a visual inspection of the cells returned CC\textsubscript{50} values of 25.94μM and 52.84μM. Thus, a therapeutic window was found although needing improvement.

This compound, another 10 of the 29 selected and compounds 77 and 78 were also tested on DENV NS5 protein with binding and enzymatic assays in the AFMB, Marseille University laboratories by the author (binding assay) or by our collaborators (enzymatic assay).
Unfortunately, compound 80 neither bound to the enzyme (with $\Delta T_m$ of $-1.13^\circ C$ in the thermal shift assay) nor inhibited the RNA polymerisation in the enzymatic assay performed with $^3$H-dATP. However, a better inhibition value was observed in the polymerase enzymatic assay using the fluorescent PicoGreen probe. The description of this assay can be found in the *INTRODUCTION* section of this thesis and the IC$_{50}$ obtained was 31.7 µM (±8).

As already mentioned, one compound (81, shown in Figure 6.8) was able to bind to NS5 full-length protein in the TSA. It gave a significant increase of Tm of 2.93°C, whilst SAM, a known binder of the MTase portion of NS5, produced a Tm shift ($\Delta T_m$) of 4.26°C.

![Figure 6.8. Chemical structure of compound 81.](image)

In order to confirm that the stabilisation of the protein and the consequent increase in melting point were correlated with compound binding, titration curves for both compound 81 and SAM were performed. The obtained melting temperature shifts are reported in Table 6.1 and plotted in Figure 6.9. The thermal shift assay does not give exact dissociation constant (K$_d$) values and this calculation was not performed for either molecules. However, titration results demonstrated a concentration-dependence in the Tm shift, supporting that both SAM and compound 81 could bind to DENV full length NS5 protein.

<table>
<thead>
<tr>
<th>Concentration (µM)</th>
<th>$\Delta T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compound 81</td>
</tr>
<tr>
<td>1000</td>
<td>2.43318</td>
</tr>
<tr>
<td>500</td>
<td>1.97781</td>
</tr>
<tr>
<td>250</td>
<td>2.44157</td>
</tr>
<tr>
<td>125</td>
<td>2.28147</td>
</tr>
<tr>
<td>62.5</td>
<td>1.78645</td>
</tr>
<tr>
<td>31.25</td>
<td>1.81042</td>
</tr>
<tr>
<td>15.625</td>
<td>1.30409</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Despite the above results, compound 81 did not exhibit inhibitory activity of the polymerisation reaction. This might be explained in two ways: either the compound bound to the enzyme without interfering with its activity or it did not bind to the RdRp domain of the
NSS full length protein, thus, not producing the desired effect. Furthermore, in the cell-based \textit{in vitro} assay, the molecule did not display any antiviral activity and proved to be very cytotoxic with a $CC_{50}$ less than 1.11µM. Given these results, it was decided not to further develop this molecule, in favour of others potentially more promising.

![Tm shifts](image)

\textbf{Figure 6.9. Titration curves for compound 81 and SAM.}

The Tm shift is reported for compound 81 (black diamonds) and SAM (grey squares) at the tested concentrations.

A third compound that gave positive results was compound 82, shown in Figure 6.10. This was the only molecule that gave a meaningful concentration-dependent inhibition of the RdRp activity in the biochemical assay. \textit{IC}_{50} values were determined for this molecule and also for compounds 80 and 81 because they were found positive in other assays. In order to increase the assay sensibility, a mini-genome RNA template (constituted of the 5’ and 3’ UTRs of DENV genome) was used instead of the polyU strand.

![Chemical structure of compound 82](image)

\textbf{Figure 6.10. Chemical structure of compound 82.}

As reported in Table 6.2, \textit{IC}_{50} determination confirmed that the only compound that significantly inhibited the polymerase activity was compound 82, with an \textit{IC}_{50} of 47.1µM (±3.3 St. Dev.). Repeating the assay using PicoGreen instead of the tritium labelled dATP and a polyU RNA strand, the obtained \textit{IC}_{50} value was 35.1µM (±1.6 St. Dev.). Unfortunately, the TSA did not show protein binding for compound 82 (with ΔTm of -1.99°C) and no antiviral activity could be detected, while cytotoxicity was observed at high concentrations (\textit{CC}_{50} of 138µM) for compound 82. The TSA should be repeated in order to
assess if this result is due to an actual lack of interaction or if the assay was not suitable to
detect the compound binding because of its hydrophobicity. The lack of activity in the cell-
based assay could be explained by many factors, like lack of good permeation in the cell of
early metabolism of the compound.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>IC\textsubscript{50} (µM ±St. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>124.7 ±22.5</td>
</tr>
<tr>
<td>82</td>
<td>47.1 ±3.3</td>
</tr>
<tr>
<td>81</td>
<td>148.2 ±34.3</td>
</tr>
</tbody>
</table>

Summarising, two promising compounds (80 and 82) were identified with the designed virtual
screening approach and consequently two classes of compounds, one for each molecule, were
further developed with the assessment of analogue molecules that were either purchased
from SPECS or synthesised and tested in vitro. For the analogue compounds, the CPE assay
(performing by our collaborators in the Rega Institute for Medical Research, K. U. Leuven) was
used to evaluate their activity in infected cells and the polymerase activity inhibition with
PicoGreen fluorescent probe assay was performed by J.-C. Guillemot group in AFMB, Marseille
University to assess their activity on the target protein.

**COMPOUND 80 ANALOGUES**

Compound 80 was selected from the SPECS database through the virtual screening approach
discussed previously. The virtual screening approach applied to both models returned this
molecule as a positive hit. Visual inspection of the binding in the two models showed that the
pose was the same and it is presented in Figure 6.11. However, only the binding to model B is
presented for a better image clarity. Using the reference areas of Figure 6.6, the molecule
spanned from the P1 to the P2 regions and the central phenyl ring was placed on top of
Trp795, giving a π-π stacking with this residue. Other hydrophobic interactions involved the
chlorine substituted phenyl ring that was placed near the lipophilic portions of the Glu415 and
Gln742 side chains. All the amino groups were involved with hydrogen bonds and the
counterparts were the backbone carbonyl oxygens of Glu414 and Thr413. The third phenyl
ring was perpendicular to Arg792, suggesting that it might interact with this residue through a
positive charge-π interaction. All but one (Thr413) of the mentioned residues are well
conserved in identity or in properties in the four DENV serotypes, complying with the
possibility of designing panserotypic inhibitors of DENV. Thr413 is a Gln in DENV4 and a Val in
DENV1. However, this was not considered an issue as the ligand binding involves the backbone of this residue.

As described above, compound 80 showed a low micromolar activity (EC$_{50}$ between 2.61µM and 3.80µM) in the inhibition of DENV infection in the CPE assay. Unfortunately, the polymerase activity inhibition assay was not clear and returned different values in two IC$_{50}$ determinations. In the first screen performed with only three different concentrations, this compound did not show a promising polymerase inhibition and the IC$_{50}$ determination with the mini-genome RNA strand and tritium labelled ATP nucleotide returned an IC$_{50}$ of 124.7µM. Conversely, the repetition of the IC$_{50}$ determination with the assay performed using a polyU template and the PicoGreen fluorescent probe gave a better inhibitory response with IC$_{50}$ of 31.7µM. Consequently, at this stage it was difficult to confirm that the cell-based activity derived from the inhibition of DENV NS5 RdRp through compound binding as suggested by the molecular modelling results.

With the aim of confirming the activity and to gain further insights on preliminary structure-activity relationships of compound 80, thirteen analogue molecules were purchased from SPECS and tested in vitro with CPE and with the polymerase activity inhibition assay, using a polyU template and PicoGreen fluorescent probe. The selection of these compounds was also based on the binding mode suggested by the molecular modelling results and shown in Figure 6.11, as a further objective was also to validate this model. Table 6.3 contains the structures and the assay results for compound 80 and the purchased analogues. Here, the indicated CC$_{50}$
and EC₅₀ values were obtained from the CPE assay and the selectivity index (SI) is also reported. The IC₅₀ values were acquired with the enzyme inhibition test. As highlighted in the table, the central core of the molecule was maintained and the analogues differed from compound 80 according to the substituents R₁, R₂ and R₃. Briefly, compound 83 was purchased to assess the importance of the phenyl moiety. Compounds from 84 to 91 were selected for the evaluation of the amide group and compounds 88 to 91 in particular were selected for the possibility of substituting the original group with a sulphonamide moiety. Compounds 92 to 95 were selected to explore the possibility of expanding the hydrophobic portion of the molecule with the addition of another phenyl ring.

Table 6.3. Structures and results for compound 80 and analogue molecules.

<table>
<thead>
<tr>
<th>Scaffold:</th>
<th>R₁</th>
<th>R₂</th>
<th>R₃</th>
<th>CC₅₀ (µM)ᵃ</th>
<th>EC₅₀ (µM)ᵇ</th>
<th>IC₅₀ (µM ± St. Dev.)ᵇ</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>benzyl</td>
<td>H</td>
<td>benzyl</td>
<td>236</td>
<td>2.7</td>
<td>31.7 ± 8</td>
</tr>
<tr>
<td>83</td>
<td>H</td>
<td>H</td>
<td>&gt;152.7</td>
<td>&gt;152.7</td>
<td>&gt;400</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>H</td>
<td>phenyl-2</td>
<td>benzyl</td>
<td>10.1</td>
<td>12.33</td>
<td>25.4 ± 6.3</td>
</tr>
<tr>
<td>85</td>
<td>H</td>
<td>phenyl-2</td>
<td>benzyl</td>
<td>&gt;130.4</td>
<td>&gt;130.4</td>
<td>283.1 ± 33.5</td>
</tr>
<tr>
<td>86</td>
<td>H</td>
<td>phenyl-2</td>
<td>benzyl</td>
<td>&gt;125.8</td>
<td>&gt;125.8</td>
<td>144.1 ± 17</td>
</tr>
<tr>
<td>87</td>
<td>H</td>
<td>piperidin</td>
<td>benzyl</td>
<td>&gt;136.1</td>
<td>61.5</td>
<td>127.4 ± 36.5</td>
</tr>
<tr>
<td>88</td>
<td>H</td>
<td>phenyl-2</td>
<td>benzyl</td>
<td>&gt;113.8</td>
<td>54.5</td>
<td>18.2 ± 1.8</td>
</tr>
<tr>
<td>89</td>
<td>H</td>
<td>phenyl-2</td>
<td>benzyl</td>
<td>&gt;106.5</td>
<td>&gt;106.5</td>
<td>4.1 ± 0.6</td>
</tr>
</tbody>
</table>
Unfortunately, none of the tested analogues improved cell-based DENV infection inhibition nor increased the selectivity index of 87 for compound 80. Conversely, some enhancements in the inhibition of the polymerase activity were achieved for some molecules. Some molecules were active in both assays, suggesting that this series of molecules might exert their antiviral function through the inhibition of the polymerase, thus supporting NS5 RdRp as a promising drug target for anti-DENV drug design. However, the lack of correspondence between some cell-based and biochemical results, suggests that other not yet identified cellular pathways might be involved in the cell-based activity of these compounds. In addition, lack of penetration or intracellular sequestration could also be implicated, but has not been investigated to date. Nevertheless, preliminary structure-activity relationships could be hypothesised on the basis of the data presented in Table 6.3 and are schematised in Figure 6.12.
Importantly, the lack of both cell infection and enzyme inhibition shown by compound 83 suggests that the presence of an aromatic group at the end of the R₁ or R₂ is essential for the activity of this class of compounds. This is in line with the docking results that demonstrated the placement of the chlorine phenyl ring in a hydrophobic area of the binding site. This essential feature is represented with an orange circle in Figure 6.12.

The removal of the amide group from R₁ or R₂ in compound 84 did not improve enzyme inhibition and increased the toxicity of the molecule, as the infection inhibition was associated with cytotoxicity. Similarly, the inversion of the amide in R₁ or in R₂ (compounds 85 to 87) reduced activity in both cell- and enzyme-based assays. Collectively, these results suggest that the amide is an important moiety for compound activity and that it must be placed in R₁ with the nitrogen linked to the central core. This complies with the proposed binding pose showed in Figure 6.11, where the amide NH group interacts with Glu414 through a hydrogen bond. This is represented in Figure 6.12 with a blue circle. Compounds 89 to 91 and compound 95 however suggested that the amide could be substituted with a sulphonamide in R₂, despite the nitrogen being not linked to the central core (as represented in Figure 6.12 by the blue circle). All of these improved the inhibition of RdRp seen with compound 80 suggesting that this could be a promising substitution. Unfortunately, this increased inhibition was at variance with the binding pose of compound 80, as the substitution would impede the interaction with Glu414 carbonyl oxygen. The difference in conformation of a sulphonamide with respect to an amide could yet produce a different binding mode for these compounds. Unfortunately, all of the compounds with the sulphonamide group had a worse selectivity index than compound 80 displaying adverse effects at high concentrations and some of them were not active against the infection.

Particularly outstanding was compound 90 that when compared to the reference molecule showed a comparable EC₅₀ value (4.87µM against 2.7µM) and an improved IC₅₀ value (3.7µM against 31.7µM). Structurally different from the other sulphonamide molecules, it contains two condensed rings in R₃, suggesting that both infection and enzymatic inhibitions could be improved by increasing the size of the hydrophobic group in the region represented by a green circle in Figure 6.12. This was supported by the group of molecules (from 92 to 95) with two phenyl rings in this region. All of them improved the inhibition of RdRp in respect to compound 80, suggesting that this could be a promising substitution. This is in accordance with the proposed binding pose showed in Figure 6.11, where the aromatic ring of compound 80 interacts with Arg792 in an area of the binding site that is large enough to accommodate a larger hydrophobic or aromatic moiety. Molecules belonging to this cluster have very different groups for the regions highlighted in orange and in blue in the scheme of Figure 6.12, but displayed comparable polymerase inhibition, making the determination of SARs more challenging. Conversely, the CPE assay demonstrated a worse selectivity index of these
compounds with respect to compound 80 as they displayed adverse effects at high concentrations and some of them were not active against the infection. Unpredictably, the cyclisation of the amide moiety in compound 93 produced a molecule that was active against DENV infection with low micromolar EC$_{50}$, suggesting a third possibility, that of the region highlighted in blue in the scheme of Figure 6.12. This could be further explored in the future. Molecule 95 was selected for evaluation as it contained both the sulphonamide and the two phenyl rings moiety with the aim of combining the two enhancement effects. Unexpectedly, a mild improvement of enzymatic inhibition and a worse cell-based activity were obtained. It was therefore hypothesised that the different conformation of a sulphonamide with respect to an amide group could lead to a different molecular conformation that allows the accommodation of two condensed phenyl rings, but not of two separate ones in the binding site. Given the high number of variables in cell-based activity assessment, it was more difficult to suggest a reason for the worse EC$_{50}$ value of this compound.

Summarising, the assessment of analogues of compound 80 suggests that this series could stop DENV-infection through NS5 RdRp inhibition, validating NS5 RdRp as a promising drug target for antiviral development as its inhibition affects cell infection. However, other unidentified cellular pathways might be involved in the mechanism of action of these molecules as cell-based and biochemical activity did not always match. Nevertheless, following the SARs that were often supported by the binding pose of compound 80, these compounds could be used for further optimisation as they have shown to be promising anti-DENV agents. As schematised in Figure 6.12, an aromatic group in the area delineated with an orange circle and an amide or sulphonamide moiety in the blue area are essential for molecule activity, which is enhanced with large aromatic substituents in R$_3$ (green area in the scheme).

Figure 6.12. Schematic representation of the SAR for compound 80 series.
The structure-activity relationships (SARs) discussed in the text are summarised here. The central core of the molecule is presented in the middle and in the circles are described suggested groups that are required or that improve activity.
COMPOUND 82 ANALOGUES

Compound 82 was selected from the SPECS database through the virtual screening approach discussed previously. In the final selection it was shown to bind in model B according to the pose presented in Figure 6.13.

![Figure 6.13. Binding pose for compound 82.](image)

The docking pose for compound 82 in model B is presented here. The protein surface is transparent and coloured in red, while the protein ribbon is shown in light blue. The discussed residues are indicated and shown with light blue carbon atoms. The ligand carbon atoms are coloured in orange. Hydrogen bonds are indicated by black dotted lines.

Using the area references of Figure 6.6, the molecule spanned from the P1 to the P2 sites and the central fluorene portion was placed on top of Trp795, giving a π-π stacking with this residue, similarly to compound 80. One of the two furan rings was placed in a hydrophobic area of the binding cleft, described by the lipophilic portion of the side chains of residues Glu415, Asn416 and Gln742. This placed the oxygen atom of the carbonyl group linked to the furan itself in an appropriate position of a hydrogen bond with Trp795 side chain indole NH. Similarly to compound 82, the other furan ring was perpendicular to Arg792, suggesting that it might interact with this residue through a positive charge-π interaction. This allowed the interaction of the oxygen atom of the carbonyl group linked to this furan ring with the Thr413 side chain hydroxyl group through a hydrogen bond. Only one of the four amino groups formed hydrogen bonds with the protein and this interaction involved Thr413 backbone carbonyl oxygen. All but one (Thr413) of the mentioned residues are well conserved or maintained their properties in the four DENV serotypes, complying with the possibility of
designing panserotypic inhibitors of DENV. In position Thr413 is a Gln in DENV4 and a Val in DENV1. Unfortunately one of the modelled interactions of compound 82 with DENV NS5 RdRp involves the side chain of this amino acid. However, as the importance of this single interaction is not known, the possibility of a panserotypic activity of compound 82 is not excluded.

Compound 82 showed a low micromolar activity (IC_{50} of 47.1µM and 35.1µM according to assay settings) in the inhibition of DENV RdRp activity in the biochemical assays with both RNA template and probe settings used. Unfortunately, the CPE assay did not show an anti-infection activity of this compound, probably due to its cytotoxicity (CC_{50} of 138µM).

With the aim of confirming the NS5 RdRp activity inhibition, improve the cell infection inhibition and to gain further insights on preliminary structure-activity relationships of compound 82, eleven analogue molecules were purchased from SPECS and four (101, 102, 103 and 109) were synthesised. All of them were tested in vitro with CPE and with the polymerase activity inhibition assay, using a polyU template and PicoGreen fluorescent probe. The selection of these compounds was also based the binding mode suggested by the molecular modelling results and showed in Figure 6.13, as a second objective was to validate this model. Table 6.4 contains the structures and the assay results for compound 82 and its analogues. Here, the indicated CC_{50} and EC_{50} values were obtained from the CPE assay and the selectivity index (SI) is reported as well. The IC_{50} values were acquired with the enzyme inhibition test. As highlighted in the table, all the compounds were symmetrical and most of them had the central fluorene scaffold A. Amongst these, 96 was the only compound that retained the sulphonylhydrazide group, which was substituted by a sulphonamide moiety in the other molecules. In particular, 100 to 108 were selected because as well as having one NH less in the sulphonylhydrazide group, they also did not have the carbonyl moiety in the chain. The position of the two nitrogens in the side chains was assessed through compounds 97 and 98, while the importance of the final aromatic ring was investigated with 99, and 106 to 108. The length of the side chains of the central scaffold was also investigated with compounds 104 to 107. In order to confirm the results and to investigate the role of chirality in the activity of compound 100, the three compounds 101 to 103 were synthesised. The central fluorene group was also explored with compounds 109 and 110.
Table 6.4. Compound 80 and analogue molecules.

<table>
<thead>
<tr>
<th>Scaffold</th>
<th>R</th>
<th>CC&lt;sub&gt;50&lt;/sub&gt; (µM)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>EC&lt;sub&gt;50&lt;/sub&gt; (µM)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (µM)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>A</td>
<td>138</td>
<td>&gt;200</td>
<td>35.1 ±1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 0</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>A</td>
<td>&gt;88.6</td>
<td>&gt;88.6</td>
<td>89.4 ±29.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 1</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>A</td>
<td>&gt;84.6</td>
<td>&gt;84.6</td>
<td>40.3 ±5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 1</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>A</td>
<td>&gt;81.1</td>
<td>&gt;81.1</td>
<td>Insoluble</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 1</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>A</td>
<td>&gt;113.5</td>
<td>&gt;113.5</td>
<td>357.2 ±40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 1</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>A</td>
<td>5.61</td>
<td>1.31</td>
<td>2.8 ±0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 4</td>
<td></td>
</tr>
<tr>
<td>101&lt;sup&gt;c&lt;/sup&gt;</td>
<td>A</td>
<td>&gt;93.9</td>
<td>0.87</td>
<td>19.0 ±2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI &gt; 107</td>
<td></td>
</tr>
<tr>
<td>102&lt;sup&gt;c&lt;/sup&gt;</td>
<td>A</td>
<td>&gt;93.9</td>
<td>0.45</td>
<td>16.1 ±2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI &gt; 211</td>
<td></td>
</tr>
<tr>
<td>103&lt;sup&gt;c&lt;/sup&gt;</td>
<td>A</td>
<td>&gt;93.9</td>
<td>1.78</td>
<td>11.2 ±1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI &gt; 53</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>A</td>
<td>&gt;20.6</td>
<td>48.2</td>
<td>3.96 ±0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SI = 0.427</td>
<td></td>
</tr>
</tbody>
</table>
As can be observed in the table, improvement of both NS5 RdRp and cellular infection inhibition could be achieved for some of compound 82 analogues, compared to the original molecule. Some molecules were active in both assays, suggesting that this series of molecules might exert their antiviral function through the inhibition of the polymerase, thus validating the NS5 RdRp as a promising drug target for anti-DENV drug design. Unfortunately, the compounds that were active in the CPE assay also displayed adverse effects at higher concentrations, implying that this class of compounds still needs improvement. Even if data were more consistent with respect to compound 80 series, the lack of correspondence between some cell-based and biochemical results, might imply that cellular pathways not yet explored could be involved in the cell-based activity of these compounds. In addition, lack of penetration or intracellular sequestration could also be implicated, but has not been investigated to date. Nevertheless, preliminary structure-activity relationships could be hypothesised on the base of the data presented in Table 6.4 and are schematised in Figure 6.15.

The substitution of the furan ring with a pyridine (in 96) did not improve the activity of the compound in cells and increased the IC50 value. Similarly, the rearrangement of R components (in 97) resulted in an activity profile very similar to the original molecule. The importance of a hydrophobic group at the end of the R chain was confirmed by 99 that was inactive in the CPE
assay and inhibited the polymerase function only at high micromolar concentration. Furthermore, the structural similar compounds 100 to 108 compounds, all displaying an aromatic of hydrophobic final moiety in the R chain, showed generally good activity in the biochemical assay. The antiviral activity of compounds 100 and 104 in cells was associated with cytotoxicity, while it was lost in the molecules with an aliphatic ring or chain at the end of R (from 106 to 108). It was consequently hypothesised that the presence of a hydrophobic group was fundamental for RdRp inhibition (as schematised in Figure 6.15 with green circles) and an aromatic moiety is preferable for the inhibition of infection in cells. This is in line with the docking pose for compound 82 suggested by the molecular modelling, in which one furan ring is placed in a hydrophobic area of the binding site and the other one interacts with Arg792.

The IC50 values for compounds from 106 to 108 were generally lower than the reference molecule, implying that substituents without the carbonyl group and with a sulphonamide substituting the sulphonylhydrazide enhance the molecule activity. This was an unexpected result as in the in silico model for the binding of compound 82 placed one carbonyl group interacting with the Thr413 side chain OH, whilst the other one hydrogen bonded with Trp795 (Figure 6.13).

Unfortunately, no insights on the chain length could be obtained. The only molecules that displayed a worse RdRp inhibition (when assessed), as well as lack of activity in cells, were the only ones with no NH groups (98 and 107). This implied that the presence of at least one NH moiety is important for activity, in line with the interaction with the backbone carbonyl oxygen of Thr413 observed in the pose in Figure 6.13. These hypotheses are collectively represented in Figure 6.15 with blue circles.

Compounds 101 to 103 were synthesised de novo with the aim of finding a possible synthetic route for the fluorene-based compounds, to confirm the activity of 100 and to evaluate the importance of the stereochemistry of this compound. Compound 101 was synthesised as a mixture of three possible stereoisomers combinations (R,R; S,S and R,S), whilst the enantiomerically pure 102 and 103 are the (R,R) and the (S,S) isomers of the molecule respectively. The three compounds displayed very similar activity, implying that the chirality is not a fundamental factor with these groups linked to the asymmetrically substituted carbon. Interestingly, all three products showed an improved cytotoxicity profile and despite adverse effects still observed at higher concentration, the selectivity index had values between 53 and 211. In order to explain this difference with 100, it was hypothesised that the purchased sample was less pure and thus had greater cytotoxicity. This structure was however confirmed to be the most active one of the whole series in the inhibition of DENV infection in cells EC50 values that reached sub-micromolar concentrations, in particular for the (R,R) isomer.
The central fluorene scaffold was also explored with modifications indicated with “scaffold B” and “scaffold C” in Table 6.4. The first one was investigated with 109, which was synthesised, as it was not commercially available. The difference from the original fluorene structure is the removal of the methylene linker between the two aromatic rings. As it was observed that the fluorene scaffold is very similar to anti-tumour telomerase inhibitors, it was hypothesised that the cytotoxicity of these compounds could have been related to this off-target mechanism. An important characteristic for the mode of action of the anti-cancer compounds is the planarity of the central core of the molecule. It is necessary for G-quadruplex (DNA structure, typical of telomere sequences) intercalation and consequent stabilisation, which inhibits the telomerase activity.[18] For this reason it was chosen to remove the methylene moiety, producing a scaffold that retains aromatic properties, but looses planarity. This molecule was inactive in infected cells and showed a lower RdRp inhibition compared to 82, complying with the molecular modelling model described in Figure 6.13. According to this binding mode, the biphenyl scaffold of 109 would be placed on Trp795, giving a π-π stacking with this residue. However, as this scaffold is less planar than a fluorene, the π-π stacking could be less important, explaining the difference in results observed in the in vitro assays. It was therefore hypothesised that the planarity of the scaffold is important for the activity of these compounds and consequently that the methylene group in the fluorene is essential.

![Compound 82](image1)

![Compound 109](image2)

Figure 6.14. CPE assay results for compounds 82 and 109.

A) Results for compound 82. Data is presented as a percentage, with 100% representing cell controls. At different compound concentrations (shown in µM), the residual metabolic activity of the cells is reported in green, while the antiviral activity of the compound is shown in red.

B) Results for compound 109. Data is presented as a percentage, with 100% representing cell controls. At different compound concentrations (shown in µg/ml), the residual metabolic activity of the cells is reported with a green line, while the antiviral activity of the compound with a red one. The green dashed line showed the residual metabolic activity of cells evaluated with microscopic evaluation.

Comparing the assay charts (Figure 6.14), the decrease of metabolic activity in the presence of 109 occurred at slightly higher concentrations (around 18.84µM instead of around 6µM). However, as this compound was not tested at a concentration higher than 94.2µM it was not possible to assess if the modification was successful in improving the cytotoxicity profile of 82.
The second central scaffold was investigated with 110, which was purchased. Also in this case the fluorene was modified at the level of the connecting methylene, but this time the CH₂ was substituted with a carbonyl group, resulting in a fluorenone core. Since a non-canonical dose-response curve was obtained in the CPE assay for this compound, with two EC₅₀ values, these results cannot be compared with other molecules. However, evaluating the RdRp inhibition data of 110 with 104, both having the same R substituent, more than a ten-fold reduction in activity was observed. This was consistent with the binding mode model discussed with Figure 6.13, as the carbonyl moiety of the fluorenone would clash with the protein Thr413-Glu415 loop. The central scaffold would consequently be shifted, weakening the π-π stacking with Trp795. Collectively, these results suggest that the methylene group is essential for compound activity, as schematised by the red circle in Figure 6.15.

Summarising, the assessment of the class of compounds related to 82 validated NS5 RdRp as a promising drug target for antiviral development as its inhibition also affects cell infection. Even if cell-based and biochemical activity did not always match, more consistency was found with respect to the 80 series, suggesting that compound 82 analogues offer a better development series to find anti-DENV agents. Further optimisation might be initiated from the observed SARs that were often supported by compound 82 binding pose obtained from the virtual screening approach. As schematised in Figure 6.15, a methylene group in the area shown with a red circle is indispensable for molecule activity. Also a hydrophobic group in the area shown with a green circle is essential for activity and an aromatic moiety is preferable, in particular for infection inhibition in cells. Additionally, a sulphonamide replacing the sulphonylhydrazide group and the removal of the carbonyl moiety (blue circle in the scheme) enhance both cell infection and RdRp inhibition.

![Figure 6.15. Schematic representation of the SAR for compound 82 series.](image)

The structure-activity relationships (SARs) discussed in the text are presented here. The fluorene scaffold of the molecule is presented in the middle and the areas with suggested groups that are required or that improve activity are described in the circles.
CHEMICAL SYNTHESIS OF COMPOUND 109

Figure 6.16. Structure of compound 109.

N\textsuperscript{4},N\textsuperscript{4'}-di(furan-2-carbonyl)\{-1,1'-biphenyl\}-4,4'-disulphonohydrazide (109, Figure 6.16) was chosen with the aim of exploring the importance of the fluorene scaffold with relationship to activity and toxicity. As it was not possible to purchase this molecule, it was prepared with a one step synthesis, according to the procedure in Scheme 6.1. In this reaction two molecules of 2-furoic hydrazide (111) react with one of biphenyl-4,4'-disulphonylchloride (112) in pyridine (Py) at room temperature (r.t.) through a coupling reaction mechanism to give the symmetric compound 109.\textsuperscript{[19,20]}


Pyridine was chosen to facilitate the coupling reaction not only because is a stronger base than the hydrazide compound 111, but also because it could be the solvent for the reaction, as compound 112 was poorly soluble in most of the common solvents. Another reason is the fact that pyridine has shown to be able to enhance acylation reactions of benzenesulphonyl chloride derivatives with a nucleophilic catalysis.\textsuperscript{[21,22]}

The reaction was monitored by TLC until 112 consumption. After the removal of Py under vacuum, the product was washed with water in order to remove the miscible Py and the water soluble compound 111. NMR analysis showed that the obtained product needed purification, but the poor solubility of 109 in the most common organic solvents made this difficult. As it did not dissolve in the eluents, column chromatography could not be applied. Crystallisation with several solvents and solvent combinations were attempted, revealing 1,4-dioxane/water mixture as the successful one. Unfortunately, the elimination of 1,4-dioxane required several washes with ethanol and diethylether, reducing the yield considerably. NMR and elemental analysis confirmed that the obtained compound was 109.
CHEMICAL SYNTHESIS OF 9H-FLUORENE-2,7-DISULPHONAMIDE COMPOUNDS (101, 102 AND 103)

*Figure 6.17. Compound 100.*

*In vitro* assays showed that 9H-fluorene-2,7-disulphonamide is a favourable scaffold for both DENV NS5 RdRp and cell-based infection inhibition. With the aim of improving this class of molecules, a synthetic pathway was developed and is discussed here. Furthermore, one of the most promising compounds (100) displayed two chiral centres, but no information about its stereochemistry was available. Thus, the synthesis of this compound was performed with the aim of confirming its activity and to explore the importance of the chirality in compound activity.

![Scheme 6.2. General procedure for the synthesis of compounds 100, 101 and 102.](image)

(a: i) Acetic Acid, 120°C, 4h; ii) NaOH; b: i) 120°C, 2h; ii) CHCl₃, reflux 60°C, 1 h; c: N,N-diisopropylethylamine, dichloromethane, r.t. 1.5h)

The synthesis required three steps, as described in Scheme 6.2, and each one of them will be discussed separately. Initially, the coupling in the third step in this procedure was performed with a racemic mixture of the amine compound in order to evaluate the possibility of obtaining and separating the three possible stereoisomers (R,R; S,S; and R,S). However, also the enantiomerically pure (R) and (S) amines were reacted with the same sulphonylchloride...
group containing precursor for the synthesis of the compounds with known chirality 102 and 103.

**Step 1: synthesis of sodium 9H-fluorene-2,7-disulphonate (113)**

Sodium 9H-fluorene-2,7-disulphonate (113) was synthesised through a double sulphonation of fluorene (114), according to the procedure reported in Scheme 6.3. In this reaction, two molecules of chlorosulphonic acid (115) react with the fluorene molecule (114) in glacial acetic acid at 120°C for 4 hours to give the symmetric compound 115 with an electrophilic aromatic substitution mechanism.[23]

![Scheme 6.3. Synthesis of compound 113. Reaction conditions: i) CH₃COOH, 120°C, 4h; ii) NaOH](image)

From this reaction scheme, it appears that 2 moles of chlorosulphonic acid are stoichiometrically needed for product formation. However, a higher quantity (almost double) was needed for the synthesis of 113. The reason behind this aspect could be the high reactivity of this molecule with water, producing sulphuric (H₂SO₄) and hydrochloric (HCl) acids, which requires anhydrous conditions. Despite the attention given to this aspect during the synthesis, probably some water was present in the reaction mixture, partially deactivating this reagent and explaining the need of an excess quantity of HSO₃Cl.

The reaction was monitored with TLC and when it was concluded, the mixture was treated with sodium hydroxide (NaOH) in brine. In this way, the excess of chlorosulphonic acid was deactivated and the sodium salt of the disulphonic acid (113) was isolated as a solid due to its poor solubility in water. The residue was then washed with brine and water. The obtainment of 113 was confirmed by NMR analysis.

**Step 2: synthesis of 9H-fluorene-2,7-disulphonyl dichloride (116)**

9H-fluorene-2,7-disulphonyl dichloride (116) was synthesised through the chlorination reaction reported in Scheme 6.4. In this reaction, sodium 9H-fluorene-2,7-disulphonate (113) reacts with phosphorus pentachloride (PCl₅) without solvent at 120°C first and then in chloroform (CHCl₃) at 60°C to give 9H-fluorene-2,7-disulphonyl dichloride (116).[24]
Scheme 6.4. Synthesis of compound 116. Reaction conditions: i) 120°C, 2h; ii) CHCl₃, reflux 60°C, 1 h

As PCl₅ reacts violently with water, giving phosphoric acid (H₃PO₄) and HCl after complete hydrolysis, all the steps of the chlorination were performed under strictly anhydrous conditions. The reaction was initially performed in absence of solvent. In order to assure the best contact between the reagents, the powders were pulverised in mortar and the smallest flask available was used. A temperature of 120°C was used at this stage as it is lower than the melting point of PCl₅ (166°C with decomposition), but higher than the boiling point of the product POCl₃ (106°C), allowing the removal of the side product, shifting the equilibrium of the reaction towards the formation of 116. With the consumption of PCl₅ and the elimination of POCl₃, the stirring of the suspension became more difficult and consequently the reaction was continued with the use of a solvent. At reaction completed, the reaction solution was washed with water in order to remove the excess of PCl₅ and the remaining POCl₃.

The product was purified with precipitation with an appropriate anti-solvent from a solution of CHCl₃. Several attempts were performed, finding petroleum ether as the best precipitating solvent. The obtainment of 116 was confirmed by NMR analysis.

**Step 3: synthesis of final compounds (101-103)**

The last step for the synthesis of compounds 101-103 was performed according the coupling reaction reported in Scheme 6.5. In this reaction two molecules of the appropriate α-methylbenzylamine (117) react with one of 9H-fluorene-2,7-disulphonyl dichloride (116) in dichloromethane (DCM) in presence of N,N-diisopropylethylamine (DIPEA) at room temperature to give the symmetric compounds 101-103. For the synthesis of N²,N⁷-bis(1-phenylethyl)-9H-fluorene-2,7-disulphonamide (101) (+)α-methylbenzylamine was used, aiming the synthesis of the mixture of all possible isomer combinations; while for N²,N⁷-bis((R)-1-phenylethyl)-9H-fluorene-2,7-disulphonamide (102) and N²,N⁷-bis((S)-1-phenylethyl)-9H-fluorene-2,7-disulphonamide (103) (R)-(−)-α-methylbenzylamine and (S)-(−)-α-methylbenzylamine were used respectively. Differently from the coupling reaction performed for the synthesis of 109, a stronger base had to be used in this case because Py is not basic enough to facilitate the coupling reaction.
Scheme 6.5. Synthesis of compounds 101-103.

The reaction was monitored with TLC and when completed the solvent was removed and the solid treated with HCl 1N, obtaining the final product as a water insoluble solid and the residual basic compounds in solution. Column chromatography was used to purify the final product, obtaining compounds 101, 102 or 103 according to the amine used, as confirmed from NMR and elemental analysis. Unfortunately, in the mass spectroscopy analysis of all three the compounds, the molecular signal given by the addition of a H⁺ (M-H⁺) was not observed, while the one given by the addition of Na⁺ (M-Na⁺) was present.

The M-Na⁺ signal in the mass spectra and the elemental analysis were equal or very similar between the three compounds, confirming that they differ only by stereochemistry. The (R,R) and the (S,S) isomers (as in 102 and 103 respectively) are enantiomers, as they are one the mirror image of the other. For this reason, it is not surprising that they present equivalent ¹H-NMR and ¹³C-NMR spectra and a mixture of the two would not be distinguishable with NMR.

The third supposed component of the 101 mixture, the (R,S) isomer. It contains a plane of symmetry and thus is a meso compound and consequently it is achiral, despite the presence of two chiral centres.

Figure 6.18. Three possible stereoisomers of compound 100.
A) (R,R) enantiomer (as in 102). B) (S,S) enantiomer (as in 103). C) (R,S) isomer or meso compound.
The three stereoisomers are shown in Figure 6.18. Experimental data showed that also 101 had equivalent $^1$H-NMR and $^{13}$C-NMR spectra, implying that also the meso isomer (R,S) could have the same spectra as the other species. The high range in the melting temperature of 101 supports that this product is a mixture of compounds, but unfortunately the other experiments did not help in the determination of the nature and of the relative quantity of the single components.

CONCLUDING REMARKS

A virtual screening approach has been applied for the design of non-nucleoside DENV polymerase inhibitors (NNI). Two very similar homology models of the protein were built and used for compound selection and the binding site (called “site D” in this thesis) was chosen on the basis of a model for which Trp795 plays an essential role in the RNA synthesis initiation step. A filtered selection of small drug-like molecules belonging to the SPECS database chose to 29 final compounds from around 209,000 using a virtual screening methodology that, to the best of knowledge, has never been applied to this molecular target. This molecular modelling approach was successful in identifying two compounds (80 and 82) with different chemical scaffolds to other DENV RdRp NNIs reported in the literature. These two showed promise in their evaluation in vitro with cell- and polymerase-based assays. In both cases, analogue molecules were purchased or synthesised and assessed for activity. Several compounds of the two series were found to be active also at very low micromolar concentrations, making them potential leads for anti-DENV agents, despite cytotoxicity seen at higher concentrations that means that further improvement is needed. Indeed, the possible methods of synthesis used for compound 100 could be a starting point for the design and assessment of other 82 analogues. Importantly, the fact that some molecules in both series were active in both assays, the NS5 RdRp was validated as a promising drug target for anti-DENV drug design, as its inhibition affects cell infection. This also suggested that these series of molecules might exert their antiviral function through the inhibition of the polymerase. However, the inhibition of DENV infection in cells and of the activity of DENV RdRp did not always correlate well, in particular in the series of 80 analogues, whilst greater consistency was seen with the 82 analogues. Consequently, it was hypothesised that other not yet explored molecule or cell-based factors might be involved in the cellular activity of these compounds. However, cell-based results also need to be repeated to obtain stronger statistical strength of the results. Nonetheless, preliminary structure-activity relationships could be hypothesised for both series. In the case of compound 82 analogues, this was aided also by the four molecules that were successfully synthesised and tested, as they not only confirmed the activity of one compound (100), but
they also provided important knowledge in constructing structure-activity relationships. Furthermore, the SARs were often supported by the binding modes of compounds 80 and 82, suggesting that these could indeed bind site D of the RdRp. The docking binding modes were not always consistent with the inhibition data, but it is reasonable to hypothesise that, being allosteric inhibitors, the compounds could induce small conformational changes that cannot be predicted with semi-flexible docking simulations where the protein is considered a rigid body. With the aim of validating the model for which these series of compounds inhibit DENV polymerase by binding it in site D, collaborators in Marseille University attempted the co-crystallisation of compounds 80 and 82, but unfortunately the soaking experiments failed. Even if the reference protein structure belonged to DENV serotype 3, particular attention to the conservation of the interacting protein residues was given in the molecular modelling approach to compound selection. Unfortunately, the in vitro assays have not been repeated on more than one serotype to date. Both CPE and the polymerase activity inhibition assays were performed with DENV serotype 2 virus and NS5 RdRp respectively. Thus, the cross serotype activity of these compounds has not yet been determined.

In conclusion, the molecular modelling approach designed in this study successfully identified two novel promising anti-DENV agents that were further developed with a classical Medicinal Chemistry approach and the establishment of a route for the synthesis of compound 82-like molecules. Improvements of these two series of compounds are still needed but could be achieved using the built SAR models and the synthetic route discussed above, turning these leads into potentially good candidates for the treatment of DENV infection.
6.2.2 ALLOSTERIC INHIBITORS BINDING SITE C

Essential for RNA replication and a validated target also for other viruses, DENV NS5 RdRp is probably the most attractive molecular target for antiviral drug design. One of the most active developers of both nucleoside (NI) and non-nucleoside inhibitors (NNI) in the last years has been the Novartis Institute for Tropical Diseases (NITD). Through high throughput screening (HTS) of compounds with an enzymatic assay, NITD has identified a N-sulphonylanthranilic acid scaffold with RdRp inhibitory activity that was developed with a series of analogue molecules.\(^{[10,11]}\) As the mechanism of interaction of these molecules with RdRp was not clear, a new crystal structure of NS5 RdRp domain was published in complex with an inhibitor belonging to this NITD series.\(^{[7]}\) However, in this complex, the ligand (NITD-107) is located in RdRp site C, in a binding mode that does not allow immediate correlation between the activity values of the compound series and the structural information.

With this premise, the putative binding site of the NITD series has been studied with the aim of identifying structural information that could correlate with the activity data published in the literature. The final objective would be to use the information obtained as the basis for a virtual screening approach to identify novel DENV RdRp NNIs.

PROTEIN MODEL

Two crystal structures of the DENV3 RdRp domain have been published after the virtual screening study described in chapter 6.2.1. These structures are at high resolution and show more resolved residues with respect to the previously published ones, but they are still incomplete.\(^{[6,7]}\) In particular, comparing the structure of the enzyme in complex with the inhibitor NITD-107 (PDB ID: 3VWS\(^{[7]}\)) with the apo protein (PDB ID 2J7U\(^{[6]}\)), it appears that the sequence between Ala406 and Asp419 is solved in the more recent structure, whilst the other two not present portions (Val310-Ala316 and Tyr451-Ala472) are still missing. Furthermore, several conformational changes in the protein upon compound binding seem to trigger the stabilisation of loop 3 (L3) and helix α5 that appear highly disordered in the apo form of the protein.

3WVS was chosen as a target RdRp structure because being the structure co-crystallised with NITD-107, it complied with the aim of this study. However, as already mentioned, it is incomplete and despite the fact that the missing loops were far from NITD-107 binding site, a model of the protein was built using 3VWS as a template and homology modelling. As done in the previous study, only the missing fragments (Glu309-Ser317 and Cys449-Ala472) were modelled according to the procedure described in the METHODS chapter and the rest of the
protein was maintained as in the crystal structure. The amino acids at the boundaries of the missing fragments were modelled as well with the aim of obtaining the most reasonable structure possible. As no other RdRp structures with similar conformation to 3VWS could be found, the missing fragments were modelled on 3VWS, using only the force field (AMBER99) parameters. The obtained model superposed very well with the original structure, with a RMSD value of 0.08Å. The quality of the model was assessed through Ramachandran plot evaluation that showed only one extra outliner compared to the original structure. The outliners belonged to a protein region far from the selected target site and were however very close to the allowed region of the psi-phi plot. The model was consequently considered suitable to be used for the determination of the putative binding site of the known inhibitors and for the subsequent virtual screening approach.

**NITD SERIES OF COMPOUNDS**

The published NITD series of N-sulphonylanthraniolic acid derivatives exhibited polymerase activity in the low micromolar range, but no potent activity in cell-based assay.\[^{3,10}\] Compounds NITD-1 and NITD-107 are shown in Figure 6.19 and the other compounds’ structures and polymerase IC\(_{50}\) values are reported in Table 6.5.

![NITD-1 (IC\(_{50}\)=7.2μM)](image1.png)  ![NITD-107 (IC\(_{50}\)=113μM)](image2.png)

*Figure 6.19. Structure and IC\(_{50}\) values for NITD-1 and NITD-107.*

The inhibitor in complex with the RdRp protein in the newly published crystal structure of DENV RdRp is NITD-107 (Figure 6.19). This compound is commercially available and it exhibited an IC\(_{50}\) of 113μM in DENV4 enzymatic assay and EC\(_{50}\) of 100μM in the DENV2 cell-based assay. It is therefore weakly active against DENV RdRp, but it has been shown not to be cytotoxic and to be specific for the enzyme. It bound full length NS5 with a dissociation constant (K\(_d\)) of 173μM and for the RdRp domain with a K\(_d\) of 225μM.\[^{7}\] This molecule is structurally similar to the other NITD molecules, but amongst the whole set of compounds it is the one that differs the most from the others (Table 6.5).
Table 6.5. NITD series of compounds.\(^{[10]}\)

<table>
<thead>
<tr>
<th>NITD-n</th>
<th>R(_1)</th>
<th>R(_2)</th>
<th>R(_3)</th>
<th>IC(_{50}) (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2-COOH</td>
<td>H</td>
<td>Benzyl</td>
<td>0.7</td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>H</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>7</td>
<td>2-COOMe</td>
<td>H</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>8</td>
<td>2-COONH(_2)</td>
<td>H</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>9</td>
<td>3-COOH</td>
<td>H</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>10</td>
<td>4-COOH</td>
<td>H</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>11</td>
<td>2-COOH</td>
<td>6-Cl</td>
<td>Benzyl</td>
<td>18.02</td>
</tr>
<tr>
<td>12</td>
<td>2-COOH</td>
<td>5-Cl</td>
<td>Benzyl</td>
<td>0.77</td>
</tr>
<tr>
<td>13</td>
<td>2-COOH</td>
<td>4-Cl</td>
<td>Benzyl</td>
<td>2.56</td>
</tr>
<tr>
<td>14</td>
<td>2-COOH</td>
<td>3-Cl</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>15</td>
<td>2-COOH</td>
<td>5-F</td>
<td>Benzyl</td>
<td>0.37</td>
</tr>
<tr>
<td>16</td>
<td>2-COOH</td>
<td>5-methoxy</td>
<td>Benzyl</td>
<td>9.99</td>
</tr>
<tr>
<td>17</td>
<td>2-COOH</td>
<td>4,5-dimethoxy</td>
<td>Benzyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>22</td>
<td>2-COOH</td>
<td>H</td>
<td>H</td>
<td>&gt;20</td>
</tr>
<tr>
<td>23</td>
<td>2-COOH</td>
<td>H</td>
<td>Me</td>
<td>&gt;20</td>
</tr>
<tr>
<td>24</td>
<td>2-COOH</td>
<td>H</td>
<td>Isobutyl</td>
<td>&gt;20</td>
</tr>
<tr>
<td>25</td>
<td>2-COOH</td>
<td>H</td>
<td>Isopentyl</td>
<td>6.77</td>
</tr>
<tr>
<td>26</td>
<td>2-COOH</td>
<td>H</td>
<td>Ethilmorpholine</td>
<td>&gt;20</td>
</tr>
<tr>
<td>27</td>
<td>2-COOH</td>
<td>H</td>
<td>Methyl-2-naphthalene</td>
<td>0.37</td>
</tr>
<tr>
<td>28</td>
<td>2-COOH</td>
<td>5-F</td>
<td>Methyl-2-naphthalene</td>
<td>0.26</td>
</tr>
</tbody>
</table>

In the crystallised complex, two molecules of NITD-107 are interacting between each other and with the protein, as displayed in Figure 6.20. As shown, the binding site is located on the opposite side of Trp795 with respect to Arg737. Thus, it does not coincide with the nascent RNA binding groove, but rather with the site for the RNA template binding. The phenyl group of one NITD-107 molecule forms a stacking interaction with the benzofuran moiety of the other one and the main contacts with the protein involve the stabilised loop L3 residues in
particular.\textsuperscript{[7]} Unfortunately, the displayed contacts between the two ligand molecules and the protein are not such to explain the activity patterns reported in Table 6.5. For instance, more active molecules have a highly hydrophobic large moiety in the R₃ position; whilst NITD-107 only has a chlorine atom. Observing the conformation of NITD-107 molecules in the cleft, there is no space for a larger group in this position and the activity cannot be enhanced with this type of substitution according to this binding model.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{binding.png}
\caption{NITD-107 complex with the protein (PDB: 3VWS).}
\begin{subfigure}{0.5\textwidth}
\centering
\includegraphics[width=\textwidth]{binding_a.png}
\caption{A). Binding pose of NITD-107. Here, the carbon atoms of the two NITD-107 molecules are in orange and the interacting RdRp residues are highlighted and indicated. Hydrogen bond interactions are shown with blue dotted lines.}
\end{subfigure}
\begin{subfigure}{0.5\textwidth}
\centering
\includegraphics[width=\textwidth]{binding_b.png}
\caption{B). Ligand interaction scheme for the two NITD-107 molecules. In this schematic representation of the binding mode the two molecules are showed in red and in green. The interactions are shown with dotted lines which colour recalls the corresponding NITD-107 molecule.}
\end{subfigure}
\end{figure}

In order to have a better understanding of the real binding mode of this series of molecules, docking studies were performed. Attempts were conducted with three docking programs (Maestro Glide, PLANTS and LeadIT FlexX) for all the NITD series of compounds, using different docking settings as well as exploring different protonation states for the ligands. Only one docking simulation, performed with LeadIT FlexX program, gave relevant results that could match the activity data published in the literature. Most intriguing is the fact that NITD-107 bound to the enzyme superposes well with one of the crystallised ligands in 3VWS (Figure 6.21), while most of the molecules were placed in another area of site C. The main difference between the crystallised conformation and the docking pose of NITD-107 was the position of the phenyl group. However, this does not appear to be involved in important interactions with the protein. This gave confidence on the docking simulation reliability.
6.2 RESULTS AND DISCUSSION

The best scoring molecules were also the most active compounds of the series (NITD-28 and NITD-27). All molecules with IC\textsubscript{50} values lower than 20μM displayed a similar binding mode to that given by NITD-28 and shown in Figure 6.22 and the interactions formed with the RdRp were consistent with the activity data. Compounds with an IC\textsubscript{50} higher than 20μM (considered “less active”) not always resembled the same binding as the “active” ones.

In the binding mode presented in Figure 6.22, the hydrophobic R\textsubscript{3} group is placed in a hydrophobic pocket (green residues in the figure) given by Val411, Phe412, Val402, Lys401, Phe398 and Phe485. Larger hydrophobic groups are able to better occupy this pocket,
contributing to an improved binding and this could explain why compounds with larger hydrophobic groups in R₃ display higher inhibition of the enzyme. Furthermore, compounds with IC₅₀ values lower than 1μM, generally displayed an additional hydrophobic interaction between the pyrazole ring and Val603, while compounds with higher activity placed the aromatic ring away from this residue, near Val411.

Several R₁ and R₂ substituents were used in the compound series and the activity data indicated that the presence of a carboxylic moiety in position 2 of the ring was fundamental for the activity and that groups that are able to increase its acidity were correlated with higher polymerase inhibition. [10] According to the docking model, the importance derived from the fact that the carboxylic acid is involved in two hydrogen bonds (with the Trp477 indole NH and with the Asn452 backbone NH) and with an electrostatic interaction with the positively charged Arg481. The position of substituents (e.g. halogens) that enhance the acidity of the COOH moiety in this ring is related to the electron-withdrawing effect and therefore ortho- and para- positions in respect to the carboxylate group would be preferred. However, molecules with a halogen in the ortho- position did not give good inhibition activity. This could be explained by the binding mode model presented here, in which a substituent in this position produces a steric clash with the protein.

Almost all of the residues discussed in this binding mode are well conserved in identity or in physical properties amongst the four DENV serotypes, suggesting that inhibitors that target this pocket could be serotype-unspecific, in accordance with the aims of this study. Only Thr413 is not conserved, as it is a Val in DENV1 and a Gln in DENV4. However, as displayed in Figure 6.22, the compounds interact with the backbone amine group of this residue, making the conservation of this amino acid less important for panserotypic drug design.

**IDENTIFICATION OF ALLOSTERIC INHIBITORS THROUGH VIRTUAL SCREENING**

The protein homology model and the binding mode hypothesis obtained from the docking simulation of the known N-sulphonylanthranilic acid derivatives were used for the design of a virtual screening approach that could aid the identification of novel antiviral agents within the drug-like small molecules SPECS[12] database (updated to February 2012). The complex of the most active molecule (NITD-28) with the protein was used as a starting point for this study. A scheme with the general virtual screening workflow is reported in Figure 6.23, while details about the single procedures are reported in the METHODS chapter.
Figure 6.23. Schematic representation of the virtual screening workflow.

The main steps are indicated and the numbers are referred to the number of molecules at every step. The homology model of the polymerase is reported with RdRp. Molecular modelling programs and functions are abbreviated as follows: FX= LeadIT FlexX; PL= PLANTS; MOE= MOE; CS= consensus score; NCS= normalised consensus score.

As explained in APPENDIX 2, the interactions between a protein and a ligand occur at the level of their surfaces and therefore the shape complementarity between the two molecules, as well as their surface properties, is fundamental for the strength of the interaction between...
them.\textsuperscript{[26]} For this reason and as the pose of the NITD-28 was considered to be related to the compound activity, its shape was used as a first filter of the more than 209,000 molecules belonging to the database. On the basis of the NITD-28 docking conformation a shape query was built with OpenEye ROCS program. As the main interactions that are linked to the inhibitory response are known, “color” features were defined in the query as well as the molecule shape. A total of seven “color” features were used: two ring features (in green) on the naphthalene portion; one anion feature (in pink) on the carboxylate moiety; one hydrogen bond donor feature (in blue) on the sulphonamide NH group; three hydrogen bond features (in red), one on a sulphonamide oxygen and one for each carboxylate oxygen atoms. Figure 6.24 shows the query built from NITD-28 shape and pharmacophoric sites and used for the first filter in the virtual screening.

OpenEye ROCS is a program that runs a rigid superposition of molecules to the given queries. As the shape is a conformation-dependent descriptor, a conformational search was applied to the SPECS database for an appropriate shape comparison. Hydrogen atoms were also added to the structures, as described in the \textit{METHODS} chapter. Approximately 7,130,000 conformations were obtained and screened against the described shape query, saving the 5,000 best scoring hits.

After this filtering, redundant molecules were removed and the NITD-28 compound was added with the aim of identifying compounds that could bind the polymerase better than NITD-28 and therefore be more active against the enzyme. A total of 2,201 structures were docked in the RdRp homology model. The docking was implemented with LeadIT FlexX program, using the NITD-28 molecule for the selection of the receptor site. The 5 poses per
ligand produced by LeadIT FlexX were then rescored using PLANTS and MOE programs. Similarly to that done in previous virtual screening studies of this PhD, the scores obtained from the three programs were used for the calculation of a consensus and normalised consensus score, with the functions reported in the METHODS section. Of all the docked molecules, only 342 scored in the best 25% of the results in all three programs and only six of them scored better than NITD-28.

All the poses were then visually inspected for final selection, considering docking scores, chemical clustering and the displayed interactions with the polymerase. In line with the aims of the study, particular attention was given to the six molecules that scored better than the NITD series most active molecule (NITD-28).

Three molecules (compounds 118, 119 and 120 in APPENDIX 9) were selected and purchased from SPECS.[12]

**BIOLOGICAL EVALUATION OF THE SELECTED COMPOUNDS**

All the purchased compounds were tested *in vitro* with the CPE, the TSA and the polymerase activity inhibition (with radiolabelled probes) assays. Unfortunately, none of the selected molecule gave positive results to the performed tests.

**CONCLUDING REMARKS**

With the aim of discovering novel RdRp NNIs, 3VWS structure was completed with a homology model of the protein and it was used as a molecular target. Docking simulations were performed on this structure with published N-sulphonylanthranilic acid derivatives and their activity data was used for the identification of a promising allosteric site (site C) for RdRp inhibition. As most of the residues of this pocket are conserved amongst the four DENV serotypes, compounds that bind to this site could potentially inhibit the polymerase of all four DENV serotypes. On the basis of these results, a virtual screening approach was designed to select putative novel RdRp inhibitors from the SPECS[12] database. The methodology combined shape screening with docking and rescoring techniques and allowed the selection of compounds, using NITD-28 as a reference. The final three selected compounds were purchased and tested with cell- and polymerase-based assays, but unfortunately none of them were active in the performed assays.

The activity data suggested that the designed virtual screening approach was not successful in identifying anti-DENV agents. Several explanations could be hypothesised for this. For instance, probably the original database did not contain molecular types that could match the
binding features discovered with the structure-activity study. The fact that very few molecules scored better than NITD-28 could suggest that this is possible. It would be interesting to employ again the methodology in order to better assess its prediction power, in particular with different starting databases to assess the validity of this hypothesis.

Despite the inability to find new DENV inhibitors, the docking simulations on the published N-sulphophyanthranilic acid derivatives and the building of structure-activity relationships suggest that the discussed binding mode could indeed be the active one. This represents an important novel piece of information that supports the polymerase site C as a promising molecular target for the design of novel anti-DENV antiviral agents. As it is well conserved in identity of properties amongst the four serotypes, it could be further explored in future drug design studies for the development of serotype-unspecific inhibitors of DENV virus.
6.3 CONCLUSION

RNA-dependent RNA polymerase (RdRp) is an essential enzyme for DENV replication as it compensates for the lack of RNA polymerases in host cells, allowing the transcription of the viral genome. Therefore, it is an attractive drug target and it has been used for the design of antiviral agents against DENV and other RNA viruses.

In this study the objective was to select potential RdRp inhibitors from a small molecule drug-like database (SPECS) through two virtual screening approaches. The first was aimed at targeting a polymerase area (site D) that involved the mechanistically relevant residues Trp795 and Arg737. Twenty-nine molecules were purchased and their antiviral behaviour was assessed in vitro and 10 of them were also tested on DENV RdRp with binding and enzymatic assays. The second virtual screening approach was based on the knowledge acquired from polymerase inhibitors reported in the literature with the objective of finding compounds that could bind the enzyme in site C. Three molecules were purchased and tested in vitro with both cell- and polymerase-based assays for anti-DENV activity. Given the advantage of targeting all four DENV serotype at the same time, particular attention to the conservation of the binding site residues was given in both molecular modelling approaches of compounds selection. Unfortunately, the in vitro assays have not been repeated on more than one serotype to date as both CPE and the polymerase activity inhibition assays were performed with DENV serotype 2 virus and NS5 RdRp respectively. This has not yet allowed the testing of the hypothesis of panserotypic inhibition activity. Further studies should be performed using the other three DENV serotypes.

The first virtual screening approach was successful in identifying two compounds (80 and 82) with a novel chemical scaffold for DENV NNIs and with promising in vitro activity, with different profiles. Compounds 80 and 82 analogues were consequently purchased or synthesised and assessed for activity. Several compounds of the two series were found to be active also at very low micromolar concentrations, making them promising anti-DENV agents, despite their cytotoxicity adverse effects that imply they still need improvement. Activity profiles were not always justified by the in silico model of compounds 80 and 82 protein binding. However, it cannot be excluded that the model does not consider small protein conformational changes, typical of allostERIC inhibitors that could better explain the biological results. Four 82 analogues were successfully synthesised and tested. The activity of one compound (100) was confirmed and important knowledge on structure-activity relationships could be achieved. Furthermore, the method of synthesis that was developed could be used for the design and assessment of additional, novel 82 analogues for the needed improvements.
Unfortunately, the second virtual screening approach failed in the identification of promising novel antiviral agents. However, this study allowed the identification of the putative binding pocket for the published N-sulphonylanthranilic acid derivatives RdRp allosteric inhibitors, returning an important novel piece of information that supports the polymerase site C as a promising molecular target for the design of novel anti-DENV antiviral agents.

Collectively, these studies support that RdRp is a promising drug design target for the development of anti-DENV inhibitors. Furthermore, the molecular modelling suggested that both the novel identified allosteric sites (site C and D) are suitable molecular targets for the design of RdRp NNIs.
6.4 BIOCHEMICAL ASSAYS ON DENV NS5 POLYMERASE

As part of a secondment period in Marseille University, the biochemical assays involving NS5 RdRp reported in Table 6.6 were performed with the aim of assessing in vitro the activity of compounds designed with the three drug design approaches described in chapters within this section. Unfortunately, as the polymerase activity assay required the use of radiolabelled material for which an appropriate training is required, a Marseille University researcher carried out some of the steps in the performance of this assay. Fifteen molecules (reported in APPENDIX 9) were tested, 12 from the virtual screening performed on site D (discussed in chapter 6.2.1) and 3 from the one performed on site C (discussed in chapter 6.2.2). The choice of the compounds among the entire selected small molecules was mainly based on chemical diversity. All tested compounds were dissolved in 100% dimethylsulphoxide (DMSO). As indicated in Table 6.6, two different proteins were used in the two assays: DENV 3 NS5 full-length protein and DENV 2 RdRp domain. An introduction to these assays can be found in the INTRODUCTION section, whilst general procedures, results and test calibrations are reported here. Further details of compound evaluation are discussed in the RESULTS AND DISCUSSION chapter of this section, in the corresponding paragraph for each drug design study. Another in vitro test (cytopathic effect, CPE, inhibition) was performed by our collaborators at the Rega Institute for Medical Research (KU Leuven) and consequently is not discussed in this chapter.

Table 6.6. Summary of performed assays for DENV helicase.

<table>
<thead>
<tr>
<th>Name</th>
<th>What it measures</th>
<th>Enzyme(s) used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal Shift Assay  (TSA)</td>
<td>Protein melting temperature (Tm) shift due to compound binding.</td>
<td>DENV3 full length NS5</td>
</tr>
<tr>
<td>Polymerase activity assay</td>
<td>Compound inhibition of RNA polymerisation, exploiting $^3$H-NTPs.</td>
<td>DENV2 NS5 RdRp</td>
</tr>
</tbody>
</table>

**Abbreviations:** TSA = Thermal Shift Assay; Tm = Melting Temperature; DENV3 = Dengue Serotype 3; NS5 = Non-Structural Protein 5; RNA = Ribonucleic Acid; $^3$H-NTP = Tritium Labelled Nucleoside Triphosphate; DENV2 = Dengue Serotype 2; RdRp = RNA-Dependent RNA Polymerase
THERMAL SHIFT ASSAY (TSA)

The thermal shift assay (TSA) was performed on DENV3 full length NS5 protein with the aim of determining the binding capacity of all fifteen compounds. Differently to the helicase, there are no known compounds that bind to the RdRp domain of NS5 and that are suitable for the assay. However, SAM is known to bind to the MTase portion of NS5 and to give relevant melting point shifts in the TSA. For this reason the full length NS5 protein was used for the assay and SAM was chosen as a reference molecule.

All the compounds were screened at a concentration of 1mM (or 0.5mM for a less soluble compound), while SAM was screened at 0.5mM. The final content of DMSO was equal to 5%, and the Tm were compared to the Tm of the protein in buffer and in buffer containing 5% DMSO. In the assay conditions one compound precipitated and gave a non-canonical melting curve. Unfortunately, it has not been possible to test the precipitated compounds at lower concentration. Also other compounds did not give canonical curves and could not be analysed with this assay. Nine remaining molecules gave canonical curves that could be analysed. Tm shifts (ΔTm) were calculated by subtracting the melting temperature of the protein in buffer with 5% DMSO to the one of the assessed molecular system.

One of the tested compounds (81) gave a significant increase of Tm (2.93°C) as well as SAM, the reference molecule, that gave a melting point increase of 4.26°C. All the other compounds gave a negative shift in melting point, often not significant. In order to confirm that the stabilisation of the protein and the consequent increase in melting point were correlated with compound binding, titration curves for this compound and SAM were performed. Eight concentrations with 2X dilutions were tested, from 1mM, including the blank. For each concentration, the Tm shift was calculated in the same way as for the compound screening.

The above compound that bound to the full length NS5 protein had been selected with the virtual screening study performed on site D of NS5 RdRp and was further discussed in chapter 6.2.1.

POLYMERASE ACTIVITY INHIBITION ASSAY

The activity of the polymerase to synthesise a novel strand of RNA in the presence of a RNA template and of suitable nucleotides was analysed in order to assess the fifteen compounds’ inhibition activity of this process. The protein used for this assay was the RdRp domain of DENV 2 NS5 protein. An initial screen in triplicate at three concentrations (10µM, 20µM and 100µM) was performed for all 15 molecules and the results were analysed considering four controls. The output results were given in percentage of inhibition. The positive control was
the protein in the assay buffer and it represented the 100% activity of the protein. The background noise was also evaluated with another control given by only the RNA template and nucleotides, in the absence of the protein. Two negative controls were used for assay validation. In one case 3’dATP was used, as it is an ATP analogue that, lacking the 3’ OH, interrupts RNA synthesis. An ATP analogue was chosen because the RNA template in the compound screening was a polyU oligonucleotide. In the second negative control, EDTA was added to the protein buffer. Its role was to sequester the catalytic ions of the enzyme and consequently impede the polymerisation reaction.

From the initial screening only one compound (82) showed a meaningful concentration-dependent inhibition of the RdRp activity. The same assay was repeated for this and another two compounds (80 and 81), testing more concentrations for IC₅₀ determination. In order to increase the assay sensitivity, a mini-genome RNA template (constituted of the 5’ and 3’ UTRs of the DENV genome) was used instead of the polyU strand.

The compound that was shown to inhibit the NS5 RdRp together with the two molecules picked for IC₅₀ determination were chosen with the virtual screening study performed on site D of NS5 RdRp and was further discussed in chapter 6.2.1.
6.5 COMPUTATIONAL METHODS

The computational methods for all the studies discussed in chapters 6.2.1 and 6.2.2 are presented here. A description of the molecular modelling programs used can be found in APPENDIX 2.

HARDWARE DETAILS

All the calculations were performed on a 8 core computer with Inter Xeon 2.80 GHz E5462 CPUs.

HOMOLOGY MODELLING

For both studies the missing fragments of the DENV RdRp structure were modelled with homology modelling. All needed template crystal structures and FASTA sequences were downloaded from the Protein Data Bank and GenBank sites respectively. Protein crystal structures were then prepared by removing water molecules and unwanted ligands and ions. The homology model was computed with the Homology Model tool of MOE 2010.10 suite. The complete FASTA sequence of the protein was modelled on the template crystal structure. The protein was protonated during the homology modelling procedure that was performed at a simulated temperature of 300K. All models’ pdb files were modified in order to maintain the correct residues numbering system. Ramachandran plots were computed for model evaluation with MOE 2010.10.

For the study applied to site D, the FASTA sequence (Accession: 2J7U_A) was modelled on DENV NS5 RdRp (PDB ID: 2J7U) template for the whole protein and on WNV polymerase (PDB ID: 2HFZ) for the three missing loops (Tyr308-Asn321; Lys401-Ala421; Cys446-Trp474). The final model was refined with 0.5 RMS gradient and AMBER99 force field. From the obtained homology model (model A) a second one (model B) was built by changing the Thr413 side chain conformation with MOE 2010.10 Rotamer Explorer tool. The modified Thr413 was then energy minimised with MOE 2010.10 using AMBER99 force field and 0.001 gradient.

For the study applied to site C, the FASTA sequence (Accession: 2J7U_A) was modelled on DENV NS5 RdRp (PDB ID: 3VWS) template for the reconstruction of missing fragments.
Neither the intermediates models, nor the final one were refined. The model was saved with the protein in complex with one of the co-crystallised ligands (VWS-1003-A) present in 3VWS crystal structure for an easier site centre definition in the docking simulations.

**SITE D VIRTUAL SCREENING INPUT PREPARATION**

In the case of the molecular modelling study performed on RdRp site D, the homology models were pre-processed with Schrödinger Maestro 9.3\(^{[29]}\) Protein Preparation tool and used for grid generation, adopting the same criteria for both model A and B. Trp795 was chosen for the grid centre and the length was set to 9Å.

The ligand database collected the molecules from the SPECS\(^{[12]}\) database (updated to February 2012). The 2D molecule structures were prepared with Maestro 9.3\(^{[29]}\) LigPrep tool, using OPLS_2005 force field. For each molecule, tautomers were designed considering all possible ionisation states at pH 7 ± 2 with Epik and maintaining the input chiralities. A filter was also applied to remove molecules with molecular weight higher than 600Da, more than 9 rings, more than 4 chiral centres and more than 12 hydrogen bond donor or acceptor groups.

**MAESTRO GLIDE DOCKING SIMULATIONS**

In the case of the molecular modelling study performed on RdRp site D, three different docking simulations using Schrödinger Maestro 9.3\(^{[28]}\) Glide applications were performed in three different moments of the virtual screening approach that was applied in the same way to both homology models. All of the docking simulations were performed using the same protein grids and in a semi-flexible way with the sample ring conformation option for an exhaustive conformational search for the ligands. Ligands with more than 300 atoms or more than 50 rotatable bonds were automatically excluded. The non-bonded interactions were calculated with a 0.8nm van der Waals radii scaling factor and a 0.15nm partial charge cut-off. Different precision modes and different numbers of conformations were saved in the three docking simulations.

The initial high throughput virtual screenings were computed in the HTVS mode, using the prepared SPECS database (see above). Only a maximum of 2 conformations were kept for each ligand and they were energy minimised during the docking simulation. The resulting poses were filtered with a single featured pharmacophore built with the United scheme in MOE 2010.10\(^{[28]}\) and ranked according to the docking score. No rotation or translation was
allowed for the pharmacophore search. One conformation for each molecule that had a negative docking score was saved for further analysis.

The second docking simulation was performed on these filtered molecules with SP precision mode. Only a maximum of 5 conformations were kept for each ligand and they were not energy minimised as a refinement procedure was applied on the docking output. These conformations were then rescored as reported later in the METHODS chapter.

After rescoring, molecules were docked for the third time with Maestro 9.3[29] Glide with XP precision mode. Only a maximum of 3 conformations were kept for each ligand and they were not energy minimised as a refinement procedure was applied on the docking output. Molecules were then visually inspected with MOE 2010.10[28] with particular attention to the molecules that were present in the results obtained for both homology models.

**FLEXX AND PLANTS RESCORING**

In the case of the molecular modelling study performed on RdRp site D, the results of the docking performed with Maestro 9.3[29] Glide with SP precision results were rescored with other two programs: LeadIT FlexX 2.1.3[30] and PLANTS.[31] In both cases no molecule placement was performed and poses were scored in their input conformation and position within the binding pocket.

Both models were used in the same way for the receptor binding site preparation with the LeadIT software. Trp795 was chosen as the centre of the receptor site and the grid sphere had a radius equal to 16Å.

The RdRp models were also used as input files for PLANTS rescoring. The same grid centres as in the Glide docking procedure were used and the binding site radius was set to 12Å. The search algorithm used 20 ants and CHEMPLP scoring function was used.[32]

**CONSENSUS SCORE**

In both virtual screening approaches, part of the procedure required the docking of molecules with one molecular modelling software and the rescoring of the obtained results with other two programs. Then, the scores were processed with a scoring function. Different programs were used in the two studies, but the same previously developed consensus scoring function was used.

For each set of data, the value corresponding to the lowest 25% of the scores (first or lower quartile, $Q_1$) was calculated. Then, molecules that had a score lower than $Q_1$ (representing the
best 25% of poses) were rewarded with a “+1” score, while the ones with score equal or higher than $Q_1$ were assessed with a “0” or “-1” score respectively. Then, the total consensus score (CS) was obtained from the sum of the given scores for each set of data. Equation 6.1 shows the mathematical expression of the consensus score function. Here the three molecular modelling programs are indicated with the letters A, B and C. In the case of the virtual screening approach performed on RdRp site D, the three programs were Maestro 9.3\textsuperscript{29} in SP mode, LeadIT FlexX 2.1.3\textsuperscript{30} and PLANTS.\textsuperscript{31} In the case of the virtual screening approach designed for RdRp site C, the three programs were LeadIT FlexX 2.1.3\textsuperscript{30}, MOE 2010.10\textsuperscript{28} and PLANTS.\textsuperscript{31}

$$CS_i = \text{sign}(Q_{1,A} - x_{A,i}) + \text{sign}(Q_{1,B} - x_{B,i}) + \text{sign}(Q_{1,C} - x_{C,i})$$

*Equation 6.1. Consensus scoring function.*

CS= consensus score; i= pose; $Q_{1,\text{name}}$ first quartile of the set of data corresponding to that letter (see text); $x_{\text{name},i}$ score obtained by pose “i” with the software corresponding to that letter (see text)

A normalised consensus score (NCS) was also calculated, using Equation 6.2. The same letter-code was applied as in Equation 6.1.

$$NCS_i = \frac{x_{A,i}}{\text{Min}_A} + \frac{x_{B,i}}{\text{Min}_B} + \frac{x_{C,i}}{\text{Min}_C}$$

*Equation 6.2. Normalised consensus scoring function.*

NCS= normalised consensus score; i= pose; Min$_{\text{name}}$ lowest value of the set of data corresponding to that letter (see text); $x_{\text{name},i}$ score obtained by pose “i” with the corresponding to that letter (see text)

In both studies Molecules were ranked according to the CS and the NCS in descending order and only molecules that had CS equal to 3 (and therefore performed in the best 25% with all the modelling programs) were kept, in single conformation for further analysis.

The saved molecules from all three docking and rescoring procedures were merged and two conformations for each compound that had CS1 = 3 in at least two of the three target molecular systems were kept for further proceedings.
N-SULPHONYLANTHRANILIC ACID DERIVATIVES SAR STUDY

The molecules belonging to the N-sulphonylanthranilic acid derivatives series published in the literature[10,11] were designed with MOE 2010.10[28] Builder tool and minimised with MMFF94X force field and 0.001 gradient.

The model built on the 3VWS template was used for the receptor binding site preparation with the LeadIT 2.1.3 software. The ligand saved in the protein model (VWS-1003-A) was chosen as the centre of the receptor site and the grid sphere had a radius equal to 9Å.

The designed database of molecules was used for the docking procedure operated with LeadIT 2.1.3 FlexX[30], where original protonation states were maintained and 100 poses for each molecule were saved.

The results were then visually inspected both in LeadIT and in MOE environments and the structure of the RdRp enzyme in complex with the most active compound (NITD-28) was used for the virtual screening procedure.

The 10 conformations saved for each of the NITD series molecules were then correlated with their reported IC$_{50}$ values.[10]

SITE C VIRTUAL SCREENING INPUT PREPARATION

In both studies, the virtual screening approaches were performed using the molecules belonging to the SPECS[12] database (updated to February 2012). In the case of the molecular modelling study performed on RdRp site C, the ligand preparation included a conformational search using MOE 2010.10[28] Import Conformations tool with the following settings: 100 conformations explored per molecule, stochastic strain limit fixed to 7, superposed RMSD test to 0.15, refinement conformation limit to 300, stochastic search failure limit to 30, stochastic search limit to 300, energy minimisation iteration limit to 200, energy minimisation gradient test to 0.5. Compounds with transition metals, more than 8 rings, d-hybridized atoms, molecular weight over 600Da, more than 12 donor/acceptor groups, more than 4 chiral centres, logP under -4 and over 8, over 3 unconstrained chiral centres, more than 7 rotatable bonds and single bond chain length over 6, were omitted. Ligands were subsequently also protonated.
SITE C VIRTUAL SCREENING SHAPE FILTER

In the case of the molecular modelling study performed on RdRp site C, the first filter in the virtual screening approach was a shape screening performed with OpenEye 3.1.2 ROCS. From the N-sulphonylanthranilic acid derivatives SAR study, the binding conformation of the most active molecule (NITD-28) was used as a query. Seven “color” features were also assigned in the query generation. All the conformations of the molecules belonging to the prepared SPECS database (see above) were treated as separate molecules.

Molecules were aligned and scored according to the shape overlap and to the chemistry features match with the assigned “color” pharmacophore. Implicit Mills Dean force field was used for chemistry evaluation of the compound database. Only the best 5,000 hits were then saved in the output file.

The only one conformation for each molecule was then kept for docking and rescoring procedures.

LEADIT FLEXX DOCKING SIMULATIONS

In the case of the molecular modelling study performed on RdRp site C, the homology model prepared on 3VWS template in complex with NITD-28 was used for the docking and rescoring processes described here. The receptor site was prepared with LeadIT 2.1.3 software, using the ligand as the centre and a grid sphere with radius equal to 6.5Å.

NITD-28 structure was added to the shape filtering results and all of these molecules were docked with LeadIT 2.1.3 FlexX, using the original protonation states. Five poses for each molecule were saved.

MOE AND PLANTS RESCORING

In the case of the molecular modelling study performed on RdRp site C, the results of the docking performed with LeadIT FlexX 2.1.3 were rescoring with the other two programs: MOE 2010.10 and PLANTS. In both cases no molecule placement was performed and poses were scored in their input conformation and position within the binding pocket.

The first rescoring was performed with MOE 2010.10 Docking tool. In the homology model, the ligand (NITD-28) was selected for the binding site determination and the protein was set
as the receptor. Neither placement nor refinements were performed and molecules were rescored with London dG scoring function. All conformations were retained.

The model was also used as input files for PLANTS\textsuperscript{[31]} rescoring. The same grid centre as in the LeadIT FlexX docking procedure was used and the binding site radius was set to 16Å. The search algorithm used 20 ants and CHEMPLP scoring function was used.\textsuperscript{[32]}
6.6 EXPERIMENTAL SECTION

6.6.1 BIOCHEMICAL ASSAYS

COMPOUND PREPARATION

Of the 15 tested compounds, 13 were purchased from SPECS\textsuperscript{[12]} and two (compounds 77 and 78) were synthesised by a student in our laboratory as part of a structure-activity study of one of the compounds selected for the RdRp allosteric site D.\textsuperscript{[17]}

For each compound, a mother solution of 20mM in DMSO was prepared, starting from approximately 5mg of compound. One compound (121) did not dissolve at this concentration and therefore it was diluted to 10mM in DMSO. These solutions were then used as starting material for the biochemical assays.

THERMAL SHIFT ASSAY (TSA)

The TSA assay was performed on one DENV full length NS5 protein from serotype 3 (EVA316, 103.359kDa). The enzyme was previously prepared and purified in the laboratory. The buffer used for the assay was constituted of Tris (pH 7.5) 50mM, NaCl 150mM, TEMED 200mM and MgCl$_2$ 1mM in water. SYPRO ORANGE (SO, with excitation wavelengths at 300nm and 470nm and emission wavelength at 570nm) was used as fluorescent probe.

PCR plates with 96 wells were used and each well was prepared with 25μL of total solution (20μL of protein, 1.25μL of compound 20mM solution and 3.5μL of SO solution). The final concentration of the protein was 0.40mg/ml, while the final concentration of the compounds was of 1mM in each well, except for the poorly soluble compound (121) that was at 0.5mM. Three controls were also prepared for each protein: one with 0.5mM of SAM as it is known to bind the enzyme, one with 5% DMSO and one with the protein alone.

Before performing the assay, the wells were visually inspected in order to detect the presence of eventual precipitated compounds. The assay was performed in a modified PCR machine that reads the fluorescence in function of the temperature. It was set for 360 cycles of 12s each with a temperature increase of 0.2°C, starting from a temperature of 20°C.

The resulting charts of fluorescence on temperature were analysed with ORIGIN. For each protein, a temperature window was chosen in order to include the reported thermal shift and
the curves were fitted with a Boltzmann distribution. Melting points and fitting errors were, in this way, determined.

The titration curves for compound 81 were performed with the same protocol as the screening, but with different concentrations of the molecule (1mM, 500µM, 250µM, 125µM, 62µM, 31µM, 16µM and 0µM). As positive control, the same approach was applied to SAM with concentrations of 500µM, 250µM, 125µM, 62µM, 31µM, 16µM and 8µM.

**POLYMERASE ACTIVITY INHIBITION ASSAY**

The inhibition activity of the 15 selected compounds was assessed by measuring the incorporation of radiolabeled nucleotides on behalf of the polymerase enzyme in presence of a RNA template. For this assay, an active form of DENV serotype 2 RdRp domain of NS5 protein was used. The protein was previously prepared and purified in AFMB, Marseille University laboratories. The reaction buffer was constituted of Hepes pH 8.0 50mM, KCl 10mM, MgCl₂ 1mM, MnCl₂ 2mM and DTT 10mM in water. Therefore, the protein mix contained the polymerase domain and the RNA template in the reaction buffer.

The compound solutions were organized in 96-wells plates and concentrations were diluted in different way with water according to the final desired compounds concentration, but maintaining a constant quantity (5%) of DMSO for all the tests. The compounds distribution was performed by Biomek NX (Beckman) pipetting robot. Protein mix was then put in contact with the prepared compound concentrations. Controls were also used for assay validation. The positive control was composed by the protein mix, while the background control comprehended the RNA template, but not the protein, in the reaction buffer. In addition to these, two negative controls were set up, both containing the protein mix, as well as EDTA (100mM) in one case and the inhibitor 3’dATP (100mM) in the other one. The distribution of the controls and of the protein mix in the plates was performed with the Biomek 3000 workstation.

The polymerase reaction was started by the addition of tritium-labeled nucleotides, distributed in the plates with Biomek 3000. Then, the plates were closed and incubated at 30°C for 10 minutes.

The reaction was stopped with the addition of EDTA 100mM, distributed by Biomek 3000. The products were transferred to a membrane filter with a Packard Filtermate Harvester. The filters were washed with ammonium formate (pH 8.0), EtOH and dried. The bound radioactivity on the membranes was determined with liquid scintillation counting and inhibition percentage was calculated for each compound.

As the assay requires the use of tritium-labelled products, the procedure here described was applied by a trained scientist in Marseille University.
In the initial screening, the compounds were tested at three final concentrations (10, 20 and 100µM) according to the procedure reported above, with a RdRp final concentration of 100nM. The RNA template used was a polyU sequence of around 200 nucleotides at a final concentration of 100nM. ATP (200µM, 2% [³H]-ATP) was used for polymerization monitoring. The assay was performed in triplicates.

Selected compounds were tested again at different concentrations, starting from 100µM, for IC₅₀ determination according to the procedure described above. Also in this case the assay was repeated in triplicates, but the RNA template was not a polyU sequence, but with a minigenome, constituted of the two UTR regions of DENV genome.
6.6.2 CHEMICAL SYNTHESIS

General information

All chemicals, reagents and solvents were purchased from Aldrich or purified by standard techniques.

Thin Layer Chromatography (TLC)

Silica gel aluminium backed plates (Merck Kieselgel 60F$_{254}$, 0.2 thickness) were used. They were developed by the ascending method. After solvent evaporation, compounds were visualised by irradiation with ultraviolet (UV) light at 254nm and 366nm.

Column Chromatography

Glass columns were slurry packed in the appropriate eluent under gravity, with gel silica (40-60µm) from Merck. Samples were applied as a concentrated solution in the same eluent. Fractions containing the product were identified by TLC, combined and the solvent removed in vacuo.

Melting point

Melting points were determined using Griffin Melting Point Apparatus.

NMR Spectroscopy

$^1$H and $^{13}$C NMR spectra were recorded on a Bruker ADVANCE 500 spectrometer (500MHz and 125MHz respectively) at 25°C. Spectra were auto calibrated to the residual signal of the deuterated solvent used. Chemical shifts (δ) are given in parts per million (ppm) and the coupling constants (J) are given in Hertz (Hz).

Abbreviations: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet).

Mass Spectrometry

Low resolution mass spectra were recorded on a Bruker Daltonics microTOF-LC Mass Spectrometer. Atmospheric pressure electrospray ionisation method (ESI) in positive mode
and Time-of-Flight mass analyser with Exact Mass measurements capability were used. Mass-charge ratio (m/z) is given in Da. Abbreviations: M-H⁺ (molecular ion signal given by the addition of a H⁺ ion), M-Na⁺ (molecular ion signal given by the addition of a Na⁺ ion).

**Elemental analysis**

The elemental analysis was performed by MEDAC Ltd, Analytical and chemical consultancy services. The values are reported in percentage.
Synthesis of \( N^{4},N^{4'}\)-di(furan-2-carbonyl)-[1,1'-biphenyl]-4,4'-disulphonohydrazide (109)

Scheme:

![Scheme](image)

Reaction conditions: Pyridine, 0°C to r.t., 4-5 h

The 2-Furoic hydrazine (111, 1.58g, 12.530mmol) was dissolved in pyridine (30ml) and the solution was cooled to 0°C. A solution of the biphenyl-4,4'-disulphonyl chloride (112, 2.00g, 5.690mmol) in pyridine (50ml) was added to the first solution and the reaction was stirred at room temperature for 4-5 hours, monitoring it with TLC (EtOAc). The solvent was then removed under vacuum and the residue was washed with water. The solid was crystallised in dioxane/water and washed with ethanol and diethylether, obtaining a 690mg (1.300mol) of compound 109 \( (N^{4},N^{4'}\)-di(furan-2-carbonyl)-[1,1'-biphenyl]-4,4'-disulphonohydrazide) as a white solid.

**Yield:** 23%

Chemical formula: \( C_{22}H_{18}N_{4}O_{8}S_{2} \)
Molecular Weight: 530.53
Monoisotopic mass: 530.06

**Aspect:** White powder

**m.p.:** 228-230°C
**1H NMR** (500MHz, DMSO), 6: 10.63 (s, 2H), 10.18 (s, 2H), 7.95-7.93 (m, 8H), 7.87-7.86 (Brs,2H), 7.20 (d, J=3.4, 2H), 6.62 (dd, J1=3.4, J2=1.5, 2H)

**13C NMR** (125MHz, DMSO), 6: 156.90 (CH, aromatic), 146.06 (CH, aromatic), 145.34 (C, aromatic), 142.41 (C, aromatic), 139.20 (C, aromatic), 128.35 (CH, aromatic), 127.42 (CH, aromatic), 115.03 (CH, aromatic), 111.83 (CH, aromatic)

**ESI-MS** (MeOH) m/z: Insoluble

**Calculated elemental analysis (%)**: C, 49.81; H, 3.42; N, 10.56; S, 12.09

**Experimental elemental analysis (%)**: C, 48.91; H, 3.57; N, 10.10; S, 11.44
Synthesis of sodium 9H-fluorene-2,7-disulphonate (113)

Scheme:

Fluorene (114, 10.0g, 60mmol) was added to glacial acetic acid (CH₃COOH, 60ml) and the system was cooled to 0°C. After drop wise addition of chlorosulphonic acid (115, 8ml, 120mmol), the reaction mixture was refluxed at 120°C under controlled atmosphere and monitored with TLC (CHCl₃, MeOH, CH₃COOH). Other 4ml of chlorosulphonic acid (60mmol) were added after 2 and 3 hours and the reaction reached completion after 4 hours. The reaction mixture was poured into 240ml of a solution of sodium hydroxide (10g) in brine. The precipitate was filtered, washed three times with 50ml of brine and twice with distilled water and it was dried over night at 40°C under vacuum, obtaining 20.87g (56.3mmol) of compound 113 (sodium 9H-fluorene-2,7-disulphonate) as a beige powder.

Yield: 94%

Chemical formula: C₁₃H₈Na₂O₆S₂
Molecular Weight: 370.31
Monoisotopic mass: 369.96

Aspect: beige powder

¹H NMR (500MHz, D₂O), δ: 7.73 (s, 2H), 7.63 (d, J=8.0, 2H), 7.51 (d, J=8.0, 2H), 3.43 (s, 2H)

¹³C NMR (125MHz, D₂O), δ: 144.71 (C, aromatic), 142.59 (C, aromatic), 141.41 (C, aromatic), 124.35 (CH, aromatic), 122.08 (CH, aromatic), 120.78 (CH, aromatic), 36.43 (CH₂)
Synthesis of 9H-fluorene-2,7-disulphonyl dichloride (116)

Scheme:

\[
\begin{align*}
\text{Na}^+ & \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \Quad
General procedure for the synthesis of compounds with a 9H-fluorene-2,7-disulphonylamide scaffold (101-103)

Scheme:

\[
\begin{align*}
\text{SO}_2\text{Cl}_2 \quad 116 \quad + \quad \text{DIPEA, DCM, r.t., 1.5h} \quad 101-103 \\
\text{NH}_2 \quad 117 \\
\end{align*}
\]

Reaction conditions: DIPEA, DCM, r.t., 1.5h

9H-fluorene-2,7-disulphonyl dichloride (116, 200mg, 0.55mmol) was dissolved in dichloromethane (DCM, 17ml) and the solution was cooled to -50°C. Then, the appropriate amine (117, 0.2ml, 1.55mmol) and N,N-diisopropylethilamine (DIPEA, 0.2ml, 1.15mmol) were added to the solution in this order. The reaction mixture was stirred at room temperature and monitored with TLC (DCM/EA) for 1.5 hours. The solvent was then evaporated under vacuum and the residue was treated with hydrochloric acid 1N. The solid was filtrated, washed with distilled water and dried over night at 40°C under vacuum. The powder was then purified by column chromatography (silica gel, DCM/EtOAc), obtaining the final product.
Synthesis of $N^2,N^7$-bis(1-phenylethyl)-9H-fluorene-2,7-disulphonamide (101)

Chemical formula: $C_{29}H_{28}N_2O_4S_2$
Molecular Weight: 532.67
Monoisotopic mass: 532.15

Yield: 66% (194mg, 0.36mmol)

Aspect: pale yellow powder

m.p.: 178-190°C

$^1H$ NMR (500MHz, CDCl$_3$), δ: 7.84-7.79 (m, 6H), 7.17-7.10 (m, 10H), 4.94 (d, J=6.9, 2H), 4.62-4.57 (m, 2H), 3.79 (s, 2H), 1.50 (d, J=6.9, 6H)

$^{13}C$ NMR (125MHz, CDCl$_3$), δ: 144.28 (C, aromatic), 143.74 (C, aromatic), 141.79 (C, aromatic), 140.10 (C, aromatic), 128.48 (CH, aromatic), 127.50 (CH, aromatic), 126.36 (CH, aromatic), 126.16 (CH, aromatic), 124.12 (CH, aromatic), 120.86 (CH, aromatic), 53.88 (CH, aliphatic), 36.75 (CH$_3$, aliphatic), 23.70 (CH$_3$, aliphatic)

ESI-MS (MeOH) m/z: 555.1 (M-Na$^+$)

Calculated elemental analysis (%): C, 65.39; H, 5.30; N, 5.26; S, 12.04
Experimental elemental analysis (%): C, 65.29; H, 5.38; N, 5.35; S, 11.84
**Synthesis of N²,N⁷-bis((R)-1-phenylethyl)-9H-fluorene-2,7-disulphonamide (102)**

Chemical formula: C_{29}H_{28}N_{2}O_{4}S_{2}

Molecular Weight: 532.67

Monoisotopic mass: 532.15

![Chemical structure of the compound](image)

**Yield:** 56% (165mg, 0.31mmol)

**Aspect:** pale yellow powder

**m.p.:** 177-180°C

**¹H NMR** (500MHz, CDCl₃), δ: 7.84-7.79 (m, 6H), 7.17-7.09 (m, 10H), 5.10 (d, J=7.0, 2H), 4.62-4.56 (m, 2H), 3.77 (s, 2H), 1.49 (d, J=7.0, 6H)

**¹³C NMR** (125MHz, CDCl₃), δ: 144.28 (C, aromatic), 143.74 (C, aromatic), 141.82 (C, aromatic), 140.11 (C, aromatic), 128.47 (CH, aromatic), 127.48 (CH, aromatic), 126.35 (CH, aromatic), 126.16 (CH, aromatic), 124.12 (CH, aromatic), 120.86 (CH, aromatic), 53.89 (CH, aliphatic), 36.75 (CH₃, aliphatic), 23.70 (CH₃, aliphatic)

**ESI-MS** (MeOH) m/z: 555.1 (M-Na⁺)

**Calculated elemental analysis (%)**:
C, 65.39; H, 5.30; N, 5.26; S, 12.04

**Experimental elemental analysis (%)**:
C, 65.33; H, 5.24; N, 5.21; S, 11.91
Synthesis of $N^2,N^7$-bis((S)-1-phenylethyl)-9H-fluorene-2,7-disulphonamide (103)

Chemical formula: C$_{29}$H$_{28}$N$_2$O$_4$S$_2$

Molecular Weight: 532.67

Monoisotopic mass: 532.15

Yield: 73% (215mg, 0.40mmol)

Aspect: pale yellow powder

m.p.: 173-176°C

$^1$H NMR (500MHz, CDCl$_3$), δ: 7.84-7.79 (m, 6H), 7.16-7.08 (m, 10H), 5.06 (d, J=7.0, 2H), 4.62-4.57 (m, 2H), 3.78 (s, 2H), 1.49 (d, J=7.0, 6H)

$^{13}$C NMR (125MHz, CDCl$_3$), δ: 144.28 (C, aromatic), 143.74 (C, aromatic), 141.82 (C, aromatic), 140.11 (C, aromatic), 128.46 (CH, aromatic), 127.47 (CH, aromatic), 126.35 (CH, aromatic), 126.17 (CH, aromatic), 124.12 (CH, aromatic), 120.86 (CH, aromatic), 53.89 (CH, aliphatic), 36.75 (CH$_3$, aliphatic), 23.70 (CH$_3$, aliphatic)

ESI-MS (MeOH) m/z: 555.1 (M-Na$^+$)

Calculated elemental analysis (%): C, 65.39; H, 5.30; N, 5.26; S, 12.04

Experimental elemental analysis (%): C, 65.46; H, 5.33; N, 5.37; S, 11.88
6.7 BIBLIOGRAPHY


[17]: H. Trinh, Design and Synthesis of Novel Allosteric Inhibitors of Dengue Virus, Cardiff School of Pharmacy and Pharmaceutical Sciences, Cardiff University (2013)


Section 7: GENERAL CONCLUSIONS
Drug discovery on DENV has been shown to be challenging in the last decades, and the development of novel therapeutic agents is still an unmet need to date. In order to meet this need, this PhD study had the overarching aim to design novel molecules as potential antiviral agents against DENV. This main aim was subdivided in three parts: to identify suitable drug targets, to generate knowledge on these targets and to design novel antiviral agents against DENV infection.

With the aim of identifying suitable drug targets, four promising strategies to target viral proteins were identified. The proteins were: the capsid structural protein (C), the NTPase/helicase portion of NS3 (NS3hel), the methyltransferase domain of NS5 (NS5 MTase) and the RNA-dependent RNA polymerase part of NS5 (NS5 RdRp). All of them are essential for DENV replication and are not present in humans, aiding the design of potential selective anti-DENV agents.

With the aim of improving the understanding of the molecular targets, this work has successfully uncovered important and novel knowledge about DENV, through two basic research studies.

As discussed in section 2, molecular modelling and docking simulations have shown a promising anti-DENV compound (ST-148) active against the virus both in vitro and in vivo to inhibit the capsid protein (C) through the stabilisation of higher-order oligomeric C structures, confirming that protein-protein interaction stabilisation is an achievable and valid tool for antiviral drug discovery. The understanding of the mechanism of action of this compound is an important discovery that should allow the development of ST-148-like molecules that might succeed in the discovery of a medicine able to reach the market for DENV infection therapy.

In another study presented in section 4, a novel approach that included classical and accelerated MD simulations, principal component analysis and binding free energy calculations has produced relevant results for understanding the mechanism of NS3hel translocation along the RNA strand that other studies were unable to reveal. Thus, for the first time it has been possible to formulate a hypothesis according to which the 3'-5' translocation of NS3hel involves both large scale domain movements and small residue conformation changes and occurs in a NTP-driven “ratchet-like” manner, similarly to HCV. Residues particularly important in the translocation mechanism and directionality were identified and could be used for future drug design approaches.

With the aim of discovering novel antiviral agents, drug discovery approaches have also been applied using the interaction of different disciplines. Several molecular modelling techniques, classical Medicinal Chemistry approaches, chemical synthesis and in vitro assays were combined and have been applied to the four viral-encoded proteins identified as promising drug targets. Even if challenges were encountered in all the fields, problems were found with
the *in vitro* assays that did not always allow an understanding of the activity of the designed compounds, in particular for NS3hel and NS5 RdRp inhibitors (discussed in sections 3 and 6 respectively). For instance, some assays were not returning consistent results when repeated with the same or slightly different conditions. For example, the fluorescence polarisation assay with GTP bodipy did not return consistent results and changing the polymerase activity inhibition assay from using $\text{H}^3$-labelled nucleotides to the fluorescent probe PicoGreen changed the IC$_{50}$ values for compound 80. In general, in the structure-activity relationships definition the cell activity assays was less informative than in the biochemical assays, suggesting that the latter should be considered the primary activity assays. Many factors might explain this, including difficulty to cross the cell membrane, intracellular molecule sequestration, interaction etc. Consequently, the reliability of *in vitro* assays has been confirmed to be an issue that has slowed down anti-DENV drug discovery and that has yet to be resolved. However, the cellular cytotoxicity seen, especially at high concentrations did raise issues with some chemical scaffolds and aided in selecting molecules to study further. Furthermore, the assays have not generally been repeated in more than one DENV serotype and consequently the attention given in drug design to the potential serotype specificity of the compounds could not be assessed.

Despite these issues, the molecular modelling-based drug discovery studies returned molecules with IC$_{50}$S in the low µM range, making them good leads for further development and compatible with published DENV inhibitors.$^{[3,4]}$

Based on co-crystallised fragments that inhibited DENV NS5 MTase, two fragment-based drug discovery (FBDD) approaches were applied to this enzyme and were discussed in section 5. Even if very few iterations of compound design and evaluation have been performed, activity improvement (sometimes over 100-fold for 2'-O methylation) was achieved for one scaffold, indicating that the designed drug design method is moving in the right direction and confirming FBDD as a suitable multidisciplinary approach for the development of NS5 MTase inhibitors. Future improvements of the designed molecules should lead to the development of promising drugs against DENV infection. The activity pattern was also explored with molecular modelling for one of the co-crystallised fragments and it was hypothesised that it could act as a S-adenosyl methionine (SAM) competitive inhibitor.

In sections 3 and 6 several virtual screening campaigns based on different general methodologies were applied to NS3hel and NS5 RdRp with the aim of selecting putative inhibitors from a database of drug-like small molecules (generally SPECS) were discussed. Identified hit molecules were further progressed through the assessment of analogue compounds that were purchased or synthesised and for some series preliminary structure-activity relationships were suggested.
One of the virtual screening campaigns was a ligand-based shape screening, as discussed in section 3. This was grounded on the hypothesis that shape complementarity between two molecules, as well as their surface properties, is fundamental for the strength of the interaction between them. Ouabain, a natural product that was found to inhibit NS3hel, was used as a reference and compounds with a similar shape were searched. Thus, a novel series of NS3hel inhibitors have been identified, validating NS3hel-RNA complex disruption as an anti-DENV drug development strategy and supporting the study’s basic hypothesis.

All the other virtual screening campaigns were structure-based. Unfortunately, the two designed to target the NTP and the RNA binding site confirmed that NS3hel is a challenging drug target and did not return promising inhibitors (section 3). However, as discussed in section 6, one of the two designed approaches that aimed the inhibition of NS5 RdRp through an allosteric inhibition approach was successful in identifying two novel chemical scaffolds with different in vitro activity profiles that probably bind to the polymerase site D. Thus, DENV NS5 RdRp was validated as an antiviral target and the design of non-nucleoside inhibitors (NNIs) was confirmed as a valid drug development strategy. Four compounds were also synthesised, confirming the activity of 100 and identifying a synthesis pathway for 9H-fluorene-2,7-disulphonamide-based compounds. Even if cell- and enzyme-based assay results were not always consistent, preliminary structure-activity relationships (SARs) could be established. Aiming to improve these compounds’ activity, and cytotoxicity, these scaffolds could be further developed in the future, through constructed SARs and suggested chemical synthetic approach. Although no promising inhibitors were identified with the other RdRp-based virtual screening approach presented in section 6, this study identified the putative binding pocket for the published N-sulphonylanthranilic acid derivative RdRp allosteric inhibitors, returning an important novel piece of information that corroborates polymerase site C as a potential molecular target for the design of novel anti-DENV antiviral agents.

In summary, this PhD study has made a positive step forward in anti-DENV research: an increase in the basic understanding on DENV, in particular about the viral enzymes and the approaches that can be adopted for the development of anti-DENV drugs was achieved, novel promising DENV inhibitor scaffolds have been discovered and developed and the chemical synthesis of one of them was suggested.
Section 7: GENERAL CONCLUSIONS


Section 8: APPENDIXES
APPENDIX 1: VACCINES AND DRUGS IN DEVELOPMENT

The vaccines and drugs in development for DENV are reported in three tables: one for the vaccines in clinical trials, one for the drugs developed to inhibit viral targets and one for the drugs designed to target host proteins. References about the reported molecules are at the end of the appendix.

Table A1.1. Status of DENV vaccine development on-going in clinical trials.

<table>
<thead>
<tr>
<th>Vaccine type</th>
<th>Developers</th>
<th>Description</th>
<th>Doses</th>
<th>Status</th>
</tr>
</thead>
</table>
| LIVE ATTENUATED VIRUS VACCINES (LAV) | WRAIR and GSK                  | DENV attenuated by several passages in dog kidney cells; Tetravalent formulation | 2     | Phase II completed\([1]\)  
Proven to be safe and immunogenic but protective efficacy needs to be evaluated (delayed) |
|                              | NIH, NIAID, LID, Biological E. (Panacea), Biotec (Butantan), Vabiotech | Reverse genetics application: deletion of 30 nucleotides in the 3’ untranslated region (UTR) of the genome of all four DENV serotypes (TV-003) | 2     | Phase II\([2]\)                           |
| LIVE CHIMERIC VIRUS VACCINES | CDC, Takeda                     | Structural genes from DENV1, 3 and 4 are inserted in a DENV2 attenuated genome (replacement of the 3’ loop structure). (DENVax) | 2     | Phase II\([3]\)                           |
|                              | Sanofi Pasteur                  | The licensed YFV 17D vaccine is used as a backbone for the expression of all four DENV prM and E proteins. (CYD-TDV) | 3     | Phase III\([4]\)  
but Phase IIb showed only 30% effectiveness and efficacy only on DENV1, 3 and 4. |
| PURIFIED INACTIVATED VIRUSES (PIV) | GSK, WRAIR, FIOCRUZ            | Whole inactivated DENV1, 2, 3 and 4 with a GSK adjuvant (DPIV-001)          | 2     | Phase I\([5]\)                           |
### Section 8: APPENDIXES

#### APPENDIX 1

<table>
<thead>
<tr>
<th>DNA VACCINE</th>
<th>NMRC (USA)</th>
<th>All serotypes DENV prM and E genes inserted in a non-replicating adenovirus vector; With/without adjuvant (TVDV)</th>
<th>3</th>
<th>Phase I[^6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECOMBINANT SUBUNITS</td>
<td>Hawaii Biotech, Merck</td>
<td>Recombinant truncated (80% of protein) DENV1, 2, 3 and 4 E proteins with native-like conformation, cultured in Drosophyla cells; With/without adjuvant (V180)</td>
<td>3</td>
<td>Phase I[^7]</td>
</tr>
</tbody>
</table>

**Abbreviations:** WRAIR, Walter Reed Army Research Institute; GSK, GlaxoSmithKline; NIH, National Institute of Health; NIAID, National Institute of Allergy and Infectious Diseases; LID, Laboratory of Infectious Diseases; CDC, Center for Disease Control; FIOCRUZ, Oswaldo Cruz Foundation; NMRC, Naval Medical Research Center


<table>
<thead>
<tr>
<th>Name</th>
<th>Drug Discovery Method</th>
<th>Target</th>
<th>In vitro activity</th>
<th>In vivo activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>NITD448\textsuperscript{[8]} (s.m.)</td>
<td>\textit{In silico} virtual screening and \textit{in vitro} validation</td>
<td>E (βOG pocket)</td>
<td>Inhibits membrane fusion and viral entry</td>
<td>ND</td>
<td>Interrupted (poor selectivity and Pk)</td>
</tr>
<tr>
<td>“Thiophene quinazoline”\textsuperscript{[9]} (s.m.)</td>
<td>\textit{In silico} virtual screening and \textit{in vitro} validation</td>
<td>E (βOG pocket)</td>
<td>Interacts with E protein and blocks entry after internalisation in endosomes. Active also against YFV, WNV, JEV.</td>
<td>ND</td>
<td>Interrupted (precipitation in mouse model GI tract)</td>
</tr>
<tr>
<td>PO2\textsuperscript{[10]} (s.m.)</td>
<td>\textit{In silico} virtual screening and \textit{in vitro} validation</td>
<td>E (βOG pocket)</td>
<td>Interacts with E protein (competition with βOG assessed) and inhibits DENV and YFV.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>A5\textsuperscript{[11]} (s.m.)</td>
<td>\textit{In silico} virtual screening and \textit{in vitro} validation</td>
<td>E (βOG pocket)</td>
<td>Inhibits E-mediated fusion and is active against DENV, YFV and WNV. Not cytotoxic.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>1662G07\textsuperscript{[12]} (s.m.)</td>
<td>HTS, using FP based on competition with a binding peptide</td>
<td>E (βOG pocket)</td>
<td>Inhibits DENV2 fusion and interacts with both pre- and post-fusion E</td>
<td>ND</td>
<td>Interrupted (No sufficiently broad spectrum)</td>
</tr>
<tr>
<td>DNS5\textsuperscript{[13,14]} (p.)</td>
<td>Synthetic peptides (portions of E) tested with plaque formation inhibition assay</td>
<td>E (sequence: conserved stem region)</td>
<td>Broad spectrum flavivirus inhibitor (DENV, WNV, JEV, TBEV), inducing the production of empty virions. It binds to lipid bilayers and trimeric E proteins</td>
<td>ND</td>
<td>Further studies needed (peptide-related issues for drug discovery)</td>
</tr>
<tr>
<td>1OAN1\textsuperscript{[14,15]} (p.)</td>
<td>Peptides with \textit{in silico} optimised sequences</td>
<td>E (sequence: mimic of β-sheet connecting domains I and II)</td>
<td>Binds to DENV2 soluble E and inhibits DENV2 binding to target cells</td>
<td>ND</td>
<td>Further studies needed (but it has peptide-related issues for drug discovery)</td>
</tr>
<tr>
<td>Antibiotic SA-1\textsuperscript{[16,17]} (s.m.)</td>
<td>CPE on DENV2 infected cells of doxorubicin analogues</td>
<td>E (βOG pocket, according to docking simulation)</td>
<td>Broad spectrum inhibition: DENV (1, 2 and 3), YFV. It binds directly to viral particle, blocking the entry in host cells.</td>
<td>ND</td>
<td>On-going preclinical</td>
</tr>
</tbody>
</table>

\textbf{Table A1.2. Potential inhibitors of viral targets.}
<table>
<thead>
<tr>
<th><strong>Antibiotic/LCTA-949</strong>&lt;sup&gt;[24]&lt;/sup&gt;</th>
<th>CPE on DENV2 infected cells of teicoplanin aglycon derivatives</th>
<th><strong>Entry</strong> (unknown target)</th>
<th>Broad spectrum entry inhibition: DENV, YFV, WNV, JEV, TBEV</th>
<th>ND</th>
<th>No follow up studies reported</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3-O-sulphated glucuronide derivatives</strong>&lt;sup&gt;[19]&lt;/sup&gt; (s.m.)</td>
<td>Design and synthesis of compounds, based on carbohydrate structures of active glycans</td>
<td><strong>E</strong> (probably absorption binding site)</td>
<td>Inhibits cell infection and virion binding to host cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>ST-148</strong>&lt;sup&gt;[20,21]&lt;/sup&gt; (s.m.)</td>
<td>HTS of compounds on DENV-infected cells</td>
<td><strong>C</strong> (C-C proteins interface)</td>
<td>Inhibits cell infection, reducing both virion assembly and disassembly. It binds to C and enhances C-self aggregation.</td>
<td>Efficacious in AG-129 mouse model.</td>
<td>On-going preclinical</td>
</tr>
<tr>
<td><strong>Retro tripeptide</strong>&lt;sup&gt;[22]&lt;/sup&gt; (p.)</td>
<td>Substrate mimetics (with aromatic N-termini) screening</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV and WNV proteases.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>Cyclic peptide</strong>&lt;sup&gt;[23]&lt;/sup&gt; (p.)</td>
<td>Screening of natural peptides and optimisation by cyclisation</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV protease, probably binding to the active site (docking). Permeates cells, but no EC&lt;sub&gt;50&lt;/sub&gt; reported.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>Tetrapeptide</strong>&lt;sup&gt;[24]&lt;/sup&gt; (p.)</td>
<td>Screening of aldehyde tetrapeptides</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV protease, probably binding to the active site (docking).</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>α-ketoamides</strong>&lt;sup&gt;[25]&lt;/sup&gt; (s.m.)</td>
<td>HTS and SAR study</td>
<td><strong>NS3-2B protease</strong></td>
<td>Active on DENV and WNV proteases and on DENV infected cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>Phthalazine derivatives</strong>&lt;sup&gt;[26]&lt;/sup&gt; (s.m.)</td>
<td>HTS on protease and SAR study</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV protease, but not active on cells (permeability issues).</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td><strong>166347</strong>&lt;sup&gt;[27]&lt;/sup&gt; (s.m.)</td>
<td>HTS on protease and SAR study</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV and WNV proteases, but not active on cells (permeability issues).</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Anthracene based&lt;sup&gt;[28,29]&lt;/sup&gt; (s.m.)</td>
<td>Virtual screening of small molecules and SAR studies</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV protease and in infected cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Aminobenzamide scaffold&lt;sup&gt;[30]&lt;/sup&gt; (s.m.)</td>
<td>Rational design of aminobenzamide derived compounds.</td>
<td><strong>NS3-2B protease</strong></td>
<td>Active on DENV and WNV proteases, but not active on cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Quinoline derivatives&lt;sup&gt;[31]&lt;/sup&gt; (s.m.)</td>
<td>Virtual screening and scaffold hopping.</td>
<td><strong>NS3-2B protease</strong></td>
<td>Active on DENV protease and in DENV replicon assay</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>1,3,4-oxadiazole derivatives&lt;sup&gt;[32]&lt;/sup&gt; (s.m.)</td>
<td>Screening of a focused library of synthesised compounds.</td>
<td><strong>NS3-2B protease</strong></td>
<td>Active on DENV and WNV proteases. Unknown activity on cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>8-hydroxyquinoline derivatives&lt;sup&gt;[33]&lt;/sup&gt; (s.m.)</td>
<td>SAR study on previously identified WNV protease inhibitors</td>
<td><strong>NS3-2B protease</strong> (competitive inhibitor)</td>
<td>Active on DENV and WNV proteases. Unknown activity on cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>BP2109&lt;sup&gt;[34]&lt;/sup&gt; (s.m.)</td>
<td>HTS on infected cells</td>
<td><strong>NS3-2B protease</strong></td>
<td>Active on DENV infected cells and causes mutations in NS3 protease in resistant viruses.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Ivermectin&lt;sup&gt;[35,36,37]&lt;/sup&gt; (s.m.)</td>
<td>Docking and enzymatic assay</td>
<td><strong>NS3hel</strong> (RNA binding site competitive inhibitor)</td>
<td>Broad spectrum (DENV, WNV, YFV) inhibition of dsRNA unwinding and active on DENV NS3-2B protease. More active for WNV and YFV in cells.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Ouabain, Avermectin&lt;sup&gt;[37]&lt;/sup&gt; (s.m.)</td>
<td>Docking and enzymatic assay</td>
<td><strong>NS3hel</strong> (RNA binding site competitive inhibitor)</td>
<td>Broad spectrum (DENV, WNV, YFV, TBEV) inhibition of unwinding.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>ST-610&lt;sup&gt;[38]&lt;/sup&gt; (s.m.)</td>
<td>HTS on infected cells</td>
<td><strong>NS3hel</strong> (RNA binding site)</td>
<td>Inhibits cell infection in all DENV serotypes and a mutation on NS3hel RNA binding pocket (A263T) causes resistance. It inhibits unwinding activity.</td>
<td>Efficacious in AG-129 mouse model.</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Drug/Compound</td>
<td>Methodology</td>
<td>Inhibitory Effect</td>
<td>Comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NITD-618[^39] (s.m.)</td>
<td>HTS on DENV2 replicon</td>
<td>NS4B</td>
<td>Selective DENV (1-4) inhibitor. Resistant viruses present A119T and P104L mutations. The latter is important for NS4B-NS3hel interactions. Poor Pk for testing</td>
<td>No follow up studies reported</td>
<td></td>
</tr>
<tr>
<td>Lycorine[^40] (s.m.)</td>
<td>HTS on WNV and DENV1 replicons and SAR study</td>
<td>NS4B</td>
<td>Selective inhibitor of DENV and WNV RNA replication. Resistant virus mutations (V9M) suggest it interacts with NS4B.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>SDM25M[^41] (s.m.)</td>
<td>HTS on DENV2 replicon</td>
<td>NS4B</td>
<td>Inhibitor of DENV. Resistant viruses develop a mutation (F164L) on NS4B. Also the reported P104L mutation on NS4B confers resistance.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>SAH analogue[^42] (s.m.)</td>
<td>Structure-based drug design and SAR</td>
<td>NS5 MTase (SAM binding pocket)</td>
<td>It is a competitive inhibitor, selective for DENV NS5 MTase over human MTases. Cell permeability problems.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Aadamantan ureidopropanoic acid derivative[^43] (s.m.)</td>
<td>HTVS</td>
<td>NS5 MTase (2’O MTase, putative SAM binding pocket)</td>
<td>It inhibits 2’O methylation. Docking predicted competition with SAM.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>ATA[^44] (s.m.)</td>
<td>VS</td>
<td>NS5 MTase (putative RNA binding site)</td>
<td>It inhibits both N7 and 2’O methylation of DENV and WNV.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Ribavirin[^45] (s.m.)</td>
<td>Binding and crystal structure</td>
<td>NS5 MTase (GTP site)</td>
<td>Inhibits NS5 MTase. Not efficacious on humans</td>
<td>ND</td>
<td>Interrupted</td>
</tr>
<tr>
<td>Sinefungin[^45] (s.m.)</td>
<td>Binding</td>
<td>NS5 MTase (SAM site)</td>
<td>Inhibits DENV NS5 MTas but it interacts also with host factors</td>
<td>ND</td>
<td>Interrupted</td>
</tr>
</tbody>
</table>
### APPENDIX 1

| Design of adenosine-based nucleoside analogues | NS5 RdRp (NI) | Broad spectrum inhibitor of several flaviviridae (DENV1-4, WNv, YFV, PV, HCV). Triphosphate inhibits NS5 RdRp through chain termination. | Efficacious in AG-129 mouse model. | Interrupted (2 weeks toxicology problems) |
| NITD008 | NS5 RdRp (NI) | Broad spectrum inhibitor of several flaviviridae (DENV1-4, WNv, YFV, PV, HCV). It inhibits RNA synthesis in DENV replicon. | Efficacious in AG-129 mouse model. Prodrug was necessary for Pk improvement. | Halted (2 weeks toxicology problems) |
| NITD203 | NS5 RdRp (NI) | Active in DENV replicon assay. | Phase II clinical trials did not show neither clinical nor virological improvement. Probably due to decreased phosphorilation. | Interrupted |
| Cytidine analog prodrug developed for HCV | NS5 RdRp (NI) | Inhibits DENV NS5 RdRp and binds in the RNA tunnel. Poor activity in CPE. | ND | No follow up studies reported |

**Abbreviations:** s.m. = small molecule; E = envelope protein; BOG = β-N-octyl-glucoside; ND = not determined; Pk = pharmacokinetics; GI = gastrointestinal; YFV = Yellow Fever Virus; WNv = West Nile Virus; JEV = Japanese Encephalitis Virus; DENV = Dengue Virus; HTS = High Throughput Screening; FP = Fluorescence Polarisation assay; p. = peptide; TBEV = Tick-Borne Encephalitis Virus; CPE = Cytopathic Effect inhibition assay; C = capsid; SAR = Structure Activity Relationships; NS3hel = NS3 helicase/NTPase; NS5 Mtase = NS5 methyltransferase; Mtase = methyltransferase; HTVS = High Throughput Virtual Screening; VS = Virtual Screening; GTP = Guanosine triphosphate; NS5 RdRp = NS5 RNA-dependent RNA polymerase; NI = Nucleoside Inhibitor; PV = Powassan Virus;
### Table A1.3. Potential host target inhibitors.

<table>
<thead>
<tr>
<th>Name</th>
<th>Drug Discovery Method</th>
<th>Target</th>
<th>In vitro activity</th>
<th>In vivo activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzomorphane derivatives\cite{53} (s.m.)</td>
<td>CPE-based HTS and SAR evaluation</td>
<td><strong>Translation</strong> (unknown target)</td>
<td>Inhibits RNA translation through interaction with host factors. Broad flavivirus spectrum (DENV, YFV, WNV)</td>
<td>Efficacious on AG-129 mice</td>
<td>Interrupted (adverse effects at higher doses)</td>
</tr>
<tr>
<td>AP30451\cite{54} (s.m.)</td>
<td>Replicon-based HTS</td>
<td><strong>Early translation</strong> (unknown target)</td>
<td>Inhibits RNA translation through interaction with host factors. Broad flavivirus spectrum (DENV, YFV, WNV, Kunjin virus)</td>
<td>ND</td>
<td>On-going SAR studies</td>
</tr>
<tr>
<td>Isoxazole-pyrazole derivatives\cite{55} (s.m.)</td>
<td>CPE-based HTS</td>
<td>DHODH</td>
<td>Inhibits DENV in infected cells through host DHODH inhibition and consequent U depletion. Broad spectrum: flavivirus (YFV, WNV, DENV), hepacivirus (HCV), alphavirus (WEEV) rhabdovirus (VSV), retroviruses (HIV)</td>
<td>Not efficacious on AG-129 mouse model (due to U uptake from diet)</td>
<td>Interrupted (lack of efficacy in vivo)</td>
</tr>
<tr>
<td>ETAR\cite{56} (s.m.)</td>
<td>Ribavirin analogues screening</td>
<td>IMPDH</td>
<td>Inhibits DENV in infected cells through host IMPDH inhibition and consequent GMP depletion. Broad spectrum: DENV1-4, LGTV, MODV</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>AZD0530 and Dasatinib\cite{57} (s.m.)</td>
<td>Replicon-based HTS</td>
<td><strong>Fyn kinase</strong> (ATP binding site)</td>
<td>Inhibit RNA replication through inhibition of Fyn kinase. Resistant viruses present the T108I mutation on NS4B.</td>
<td>ND</td>
<td>No follow up studies reported</td>
</tr>
<tr>
<td>Celgosivir\cite{57,58,59,60} (s.m.)</td>
<td>Prodrug of a natural product with known ER glucosidase inhibition activity.</td>
<td><strong>α-glucosidase I and II</strong></td>
<td>Effective against DENV1-4. Impedes NS1 correct folding through α-glucosidases (I and II) inhibition at nM concentration.</td>
<td>Efficacious on AG-129 mouse model. Phase Ib clinical trial: safe compound; non-significant reduction of viral load and fever.</td>
<td>On-going (Next clinical trial will be improved with dose frequency or multi-therapy adjustments)</td>
</tr>
<tr>
<td>Compound</td>
<td>Description</td>
<td>Effect</td>
<td>ND</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------</td>
<td>----</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>U18666A&lt;sup&gt;[61]&lt;/sup&gt; (s.m.)</td>
<td>Known compound that inhibits intracellular trafficking of cholesterol.</td>
<td>Blocks DENV entry and intracellular trafficking (in cholesterol-loaded late endosomes) and inhibits de novo sterols synthesis, producing antiviral effect.</td>
<td>ND</td>
<td>No following studies reported</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** s.m. = small molecule; CPE = Cytopathic Effect inhibition assay; HTS = High Throughput Screening; SAR = Structure Activity Relationships; DENV = Dengue Virus; YFV = Yellow Fever Virus; WNV = West Nile Virus; ND = not determined; DHODH = Dihydroorotate Dehydrogenase; HCV = Hepatitis C Virus; WEEV = Western Equine Encephalitis Virus; VSV = Vesicle Stomatitis Virus; HIV = Human Immunodeficiency Virus; IMPDH = Inosine Monophosphate Dehydrogenase; LGTV = Langat Virus; MODV = Modoc Virus.
BIBLIOGRAPHY


[11]: T. Kampmann, R. Yennamalli, P. Campbell, M. J. Stoermer, D. P. Fairlie, B. Kobe, P. R. Young, In silico screening of small molecule libraries using the dengue virus envelope E protein has identified compounds with antiviral activity against multiple flaviviruses, Antiviral Research 84 (2009), p. 234-241, DOI: 10.1016/j.antiviral.2009.09.007


APPENDIX 2: INTRODUCTION TO USED MOLECULAR MODELING PROGRAMS

SCHRÖDINGER MAESTRO GLIDE

Maestro is the unified graphical interface for all the products of Schrödinger, a provider of chemical simulation software for use in pharmaceutical, biotechnology and materials science research. This program gives the possibility to perform different types of computational chemistry analysis and Glide (Grid-based Ligand Docking with Energetics) is the algorithm used for docking calculations.[1]

The docking calculation operated with Glide can be run in rigid or semi-flexible modes. In the first case, ligands’ input conformations are placed as they are in the receptor pocket and only rotation and translation degrees of freedom are used for pose generation; while if the semi-flexible approach is chosen, the generation of poses produces several ligand conformations. The conformational search is conducted by dividing each ligand into a core and rotamer groups and the number of final conformations depended on the number of rotatable bonds, conformationally labile 5- and 6- membered rings, and asymmetric pyramidal trigonal nitrogen centres.[1]

Both in rigid and semi-flexible docking modes, the protein binding site must be input to the program for the evaluation of the interactions. Glide represents the target site with a grid, which points contain informations about the shape and the properties of the cleft. The grid is positioned by using a box centred in the binding site by the user. The box must contain all ligand atoms for correct docking, therefore the size is calibrated during the receptor grid calibration. A smaller box is also created by the program during grid preparation. This inner box has the same centre of the outer one and it describes the space region in which the small molecule’s centre (the midpoint between the two furthest atoms of the core region) is initially positioned during the placement.[1]

In all docking modes, the produced poses pass through a series of hierarchical filters for the evaluation of the interactions with the protein. In this filtering process, the first pose assessment is computed through the Schrödinger’s discretized version of an empirical scoring function in which favourable hydrophobic, hydrogen-bonding, and metal-ligation interactions are rewarded, and steric clashes are penalized. The ligand conformations can then be minimized and at the end of the process, the refined poses are re-scored with a Schrödinger’s registered GlideScore scoring function, reported in Equation A2.1.[1]
$G_{\text{Score}} = 0.065 \cdot \text{vdW} + 0.130 \cdot \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site}$

**Equation A2.1. GlideScore scoring function.**[1]

vdW = van der Waals energy; Coul = Coulomb energy; Lipo = lipophilic term; Hbond = hydrogen-bonding term; Metal = Metal-binding term; BuryP = penalty for buried polar groups; RotB = penalty for freezing rotatable bonds; Site = polar interactions in the active site

With Maestro Glide software it is possible to choose between three precision modes: high throughput virtual screening (HTVS), standard precision (SP) and extra precision (XP). The precision mode determines the number of energy components that are considered in the pose scoring and consequently the scoring function complexity and the calculation time. HTVS allows the fastest docking simulations of the three. It does not produce very accurate results, but it allows the screening of a very large number of molecules rapidly. SP mode is more accurate and it is appropriate for screening ligands of unknown quality in large numbers with a reasonable calculation time. The XP scoring function includes additional terms over SP and a more complete treatment of some of the SP terms for the offset of observed underestimation in the SP version. In detail, explicit water molecules are used as a measure of whether the complex is physically realistic; penalties are assigned to structures where statistical results suggest that one or more groups are inadequately solvated; rewards are given for occupancy of well-defined hydrophobic pockets by hydrophobic ligand groups.[1]

**SCHRÖDINGER BIOOLUMINATE**

BioLuminate is a Schrödinger product that is able to interface with Maestro. It was designed for protein modelling and has a wide range of tools that allow protein analysis (e.g. structure quality assessment), protein engineering (e.g. homology modelling or protein-protein docking) and antibody modelling (e.g. antigen-antibody complex prediction).[2]

The prediction of protein-protein interactions can be achieved with the implemented protein-protein docking procedure that uses the PIPER program.[3] With this program it is possible to dock two arbitrary proteins or one protein with itself for dimer or trimer simulation. The main difference between the two cases is that in the latter one symmetry is taken into account in the docking procedure. Furthermore, a special docking can be performed for antigen-antibody complexes, as the software is able to recognise the features of the binding site of the antibodies, giving more accurate results. In all the kind of simulation, a protein is treated as “receptor” and the other one as the “ligand”. In the case of antigen-antibody complex the antibody is always the receptor, while in the other cases it makes no difference. Protein-protein docking is performed rigidly, sampling the interactions between the protein partners through translation and rotation of the “ligand” in respect of the “receptor”. The resulting
poses are not energy minimised, but scored according to the scoring function in Equation A2.2 that estimates the binding energy.\textsuperscript{[2,3]}

\[
S = (E_{\text{attr}} + aE_{\text{rep}})_{\text{shape}} + bE_{\text{elec}} + cE_{\text{pair}}
\]

Equation A2.2. PIPER scoring function.\textsuperscript{[3]}

\(S\) = score; \((E_{\text{shape}})\) shape contribution; \(E_{\text{attr}}\) = energy related to attractive interactions; \(E_{\text{rep}}\) = energy related to repulsive interactions; \(E_{\text{elec}}\) = electrostatic contribution; \(E_{\text{pair}}\) = pairwise desolvation contribution; \(a, b\) and \(c\) = weight coefficients for terms

Poses are evaluated according to the shape complementarity, to the electrostatic contribution \((E_{\text{elec}})\) and to the desolvation contribution \((E_{\text{pair}})\). The shape complementarity is evaluated by considering both attractive \((E_{\text{attr}})\) and repulsive \((E_{\text{rep}})\) interactions. As the docking used a rigid body model that does not consider the small protein changes in conformations upon interaction, a weighting factor \((a)\) is used to allow some atomic overlaps. Electrostatic contributions are calculated with a simplified Generalised Born equation and the desolation component is given by a function that considers “receptor” and “ligand” atoms interactions in a pairwise manner. Weighting factors \((a, b\) and \(c)\) were calibrated using test cases.\textsuperscript{[3]}

The top 1000 scoring poses are clustered and the centre of the cluster is returned as result by the program.\textsuperscript{[2]} The procedure can be biased by introducing constraints that increase or decrease the attractive or the repulsive factors on specific residues. This can be useful to drive the docking for consistency with known experimental data.\textsuperscript{[2]}

**SCHRÖDINGER MAESTRO PREPARATION TOOLS**

Prior either small molecule or protein-protein docking, it is advisable to prepare the molecules. For docking simulations performed with Schrödinger programs, this can be achieved with two programs within the Mestro interface: Protein Preparation Wizard (for protein preparation) and LigPrep (for small molecule preparation).\textsuperscript{[4,5]}

Protein structures, in particular X-ray data, might contain: unwanted co-crystallised ligands, more than one protein chain, water molecules, cofactors, ions, buffer elements and even errors (for instance missing atoms and partial or missing residues). For this reason, before calculations, a protein should be prepared and this process is automated by the Protein Preparation Wizard. It allows to delete unwanted molecules (e.g. water or unwanted cofactors), fix incomplete protein residues, fill missing loops, find atoms that overlap, create disulphide bonds, fix metal coordination, check bond orders and atom types, fix the orientation of amino acids’ functional groups, protonate and charge the structure and to chose between alternate states if present.\textsuperscript{[4]}
Small molecule preparation consists mainly in the optimisation and evaluation of the chemical structures and can be performed with the LigPrep program. It consists in the addition of hydrogens according to atomic valences, generation of possible ionisation states in a pH range, generation of possible tautomers, exploration chirality (if desired), generation of low-energy ring conformation, and geometry optimisation. For each small molecule that is prepared, several structures are generated with different properties, following the above mentioned steps. These structures are then evaluated according to their chemical properties and the assigned score can be used in Glide docking simulations for a more precise pose score that considers not only the interactions between the protein and the ligand, but also the ligand’s properties in that chemical state. For this reason ligand preparation generally improves modelling results.\(^5\)

**LEADIT FLEXX**

FlexX is a computer program licensed by BioSolveIT that performs docking and scoring procedures. As in most of docking programs, the protein is considered to be rigid, so the protein must be given in a conformation similar to the bound state.\(^6\)

Prior docking or rescoring simulations, protein structure preparation and adjustments (e.g. atom protonation, unwanted chains and cofactors removal, choice between residues’ alternate states) is advisable, but usually performed with other molecular modelling programs. As for Schrödinger-based docking simulations, also in this case the protein binding site must be indicated for poses evaluations. The docking site is a spherical region of space that must contain all the atoms of the docked ligands. It is defined by the user by selecting a centre and a sphere radius.\(^6\) Ligand’s structure preparation (in particular correct hydrogen atoms additions and energy minimisation) is also advisable, but is usually performed with other molecular modelling programs.

FlexX is a fragment-based docking program, in the sense that during placement the ligands are split and a base fragment is selected. This molecule portion is then positioned in the receptor site, finding the placements that give most interactions with the protein. The base fragment is then gradually grown up to the original compound using a torsion angle database. Consequently, the new fragment placement is optimised and the best solutions are chosen. This is an iterative process that ends when the ligand is completely reconstructed and its conformational space has been explored.\(^6,7\)

Poses are scored with an empirical scoring function. This means that the contributions to the binding energy (\(\Delta G_{\text{bind}}\)) were calibrated through the analysis of a set of protein-ligand complexes. The main binding energy contributions are given by solvent effects (\(\Delta G_{\text{solv}}\)), specific protein-ligand interactions (\(\Delta G_{\text{int}}\)), the energy needed for freezing internal rotatable bonds
(△G_{rot}), conformational energy (△G_{conf}), the energy involved in vibrational modes changes (△G_{vib}) and the loss in rotational and translational free energy (△G_{t/r}), as shown in Equation A2.3.\[6\]

\[
\Delta G_{\text{bind}} = \Delta G_{\text{solv}} + \Delta G_{\text{int}} + \Delta G_{\text{rot}} + \Delta G_{\text{conf}} + \Delta G_{\text{vib}} + \Delta G_{\text{t/r}}
\]

*Equation A2.3. FlexX scoring function.*\[6\]

△G_{solv}= solvent effects energy; △G_{int}= specific protein-ligand interactions energy; △G_{rot}= energy involved in freezing internal rotatable bonds; △G_{conf}= energy involved in conformational changes; △G_{vib}= energy involved in vibrational modes changes; △G_{t/r}= loss in transational and rotational free energy

**PLANTS**

PLANTS (Protein-Ligand ANT System) is a docking algorithm.\[8\] It is based on a class of stochastic algorithms called Ant Colony Optimization (ACO), inspired by the indirect communication (in the form of pheromone) among real ants finding the shortest path between their nest and a food source. The artificial ants in PLANTS are randomized procedures that generate a complete candidate solution using a constructive mechanism. In this docking system, both the ligand and the protein are treated as flexible, taking into account the ligand’s translational, rotational and torsional degrees of freedom and the protein’s torsional degrees of freedom. At the end of the process the generated solutions are post-processed and sorted according to increasing scoring function values.\[8,9\]

There are two available scoring functions in PLANTS: the PLP scoring function (fPLP) and CHEMPLP scoring function (fCHEMPLP). fPLP is considers distance-based potentials only and has the form reported in Equation A2.4.\[9\]

\[
f_{\text{PLP}} = f_{\text{plp}} + f_{\text{tors-lig}} + f_{\text{clash-lig}} + 0.3 \cdot f_{\text{score-prot}} - 20.0
\]

*Equation A2.4. PLANTS scoring function.*\[9\]

f_{\text{plp}}= steric intercations between the protein and the ligand; f_{\text{tors-lig}}= torsional potential of the ligand; f_{\text{clash-lig}}= clash term for the ligand; f_{\text{score-prot}}= intramolecular protein interactions

fCHEMPLP introduces angle-dependent terms for hydrogen-bonding and metal binding, and intralingand interactions are taken in account, as shown in Equation A2.5.\[9\]

\[
f_{\text{CHEMPLP}} = f_{\text{plp}} + f_{\text{chem-hb}} + f_{\text{tors-lig}} + f_{\text{clash-lig}} + 0.3 \cdot f_{\text{score-prot}} - 20.0
\]

*Equation A2.5. PLANTS scoring function.*\[9\]

f_{\text{plp}}= steric intercations between the protein and the ligand; f_{\text{chem-hb}}= hydrogen bonds and metal contacts between the protein and the ligand; f_{\text{tors-lig}}= torsional potential of the ligand; f_{\text{clash-lig}}= clash term for the ligand; f_{\text{score-prot}}= intramolecular protein interactions

In both scoring functions a penalty term is added if the ligand’s reference point is outside the predefined binding site of the protein.\[8,9\]
As described in the other docking programs, prior the simulation, protein structure preparation and adjustments (e.g. atom protonation, unwanted chains and cofactors removal, choice between residues’ alternate states) is advisable, but usually performed with other molecular modelling programs. Also in the case of PLANTS, the protein binding site must be indicated for poses evaluations. The docking site is a spherical region of space that must contain all the atoms of the docked ligands. It is defined by the user by selecting a centre and a sphere radius. Ligand’s structure preparation (in particular correct hydrogen atoms additions and energy minimisation) is also advisable, but is usually performed with other molecular modelling programs.

**MOE DOCKING**

Molecular Operating Environment (MOE) is an integrated software package that allows the application of a wide range of molecular modelling techniques to molecular systems. Docking and scoring are some of the possible tools that can be used with MOE. As described in the other docking programs, prior the simulation, protein structure preparation and adjustments (e.g. atom protonation, unwanted chains and cofactors removal, choice between residues’ alternate states) is advisable and can be performed with this suite. For instance, hydrogens can be added and energy minimised in a pH and temperature-dependent manner using the Protonate3D tool, while incomplete residues can be constructed with the Mutate tool. During the simulation, the protein region in which the small molecule must be placed is indicated by the user through box size and centre. Ligand’s structure preparation (in particular correct hydrogen atoms additions and energy minimisation) is also advisable, and can be performed using the Database tools.

The MOE docking and scoring application works through six steps that can be switched on or off according to the purpose of the study. The first stage in the docking algorithm is the 3D conformational search of the ligands. It is a systematic search that employs a set of preferred torsion angles to routable bonds and saves a maximum of 5,000 conformers for each ligand. As this conformational search does not change bond lengths and angles and it does not explore ring conformations. Therefore, it is important that the starting conformations are reasonable.

The second step (placement) a set of poses are generated and scored for each ligand. If the binding mode of reference ligands is known, then it is possible to use this information: a 3D pharmacophore can be built and used for constraining the generation of the pose on the base of the given features.

The third stage of the docking algorithm is the first rescoring that evaluates the generated poses on the base of the ligand-protein contacts. The default London dG scoring function
estimates the binding free energy ($\Delta G$) of the poses on the base of the gain/loss of the ligand rotational and translational entropy ($s$); the energy correlated to the ligand’s loss of flexibility ($E_{\text{flex}}$); the energy associated with hydrogen bonds ($E_{\text{h-bond}}$); the energy of metal interactions ($E_{\text{metal}}$); the desolvation energy for each atom ($E_{\text{des}}$). The functional form of the scoring function is reported in Equation A2.6:

$$
\Delta G = s + E_{\text{flex}} + E_{\text{h-bond}} + E_{\text{metal}} + E_{\text{des}}
$$

Equation A2.6. London $dG$ scoring function. (modified from$^{[10]}$)

$\Delta G$=binding free energy, $s$=gain/loss of ligand rotational and translational entropy, $E_{\text{flex}}$=energy correlated to the ligand’s loss in flexibility, $E_{\text{h-bond}}$=hydrogen bond energy, $E_{\text{metal}}$=metal interactions energy, $E_{\text{des}}$=desolvation energy

Another scoring function is GBVI/WSA $dG$. It is an empirical scoring function that estimates the binding free energy ($\Delta G$) through Equation A2.7:

$$
\Delta G \approx c + \alpha \left[ E_{\text{coul}} + \Delta E_{\text{sol}} \right] + \beta \Delta E_{\text{vdw}} + \beta \Delta S_{\text{weighted}}
$$

Equation A2.7. GBVI/WSA $dG$ scoring function.$^{[10]}

$\Delta G$=binding free energy, $c$=gain/loss of ligand rotational and translational entropy, $\alpha$=AMBER force field-dependent constant, $\beta$=MMFF94x force field-dependent constant, $E_{\text{coul}}$=electrostatic interactions, $\Delta E_{\text{sol}}$=solvation electrostatic term, $\Delta E_{\text{vdw}}$=van der Waals interactions, $\Delta S_{\text{weighted}}$=weighted surface area

The binding free energy estimation is calculated through the calculation of several ligand or binding parameters and force field-dependent constants ($\alpha$ with AMBER and $\beta$ with MMFF94x force fields) that were calibrated during the force field optimisation on 99 protein-ligand complexes. More in detail, the calculated main components to the binding energy are: the gain or loss of the ligand’s rotational and translational entropy ($c$); the electrostatic component ($\Delta E_{\text{coul}}$) calculated through Coulomb’s equations; the solvation electrostatic term ($\Delta E_{\text{sol}}$) computed through the Generalised Born equation, the van der Waals contribution ($\Delta E_{\text{vdw}}$); the penalization for the solvent exposed surface area of the ligand ($\Delta S_{\text{weighted}}$).$^{[10]}

The fourth step of the docking procedure is the molecular mechanics force field or grid refinement that has the aim of optimising the protein-ligand interactions.$^{[10]}

The fifth stage of the docking algorithm is the pharmacophore constraint that induces the poses refinement to comply the formation of the known interactions according the given features. It also works as a filter of the poses, eliminating the poses that do not satisfy the given pharmacophore.$^{[10]}

The sixth stage of the algorithm is the second rescoring of the poses with one of the scoring methods. Possible rescoring functions are the same as in the first and the default one is GBVI/WSA $dG$ scoring function.$^{[10]}
MOE FRAGMENT-BASED DRUG DESIGN TOOLS

Also *de novo* drug design can be carried out with MOE tools, with the modification of existing structures for the achievement of more active molecules.

MedChem Trasformations is one of these tools and it allows alterations of existing ligands or fragments in order to build novel molecules in a rational manner. The general methodology of the program consists of a series of iterations in which the selected regions of the molecule are transformed. For each cycle, the resulting new molecules are then filtered (according to molecular properties, QSAR models, fingerprints or pharmacophores), synthetically scored, energy minimised and evaluated according scoring functions.\textsuperscript{[10]}

The transformations can be of two types: those that can preserve or those that can change properties. From medicinal chemistry studies available in literature, the most common transformations have been collected in MOE databases. They are written as transformation rules that appear like chemical reactions: on the left of an arrow there is the initial chemical substructure, while on the right of it there is the new substructure.\textsuperscript{[10]} An example is reported in Figure A2.1.

![Figure A2.1. Example of transformation.](image)

In this example hydrogen is transformed in a chloride atom. On the left of the arrow there is the original structure and on the right there is the new one.

After two short energy minimisations (one without the receptor and one with the receptor), the new coordinates are checked and structures presenting bond lengths or angles with values outside the normal range are ruled out.\textsuperscript{[10]}

As mentioned above, the newly designed molecules are marked with a synthetic feasibility score that ranges from 0 (not feasible synthesis of the molecule) to 1 (molecule fully synthesizable). This is a recent development of the software, aiming the improvement of the major problem of *de novo* drug design approaches: the generation of non-synthesisable molecules. The score is calculated through three steps. At the beginning, a retrosynthetic bond disconnection approach is adopted until no more bonds can be disconnected. All the resulting fragments are then compared to the molecules present in a software-implemented database that contains possible starting materials. The heavy atoms of these fragments are then “marked”. The fraction of the “marked” heavy atoms over the total number of heavy atoms of the molecule constitutes the synthetic feasibility score.\textsuperscript{[10]}

At the end of the transformation cycle, if the structure of the receptor is present, a refinement is usually applied. This allows the energy minimisation of the new molecules within the
binding pocket, using the chosen force field functions. If the receptor is not present, a tethered minimisation can be applied to the molecule in order to relax the structure.\textsuperscript{10}

Another MOE tool that allows modification of a ligand is the Combinatorial Builder tool. This program generates all the possible products that can be built keeping a common scaffold and applying a set of R-groups described in a database. In the same manner as the MedChem Transformations, the generated molecules are filtered with molecular descriptors and possibly with user-defined QSAR models, fingerprints or pharmacophores. Then, they are synthetically scored and refined.\textsuperscript{10}

For the generation of the new molecules, the user must select the attachment points and the software inserts the R-groups in these locations. The new molecules are first filtered in order to eliminate undesirable chemical groups (e.g. peroxides). Then, all of the possible rotamers of the R-groups are generated through an application that samples torsion angles related to rotatable bonds. If the receptor is present, the conformations that give clashes with it are automatically removed. In the end, the duplicate and the high-energy structures are removed.\textsuperscript{10}

As well as the “growing” approach applied through the tools described above, it is also possible to link close fragments to form a ligand that potentially binds the protein better with the Link Multiple Fragment tool. The program connects two or more independent fragments via inserting linkers between user-defined sites in the starting fragments.\textsuperscript{10}

The construction of the new molecules is operated exploiting built-in databases of several conformations of possible linkers. The user-defined hydrogen atoms are substituted by the heavy atoms of the linkers and the new joined structures are subsequently protonated before their evaluation in terms of bond order, correct geometries and clash energies.\textsuperscript{10}

In a similar manner to the other MOE de novo drug design tools, the new structures are filtered through molecular properties and user-defined QSAR models, fingerprints and pharmacophores; as well as being synthetically scored. In the presence of the receptor, it is also possible to refine and score the structures through a force field refinement with pharmacophore restraints, as for the other tools.\textsuperscript{10}

**LIGBUILDER**

LigBuilder is a structure-based drug design software that computes two de novo approaches: molecule growing and molecule linking.\textsuperscript{11} The main program’s goal is the construction of novel chemical ligands within the constraints of the target molecule and consequently the structure of the receptor is necessary for the application of this methodology. In the main software procedure, before the construction of novel chemical entities, the binding pocket is analysed and a pharmacophore is built. The new ligand molecules are then evolved through a
genetic algorithm and scored according to their chemical viability and binding affinity. After this, the molecules are also evaluated for their drug-likeness, chemical stability and potential toxicity.\textsuperscript{[11]}

The receptor analysis is operated by the LigBuilder POCKET tool. It requires the presence of the protein’s structure alone and the structure of a docked reference ligand for binding pocket localisation. The program recognises the ligand and builds a grid box that covers the small molecule and the protein in that region. Through a hydrogen probe, it then calculates the accessibility of the pocket. Key interaction sites are identified with different probes in the vacant places, allowing to build a pharmacophore with the following possible features: hydrogen bond donor, hydrogen bond acceptor and hydrophobic regions.\textsuperscript{[11]}

The outputs from POCKET are then used for the “building-up method”. Molecules are constructed in a fragment-based manner, meaning that an initial starting structure is necessary for the construction of novel chemical entities. The construction of these molecules is obtained through three main operations: growing, linking and mutation. In the growing operation, the growing sites-selected hydrogen atoms are substituted with new fragments that are contained in a building block library, maintaining the hydrogen-heavy atom bond orientation. As the new constructed bond is rotatable, a rotamer exploration is computed in order to evaluate the conformation at the energy minimum within the protein’s binding pocket. If more than one energy minimum are recognised, the different conformations at the minima are kept and considered as different molecules. The linking operation adopts three types of algorithms according to the distance of the atoms to link. If two structures collide through a pair of hydrogen atoms, a CH\textsubscript{2} group is used to substitute the two hydrogens and link the molecules. If the collision happens between two heavy atoms, the program will create a new bond between them, but if the two heavy atoms overlap, one of them is deleted and the other one is used to connect the two structures. The mutate operation allows C, N, O atoms with the same hybridization to mutate to each other. All the heavy atoms are evaluated in respect of the pharmacophore that was built with POCKET and if it does not fit the given feature it will be mutated to a more appropriate heavy atom. At the end of these operations all the new hydrogen atoms are labelled as growing sites, unless they are too close to the receptor.\textsuperscript{[11]}

The evolution of the new molecules is computed through a genetic algorithm approach that is a search that mimics the natural process of evolution. In LigBuilder, the initial ligand structure (or seed structure) is used to generate a first population of new molecules obtained through the growing, linking, mutate operations. Each molecule of this population is then scored with a fitness value. With a fitness value-weighted random selection of the molecules a mating pool of parent molecules that are structurally different is built. Importantly, during the selection, the best ranked molecules are directly copied from the old population in order to
ensure the enhancement of the trend in the fitness value. The mating pool is then used for new growing, linking and mutate operations, for the construction of the next generation of population of molecules. The whole process continues in an iterative manner. Each iteration is called generation and the number of generations is decided by the user prior the beginning of the methodology. When the number of generations has been reached and the last population has been formed, the molecules are ranked according to the fitness value.\textsuperscript{[11]}

The fitness value calculation is operated through the consideration of two factors: the binding affinity and the bioavailability. The SCORE algorithm employs a linear empirical scoring function that estimates binding free energy, computing the binding affinity.\textsuperscript{[11]} It was built on the bases of the study of 170 protein-ligand complexes and it can be described with Equation A2.8.

\[
\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{hydrophobic}} + \Delta G_{\text{rotor}} + \Delta G_0
\]

Equation A2.8. LigBuilder SCORE scoring function.\textsuperscript{[11]} 
\(\Delta G_{\text{bind}}\) = binding free energy, \(\Delta G_{\text{vdw}}\) = van der Waals interaction, \(\Delta G_{\text{H-bond}}\) = hydrogen bonding energy, \(\Delta G_{\text{hydrophobic}}\) = hydrophobic interaction, \(\Delta G_{\text{rotor}}\) = entropy loss for rotatable bonds freezing, \(\Delta G_0\) = empirical constant

The binding free energy is calculated through the sum of various energy components associated to molecule binding: van der Waals interactions (\(\Delta G_{\text{vdw}}\)), hydrogen bond interactions (\(\Delta G_{\text{H-bond}}\)), hydrophobic interactions (\(\Delta G_{\text{hydrophobic}}\)) and the loss of entropy linked to the rotatable bonds freezing (\(\Delta G_{\text{rotor}}\)). The bio-availability component of the fitness value consists in the application of Lipinski’s rules for the design of drug-like molecules to be delivered orally. The score is based on the violation of these rules: violation of these rules produces a proportional decrease in the bio-availability score.\textsuperscript{[11]}

On top of the fitness value, molecules are also evaluated through chemical rules and toxicity parameters. The chemical rules are applied in order to rule out molecules that might not be chemically stable. This is achieved through the identification of unacceptable (e.g. heteroatoms bound to each other or too many heteroatoms bound to the same carbon) or forbidden structures. Similarly, the toxicity of the molecules is predicted through the individuation of particular chemical structures (e.g. epoxide groups) that have been associated with drug toxicity.\textsuperscript{[11]}

**OPENEYE ROCS**

The interactions between a protein and a ligand occur at the level of their surfaces and therefore the shape complementarity between the two molecules, as well as their surface properties, is fundamental for the strength of the interaction between them.\textsuperscript{[12]} The shape of a molecule can be defined in different ways, but in a very broad sense it can be indicated as the
part of space that is delimited by the object’s surface and occupied by the object itself.[13] On the base of the principle that “if two molecules have a similar shape, perhaps they have similar properties”, shape similarity searches have been widely used in molecular modelling approaches, in particular for ligand-based drug design approaches.[13,14]

OpenEye ROCS is a software that allows to compare molecules and determine their similarity with a shape-based superimposition method.[15] In order to have the same shape, two molecules must have the same volume and this is why in this method the volume calculation is fundamental. Nevertheless, the volume is a scalar field as it can be expressed with a number and therefore the molecule’s conformation is not specified with this descriptor. Hence, two molecules that have the same volume do not necessarily have the same shape. For this reason, even if volume and shape are closely related, the only volume calculation is not sufficient for an appropriate shape screening.[15] On the base of this, ROCS’s fundamental equation for shape comparison can be written as follows:

\[ S_{f,g} = I_f + I_g - 2O_{f,g} \]

*Equation A2.9. ROCS fundamental equation for shape comparison.*[15]

\( S \) = shape similarity, \( I \) = self-volume overlaps, \( O \) = overlap between the 2 functions, \( f,g \) = molecules

The shape similarity \((S_{f,g})\) between two molecules \(f\) and \(g\) is calculated through the molecules’ volumes (self-volume overlaps, \(I_f\) and \(I_g\)) that are independent of orientation and the orientation-dependent overlap between the two functions \((O_{f,g})\). If the two molecules have the same shape, \(I_f\) and \(I_g\) will be equal and correspondent to double of the value of overlap \(O_{f,g}\) and therefore \(S_{f,g}\) will be equal to 0. Otherwise, if the overlap is not complete, \(S_{f,g}\) will be a positive number.[15]

For a better shape comparison, ROCS applies an orientation search on the molecules in order to find the best overlap that outputs the lower \(S_{f,g}\) value and then it evaluates the shape similarity according to the chosen scoring method. The most common scoring method used in shape comparison is the Tanimoto index that is calculated through Equation A2.10.

\[ T_{f,g} = \frac{O_{f,g}}{I_f + I_g - O_{f,g}} \]

*Equation A2.10. Tanimoto index equation.*[15]

\( T_{f,g} \) = Tanimoto index, \( I \) = self-volume overlaps, \( O \) = overlap between the 2 functions, \( f,g \) = molecules

If two molecules have the same shape, the overlap of the volumes \((O_{f,g})\) will be equal to the volume of one molecule \((I)\) and the two volumes \((I_f\) and \(I_g\)) will be equal. This means that the Tanimoto index \((T_{f,g})\) will be equal to 1. Otherwise the score will be inferior to 1.[15]
Not only molecular volume and orientation are important in shape comparison, but also the molecule’s conformation is fundamental. ROCS does not compute conformational searches, therefore conformations need to be prepared prior the simulation.\textsuperscript{[15]}

As previously stated, shape is not the only 3D feature that is essential for protein-ligand recognition and binding. Physical-chemical properties and their distribution of the molecule’s surface have also a key role.\textsuperscript{[12,14]} One way of representing the space distribution of chemical features is the pharmacophore.\textsuperscript{[16,14]} ROCS allows, optionally, to consider these features as well as the molecules’ shapes for the similarity screening through the use of “color features”. In this way, the position of hydrogen bond donors, hydrogen bond acceptors, anion, cation, ring or hydrophobic groups can be used for a more accurate molecule superimposition and/or scoring. ROCS assesses chemical similarity between the query and the screened molecules with one of the two color force fields at disposition (Implicit Mills Dean and Explicit Mills Dean). Both of them have the same color atom types described above, the same list of chemical functional groups and a weight term that describes the strength of the interaction. The main difference between the two force fields is the way that ionisable groups are treated. Explicit Mills Dean force field does not have a pKa model and consequently uses the atoms’ protonation states given in input. On the other hand, Implicit Mills Dean force field has a simple pKa model that allows it to define the ionisation state (considering pH 7) of the atoms, irrespective of the input structures. The choice between the two force fields is consequently based on the preparation of the molecules.\textsuperscript{[15]}

\textbf{GROMACS}

GROMACS (GROningen Machine for Chemical Simulations) is a freeware software, originally developed in the University of Groningen (Netherlands), that gives the possibility to perform energy minimization, molecular dynamics (MD) and peculiar MD procedures.\textsuperscript{[17,18]} The MD simulation itself is only part of the entire procedure, which is composed by the following main steps: generation of a topology file, generation and solvation of the box, neutralization of the system, energy minimisation, equilibration and proper MD simulation.

\textit{Topology file generation}

GROMACS software needs a topology input file to run the simulation. A topology file contains a list of the constant attributes of each atom of the system. Furthermore, it contains also the instructions on how these parameters should be applied in the energy functions. This file is necessary for all the steps of the simulation and every time GROMACS uses the topology file it updates it with the changes made to the system (e.g. addition of water molecules). As the
information format varies from one force field to another, it is at this point that the user-chosen force field and the preferred water model have to be specified. A water model is defined as the geometry, together with other parameters such as the atomic charges and Lennard-Jones parameters, of the water molecules that will constitute the environment of the system.\cite{17}

**Generation and solvation of the box**

In a MD study, the system is confined in a physical space called “box” that can have shapes with different complexity (triclinic, cubic, octahedron or dodecahedron). Confining the system in a finite volume could create some artefacts at the boundaries of this space. In order to minimize these edge effects, GROMACS uses *periodic boundary conditions* (PBC) combined with the *minimum image convention* (for which only the nearest image of each particle is considered for short range non-bonded interaction terms).\cite{17} With PBC, the space-filling box that contains the system is surrounded by translated copies of itself, as shown in Figure A2.2.

![Figure A2.2. Illustration of the periodic boundary condition concept.](image)

One box and three images box are represented. The molecular system in the box is represented by a protein (red ribbon) and small molecules (colours space-filling model representation). The movement of two small molecules out of the box is simulated with arrows that show how the periodic boundary conditions allow having always the same number of molecules in the simulation box.

With periodic boundary conditions, if a particle leaves the box during the simulation, another image particle will replace it, entering from the other side of the box. In this way, the total number of particles inside the box remains constant.\cite{17,19} Some times, PBC are not desired in a molecular system (e.g. in the study of molecular adsorption onto a surface), but with molecular systems in solution it is usually advisable to use PBC.\cite{19} If PBC conditions are used, it is important to make sure, though, that the molecular system cannot interact with itself in the neighbour box. For this reason, non-bonded interactions cut-offs are usually set to be at least double the minimum distance between the molecular system and the box.\cite{17}
Once generated, the box is then filled with randomly positioned water molecules generated according to the chosen water model (exemplified in Figure A2.3).

![Figure A2.3. Image of the system box filled with water molecules. The box edges are in blue and the water molecules are represented by red dots. The molecular system (a protein) is represented in lines.](image)

**Neutralization of the system**

Due to the use of PBC, if the system has a net charge, even if this is very small, this will be multiplied by the infinite number of boxes created, generating an infinitive net charge. In order to avoid this issue, the system is neutralised by replacing random solvent molecules with monoatomic ions in random positions.\[17\]

**Energy minimisation (EM)**

In a MD simulation the initial state of the simulation is important because it can determine the success or the failure of the simulation itself. For example a system with high-energy interactions can give simulation instability. For this reason, it is important that the system structure is prepared (e.g. by adding hydrogen atoms, completing incomplete residues, or deleting unwanted factors) prior to the simulation, as explained also in the docking programs chapters. Additionally, it is wise to operate an energy minimization (EM) prior the system equilibration and the MD production steps, in order to improve the mentioned simulation instability issues.\[19\]

In GROMACS, EM can be operated with several integrators, according to the system size, calculation time and desired precision. The most used integrators are: steepest descendent and conjugated gradient. Steepest descendent integrator calculates the force on each atom...
and for the following time steps it moves the atoms according the direction of the force gradient. At the beginning of the minimisation it quickly reaches a conformation near the energy minima, but then it is very slow in reaching the minima itself. On the other side, conjugate gradient method, through the use of a vector orthogonal to the force gradient, is slower in the early stages of energy minimisation, but more efficient near the energy minimum compared to steepest descendent algorithm.[17]

**System equilibration**

The system is in a water environment, but water molecules are added in random positions. Therefore, they do not “solvate” correctly with the molecular system. Furthermore, the system at this stage is not described by important physical parameters, as temperature and pressure. Therefore, before the proper MD simulation is undertaken it is useful to run a position restrained molecular dynamics (PR-MD). With PR-MD the atom positions of the macromolecule are restrained (kept still, or “frozen”), while the solvent molecules are left free to move and to “soak” into the macromolecule. Given that the relaxation time (the time needed for a perturbed system to return to equilibrium) of water is more or less 10ps, the PR-MD has to be longer than this time, at least an order of magnitude bigger. Moreover, during the equilibration period, the system is coupled to a preferred temperature and/or pressure and/or volume of the box. Most commonly, the system is treated in NVT (constant number of atoms N, volume V and temperature T) or in NPT (constant number of atoms N, pressure P and temperature T) conditions. The difference between the two is that in the first case the volume of the box is kept constant, while in the second case the pressure is maintained constant.[17,19]

The control of the temperature can be achieved with several methods. For instance, velocity rescaling temperature coupling corrects temperature deviations acting on the velocity of random atoms, but ensuring that the kinetic energy is equally distributed between “solute” and solvent. To force a constant temperature is necessary because the truncation on the force calculation or integration errors that can occur during a MD simulation can produce small drifts in this descriptor that are not related to the physical behaviour of the system. Similarly, also the pressure of the system can be controlled through several methods. For instance, the Berendsen pressure coupling corrects pressure drifts by acting on the box volume.[17]

**Molecular dynamics simulation**

This is the proper production step. Longer MD simulations should resemble better molecular mechanisms, but this implies also longer computational time. Consequently, it is common to
select the total simulation time on the base of a compromise between accuracy and computational time.

Trajectory analysis

MD trajectories can be analysed with GROMACS with several tools. Most of the time, the main informations that are achieved from a molecular dynamics simulation are the monitoring of the system total, potential, kinetic energies, the system’s temperature and pressure in order to verify that these elements are constant during the simulation, as well as the visual inspection of the trajectory.\textsuperscript{[17]}

The equilibrium status of the system can be investigated through the calculation of the root mean square deviation (RMSD) of the atoms for each time step. The RMSD is a statistical tool that measures the average distance between the atoms in a molecule with respect to a reference structure. The calculation of RMSD values is performed applying Equation A2.11.

\[
RMSD(t_1, t_2) = \sqrt{\frac{\sum_{i=1}^{N} m_i (r_i(t_1) - r_i(t_2))^2}{M}}
\]

\textit{Equation A2.11. RMSD calculation.}\textsuperscript{[17]}

\(m_i\) = mass of atom \(i\), \(r_i\)=position of atom \(i\) at time \(t\), \(M\)=mass of the \(N\) atoms, \(t_1\)=time, \(t_2\)=reference time (usually 0)

On the base of RMSD values, structures clustering can be performed and the ensemble of conformations that the molecular system adopts during the simulation can be assessed.

It is also possible to inspect the time variation of the distance of atoms and to monitor the formation/brake of hydrogen bonds. In the latter case, the program recognises the presence of groups that can interact with this non-covalent bond and if distance (less than 0.35nm) and angle (angle with the vertex on the hydrogen between 0° and 30°) features are respected then the hydrogen bond is recognised.

AMBER

Assisted Model Building with Energy Refinement (AMBER) is a suite of programs that allows to perform MD simulations and to make calculations on the obtained trajectories.\textsuperscript{[20,21]} MD simulations are performed by a tool named sander and similarly to GROMACS, precise preparation steps before the production simulation itself must be performed because a proper system preparation is crucial for the outcome of the production simulation itself. These steps are the preparation of the molecular system with topology determination, simulation box preparation and (generally) solvation, system neutralisation, energy minimisation and equilibration. Furthermore, as already explained, the choice of force field
and water model are also important and depend on the molecular system that has been studied. Like GROMACS, energy minimisations can be computed with different methods (e.g. steepest descendent or conjugate gradient) and MD simulations follow the Leap Frog algorithm described in the main INTRODUCTION section of this thesis.

As well as for common analysis, with AMBER MD trajectories can be used for the application of the Molecular Mechanics Poisson-Boltzmann (and Generalized Born) Surface Area (MMPB(GB)SA) approach. It is computed by MMPBSA program within this suite and allows the calculation of binding free energy ($\Delta G_{\text{bind}}$) of molecular complexes.\cite{1}

The calculation is performed on snapshots taken from MD trajectories because in the average of the obtained values give a more accurate estimation of the interaction energy. MMPBSA estimates the $\Delta G_{\text{bind}}$ values according Equation A2.12.

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}}$$

*Equation A2.12. General equation for the estimation of the binding free energy.*\cite{2,3}

$\Delta G_{\text{bind}}$=binding free energy, $G_{\text{complex}}$=free energy of the complex, $G_{\text{receptor}}$=free energy of the receptor, $G_{\text{ligand}}$=free energy of the ligand

This calculation requires the free energies of the complex, of the ligand and of the receptors that are averaged upon the set intervals of time in the MD simulation. The three free energy components are computed on the base of the molecular mechanics energy ($E_{\text{MM}}$), the solvation energy ($E_{\text{solv}}$) and the solute entropy ($S$), as reported in Equation A2.13.\cite{4,2,3}

$$G = E_{\text{MM}} + E_{\text{solv}} - TS$$

*Equation A2.13. Free energy calculation.*\cite{4,2,3}

$G$=free energy, $E_{\text{MM}}$= molecular mechanics energy, $E_{\text{solv}}$=solvation energy, $T$=temperature, $S$=solute entropy

The molecular mechanics energy ($E_{\text{MM}}$) is force field based and it follows the general force field function reported in Equations 1.1 and 1.2 (see main INTRODUCTION section).\cite{4,2,3} This is the molecular mechanics component of the MMPB(GB)SA approach.

The solvation energy is estimated through its decomposition in polar and nonpolar contributions. The first one applies the implicit solvent approach, in which the solvent is not considered as a group of molecules, but as a continuous medium. In this case, the solvent interactions are measured exploiting its dielectric constant. The polar contribution is computed through two different equations: Poisson-Boltzman (PB) or Generalized Born (GB).\cite{4,2,3} The PB equation is a second-order differential equation that correlates the distribution of the charge throughout the system (charge density) and the dielectric constant to the electrostatic potential that is used for the calculation of the polar solvent contribution. With this method the solute is considered to be a body with constant low dielectric value, while the solvent has a high dielectric. In the GB model, the solute is considered as a set of
spheres with charges and different dielectric constant to the continuum solvent and it recalls Coulomb’s equations. This is the PB or GB part of the MMPB(GB)SA approach. The nonpolar solvation contribution is computed through the solvent accessible surface area (SASA) method that probes the solute’s surface with a water molecule according to the following equation:

\[ E_{sol, np} = \gamma \cdot SASA + o \]


Here, the non-polar solvation energy is correlated to SASA using a surface tension coefficient (\( \gamma \)) and an offset constant (\( o \)).[^19,21,24]
BIBLIOGRAPHY

[1]: Glide 5.7 User Manual Copyright © 2011 Schrödinger, LLC.

[2]: BioLuminate User Manual Copyright © 2013 Schrödinger, LLC.


[4]: Protein Preparation Guide Copyright © 2009 Schrödinger, LLC.

[5]: LigPrep User Manual Copyright © 2013 Schrödinger, LLC.


[10]: Molecular Operating Environment (MOE), 2011.10; online user tutorials, http://www.chemcomp.com/


APPENDIX 3: SHAPE-BASED VIRTUAL SCREENING SELECTED MOLECULES

<table>
<thead>
<tr>
<th>Compound</th>
<th>Molecule structure</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Molecule 1" /></td>
<td>4-((2,4-dichlorobenzyl)piperazin-1-yl)-1-phenyl-1H-pyrazolo[3,4-d]pyrimidine</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Molecule 2" /></td>
<td>4-((1-(3,4-dichlorobenzyl)-1H-indol-3-yl)methylene)-1-phenylpyrazolidine-3,5-dione</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Molecule 3" /></td>
<td>2-bromo-4-((3,5-dioxo-1-phenylpyrazolidin-4-yldene)methyl)-6-methoxyphenyl adamantane-1-carboxylate</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Molecule 4" /></td>
<td>naphthalene-2,7-diyl bis(azepane-1-carboxylate)</td>
</tr>
<tr>
<td>123</td>
<td><img src="image123.png" alt="Molecule 123" /></td>
<td>1-(5-methoxy-2-methylphenyl)-3-(4-(3-methoxyphenyl)piperazin-1-yl)pyrrolidine-2,5-dione</td>
</tr>
<tr>
<td>124</td>
<td><img src="image124.png" alt="Molecule 124" /></td>
<td>7-(2-cyclohexyl-2-oxoethyl)-3-(naphthalen-2-yl oxy)-4H-chromen-4-one</td>
</tr>
<tr>
<td>125</td>
<td><img src="image125.png" alt="Molecule 125" /></td>
<td>3-benzyl-7-((4-methyl-5-phenyl-4H-1,2,4-triazol-3-yl)thio)-3H-[1,2,3]triazolo[4,5-d]pyrimidine</td>
</tr>
<tr>
<td>126</td>
<td><img src="image126.png" alt="Molecule 126" /></td>
<td>1-benzyl-4-((4-cyclohexylyphenyl)sulfonyl)piperazine</td>
</tr>
<tr>
<td>127</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>4-(3-(2-(2,4-dichlorophenyl)-8,9-dimethyl-7H-pyrrolo[3,2-e][1,2,4]triazolo[1,5-c]pyrimidin-7-yl)propyl)morpholine</td>
</tr>
<tr>
<td>128</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>2-(benzo[d][1,3]dioxol-5-ylmethylene)-3-oxo-2,3-dihydrobenzofuran-6-yl 2-fluorobenzoate</td>
</tr>
<tr>
<td>129</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>2-(4,9-dimethyl-7-oxo-3-phenyl-7H-furo[2,3-f]chromen-8-yl)-N-(1,1-dioxidotetrahydrothiophen-3-yl)-N-methylacetamide</td>
</tr>
</tbody>
</table>
# APPENDIX 4: NS3HEL NTP SITE VIRTUAL SCREENING SELECTED MOLECULES

<table>
<thead>
<tr>
<th>Compound</th>
<th>Molecule structure</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td><img src="image1" alt="Molecule Structure" /></td>
<td>2-((4-((2-(p-tolylcarbamoyl)hydrazono)methyl)p-henoxoxy)acetic acid</td>
</tr>
<tr>
<td>17</td>
<td><img src="image2" alt="Molecule Structure" /></td>
<td>4-((4-chloro-2-((4,6-dioxo-2-thioxotetrahydropyrimidin-5(2H)-ylidene)methyl)phenoxy)methyl)benzoic acid</td>
</tr>
<tr>
<td>130</td>
<td><img src="image3" alt="Molecule Structure" /></td>
<td>2-((1,8-dioxo-1,2,3,4,5,6,7,8,9,10-decahydroacridin-9-yl)2-methoxyphenoxy)acetic acid</td>
</tr>
<tr>
<td>131</td>
<td><img src="image4" alt="Molecule Structure" /></td>
<td>2-((2-((3-methoxy-5-(1H-tetrazol-1-yl)phenyl)amino)-2-oxoethyl)thio)acetic acid</td>
</tr>
<tr>
<td>132</td>
<td><img src="image5" alt="Molecule Structure" /></td>
<td>3-(benzo[d][1,3]dioxol-5-ylcarbamoyl)cyclopentanecarboxylic acid</td>
</tr>
<tr>
<td>133</td>
<td><img src="image6" alt="Molecule Structure" /></td>
<td>3-((5-(4-methoxybenzylidene)-2,4-dioxothiazolidin-3-yl)methyl)benzoic acid</td>
</tr>
<tr>
<td>134</td>
<td><img src="image7" alt="Molecule Structure" /></td>
<td>4-((4-oxo-5-(4-propoxyphenyl)thieno[2,3-d]pyrimidin-3(4H)-yl)methyl)benzoic acid</td>
</tr>
<tr>
<td>135</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
<td>5-{5-((2-(allylcarbamothioyl)hydrazono)methyl)furan-2-yl}isophthalic acid</td>
</tr>
<tr>
<td>136</td>
<td><img src="image2.png" alt="Chemical Structure" /></td>
<td>3-{5-((2,4-dioxo-1,5-dioxaspiro[5.5]undecan-3-ylidene)methyl)furan-2-yl}benzoic acid</td>
</tr>
<tr>
<td>137</td>
<td><img src="image3.png" alt="Chemical Structure" /></td>
<td>3-{(1-carbamothioyl-3-methyl-5-oxopyrazolidin-4-yl)diazeny]benzoic acid</td>
</tr>
</tbody>
</table>
APPENDIX 5: NS3HEL POTENTIAL INHIBITORS TESTED DURING THE SECONDMENT PERIOD

The following table reports the compound number, its structure, the virtual screening approach used for its identification and the maximum concentration at which it could be solubilised in water, starting from a 20mM (10mM for compounds 1 and 15) solution in 100% DMSO.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>Design approach</th>
<th>Soluble at</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure" /> 4-((2,4-dichlorobenzyl)piperazin-1-yl)-1-phenyl-1H-pyrazolo[3,4-d]pyrimidine</td>
<td></td>
<td>100μM</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure" /> 4-(((3,4-dichlorobenzyl)-1H-indol-3-yl)methylene)-1-phenylpyrazolidine-3,5-dione</td>
<td></td>
<td>100μM</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure" /> 2-bromo-4-((3,5-dioxo-1-phenylpyrazolidin-4-ylidene)methyl)-6-methoxyphenyl adamantane-1-carboxylate</td>
<td>Ligand-based virtual screening (chapter 3.2.2)</td>
<td>500μM</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Structure" /> naphthalene-2,7-diyl bis(azepane-1-carboxylate)</td>
<td></td>
<td>100μM</td>
</tr>
<tr>
<td>123</td>
<td><img src="image5.png" alt="Structure" /> 1-(5-methoxy-2-methylphenyl)-3-{{3-(4-(3-methoxyphenyl)piperazin-1-yl)pyrrolidine-2,5-dione</td>
<td></td>
<td>500μM</td>
</tr>
<tr>
<td>11</td>
<td><img src="image.png" alt="Structure" /></td>
<td>2mM</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>ethyl 4-((S)-2,5-dichlorophenyl)furan-2-yl)-6-methyl-2-oxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate</td>
<td>500μM</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1-(4-oxo-6-((p-tolylthio)methyl)-1,4-dihydropyrimidin-2-yl)-3-phenylguanidine</td>
<td>250μM</td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>2-methoxyethyl 4-((3-((4-chlorophenyl)-1-phenyl-1H-pyrazol-4-yl)-6-methyl-2-oxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate</td>
<td>250μM</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>3-benzy-4-((2-(piperazin-1-yl)ethyl)amino)-2H-chromen-2-one</td>
<td>2mM</td>
<td></td>
</tr>
</tbody>
</table>

Structure-based virtual screening on RNA binding site (chapter 3.2.3)
<table>
<thead>
<tr>
<th>No.</th>
<th>Chemical Structure</th>
<th>Formula</th>
<th>NTP Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td><img src="image1" alt="Structure" /></td>
<td>2-{(2-tolylcarbamoyl)hydrazono}methyl)</td>
<td>phenoxy</td>
</tr>
<tr>
<td>17</td>
<td><img src="image2" alt="Structure" /></td>
<td>4-{(4-chloro-2-{(4,6-dioxo-2-thioxotetrahydropyrimidin-5(2H)-ylidene}methyl)</td>
<td>phenoxy</td>
</tr>
<tr>
<td>130</td>
<td><img src="image3" alt="Structure" /></td>
<td>2-{(4,1,8-dioxo-1,2,3,4,5,6,7,8,9,10-decahydroacridin-9-yl}-2-methoxyphenoxy</td>
<td>acetic acid</td>
</tr>
<tr>
<td>131</td>
<td><img src="image4" alt="Structure" /></td>
<td>2-{(2-(3-methoxy-5-(1H-tetrazol-1-yl)</td>
<td>phenyl</td>
</tr>
<tr>
<td>132</td>
<td><img src="image5" alt="Structure" /></td>
<td>3-(benzo[d][1,3]dioxol-5-ylcarbamoyl)cyclopentane</td>
<td>carboxylic acid</td>
</tr>
</tbody>
</table>
# APPENDIX 6: FRAGMENTS SOLVED IN COMPLEX WITH NS5 METHYLTRANSFERASE

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Structure</th>
<th>$IC_{50}$ 2’O-methylation</th>
<th>$IC_{50}$ N-7 methylation</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A2</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>&gt;10 mM</td>
<td>7 mM</td>
<td>Pyridine-4-carboxylic acid</td>
</tr>
<tr>
<td>2A4</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>&gt;10 mM</td>
<td>&gt;10 mM</td>
<td>7-aminothieno[2,3-b]pyrazine</td>
</tr>
<tr>
<td>2C3</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>4.4 mM</td>
<td>6.5 mM</td>
<td>2-amino-4-methylbenzoic acid</td>
</tr>
<tr>
<td>2E11</td>
<td><img src="image4.png" alt="Structure" /></td>
<td>4.2 mM</td>
<td>&gt;10 mM</td>
<td>4-chloro-5-methylbenzene-1,2-diamine</td>
</tr>
<tr>
<td>2G3</td>
<td><img src="image5.png" alt="Structure" /></td>
<td>0.3 mM</td>
<td>2 mM</td>
<td>2-(2,5-dimethylpirrol-1-yl)benzoic acid</td>
</tr>
<tr>
<td>3A9</td>
<td><img src="image6.png" alt="Structure" /></td>
<td>4 mM</td>
<td>4 mM</td>
<td>2,3-dihydro-1-benzofuran-5-carboxylic acid</td>
</tr>
<tr>
<td>3C2</td>
<td><img src="image7.png" alt="Structure" /></td>
<td>&gt;10 mM</td>
<td>&gt;10 mM</td>
<td>4-(trifluoromethyl)benzene-1,2-diamine</td>
</tr>
</tbody>
</table>
## APPENDIX 7: COMPOUNDS DESIGNED FROM COMPOUND 44 MODIFICATIONS

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>Binding pose</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td><img src="image1.png" alt="Structure Image" /></td>
<td><img src="image2.png" alt="Binding Pose Image" /></td>
</tr>
<tr>
<td>54</td>
<td><img src="image3.png" alt="Structure Image" /></td>
<td><img src="image4.png" alt="Binding Pose Image" /></td>
</tr>
<tr>
<td>140</td>
<td><img src="image5.png" alt="Structure Image" /></td>
<td><img src="image6.png" alt="Binding Pose Image" /></td>
</tr>
<tr>
<td>141</td>
<td><img src="image7.png" alt="Structure Image" /></td>
<td><img src="image8.png" alt="Binding Pose Image" /></td>
</tr>
<tr>
<td>142</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
<td><img src="image2.png" alt="3D Structure" /></td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>143</td>
<td><img src="image3.png" alt="Chemical Structure" /></td>
<td><img src="image4.png" alt="3D Structure" /></td>
</tr>
</tbody>
</table>
## APPENDIX 8: SELECTED MOLECULES FROM RDRP SITE D VIRTUAL SCREENING STUDY

<table>
<thead>
<tr>
<th>Compound</th>
<th>Molecule structure</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td><img src="image1.png" alt="Molecule Structure" /></td>
<td>N-(3,5-dimethylphenyl)-3-ethoxy-4-fluorobenzenesulfonamide</td>
</tr>
<tr>
<td>80</td>
<td><img src="image2.png" alt="Molecule Structure" /></td>
<td>2-chloro-N-(3-(3-(2-phenylacetyl)thioureido)phenyl)benzamide</td>
</tr>
<tr>
<td>81</td>
<td><img src="image3.png" alt="Molecule Structure" /></td>
<td>4-bromo-2-((2-(4-morpholino-6-((4-(trifluoromethyl)phenyl)amino)-1,3,5-triazin-2-yl)hydrazono)methyl)phenol</td>
</tr>
<tr>
<td>82</td>
<td><img src="image4.png" alt="Molecule Structure" /></td>
<td>N’2,N’7-di(furan-2-carbonyl)-9H-fluorene-2,7-disulfonohydrazide</td>
</tr>
<tr>
<td>121</td>
<td><img src="image5.png" alt="Molecule Structure" /></td>
<td>quinolin-8-yl 1-methyl-2-oxo-1,2-dihydrobenzo[cd]indole-6-sulfonate</td>
</tr>
<tr>
<td>144</td>
<td><img src="image6.png" alt="Molecule Structure" /></td>
<td>N1,N5-bis(2-carbamoylphenyl)glutaramide</td>
</tr>
<tr>
<td>145</td>
<td><img src="image7.png" alt="Molecule Structure" /></td>
<td>2-((5-(3-fluorophenyl)-4-phenyl-4H-1,2,4-triazol-3-yl)thio)-N-(1-phenylethyl)acetamide</td>
</tr>
<tr>
<td>No.</td>
<td>Chemical Structure</td>
<td>Chemical Formula</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>146</td>
<td><img src="image1.png" alt="Image" /></td>
<td>2-((4-fluorophenyl)[5-hydroxy-3-phenyl-1H-pyrazol-4-yl]methyl)malononitrile</td>
</tr>
<tr>
<td>147</td>
<td><img src="image2.png" alt="Image" /></td>
<td>4-(2-chlorophenyl)-2-cyclopropyl-5-(1-(p-tolyl)-1H-tetrazol-5-yl)pyrimidine</td>
</tr>
<tr>
<td>148</td>
<td><img src="image3.png" alt="Image" /></td>
<td>6-(4-nitrobenzyl)benzo[b]pyrido[3,2-f][1,4]oxazepin-5(6H)-one</td>
</tr>
<tr>
<td>149</td>
<td><img src="image4.png" alt="Image" /></td>
<td>2-(4-chloro-2-methylphenoxy)-N-(4-(pyrrolidine-1-carbonyl)phenyl)acetamide</td>
</tr>
<tr>
<td>150</td>
<td><img src="image5.png" alt="Image" /></td>
<td>2-(4-methoxyphenoxy)-N-(4-((4-methylpiperazin-1-yl)sulfonyl)phenyl)acetamide</td>
</tr>
<tr>
<td>151</td>
<td><img src="image6.png" alt="Image" /></td>
<td>3-(furan-2-yl)-2-phenyl-5-(p-tolyl)dihydro-2H-pyrrolo[3,4-d]isoxazole-4,6(5H,6aH)-dione</td>
</tr>
<tr>
<td>152</td>
<td><img src="image7.png" alt="Image" /></td>
<td>5-((4-bromo-1H-pyrazol-1-yl)methyl)-N-(2-chloro-5-(trifluoromethyl)phenyl)furan-2-carboxamide</td>
</tr>
<tr>
<td>153</td>
<td><img src="image8.png" alt="Image" /></td>
<td>N-(3-acetylphenyl)-4-((tetrahydrofuran-2-yl)methoxy)benzamide</td>
</tr>
<tr>
<td>154</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
<td>4-(((3-(2-tert-butoxyamino)-2-oxoethoxy)benzyl)amino)ethyl)benzoic acid</td>
</tr>
<tr>
<td>155</td>
<td><img src="image2.png" alt="Chemical Structure" /></td>
<td>1-([1,1'-biphenyl]-4-yloxy)-3-(3,5-dimethyl-1H-1,2,4-triazol-1-yl)propan-2-ol</td>
</tr>
<tr>
<td>156</td>
<td><img src="image3.png" alt="Chemical Structure" /></td>
<td>5-((4-(2,4-dichlorophenoxy)phenyl)amino)-5-oxopentanoic acid</td>
</tr>
<tr>
<td>157</td>
<td><img src="image4.png" alt="Chemical Structure" /></td>
<td>2-((3-hydroxyphenyl)amino)-2-oxoethyl 4-((2-methoxy-2-oxoethyl)thio)-1H-tetrazol-1-yl)benzoate</td>
</tr>
<tr>
<td>158</td>
<td><img src="image5.png" alt="Chemical Structure" /></td>
<td>2-((4-chlorophenoxy)methyl)-5-((4-fluorobenzyl)amino)oxazole-4-carbonitrile</td>
</tr>
<tr>
<td>159</td>
<td><img src="image6.png" alt="Chemical Structure" /></td>
<td>3-(4-chloro-3-methylphenoxy)-5-((2-(4-nitro-1H-pyrazol-1-yl)acetamido)benzoic acid</td>
</tr>
<tr>
<td>160</td>
<td><img src="image7.png" alt="Chemical Structure" /></td>
<td>N-[4-([1H-imidazol-1-yl)methyl]phenyl)-3-(benzo[d][1,3)dioxol-5-yl)propanamide</td>
</tr>
<tr>
<td>161</td>
<td><img src="image8.png" alt="Chemical Structure" /></td>
<td>5-amino-3-methyl-1-(6-&lt;p-toloyloxy)pyrimidin-4-yl)-1H-pyrazole-4-carbonitrile</td>
</tr>
<tr>
<td>162</td>
<td><img src="image9.png" alt="Chemical Structure" /></td>
<td>2-(3,4-dimethoxyphenyl)-N-(4-ethylphenyl)acetamide</td>
</tr>
<tr>
<td>163</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td>N-(2-carbamoylphenyl)-5-((3-chlorophenoxy)methyl)furan-2-carboxamide</td>
</tr>
<tr>
<td>164</td>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>2-(4-methoxyphenyl)-N-(4-((3-methyl-1,2,4-oxadiazol-5-yl)methoxy)benzyl)ethanamine</td>
</tr>
<tr>
<td>165</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td>5-((1-methyl-1H-indol-3-yl)methylene)-3-(3-methylbenzyl)imidazolidine-2,4-dione</td>
</tr>
<tr>
<td>166</td>
<td><img src="image4" alt="Chemical Structure" /></td>
<td>3-(4-chlorophenyl)-1-((4-chlorophenyl)sulfonyl)-1H-1,2,4-triazol-5-amine</td>
</tr>
<tr>
<td>167</td>
<td><img src="image5" alt="Chemical Structure" /></td>
<td>3-benzyl-2-(pyridin-2-yl)-2,3-dihydroquinazolin-4(1H)-one</td>
</tr>
</tbody>
</table>
APPENDIX 9: NS5 RDRP POTENTIAL INHIBITORS TESTED DURING THE SECONDMENT PERIOD

The following table reports the compound number, its structure, the virtual screening approach used for its identification and the maximum concentration at which it could be solubilised in water, starting from a 20mM (10mM for compound 121) solution in 100% DMSO.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>Design approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td><img src="image" alt="Structure 77" /></td>
<td>N-(3,5-dimethylphenyl)-4-methoxybenzenesulfonamide</td>
</tr>
<tr>
<td>78</td>
<td><img src="image" alt="Structure 78" /></td>
<td>3-(N-(3,5-dimethylphenyl)sulfamoyl)benzoic acid</td>
</tr>
<tr>
<td>80</td>
<td><img src="image" alt="Structure 80" /></td>
<td>2-chloro-N-[(3-[[2-phenylacetyl]thiourea]phenyl)benzamide</td>
</tr>
<tr>
<td>81</td>
<td><img src="image" alt="Structure 81" /></td>
<td>4-bromo-2-((2-(4-morpholino-6-((4-(trifluoromethyl)phenyl)amino)-1,3,5-triazin-2-yl)hydrazono)methyl)phenol</td>
</tr>
</tbody>
</table>

Allosteric site D (chapter 5.2.2)
82. N’2,N’7-di(furan-2-carbonyl)-9H-fluorene-2,7-disulfonohydrazide

121. quinolin-8-yl 1-methyl-2-oxo-1,2-dihydrobenzo[cd]indole-6-sulfonate

144. N1,N5-bis(2-carbamoylphenyl)glutaramide

145. 2-((5-(3-fluorophenyl)-4-phenyl-4H-1,2,4-triazol-3-yl)thio)-N-(1-phenylethyl)acetamide

146. 2-((4-fluorophenyl)[5-hydroxy-3-phenyl-1H-pyrazol-4-yl]methyl)malononitrile

147. 4-(2-chlorophenyl)-2-cyclopropyl-5-{1-(p-tolyl)-1H-tetrazol-5-yl}pyrimidine
<table>
<thead>
<tr>
<th>148</th>
<th><img src="image1.png" alt="Chemical Structure" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6-(4-nitrobenzyl)benzo[b]pyrido[3,2-f][1,4]oxazepin-5(6H)-one</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>149</th>
<th><img src="image2.png" alt="Chemical Structure" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-(4-chloro-2-methylphenoxy)-N-(4-(pyrrolidine-1-carbonyl)phenyl)acetamide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>118</th>
<th><img src="image3.png" alt="Chemical Structure" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-((4-(1,1′-biphenyl)-2-xyloxyphenyl)amino)-4-oxobutanoic acid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>119</th>
<th><img src="image4.png" alt="Chemical Structure" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-(((1-phenyl-1H-tetrazol-5-yl)thio)methyl)-N-(thiazol-2-yl)benzamide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>120</th>
<th><img src="image5.png" alt="Chemical Structure" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-(((2-methoxyethyl)carbamoyl)phenyl 2-(naphthalen-1-yl)acetate</td>
</tr>
</tbody>
</table>

Allosteric site C (chapter 5.2.3)
APPENDIX 10: PUBLICATIONS


Characterization of the Mode of Action of a Potent Dengue Virus Capsid Inhibitor

Pietro Scaturro, Iuni Margaret Laura Trist, David Paul, Anil Kumar, Eliana G. Acosta, Chelsea M. Byrd, Robert Jordan, Andrea Brancale and Ralf Bartenschlager


Updated information and services can be found at: http://jvi.asm.org/content/88/19/11540

**SUPPLEMENTAL MATERIAL**

Supplemental material

**REFERENCES**

This article cites 66 articles, 31 of which can be accessed free at: http://jvi.asm.org/content/88/19/11540#ref-list-1

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml

To subscribe to to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Characterization of the Mode of Action of a Potent Dengue Virus Capsid Inhibitor

Pietro Scaturro, a Iuni Margaret Laura Trist, b David Paul, a Anil Kumar, c* Eliana G. Acosta, a Chelsea M. Byrd, c* Robert Jordan, c* Andrea Brancale, d Ralf Bartenschlager a,d

ABSTRACT

Dengue viruses (DV) represent a significant global health burden, with up to 400 million infections every year and around 500,000 infected individuals developing life-threatening disease. In spite of attempts to develop vaccine candidates and antiviral drugs, there is a lack of approved therapeutics for the treatment of DV infection. We have previously reported the identification of ST-148, a small-molecule inhibitor exhibiting broad and potent antiviral activity against DV in vitro and in vivo (C. M. Byrd et al., Antimicrob. Agents Chemother. 57:15–25, 2013, doi:10.1128/AAC.01429-12). In the present study, we investigated the mode of action of this promising compound by using a combination of biochemical, virological, and imaging-based techniques. We confirmed that ST-148 targets the capsid protein and obtained evidence of bimodal antiviral activity affecting both assembly/release and entry of infectious DV particles. Importantly, by using a robust bioluminescence resonance energy transfer-based assay, we observed an ST-148-dependent increase of capsid self-interaction. These results were corroborated by molecular modeling studies that also revealed a plausible model for compound binding to capsid protein and inhibition by a distinct resistance mutation. These results suggest that ST-148-enhanced capsid protein self-interaction perturbs assembly and disassembly of DV nucleocapsids, probably by inducing structural rigidity. Thus, as previously reported for other enveloped viruses, stabilization of capsid protein structure is an attractive therapeutic concept that also is applicable to flaviviruses.

IMPORTANCE

Dengue viruses are arthropod-borne viruses representing a significant global health burden. They infect up to 400 million people and are endemic to subtropical and tropical areas of the world. Currently, there are neither vaccines nor approved therapeutics for the prophylaxis or treatment of DV infections, respectively. This study reports the characterization of the mode of action of ST-148, a small-molecule capsid inhibitor with potent antiviral activity against all DV serotypes. Our results demonstrate that ST-148 stabilizes capsid protein self-interaction, thereby likely perturbing assembly and disassembly of viral nucleocapsids by inducing structural rigidity. This, in turn, might interfere with the release of viral RNA from incoming nucleocapsids (uncoating) as well as assembly of progeny virus particles. As previously reported for other enveloped viruses, we propose the capsid as a novel tractable target for flavivirus inhibitors.

Dengue virus (DV) belongs to the genus Flavivirus, a large group of emerging pathogens capable of causing severe human diseases. Among these, DV represents the most prevalent mosquito-borne human-pathogenic virus worldwide, comprising 4 serotypes that are transmitted mainly by infected *Aedes* mosquitoes during a blood meal. DV infections can lead to a wide range of clinical manifestations, ranging from asymptomatic infections to life-threatening dengue hemorrhagic fever and shock syndrome. A recent study estimated around 390 million DV infections each year, resulting in approximately 100 million symptomatic cases and around 25,000 deaths (1). Despite intense efforts and growing public interest, no licensed antiviral drug against DV infection is available, and the most advanced DV vaccine candidate did not meet expectations in a recent large clinical trial (2).

DV has a single-stranded RNA genome of positive polarity that codes for a polyprotein, which is co- and posttranslationally processed into three structural proteins (capsid, prM, and envelope) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (3). The virus enters mammalian cells via receptor-mediated endocytosis. In the endosomal compartment, the low pH induces a conformational change in the envelope (E) protein, triggering membrane fusion and nucleocapsid release into the cytoplasm (4, 5). Disassembly of the nucleocapsid occurs by a poorly understood mechanism leading to the release of viral RNA into the cytoplasm of infected cells. Upon synthesis of viral proteins, massive intracellular membrane remodeling events occur, which is a conserved feature among plus-strand RNA viruses (6, 7). These rearrangements include membrane invaginations into the endoplasmic reticulum (ER), which are the assumed sites of...
flavivirus genome replication, and can be observed in both mammalian and arthropod cells (8, 9). Nucleocapsid formation is thought to occur in close proximity to replication sites (9). The envelope is acquired by budding through the ER membrane into which the envelope proteins E and prM have been inserted. Assembled virions, stored within ER stacks in highly ordered arrays, are released from the cell via the conventional secretory pathway, where cleavage of the prM protein by furin, a protease residing in the trans-Golgi network (TGN), occurs. This cleavage is required to render DV particles infectious.

The mature capsid protein (C) is a highly basic protein of 12 kDa that forms homodimers in solution (10, 11). The three-dimensional structure of DV C protein was solved by nuclear magnetic resonance (NMR) (12), indicating that the monomer consists of four alpha helices (α1 to α4). The longest helix, α4, extends away from the core of the protein. It has a high density of basic amino acid residues on the solvent-accessible surface that were proposed to interact with the viral RNA. On the opposite side of the molecule, the surface, formed primarily by α2–α2′ and α1–α1′, is largely uncharged and proposed to interact with membranes (12). Although the C protein is the least conserved among flavivirus proteins, both its charge distribution and structural properties are well conserved.

We have recently reported the identification of a small-molecule inhibitor, ST-148, inhibiting replication of all DV serotypes with high potency (13). Intraperitoneal and intravenous administrations of ST-148 in the AG129 mouse model revealed good bioavailability of the compound, which significantly reduced plasma viremia and viral load in spleen and liver of DV-infected mice. Selection of ST-148-resistant viruses led to the identification of a single S34L amino acid change in C protein which was shown to confer resistance to ST-148. Additionally, in vitro binding studies of ST-148 to purified C protein suggested that the compound bound equally well to wild-type (WT) and S34L-containing C proteins. Although these studies identified C protein as the primary target of ST-148, its mode of action remained unknown. In the present study, we addressed this aspect by using a combination of biochemical, virological, and imaging-based methods. We obtained evidence that ST-148 enhanced C protein self-interaction, providing an explanation for the observed impairment of DV assembly/release as well as entry competence of virus particles.

MATERIALS AND METHODS

Antibodies and sera. Mouse monoclonal antibodies recognizing human gliceraldehyde-3-phosphate dehydrogenase (GAPDH) (sc-47724/0411) and human lamin A/C (sc-7292/636) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Mouse monoclonal antibody against human ATP5B (3D5; no. ab14730) was purchased from Abcam, and mouse monoclonal antibody against human vimentin (VI-10) was obtained from GeneTex Inc. (Irvine, CA). Mouse anti-Envelope monoclonal antibody (3H5-1) was purchased from the ATCC. Mouse anti-capsid monoclonal antibody derived from hybridoma cells (6F3.1) was a kind gift of John G. Aaskov (Queensland University of Technology, Australia), and rabbit polyclonal serum against capsid was a kind gift of Andrea Gamarnik (Fundación Instituto Leloir, Argentina). J2 mouse monoclonal anti-double-stranded RNA (dsRNA) antibody was purchased from English Antibodies and sera.

MATERIALS AND METHODS

Antibodies and sera. Mouse monoclonal antibodies recognizing human gliceraldehyde-3-phosphate dehydrogenase (GAPDH) (sc-47724/0411) and human lamin A/C (sc-7292/636) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Mouse monoclonal antibody against human ATP5B (3D5; no. ab14730) was purchased from Abcam, and mouse monoclonal antibody against human vimentin (VI-10) was obtained from GeneTex Inc. (Irvine, CA). Mouse anti-Envelope monoclonal antibody (3H5-1) was purchased from the ATCC. Mouse anti-capsid monoclonal antibody derived from hybridoma cells (6F3.1) was a kind gift of John G. Aaskov (Queensland University of Technology, Australia), and rabbit polyclonal serum against capsid was a kind gift of Andrea Gamarnik (Fundación Instituto Leloir, Argentina). J2 mouse monoclonal anti-double-stranded RNA (dsRNA) antibody was purchased from English Antibodies and sera.

MATERIALS AND METHODS

Antibodies and sera. Mouse monoclonal antibodies recognizing human gliceraldehyde-3-phosphate dehydrogenase (GAPDH) (sc-47724/0411) and human lamin A/C (sc-7292/636) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Mouse monoclonal antibody against human ATP5B (3D5; no. ab14730) was purchased from Abcam, and mouse monoclonal antibody against human vimentin (VI-10) was obtained from GeneTex Inc. (Irvine, CA). Mouse anti-Envelope monoclonal antibody (3H5-1) was purchased from the ATCC. Mouse anti-capsid monoclonal antibody derived from hybridoma cells (6F3.1) was a kind gift of John G. Aaskov (Queensland University of Technology, Australia), and rabbit polyclonal serum against capsid was a kind gift of Andrea Gamarnik (Fundación Instituto Leloir, Argentina). J2 mouse monoclonal anti-double-stranded RNA (dsRNA) antibody was purchased from English Antibodies and sera.

MATERIALS AND METHODS

Antibodies and sera. Mouse monoclonal antibodies recognizing human gliceraldehyde-3-phosphate dehydrogenase (GAPDH) (sc-47724/0411) and human lamin A/C (sc-7292/636) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Mouse monoclonal antibody against human ATP5B (3D5; no. ab14730) was purchased from Abcam, and mouse monoclonal antibody against human vimentin (VI-10) was obtained from GeneTex Inc. (Irvine, CA). Mouse anti-Envelope monoclonal antibody (3H5-1) was purchased from the ATCC. Mouse anti-capsid monoclonal antibody derived from hybridoma cells (6F3.1) was a kind gift of John G. Aaskov (Queensland University of Technology, Australia), and rabbit polyclonal serum against capsid was a kind gift of Andrea Gamarnik (Fundación Instituto Leloir, Argentina). J2 mouse monoclonal anti-double-stranded RNA (dsRNA) antibody was purchased from English Antibodies and sera.
anti-capsid antibody diluted in PBS containing 3% goat serum. After extensive washes with PBS, secondary staining was carried out by a 45-min incubation with an Alexa 568-conjugated secondary antibody diluted 1:1,000 in PBS containing 3% goat serum. Nuclear DNA was stained with 4′,6-diamidino-2-phenylindole (DAPI) (Molecular Probes, Karlsruhe, Germany). For staining of LDs, Bodipy-488 was added (1:500 dilution) to each coverslip for 10 min at room temperature. Samples were mounted on glass slides with Fluoromount G (Southern Biotechnology Associates, Birmingham, AL), and images were acquired using a Leica CTR MIC fluorescence microscope or a Nikon Eclipse Ti spinning-disc confocal laser microscope. Immunofluorescence signals were quantified using ImageJ, and images were assembled with the Adobe Photoshop software package. For quantification of LD size and number, cells were stained as described above and imaged using a 63× objective, and the size and number of Bodipy-488-stained LDs was calculated by using ImageJ. For each condition, at least 10,000 LDs in more than 500 cells and 8 to 10 fields of view were examined.

**Infectivity assay.** For the determination of virus titers, HuH7 target cells were seeded into 96-well plates (104 cells/well) the day before infection. Cells were inoculated with serial dilutions of virus-containing supernatants that had been filtered through a 0.45-μm-pore-size filter. Infected cells were detected by immune staining of the E protein using the mouse anti-E antibody (3H5-1; diluted 1:500) and secondary horseradish peroxidase-conjugated antibody (1:200). Virus titers (the 50% tissue culture infective dose [TCID50]/ml) were calculated as previously reported (16).

**Determination of antiviral activity of ST-148.** HuH7 cells were seeded into 12-well plates at a density of 1 × 105 cells per well and incubated overnight. Cells were then infected with DV, DV534L, DVR2A, or DVR2A534L at an MOI of 0.1 in the presence of a range of compounds. Seven compound concentrations (25, 8, 2.5, 0.8, 0.25, and 0.008 μM) were used to determine the 50% effective concentration (EC50). After 1 h, inocula were removed and fresh medium containing various concentrations of ST-148 was added to each well. After a 48-h incubation period, supernatants were harvested and filtered through a 0.45-μm-pore-size filter, and 10-fold dilutions were used to determine virus titer by plaque assay as described previously (13). EC50 were calculated by fitting the data to a four-parameter logistic model (variable-slope, nonlinear regression model) to generate a dose-response curve by using the GraphPad Prism 5.0 software package (GraphPad Software, San Diego, CA). From this curve, the concentration of compound that reduced virus-induced cytotoxic effect (CPE) by 50% was calculated.

**In vitro transcription and RNA transfection.** In vitro transcripts were generated as previously described (17). For RNA transfection, HuH7 or 293T single-cell suspensions were prepared by trypsinization, washed with PBS, and resuspended at a concentration of 1 × 106 cells (HuH7) or 1.5 × 105 cells (293T) per ml in Cytomix, supplemented with 2 mM ATP and 5 mM glutathione. Ten μg of subgenomic or genomic in vitro transcript then was mixed with 400 μl of the cell suspension and transfected by electroporation using a Gene Pulser system (Bio-Rad) and a cuvette with a gap width of 0.4 cm (Bio-Rad) at 975 μF and 270 V. Cells were immediately diluted into 20 ml of DMEM complete and seeded in the appropriate format (1 ml/well in 24-well plates, 2 ml/well in 12-well plates, and 15 ml/dish in 15-cm-diameter dishes).

**Lentivirus production.** The pWPI-BLR capsid constructs, encoding either the capsid WT or containing the S34L mutation, and the pWPI-Puro-prM-E constructs were cotransfected into 293T cells with plasmids expressing Gag–Pol of HIV and the G protein of vesicular stomatitis virus (ratio, 3:3:1) as previously reported (16). After 48 and 72 h, supernatants were harvested, clarified through 0.45-μm-pore-size filters, pooled, and stored in aliquots at −20°C until use.

**Molecular modeling.** For protein–protein docking, the NMR structures of DV capsid dimers (PDB entry 1R6R) (12) were used. The 20 dimer models were compared using the MOE Protein Consensus tool (version 2010.10; Chemical Computing Group Inc., Montreal, QC, Canada), and the 5 most different structures were selected for rigid protein–protein docking. Each dimer model was used both with the WT sequence and with the S34L mutation, inserted with the MOE Mutate tool. All structures were prepared using the Schrödinger Maestro protein preparation wizard (version 9.5; Schrödinger, LLC, New York, NY). For each model, protein–protein docking was performed with Schrödinger BioLuminate Piper (version 1.2; Schrödinger, LLC, New York, NY) using the same dimer structure for both receptor and ligand entries. The standard mode was applied, probing 70,000 ligand rotations and giving a 0.21 bonus to the complexes that involved Ser34 in the interactions. The best tetramer model was selected by visual inspection.

All of the molecular dynamics (MD) simulations were computed with GROMACS (version 4.5.3) (18) and the AMBER99 force field. All simulations were performed on both WT and S34L mutant structures; the mutation was inserted starting from the WT structure. Periodic boundary conditions were applied, using a cubic box with 0.9- nm minimum distance between the molecular system and the box. Explicit water molecules described with the TIP3P model were added, and systems were neutralized with the addition of 72 Cl− ions. Missing ST-2148 topology parameters were computed with Antechamber, which is part of the Amber software package (version AMBER12; University of California, San Francisco, CA). Atom partial charges were obtained with the AM1-BCC charge method, while atom types and force field missing parameters were assigned with the GAFF force field. The coordinate and topology files then were converted into GROMACS-compatible format with the acyppe converter (Department of Biochemistry, University of Cambridge, United Kingdom).

Prior to the production simulation, two energy minimizations were performed: first with the steepest descendant and second with conjugate gradient method. Both energy minimizations were implemented for 3,000 steps (with a step size of 1 fs), with a force tolerance of 100 kJ mol−1 nm−1 for the steepest descent and of 10 kJ mol−1 nm−1 for the conjugated gradient. Successively, the protein atoms were constrained with a position restraint of 1,000 kJ mol−1 nm−2, and water molecules were relaxed with two 50-ps (50,000 steps, step size of 1 fs) position-restraint MDs. Coordinates, velocities, and energy values were saved every 500 steps in both stages. The first position-restraint MD was conducted under NVT conditions (constant number of atoms, N, volume, V, and temperature, T), allowing the system to be heated to 300 K with v-rescale temperature coupling and 0.1-ps time constant. The second stage then was performed under NPT conditions (constant number of atoms, N, pressure, P, and temperature, T). Both temperature (with v-rescale, temperature of 300 K and time constant of 0.1 ps) and pressure coupling (with Berendsen algorithm, 1 bar pressure and time constant of 0.5 ps) were applied in this case. A 30-ns (15,000,000 steps with 2-fs step size) production simulation was run for each molecular system under the NPT conditions described above. Coordinates, velocities, and energy values were saved every 1,000 steps. In all procedures, a 0.9-ns short-time cutoff was applied for long-range electrostatic interactions that were calculated with the particle mesh Ewald (PME) method, while short-range nonbonded interactions were calculated within a cutoff of 1.4 nm.

For ST-148 docking, each small-molecule docking simulation was performed using Schrödinger Maestro (version 9.5; Schrödinger, LLC, New York, NY). Prior to docking, the protein and ST-148 were prepared with the protein preparation wizard tool and LigPrep, respectively, generating all possible tautomers and ionization states at pH 7.0 ± 2.0. For each docking, the grid was generated with Glide, defining appropriate center and box sizes. Docking was performed with Glide standard precision (SP) mode, sampling nitrogen inversions and ring conformations and penalizing nonplanar conformations for amides. The 25 poses generated then were refined with the SP mode.

**Plasmid constructs.** Plasmids carrying the subgenomic replicon (pFK-sgDV5R2A) and the full-length genomes, pFK-DVs and pFK-DVs.R2A (carrying a Renilla luciferase reporter gene), are based on the DV-2 16681 strain and have been described recently (17) (Fig. 1A). The S34L amino acid substitution conferring resistance to ST-148 was intro-
FIG 1 Schematic representation of constructs used in this study and antiviral potency of ST-148. (A) Schematic representation of genomic and subgenomic constructs. The DV2 full-length genome is shown at the top (DV). The 5’ and 3’ N terminal regions are depicted with their putative secondary structures. Polyprotein cleavage products are separated by vertical lines and labeled as specified in the introduction. DVR2A is derived from the DV2 full-length genome by insertion of a Renilla luciferase coding sequence preceding the capsid cyclization sequence (CAE) and followed by a Tosa asigna virus 2A cleavage site. The middle panel depicts the constructs used for production of trans-complemented DV particles (DV(C)). sgDVR2A is a subgenomic reporter replicon that is packaged into virus-like particles by the capsid and prM/envelope proteins transiently expressed in trans via two different lentiviruses (pWPI capsid-BLR and pWPI prM-Puro, respectively). All constructs are derived from the DV2 16681 isolate (17). The right panel depicts the constructs used for BRET experiments. YFP and Renilla luciferase proteins were N-terminally fused to either capsid or NS4A of DV2 (strain 16681). (B) Structure of ST-148. (C) Antiviral potency of ST-148. Huh7 cells were infected at a DV or DVR2A reporter virus MOI of 0.1, with the virus containing or not containing the resistance mutation S34L, for 48 h. Virus amounts released into the supernatant were determined by PFU assay with VeroE6 cells. Data represent averages and error ranges from two independent experiments.
using the Green Dye RT-PCR master mix (PKJ GmbH, Hilden, Germany) according to the manufacturer’s instructions, with the following primers binding to a region within the DV NS5 coding sequence (primer polarities are indicated): SDV-2-9687 (5’-GCCCTCTGGTTCACACATT-3’, forward) and asDV-2-9855 (5’-CCACATTGCGGCTAAAGCT-3’, reverse). The total volume of the reaction mixture was 15 μL, and reactions were performed in two stages: stage 1, 15 min at 95°C; stage 2, 40 cycles of 15 s at 95°C and 60 s at 60°C. Quantities of DV RNA were calculated by using serial dilutions of known amounts of DV in vitro transcribed that were processed in parallel as previously described (19).

Sequence alignments. Sequence alignments were performed with the following flaviviruses (UniProtKB/Swiss-Prot accession numbers are indicated): DV-1 (Brazil/97-11/1997), P27909; DV-2 (Thailand/1681-PDK53), P29991; DV-3 (Martinique/1243/1999), Q6YMS5; DV-4 (Thailand/0348/1991), Q2YHF0; West Nile virus, P06935; yellow fever virus (Ivory Coast/1999), Q63P; Japanese encephalitis virus (SA-14), P27395; Kunjin virus (MRM61C), P14335; and St. Louis encephalitis virus (MS-7), P09732.

Statistical analyses. Statistical analyses were performed by applying the two-tailed, unequal Student’s t test. In some cases, indicated in Results, data were analyzed with one-way analysis of variance (ANOVA) and Dunnett’s post hoc test for multiple comparisons to nontreated controls or ANOVA and Tukey’s post hoc test to compare each group to other, as indicated in the figure legends. Microsoft Excel and GraphPad Prism (v. 5.0) were used for statistical analyses.

Subcellular fractionation. For subcellular fractionation of cytosolic, membrane, nuclear, and postnuclear fractions, the Qproteome cell compartment kit, purchased from Qiagen (Hilden, Germany), was used as recommended by the manufacturer. In brief, 10^5 Huh7 cells were seeded into 10-cm-diameter dishes the day before infection. Cells were infected with either DV WT or DV^S34L at an MOI of 1 for 4 h at 37°C and washed three times with PBS, and then complete medium containing 10 μM ST-148 or DMSO was added to the cells. After 48 h, cells were washed 3 times with ice-cold PBS, trypsinized, and collected by 5 min of low-speed centrifugation (700 × g). Collected fractions were acetone precipitated overnight at −20°C and resuspended in equal volumes of 2× SDS loading buffer. One-tenth of the total cell suspension was saved as the input, and equal amounts of each fraction were subjected to SDS–12% PAGE as described above. Fraction purity was evaluated by determining amounts of cytosolic (GAPDH), membrane (ATPSB), nuclear (lamin A/C), and postnuclear/cytoskeletal (vimentin) markers.

Production of trans-complemented DV particles. To produce trans-complemented DV particle (DV^TCP) stocks, 293T cells were electroporated with 10 μg of subgenomic DVR2A (sgDVR2A) replicon RNA, seeded into 15-cm-diameter dishes, and incubated at 37°C. The next day, cells were inoculated with lentiviral particles encoding capsid and prM-E for 8 h. Cells were washed carefully 3 times with PBS, fresh medium was added, and cells were cultured at 33°C for 4 to 6 days. Supernatants were pooled and filtered through a 0.45-μm-pore-size filter, and 15 μM HEPES (pH 7.2 to 7.5) was added to the medium. DV^TCP stocks were stored in aliquots at −80°C until use.

Transient replication assay and DV^TCP-based virus entry assay. Huh7 cells transfected with full-length DVR2A RNA were seeded into 12-well plates. Replication was determined by measuring luciferase activity in cell lysates 4, 24, 48, and 72 h after transfection. For harvest, cells were washed once with PBS and lysed by adding 200 μl lysis buffer as previously described (19). Cells were frozen immediately at −70°C, and after thawing, lysates were resuspended by gentle pipetting. For each well, 20 μl lystate, mixed with 400 μl assay buffer (25 mM glycyglycine, 15 mM MgSO4, 4 mM EGTA, 1 mM DTT, 2 mM ATP, 15 mM K2PO4 [pH 7.8], 1.42 μM coelenterazine H), was measured for 10 s in a luminometer (Lumat LB9507; Berthold, Freiburg, Germany). Each well was measured in duplicate. To determine the amount of infectious virus particles released into culture supernatants 72 h after electroporation, naïve Huh7 cells were inoculated with these supernatants, and 48 h later luciferase activity was determined. Kinetics of virus replication were calculated by normalizing the relative light units (RLU) measured at a given time point to the respective 4-h value. Four-hour values, reflecting transfection efficiency, were comparable among samples and biological replicates, and no statistically significant differences were observed. For DV^TCP experiments, Huh7 cells were inoculated with particles composed of WT capsid protein or containing the S34L resistance mutation for 90 min at 37°C. At different times before, during, or after inoculation, 10 μM ST-148 or 0.1% DMSO was added to the medium. Replication was measured 48 h after inoculation.

Virucidal assay. Preparations of DV and DV^S34L particles (corresponding to 3.8 × 10^7 TCID50/ml) were incubated for 1 h at 37°C in DMEM containing 10 μM ST-148 or 0.1% DMSO. Viruses were serially diluted, and the remaining infectivity was measured by TCID50 assay.

Analysis of infected cells by transmission electron microscopy. Huh7 cells (6 × 10^5) were seeded onto glass coverslips, and around 16 h later, cells were infected with either DV or DV^S34L at an MOI of 3. After 4 h, cells were washed three times with prewarmed PBS, and DMEM containing 10 μM ST-148 or 0.05% DMSO was added. After a 48-h incubation period, cells were fixed and prepared for transmission electron microscopy as described previously (20).

RESULTS
ST-148 is a potent inhibitor of DV. ST-148 (Fig. 1B) was identified by a high-throughput screen as a potent inhibitor of DV, exhibiting antiviral activity against all four serotypes and reducing viremia in vivo (13). In the first set of experiments, we confirmed the antiviral potency of ST-148 by using synthetic DV-2-based clones derived from the isolate 16681 (Fig. 1A) (17), which was used throughout the study. To determine the inhibitory potency of the compound, Huh7 cells were infected with DV or the DVR2A luciferase reporter virus, encoding either the WT or the previously identified ST-148 resistance mutation in capsid protein (S34L). Released infectivity was assessed by a PFU titration assay using culture supernatants harvested 48 h after infection. As previously observed for the DV-2 NGC strain (13), ST-148 also inhibited 16681-derived DV with EC50 in the low-nanomolar range (52 and 37 nM, respectively) (Fig. 1C and Table 1). These values were around 220- to 300-fold higher in the case of the DV^S34L- and DVR2A^S34L-resistant mutants, respectively, corroborating our previous findings and validating the antiviral potency of ST-148 independent of the DV strain used.

ST-148 does not affect viral RNA replication but inhibits DV spread. Previous time-of-addition studies have shown that ST-148 is effective when added up to 12 h postinfection, suggesting a postentry mechanism of action (13). However, the virus yield inhibition assays used allowed several cycles of viral replication; therefore, possible effects on DV entry, viral RNA replication, and assembly or release of virus particles could not be distinguished. In order to dissect the possible impact of ST-148 on different steps of the viral replication cycle, we first assessed viral RNA replication

<table>
<thead>
<tr>
<th>Virus</th>
<th>EC50 (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV2</td>
<td>0.052 ± 0.016</td>
</tr>
<tr>
<td>DVR2A</td>
<td>0.037 ± 0.028</td>
</tr>
<tr>
<td>DV^S34L</td>
<td>11.79 ± 2.34</td>
</tr>
<tr>
<td>DVR2A^S34L</td>
<td>11.14 ± 3.27</td>
</tr>
</tbody>
</table>

*Cells were infected with given viruses at an MOI of 0.1 and treated with ST-148. Culture medium was collected 48 h postinfection, and virus titers were determined by plaque assay (Huh7 cells, CC50 > 100 μM as determined in reference 13). Data represent averages and error ranges from two independent experiments.

TABLE 1 Antibiotic activity of ST-148 against DV 16681-based viruses
FIG 2 ST-148 does not affect viral RNA replication but reduces production of infectious DV particles. (A) No effect of ST-148 on replication kinetics of the DVR2A reporter virus. Huh7 cells were infected with DVR2A at an MOI of 0.1 or 1 for 4 h at 37°C, washed with PBS, and incubated with fresh medium containing 5 μM ST-148 or DMSO. At the indicated time points after infection, luciferase activity was measured in the lysates as described in Materials and Methods. Note that error bars are poorly visible due to the size of the graph. (B) Reduction of virus production by ST-148. Supernatants of infected cells treated with 5 μM ST-148 were harvested 72 h postinfection and used to infect naïve Huh7 cells. After 48 h, cells were harvested and luciferase activity was determined. (C) Comparison of replication kinetics of DVR2A and DVR2A^S34L in the presence or absence of ST-148. Huh7 cells were electroporated with 10 μg of capped in vitro transcripts of WT DVR2A or DVR2A^S34L (S34L), and 4 h later, 5 μM ST-148 or DMSO was added to the medium. At time points specified at the bottom, luciferase activity was determined. (D) Inhibition of virus production by DVR2A- and DVR2A^S34L-transfected cells upon ST-148 treatment. Supernatants of transfected and ST-148-treated cells were harvested 72 h postinfection and used to infect naïve Huh7 cells. After 48 h, cells were harvested and luciferase activity was determined. Data shown in each panel represent averages and standard deviations from three independent experiments (***, P < 0.001; *, P < 0.05).

and virus spread by using a DVR2A luciferase reporter virus and addition of ST-148 4 h postinfection, i.e., after virus entry had taken place. Virus replication in infected cells was determined 24, 48, and 72 h after infection by luciferase assay (Fig. 2A). Effects on virus production were quantified by using culture supernatants of infected and drug-treated cells harvested 72 h after inoculation, infection of naïve Huh7 cells, and measuring luciferase activity 48 h later (Fig. 2B). While viral RNA replication was not affected by ST-148, luciferase counts detectable in the reinfection assay were decreased 60 to 80%, arguing that the compound reduced virus production (Fig. 2A and B, respectively).

To corroborate that the antiviral effect of ST-148 observed was due to virus production and independent from virus entry under these conditions and to investigate the phenotype of the ST-148-resistant mutant, we next determined the replication competence of either virus in Huh7 cells that had been transfected with the corresponding DV genomes, deliberately bypassing the entry step. Under these conditions, both viruses replicated comparably regardless of ST-148 treatment (Fig. 2C). However, a significant ~80% reduction of amounts of infectious virus was observed exclusively for the WT virus (Fig. 2D), confirming a selective inhibition of the late steps of the DV replication cycle by ST-148.

ST-148 inhibits the production of infectious DV particles. In order to determine whether the reduction of virus production resulted from an impairment in virus particle release or a reduction of their specific infectivity, we next performed infectivity and viral RNA titration experiments. Huh7 cells were infected with 1 MOI of DV, and various concentrations of ST-148 were added 4 h after infection. Amounts of released viral RNA and infectious DV particles in the supernatant were determined 48 h later by qRT-PCR and TCID_{50} assay, respectively (Fig. 3A and B). These experiments revealed a dose-dependent reduction of virus titers as deduced from the up to 10-fold smaller amounts of infectivity and viral RNA observed at the highest inhibitor concentration.

We next investigated whether the observed reduction of virus titer resulted from an impairment of virus spread or from a direct block in virus particle production by using a single-cycle infection experiment. Huh7 cells were infected with DV or DV^S34L at an MOI of 1 and incubated for 48 h to allow for nearly complete infection of the cell monolayer (Fig. 3C). Released viruses then were removed by extensive washing, and medium containing DMSO, ST-148, or brefeldin A, which served as a positive control, was added to the cells. Cells were incubated for just 12 h, which is the estimated length of a single complete DV replication cycle (21 and B. Schmid, unpublished observations). Therefore, reinfesting or superinfecting virus produced within this time period would not significantly contribute to extracellular virus titer. Viral RNA contained within infected cells or released into the culture supernatant was quantified by RT-qPCR. In addition, infectivity titers were determined by limiting-dilution assay. Under these conditions, a highly significant reduction of ~90% of infectivity was observed for the drug-treated WT virus, while the S34L mutant was unaffected by ST-148 treatment (Fig. 3D). Concomitantly, ST-148 profoundly reduced secretion of viral RNA by ~70% in the case of DV WT, while no statistically significant reduction could be observed for the DV^S34L-resistant variant (Fig. 3E).
expected, in both cases BFA treatment drastically reduced titers of both viral RNA and infectivity (Fig. 3D and E; black columns), consistent with an earlier report (22). In conclusion, these results suggest that ST-148 reduces DV particle production. The concomitant reduction of extracellular infectivity and viral RNA amounts implies that ST-148 has no significant effect on specific infectivity of DV particles.

**ST-148 additionally inhibits DV entry.** Since viral RNA replication in the presence of ST-148 was unaltered and the observed inhibitory effect of the drug on assembly/release of DV particles was rather modest compared to the high inhibitory potency observed by virus yield inhibition assay, we hypothesized that ST-148 has an additional effect on virus entry. To verify this hypothesis, we developed a trans-complementation system, which allows production of infectious virus-like particles containing a subgenomic replicon that are unable to spread in cell culture (23, 24, 25, 26, 27). These particles, which we call DV trans-complemented particles (DV_TCP), in analogy to our earlier report on HCV_TCP (16), were obtained by transfecting subgenomic replicons into helper cells that provide the structural proteins in trans (Fig. 4A).

Since packaged subgenomic replicon RNAs lack the region encoding the structural proteins, DV_TCP only support a single round of infection. Moreover, DV_TCP encode the Renilla luciferase reporter gene; therefore, single-round infection can be measured as a function of luciferase activity.

Taking advantage of luciferase-encoding DV_TCP composed of WT or S34L capsids, we performed time-of-addition studies by treating cells before, during, or at different times after infection with ST-148 and measured luciferase activity 48 h later. As shown in Fig. 4B, maximal antiviral activity was achieved when cells were treated prior to or during infection, reducing viral replication up to 70% (*P < 0.01; ***P < 0.001). Significant antiviral effect was still obtained up to 120 min postinfection (i.e., 30 min after removal of virus inoculum), whereas treatment at later time points had no effect (Fig. 4B).

To exclude a general nonspecific virucidal effect of ST-148 on DV particles, we incubated high-titer virus stocks of DV or DV^S34L^ with 10 μM ST-148 for 1 h at 37°C and determined the infectivity titer after extensive dilution by TCID₅₀ assay (Fig. 4C). Under these conditions, no significant reduction of viral infectivity could be observed. Taken together, these data sug-

---

**FIG 3** Inhibition of the release of infectious DV particles by ST-148 as revealed by single-cycle experiments. (A) Production of infectious extracellular DV upon ST-148 treatment. (B) Quantification of released viral RNA. Huh7 cells were infected with DVR2A at an MOI of 1 for 4 h at 37°C, washed with PBS, and incubated with fresh medium (nontreated [NT]) or medium containing DMSO or the given concentrations of ST-148. After 48 h, clarified supernatants were used for TCID₅₀ assay to quantify infectivity titers or two-step qRT-PCR for quantification of viral RNA amounts released into the supernatant. Data were analyzed using one-way ANOVA and Dunnett’s post hoc test (**, *P < 0.01; ***P < 0.001). (C) Determination of infection efficiency. Cells were infected with DV or DV^S34L^ and 48 h later were fixed and analyzed by immunofluorescence using a DV E-specific antibody. Nuclear DNA was stained with DAPI. Numbers in the lower right of the panels refer to the percentage of infected cells (means from at least 1,500 cells ± standard deviations). (D and E) Effect of ST-148 on single-round DV particle release. Huh7 cells were infected with DV or DV^S34L^ for 48 h, washed extensively with PBS, and incubated for 12 h with medium containing 0.1% DMSO, 1 μg/ml brefeldin A (BFA), or 10 μM ST-148. Released infectivity and viral RNA were quantified by TCID₅₀ assay and qRT-PCR, respectively. Results represent means and standard deviations from at least 3 independent experiments. (ns, nonsignificant; *, *P < 0.05; **, ***P < 0.001).
gested that ST-148 specifically blocks DV entry, which is not due to a virucidal effect.

**ST-148 does not affect colocalization of capsid with lipid droplets and dsRNA but increases clustering of virus particles in the ER.** In DV-infected cells, C protein resides both in the cytoplasm and nucleus \( (28, 29, 30) \). Cytosolic capsid protein was shown to localize around LDs \( (15) \) and in close association with vesicle packets, which are sites of RNA replication \( (30, 31, 32) \), while nuclear capsid protein accumulates in nucleoli and was reported to interact with nuclear proteins \( (33, 34, 35) \). In order to investigate whether ST-148 would affect the recruitment of capsid to LDs and its colocalization with dsRNA, we performed colocalization studies using DV-infected cells treated with ST-148. Treatment was started 4 h postinfection until cells were fixed by using paraformaldehyde and a low concentration of Triton X-100 as reported earlier \( (15) \). As shown in Fig. 5A, we could not observe a difference between inhibitor- and DMSO-treated cells, and colocalization of capsid around LDs or with dsRNA appeared unaltered. To rule out possible effects of ST-148 on the ultrastructural morphology of DV-infected cells treated with ST-148 by using transmission electron microscopy (Fig. 5D). Both DV- and DV\(^{S34L}\)-infected cells contained the characteristic membrane invaginations, which have been proposed to represent viral replication factories (vRFs) \( (9) \), and virions forming regular arrays within the ER. Treatment of DV-infected cells with ST-148 did not affect the morphology of vRFs, consistent with a lack of inhibition of viral RNA replication, whereas the overall amount of virions was drastically reduced. Interestingly, in the case of the WT virus, the fewer virions that could be found intracellularly in ST-148-treated cells appeared to form larger arrays than control cells (Fig. 5E). This phenotype was specific and not found with ST-148-treated cells that had been infected with the DV\(^{S34L}\)-resistant mutant (Fig. 5E).
ST-148 alters intracellular distribution of capsid protein. Several lines of evidence point toward a functional role for flavivirus capsid within the nucleus of infected cells. For instance, West Nile virus capsid protein has been proposed to dynamically shuttle between the cytoplasm and the nucleus, and the temporal regulation of this process was suggested to prevent premature encapsidation of viral RNA at early stages of infection (33, 36). For Japanese encephalitis virus, it was shown that disrupting nuclear localization affected virus replication and pathogenesis in mice (37). In the case of DV, several studies reported the presence of C protein in the nucleoli of infected cells (28, 33, 34, 38). Moreover, DV C protein has been shown to interact with a number of nuclear proteins, including hnRNP-K and histones (39, 40, 41); however, the relevance of nuclear localization of DV C protein with respect to viral replication is unclear (28). In light of these reports, we investigated the influence of ST-148 on the nuclear accumulation of C protein by using an immunofluorescence protocol that includes the permeabilization of the nucleus. Consistent with earlier reports (15), under these conditions we detected capsid both in the cytosol and within the nucleoli of infected cells (Fig. 6A and B). In the case of DV WT-infected cells, ST-148 treatment increased capsid accumulation within the nucleoli compared to that of DMSO-treated control cells. In contrast, although the S34L-containing capsid protein exhibited enhanced nuclear localization, this subcellular distribution was not affected by ST-148 treatment. To further investigate the intracellular distribution of C protein upon ST-148 treatment, we generated subcellular fractions of DV- or DV S34L-infected cells by using a method that allows separation of cytosolic, membrane, nuclear, and insoluble postnuclear proteins (Fig. 6C). This approach revealed that ST-148 treatment reduced the abundance of cytosolic WT C protein, concomitant with an increase of this protein in the postnuclear, insoluble fraction (Fig. 6C). In contrast, subcellular distribution of the S34L capsid protein was not affected by the treatment of infected cells with the compound (Fig. 6C). In conclusion, these results show that ST-148 reduces the abundance of cytosolic capsid concomitant with an accumulation of the protein in nuclear, detergent-resistant aggregates, which might contribute to the observed reduction of virus particle production.

ST-148 increases capsid self-interaction. DV capsid forms dimers (10) and, eventually, oligomers during nucleocapsid assembly. It is plausible to assume that the stability of capsid oligom-
ers is important for assembly of infectious virus particles and for uncoating during virus entry. Since ST-148 appeared to affect both the efficiency of virus production and entry properties of DV particles, we assumed that this compound affects capsid self-interaction. To address this assumption, we used a BRET-based assay, which allows investigating the protein of interest within its physiologically relevant subcellular compartment in live cells (42). BRET is detected when energy emitted by the Renilla luciferase (Rluc) of a donor fusion protein is transferred to an acceptor YFP fusion partner following catalytic degradation of the cell-permeable Rluc substrate coelenterazine H (Fig. 7A) (14). In the first set of experiments, we performed DSA by using capsid proteins fused N terminally with either YFP or Rluc (Fig. 1A). As shown in Fig. 7A, a nonlinear increase in BRET intensity and a rapid saturation of the BRET signal was detected when a fixed amount of the Rluc-capsid construct was coexpressed with increasing quantities of the YFP-capsid construct. This pattern clearly indicated specific capsid self-interaction. In contrast, increasing concentrations of a

FIG 6 ST-148 induces accumulation of capsid protein in nuclear insoluble aggregates. (A) Intracellular distribution of capsid protein upon ST-148 treatment. Huh7 cells were infected with DV or DV S34L at an MOI of 1, and 4 h later, DMSO or 10 μM ST-148 was added to the medium. After 48 h, cells were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100, and immune stained for capsid with a mouse monoclonal antibody. Scale bars represent 20 μm. (B) Image-based quantification of intracellular distribution of capsid protein. The extent of nuclear localization of capsid was determined by quantifying capsid-specific signals detected in the nucleus (N) and the cytoplasm (C), respectively, and calculating the ratio of mean fluorescence (F) in these two compartments [F(N/C)]. Results represent the means ± standard errors of the means (n ≥ 50). Data were analyzed using one-way ANOVA with Tukey’s post hoc test (***, P < 0.001; ns, nonsignificant). (C) Subcellular fractionation of DV proteins. Huh7 cells were infected with DV or DV S34L at an MOI of 1, and 4 h later, cells were treated with DMSO or 10 μM ST-148. Intracellular distribution of capsid in soluble cytosolic (Cyt), membrane-associated (Membr), nuclear (Nuc), and postnuclear insoluble (Post-Nuc) fractions was evaluated by Western blotting. GAPDH, ATP5B, lamin A/C, and vimentin served as specificity controls for the various fractions. Numbers below each lane refer to the relative amounts of C proteins as determined by densitometry. After subtraction of the background, capsid-specific signals in each fraction were normalized to the corresponding input and expressed as a fold of total signal in the four fractions. Numbers on the left refer to the positions of molecular weight standards (in thousands). One result from two independent experiments is shown.
FIG 7 ST-148 increases DV capsid protein self-interaction. (A) Self-interaction kinetics of WT and S34L capsid as determined by BRET. DSA were performed in live 293T cells cotransfected with increasing amounts of YFP-tagged and a constant amount of Rluc-tagged plasmids encoding WT or mutant (S34L) capsid or NS4A, which served as negative controls. Forty-eight hours later, energy transfer was induced by the addition of the Renilla luciferase substrate coelenterazine H. The x axis shows the ratio between the normalized fluorescence of the acceptor (netYFP) measured before coelenterazine addition and the luminescence of the donor. BRET_{50} values (netYFP/Renilla ratio at which 50% of maximal BRET is occurring), which reflect the relative affinity of the acceptor protein for the donor protein, are given in the top right of the panel. Curves represent the means ± standard deviations of results from one representative experiment carried out in triplicate. The curves were fitted using a nonlinear regression equation in which a single binding site was assumed. Representative experiment carried out in triplicate. The curves were fitted using a nonlinear regression equation in which a single binding site was assumed. (B) Effect of ST-148 on capsid self-interaction. 293T cells were transfected with 1 µg of YFP-capsid and 20 ng of Rluc capsid (WT and S34L). Prior to transfection, different concentrations of ST-148 or DMSO were added to the medium, and 48 h later BRET was measured (***, P < 0.001; *, P < 0.01).

YFP-NS4A construct (used as a negative control) coexpressed with the Rluc-capsid construct resulted in only a weak and linear increase of BRET, which is a characteristic pattern for noninteracting proteins (43, 44).

Besides providing information on the specificity of a given interaction determined by BRET, additional parameters, such as BRET_{max}, i.e., the BRET signal at which the saturation curve reaches a plateau, and BRET_{50}, i.e., the concentration of acceptor giving 50% of BRET_{max} can be extrapolated from DSA experiments. While BRET_{max} is a function of the distance and orientation between the donor and acceptor within their oligomeric complex, BRET_{50} reflects the relative affinity between the fusion proteins. As deduced from the BRET_{50} values, the S34L mutation did not alter capsid self-interaction, and no significant difference between the oligomerization profiles of WT and S34L capsid proteins could be observed (BRET_{50} of WT, 4.99 ± 1.071; BRET_{50} of the S34L mutant, 4.01 ± 0.67) (Fig. 7A).

We next evaluated the effect of increasing concentrations of ST-148 on WT and S34L capsid self-interaction under conditions of BRET_{max}, i.e., ~50-fold excess of YFP-capsid plasmid relative to the Rluc-capsid plasmid (Fig. 7B). Under these conditions, we observed a dose-dependent increase of the BRET ratio, first detectable at 156 nM ST-148 and reaching a plateau at 1.25 µM. This increase in BRET signal was found only with the WT capsid protein, whereas the oligomerization status of the S34L capsid appeared unaffected for all tested concentrations, demonstrating the specificity of the ST-148-mediated enhancement of capsid self-interaction. These results support the notion that ST-148 directly affects capsid protein. The correlation between antiviral efficacy and enhancement of capsid self-interaction argues that increased capsid oligomerization is responsible for the antiviral activity of ST-148.

In silico modeling corroborates ST-148-mediated stabilization of interdimer capsid interactions. To further understand the mode of action of ST-148 at the molecular level, we performed molecular modeling studies of WT and S34L capsid protein by using protein–protein docking, MD simulations, and small-molecule docking. These studies were based on the NMR structures reported by Ma et al. (PDB accession no. 1R6R) (12). Since capsid protein was reported to form stable dimers in solution (10, 12), they were considered fundamental units; therefore, the formation of tetramers was simulated. Since we observed the nuclear accumulation of capsid in the presence of ST-148, only tetramers that did not involve nuclear localization sequences (NLS) (28) at the interaction surfaces were considered, assuming that otherwise these NLSs were hidden. Likewise, complexes that did not involve LD-binding domains (12, 45, 46) were preferred, because no effect on capsid–LD colocalization by ST-148 was detected. Priority was given to complexes in which the protein–protein interaction surface involved Ser34, as this amino acid residue was altered in the case of the ST-148-resistant DV mutant (Fig. 8A). The symmetry of the tetramer was not considered an important element, as cryo-electron microscopy data revealed that the DV nucleocapsid is disordered (47, 48, 49).

Two 30-ns MD simulations were performed on the best tetramer complex: one on the WT and one on the same structure containing the S34L mutation. In accordance with the BRET data, the two structures behaved similarly, indicating no differences between the WT and the mutant proteins in the absence of the inhibitor. In both cases, the tetramer quaternary structure changed during the first 2 ns (see Videos S1 and S2 in the supplemental material) and was stable for the rest of the simulation, suggesting that the system reaches equilibrium.

By using the energy-minimized tetramer structures of both the WT and S34L capsid, we next docked ST-148 using a 30-Å grid centered on the two Ser34 (or Leu34 in case of the mutant) residues at the dimer–dimer interface and examined the obtained ST-148 pose for WT and the S34L mutant over a 30-ns MD simulation. Comparable to the MD simulations in the absence of ligands, a similar change in the quaternary structures of both WT and S34L capsid could be observed during the first 2 ns (see Videos S3 and S4 in the supplemental material), and then the system reached equilibrium. Figure 8B shows the final MD result structures obtained for the WT protein in absence (left) or in the presence (right) of ST-148. In the case of the WT capsid, ST-148 bound the
FIG 8 Docking model of ST-148 and DV capsid and possible stabilization of interdimer capsid protein interaction. (A) Sequence comparison of capsid amino acid sequences of DV serotypes 1 to 4 and selected flaviviruses. The position of serine 34 is highlighted in red. Sequence alignment was carried out using the ClustalW algorithm, available at the Uniprot webserver, on the following isolates: DV-2 (Thailand/16681-PDK53), DV-1 (Brazil/97-11/1997), DV-3 (Martinique/1243/1999), DV-4 (Thailand/0348/1991), West-Nile virus (WNV; H442), yellow fever virus (YFV; Ivory Coast/1999), Japanese encephalitis virus (JEV; SA-14), Kunjin virus (KUNV; MRM61C), and St. Louis encephalitis virus (SLEV; MS1-7). UniprotKB accession numbers are given in Materials and Methods. (B) Results from molecular dynamics (MD) simulations. Stable conformations of the WT capsid tetramer in the absence (left) or presence (right) of ST-148. Capsid dimers are represented in blue and orange, with each monomer given in different shades of the same color. ST-148 is shown with green sticks, and the serine residue at position 34 is represented with the Corey-Pauling-Koltun (CPK) model. Movies of the MD simulations are given in the supplemental material. (C) Results of superimposition of the MD simulation of the WT C tetramer in the absence (red ribbon) or presence (blue ribbon) of ST-148. Capsid dimers are represented in blue and orange, with each monomer given in different shades of the same color. ST-148 is shown with green sticks, and the serine residue at position 34 is represented with the Corey-Pauling-Koltun (CPK) model. Movies of the MD simulations are given in the supplemental material. (D) WT capsid dimer-dimer interface complexed with ST-148. (E) Same interface as that shown in panel D, with the serine residue superimposed on Leu34. The dashed circle highlights the steric hindrance induced by the leucine residue at position 34, clarifying how the mutation impedes this binding pose. In both panels one dimer is depicted in light blue, while the other capsid dimer is in red in the case of the WT or in yellow in the case of the S34L resistance mutant. Protein residues are displayed as lines, while Ser34 and Leu34 are drawn with stick representations.
complex in a cleft between the α1 helix of one dimer and α1 and α3 helices of the other dimer and remained stable at this site for the last 10 ns of the simulation (see Video S5 in the supplemental material). It is worth noting that ST-148 induced a shift of the α3 and α2 helices (Fig. 8C, dashed arrows), creating the ligand cleft and stabilizing further compound binding. Conversely, in the case of the S34L mutant protein, the compound did not bind a cleft at the dimer-dimer interface but explored more than one site in a solvent-exposed area of one dimer (see Video S6 in the supplemental material).

To confirm this difference in ligand binding between the two structures, we used the stable WT tetramer-ligand complex corresponding to the equilibrium state and docked ST-148 into the pocket identified during the MD by using a 20-Å grid (Fig. 8D). This docking confirmed that ST-148 is able to bind the tetramer at the dimer-dimer interface, and the same compound conformation obtained from the MD was reproduced. In the case of the WT structure, the binding site is formed primarily by residues Val26, Leu29, Arg41, and Arg68 of one dimer and Ser34 of the other dimer. As shown in Fig. 8D, the phenyl moiety of ST-148 binds Phe33 with a π-π stacking interaction, while the thieno[2,3-b]pyridine portion of ST-148 is tethered between Ser34 and Arg68, where the positive charge interacts with the electron-dense aromatic system. Furthermore, as ST-148 is highly lipophilic, its binding is enhanced by the large number of hydrophobic residues forming the pocket. As the compound is in close contact with Ser34, a leucine residue at this site creates a steric clash between the protein and the ligand (Fig. 8E), confirming that the S34L resistance mutation does not allow ST-148 to bind to the tetramer in the same way, consistent with the MD simulations. In summary, these results support the notion that ST-148 stabilizes capsid oligomerization and reveal a plausible model for compound binding to the capsid protein as well as binding inhibition by a distinct resistance mutation.

DISCUSSION
Capsid proteins of enveloped viruses have emerged as promising targets for the design of antiviral agents. In fact, the processes regulating both viral genome encapsidation and release during entry proved extremely sensitive to even subtle molecular disturbances (reviewed in reference 50). In this study, we describe the mechanism of action of ST-148, a recently identified small-molecule inhibitor that binds to and induces resistance in DV capsid (13). By using a combination of full-length reporter viruses and DV\_TCP, we identified two distinct inhibitory effects: impaired DV entry and reduced production of infectious virus particles. Based on BRET assays, in silico docking, and MD simulations, we propose that these antiviral effects are the result of stabilized capsid self-interaction.

With respect to inhibition of virus entry, the lack of virucidal effects and the observed specific inhibitory effects of ST-148 on virus entry up to 30 min postinfection point toward a block in viral RNA uncoating. Although no detailed kinetics of nucleocapsid disassembly have been reported to date, earlier studies suggested that DV is internalized within 30 min after inoculation (51, 52). While our data do not demonstrate inhibition of the uncoating process, the observed increase of capsid self-interaction makes this assumption plausible. Nevertheless, additional experiments are required to directly demonstrate that viral RNA release is blocked at the nucleocapsid disassembly/RNA uncoating step.

In addition to impaired entry, ST-148 also reduced virus production. This might occur at the level of particle assembly or their release, but likely it is not due to alteration of specific infectivity of DV particles, as suggested by the concomitant reduction in both the released infectivity and the secreted viral RNA. Consistent with this result, we note that in spite of the overall reduction of electron-dense virus particles found within DV-infected cells upon ST-148 treatment, particle morphology appeared unaltered. Interestingly, these particles accumulated in larger numbers within ER stacks. Whether this phenomenon is a consequence of altered nucleocapsid assembly or results from disturbed intracellular transport of assembled virions remains to be determined. In any case, the observed redistribution of capsid from the cytosol to the nucleoli in ST-148-treated cells might contribute to the interference with virus production by reducing the amount of capsid protein available for assembly. Although lower cytoplasmic capsid levels also were observed for the DV\_S34L mutant, its ability to traffic in and out of the nucleus likely is not compromised or is insufficient to support virion morphogenesis. Conversely, in the case of the WT capsid, the ST-148-induced nuclear retention correlated with an alteration of the protein (or protein complex) solubility, which might account for the observed reduction in virus production.

The proposed mechanism of action for both the early- and late-stage inhibition of ST-148 involves a direct effect on higher-order structures of DV capsid. This assumption is supported by the significant and dose-dependent increase in capsid self-interaction observed by BRET. Since we used conditions of BRET\_max, our results suggest that the compound induces the formation of higher-order oligomeric capsid species or stabilizes further preexisting capsid dimers. This effect was specific for the WT capsid protein and not due to pleiotropic effects caused by the compound, because the interaction profile of the ST-148-resistant mutant was unaltered.

By using MD, we generated a model of a capsid tetramer in complex with ST-148. Obtained results confirmed that ST-148 stabilizes capsid self-interaction. Thus, ST-148 can be classified as a direct protein-protein interaction (PPI) stabilizer, establishing hydrophobic and electrostatic interactions in a cleft between the capsid dimers. As shown by our MD simulations, ST-148 retains the ability to bind the S34L-containing tetramer but in a less stable conformation and involving only one dimer. Therefore, molecular modeling data are in agreement with the previously observed binding of ST-148 to both the WT and the S34L-containing capsid protein (13). Furthermore, our in silico results clarify the role of S34L mutation in drug resistance, as the steric hindrance of leucine at position 34 impedes the stable binding of the compound to the tetrameric capsid protein complex.

Drug-induced modulation of capsid protein self-interactions is not unique to DV and has been described as an applicable strategy for several other enveloped viruses. For instance, in the case of hepatitis B virus (HBV), several inhibitors targeting the core protein have been reported, and at least two classes of HBV assembly effectors, the heteroaerylidydropyrimidines (HAPs) and phenylpropenamides, have been extensively characterized (53, 54, 55, 56). For the HAPs, elegant work from Zlotnick and colleagues demonstrated that these compounds can increase the kinetics of assembly and strengthen dimer-dimer association to stabilize capsids (57, 58). Comparable to HAPs, certain phenylpropenamides accelerate assembly and stabilize capsids, and some of these com-
pounds, such as AT-130, were shown to block viral RNA packaging without altering the geometry of nucleocapsids; they still formed, but they were devoid of nucleic acids (59, 60). Interestingly, molecules belonging to either class were shown to bind to partially overlapping sites of the HBV core protein.

In the case of HIV, different series of capsid inhibitors have been identified. For instance, the benzodiazepine (BD) and the benzimiazole (BM) series were reported to inhibit virion release or prevent virus maturation, respectively (61, 62). In the case of the BD series, molecular studies, complemented by nuclear magnetic resonance and crystallography approaches, have shown that compounds from this class inhibited HIV particle release by blocking assembly of immature capsids (63). Conversely, molecules of the BM series did not affect virus budding but prevented capsid maturation. Similar to HBV, these two compound series induced different biological phenotypes, although they both shared the same binding site within the N-terminal domain of the HIV capsid protein (64).

We note that Blair and colleagues characterized another series of HIV-1 capsid-targeting compounds interfering with both viral uncoating and formation of infectious particles (65). Similar to those of ST-148, these effects appeared to be mediated by the direct binding of the compound to HIV-1 capsid protein, increasing the rate of capsid protein multimerization. Similar to the study of Blair et al., a recent study reported the identification of a novel class of HIV-1 capsid inhibitors accelerating HIV capsid-nucleocapsid assembly and stabilizing preassembled capsid-nucleocapsid tubes against dissociation (66). These reports strengthen the concept of potent inhibition of viral replication via stabilization of the viral capsid.

Comparable to what has been reported for the anti-HBV HAPs, we hypothesize that the late-stage inhibitory effect of ST-148 on DV results from the stabilization of capsid protein interdimer association, which might reduce capsid plasticity and its ability to package viral RNA. Additionally, ST-148 seems to share selected features with the anti-HIV pyrrolopyrazolones, exerting antiviral activity on the early steps of the viral replication cycle, presumably by stabilizing preexisting capsid quaternary structures. However, the mechanism by which these HIV inhibitors stabilize capsid complexes is uncertain, because the inhibitor binding site does not reside within an inter- or intrahexameric capsid interface stabilizing the capsid lattice, while in case of ST-148, the putative binding site was identified between two DV capsid dimers.

A growing list of studies supports the notion that the multifunctional role of viral capsid protein, involving different protein conformations and interactions with different protein surfaces and host factors, represents a target that can be exploited for the development of potent inhibitors. Our work supports this hypothesis and argues that inhibitors targeting viral capsid proteins also are applicable to flaviviruses. Given the lack of approved DV-specific antiviral drugs, this target should be further investigated, ideally to design combination therapies with a high genetic resistance barrier.

In conclusion, we determined the mode of action of a DV capsid inhibitor that affects both entry and assembly/release of infectious virus particles. The antiviral effect appears to be mediated by enhanced capsid protein interaction, eventually inducing structural rigidity and/or steric hindrance during nucleocapsid formation. These results, together with the first description of a robust BRET-based assay to measure capsid self-interaction in live cells, might serve as a starting point for optimization of more effective therapeutics simultaneously targeting early and late stages of the DV life cycle.

ACKNOWLEDGMENTS

We thank Andrea Gamarnik and John G. Aaskov for the gift of rabbit polyclonal anti-C-specific and mouse monoclonal anti-C-specific antibodies, respectively. We are grateful to Martin Baril for insightful discussions on the BRET assay, Mirko Cortese for critical readings of the manuscript, and Wolfgang Fischl, Laurent Chatel-Chaix, and Stefan Seitz for helpful discussions.

This work was supported by grants from the European Union (Silver [grant no. 260644] and EUVIRNA [grant no. 264286]) and the Deutsche Forschungsgemeinschaft (SFB638, Teilprojekt A5), all to R.B.

When the study was initiated, C.M.B. and R.J. were employees of SIGA Technologies, Inc., and held stock in the company. R.J. no longer holds stock in SIGA.

REFERENCES


Incorporation of tick-borne encephalitis virus replicons into virus-like particles by a packaging cell line. J. Virol.


http://jvi.asm.org/
A novel radiochemical approach to 1-(2′-deoxy-2′-[18F]fluoro-β-D-arabinofuranosyl) cytosine (18F-FAC)†

Jan-Philip Meyer,a,b Katrin C. Probst,b Iuni M. L. Trist,a Christopher McGuigan,a and Andrew D. Westwell a,*

18F-FAC (1-(2′-deoxy-2′-[18F]fluoro-β-D-arabinofuranosyl)-cytosine) is an important 2′-fluoro-nucleoside-based positron emission tomography (PET) tracer that has been used for in vivo prediction of response to the widely used cancer chemotherapy drug gemcitabine. Previously reported synthetic routes to 18F-FAC have relied on early introduction of the 18F radiolabel prior to attachment to protected cytosine base. Considering the 18F radiochemical half-life (110 min) and the technical challenges of multi-step syntheses on PET radiochemistry modular systems, late-stage radiofluorination is preferred for reproducible and reliable radiosynthesis with in vivo applications. Herein, we report the first late-stage radiosynthesis of 18F-FAC. Cytidine derivatives with leaving groups at the 2′-position are particularly prone to undergo anhydro side-product formation upon heating because of their electron density at the 2-carbonyl pyrimidine oxygen. Our rationally developed fluorination precursor showed an improved reactivity-to-stability ratio at elevated temperatures. 18F-FAC was obtained in radiochemical yields of 4.3–5.5% (n = 8, decay-corrected from end of bombardment), with purities ≥98% and specific activities ≥63 GBq/μmol. The synthesis time was 168 min.

Keywords: nucleosides; positron emission tomography; radiochemistry; radiopharmaceuticals; computational chemistry

Introduction

Nucleoside derivatives represent an important emerging class of biomarkers for positron emission tomography (PET) imaging.1 For example, the clinically approved proliferation biomarker 18F-FLT (3′-deoxy-3′-[18F]fluoro-L-thymidine) is accessible via an effective radiochemical synthesis in high radiochemical yield and purity suitable for human production under Good Manufacturing Practice guidelines.2 Hence, 18F-FLT is now commonly used for the early detection of various cancers and the evaluation of treatment response.3 Especially in high-grade tumours, it appears to be more powerful in predicting treatment response and survival than 18F-FDG, as measured by standard uptake values of 18F-FLT directly corresponding to cell proliferation shown by the Ki-67 proliferation index.4

Cytidine-based radiotracers for PET imaging such as 18F-FAC (1, Figure 1) and 1-(2′-deoxy-2′-[18F]fluoroarabinofuranosyl)-5-methylcytosine (18F-FMAC, 2, Figure 1) have been previously shown to be successful as predictors of response for nucleoside analogue-based treatment of cancer5,6 as well as local immune activation in mouse models.7 As such, 18F-FAC has been used for response prediction to the widely used chemotherapy drug gemcitabine.8 Gemcitabine is frequently used for (combination) chemotherapy of a number of solid tumours such as pancreatic, non-small cell lung, breast and bladder cancer. Importantly, gemcitabine efficacy is compromised by poor cellular uptake by nucleoside transporters, variable activity of the essential deoxycytidine kinase (dCK) activating enzyme and rapid deactivating metabolism by cytidine deaminase (CDA). In addition, serious dose-limiting toxicities are associated with this chemotherapy drug.8 Prediction of patient response to gemcitabine chemotherapy would therefore be an important milestone towards personalized patient chemotherapy in the cancer clinic.

Given the close structural resemblance to gemcitabine, 18F-FAC was found to have a comparably high affinity for deoxycytidine kinase (dCK), a rate-limiting enzyme in the nucleoside salvage pathway. Hence, in vivo imaging determined dCK-positive and dCK-negative tumours. Furthermore, an additional study showed that 18F-FMAC PET imaging provides information about cytidine deaminase (CDA) enzymatic activity in vivo, which is related to gemcitabine resistance.3 Together, cytidine-based PET probes enable further conclusions towards personalized

†School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Redwood Building, King Edward VII Avenue, Cardiff, CF10 3NB, Wales, UK

bWales Research & Diagnostic PET Imaging Centre (PETIC), Institute for Translation, Innovation, Methodology & Engagement (TIME), School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, Wales, UK

cCorrespondence to: Andrew D. Westwell, School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Redwood Building, King Edward VII Avenue, Cardiff, CF10 3NB, Wales, UK. E-mail: WestwellA@cf.ac.uk

*Additional supporting information may be found in the online version of this article at the publisher’s web-site.
chemotherapy that, considering the low (≤20%) response rates and severe side effects, could lead to significantly improved treatment efficacy.8

2′-18F]arabinocytidines such as 18F-FAC and 18F-FIAC (3, Figure 1) have been synthesized in high purities using an early stage 18F introduction strategy.9 However, this synthetic approach limits reliable and reproducible production on commercially available synthesis modules for human application.10 For instance, multiple steps after 18F introduction with an additional purification step due to the formation of the D- and L-isomers, as well as the use of toxic substances such as hydrogen bromide, make the translation to automated synthesizers difficult. These challenges go hand in hand with reproducibility issues when it comes to quality control in clinical practice according to the British/European/American Pharmacopeia. Furthermore, as the number and space in hot cells is limited, multiple-step radiosyntheses should preferably be replaced by shorter and less complex synthesis procedures. Lazari et al. recently presented a newly designed radiosynthesis module system that furnished the nucleoside analogues 18F-FMAU and 18F-FAC in good overall radiocchemical yields (46% and 31%, respectively, decay-corrected) and with acceptable reproducibility following early stage fluoride introduction.11 Overall synthesis times were reported to be 165–170 min in each case. The flexible radiosynthesizers developed are described as being suitable for both the development and production of different radiotracers for applications in PET.

The aim of this study was to develop a novel and straightforward synthetic approach towards 18F-FAC that could be applied to the radiosynthesis of related cytidine-based radiotracers. A more straightforward radiosynthesis similar to the 18F-FDG synthesis in combination with innovative radiosynthesizers would enable a faster transition from tracer development to clinical application. Late-stage 18F-introduction at the 2′-arabino position of an intact nucleoside towards nucleoside-based radiotracers was previously considered as difficult.12 However, Alauddin and co-workers recently developed a novel radiosynthesis of the uridine analogue 18F-FMAU based on a late-stage 18F-introduction approach in order to increase possibilities for making nucleoside-based PET tracer suitable for clinical applications.10,12 The key towards a stable precursor that would be less likely to undergo anhydro side-product formation (e.g. Scheme 1) due to neighbouring group participation was the introduction of a good electron-withdrawing N-protecting group(s). Because the amino group at the 4-position of cytidine derivatives likely enhances the nucleophilicity of the 2-carbonyl oxygen, potential precursor molecules containing a leaving group at the 2′-position are even more prone to undergo anhydro formation especially at high temperatures (≥75°C). Thus, a reasonable radiofluorination precursor must combine sufficient reactivity with stability towards heat as well as basic conditions. In this study, we present for the first time a late-stage radiofluorination approach to the 18F-labelled cytidine derivative 18F-FAC.

### Experimental section

#### Radiosynthetic procedures

18F-fluoride was trapped on a QMA cartridge (equilibrated with 5% NaHCO3 solution (5 mL) and water (5 mL)) before it was eluted with an aqueous solution (0.7 mL) of KHCO3 (22.6 mg/mL) and kryptofix K222 (22.6 mg/mL) in MeCN. The resulting 18F/KHCO3/K222-complex was dried by co-evaporation under reduced pressure and a stream of nitrogen. The drying process was repeated three times (3×1 mL MeCN). A solution of precursor (23) (10 mg, 14.5 μmol) in DMF (0.3 mL) was then added, and the resulting mixture was stirred at 110°C for 20 min. The crude mixture was either analysed directly for determination of 18F-incorporation (radio-TLC) or passed through an Alumina-Sep-Pak cartridge (equilibrated with water (5 mL)) for characterization (radio-HPLC). The 18F-labelled intermediate was eluted with EtOAc (2 mL), and an aliquot (0.2 mL) of the mixture was analysed by analytical HPLC using co-elution with the non-radioactive reference compound (24). Rf = 12.9 min, MeCN/H2O 9:1, 1 mL/min, ≥87%, radio-TLC, Rf = 0.6 (4% MeOH/DCM).

18F-FAC (1)

After elution of the 18F-labelled intermediate (25) with EtOAc into a new reaction V-vial, the solvent was removed under reduced pressure and a stream of nitrogen at 80°C. 2N HCl solution (0.3 mL) was added to the dried residue followed by stirring at 90°C for 15 min. The solution was then neutralized with 2N NaOH solution (0.3 mL). The resulting mixture was loaded onto a semi-preparative HPLC column via a 10-mL injection loop. The fraction containing 18F-FAC was collected after 22.5 min with 3% MeCN in H2O and was analysed by analytical HPLC using co-elution with the non-radioactive reference compound (27), Rf = 13.8 min, 3% MeCN/H2O, 1 mL/min, ≥98%. A carrier-added synthesis was performed according to the previous procedure, and the product 18F-FAC was saved for decay for 20 half-lives (>2 days) and subsequently analysed by mass spectrometry; MS (70 eV): m/z (%): 245.4 (5) [(M+H)+], 267.7 (95) [(M+Na)+].
Computational methods

MOE 2010.10 software was used to design each compound and to explore possible conformations. The searching method applied was LowModMD that generates conformations using a short run of molecular dynamics at constant temperature followed by an all-atm energy minimization. Amongst the conformations, 16 of them were selected for each compound for quantum mechanics (QM) evaluation. QM calculations were executed for the selected conformations with the GAMESS13 software. A geometry optimization was performed in vacuo, using analytical energy gradients with the restricted Hartree–Fock wave function. The 3-21G split valence basis set was used, and diffuse sp shell was added to heavy atoms. The initial hessian was guessed, and the convergence gradient tolerance was set to 0.0005 E0/bohr. The electrostatic potential was calculated at points determined by Michael Connolly algorithm on the surface of van der Waals fused spheres. Atomic charges were consequently fitted to this potential. The partial net charges of atoms of interest were gathered from each molecule conformation, and values were averaged.

Results and discussion

Precursor design

In order to develop a new precursor for radiofluorination of the 2′-arabino position of cytidine derivatives, the idea of radiofluorination of an intact nucleoside moiety was applied. The main challenge of this new approach is the rational design of precursors that combine reactivity and stability in a way that the 18F-nucleophile can be incorporated at high temperatures resulting in acceptable conversion rates whilst heat-mediated anhydro formation is kept to a minimum. A quantitative method for the characterization of precursor stabilities was desirable prior to synthesis. The electric charge density of the 2-carbonyl oxygen of the pyrimidone moiety was determined using theoretical methods for the potential density of the 2-carbonyl oxygen of the pyrimidone moiety and the 3-21G basis set.

LowModMD that generates conformations using a short run of molecular dynamics at constant temperature followed by an all-atom energy minimization. Amongst the conformations, 16 of them were included in this study, in order to obtain both a reference value as well as a measurement of how the substituents, the lower the calculated electron density Ω at the 2-carbonyl oxygen. For instance, double N4-Boc-protection (7) decreases the electron density more than single N4-Boc-protection (6) and N3-monooacetylation (5), respectively.

Compounds 5 and 7 were synthesized (see Synthesis section and Supporting information) and subsequently tested for their ability to undergo heat-mediated decomposition by forming the appropriate anhydro compounds 12 and 13. First, an NMR sample of compound 5 in d6-DMSO was heated up to 50°C. 1H-NMR spectra were recorded hourly. Figure 3 shows that almost full conversion to anhydro compound 12 occurred after 2 h (Scheme 1).

Chemical shifts of the amino proton (11.5 to 12.25 ppm) and the ribose protons clearly show the conversion into the cyclic nucleoside compound (Figure 3). Furthermore, the vanishing signal of the mesyl group at 3.5 ppm also indicates decomposition as the free mesylate is expected to have a different chemical shift that indeed could be observed at around 2.3 ppm next to the methyl-signal of the acetamide. 13C-NMR data after 1-h incubation showed 24 instead of 12 signals of pure compound 5. Additionally, ESI-MS was performed in order to show the change in the molecular mass. The m/z signal 268 [M + H] further confirmed generation of anhydro compound 12.

The double N4-Boc-protected compound (7), however, showed significantly increased thermal stability as shown in Scheme 2 and Figure 4. Even though a tiny set of additional

<table>
<thead>
<tr>
<th>No.</th>
<th>R1</th>
<th>R2</th>
<th>(\Omega) ± %</th>
<th>(\Omega) ± %</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>H</td>
<td>H</td>
<td>-0.911 ± 0.09</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>Ac</td>
<td>-0.901 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>Boc</td>
<td>-0.860 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Boc</td>
<td>Boc</td>
<td>-0.825 ± 0.06</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Phthalimide</td>
<td>Phthalimide</td>
<td>-0.814 ± 0.08</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td></td>
<td>-0.838 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Boc</td>
<td></td>
<td>-0.770 ± 0.03</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>NO2</td>
<td></td>
<td>-0.589 ± 0.10</td>
<td></td>
</tr>
</tbody>
</table>

Quantum calculated net charge at the 2-carbonyl oxygen.
peaks appeared after 2 h at 80 °C, the observed stability of compound 7 looked promising with regard to 18F-radiolabelling reactions.

Hence, the calculated and experimental data lead to three major conclusions: (1.) The electron density at the 2-carbonyl oxygen of the pyrimidine moiety can be used as a measurement for its ability to undergo intramolecular cyclization resulting in the appropriate 2,2′-anhydro compounds. (2.) Electron-withdrawing groups have a measurable/quantifiable effect on the electron density at the 2-carbonyl oxygen. (3.) Depending on the strength of its electron-withdrawing nature, a N^4(N^3)-protecting

Figure 3. Heat-mediated anhydro formation of 12. After only 2 h at 50 °C, nearly full conversion could be observed by 1H NMR analysis.

Scheme 2. Increased thermal stability of 7. No significant formation of side product (13) could be observed (in d_4-MeOD).

peaks appeared after 2 h at 80 °C, the observed stability of compound 7 looked promising with regard to 18F-radiolabelling reactions.

Hence, the calculated and experimental data lead to three major conclusions: (1.) The electron density at the 2-carbonyl oxygen of the pyrimidine moiety can be used as a measurement for its ability to undergo intramolecular cyclization resulting in the appropriate 2,2′-anhydro compounds. (2.) Electron-withdrawing groups have a measurable/quantifiable effect on the electron density at the 2-carbonyl oxygen. (3.) Depending on the strength of its electron-withdrawing nature, a N^4(N^3)-protecting

Figure 4. Recorded 1H-NMR spectra of purified compound (7) after 0, 60 and 120 min at 80 °C. The absence of additional peaks indicate no or minimal anhydro formation through increased thermal stability.
group influences the thermal stability of the nucleoside analogue. Finally, an N,N-diBoc fluorination precursor was selected for radiofluorination studies.

Synthesis

The synthesis of test compound 5 was carried out with the selective 3′- and 5′-protection of N-acetylcytidine (14) using TIPDSCl in pyridine yielding compound 15 in good yield, which was consecutively mesylated at the 2′-position using an excess of MsCl and TEA in DCM furnishing compound 16. Subsequent removal of the silica-based protecting group using TBAF in anhydrous THF at low temperatures gave test compound 5 that was purified immediately and characterized because of its high ability to undergo anhydro formation (Scheme 3).

The synthesis of precursor (23) started with commercially available cytidine (17), which was first converted into the 3′- and 5′-protected nucleoside 18 using TIPDSCl and pyridine at room temperature as previously described. Subsequent TMS-protection furnished 19 that was followed by double N4-Boc-protection using excess of Boc₂O (5–6 eq.) and DMAP that gave compound 20 in good yield. Selective removal of the TMS-group using p-TsOH was performed at −5 °C to give intermediate 21. Subsequent mesylation of the free 2′-hydroxy group was performed in high yield to form compound 22. TBAF-mediated removal of the bidentate-protecting group had to be carried out at lower temperatures (approximately −10 °C) in order to avoid significant decrease in yield probably because of single Boc-group removal. Purified compound 7 was then converted into fluorination precursor 23 using 3,4-dihydro-6H-pyran in excess. Highly purified precursor 23 was obtained in 9% overall yield from 17 (Scheme 4).

The choice of the leaving group as well as the protecting group is based on a previous study that found that 3′/5′-THP protection of uridine derivatives gave the best radiochemical yields and a 2′-mesyl group an appropriate balance between stability and leaving group abilities. Furthermore, our pilot studies showed that the appropriate tosyl-containing and nosyl-containing counterparts already decomposed during purification. Furthermore, even though double THP protection makes NMR analysis complex (mixture of diastereomers), the later removal of these groups was straightforward.

Scheme 3. Reagents and conditions: (a) TIPDSCl₂, pyridine, 0 °C – room temperature (RT), 12 h, 87%; (b) MsCl, TEA, DCM, 0 °C – RT, 2 h, 81%; (c) TBAF, THF, −10 °C, 30 min, 53%.

Scheme 4. Reaction conditions: (a) TIPDSCl₂, pyridine, 0 °C – RT, 18 h, 81%; (b) TMSCI, TEA, DCM, 0 °C – RT, 2 h, 83%; (c) Boc₂O, DMAP, DCM, RT, 5 h, 77%; (d) p-TsOH, THF, −5 °C, 66–78%; (e) MsCl, TEA, DCM, 0 °C, 90%; (f) TBAF, THF, −10 °C, 70–80%; and (g) DHP, pTsOH, DCM, 0 °C – RT, 6–8 h, 45–56%.

Scheme 5. Reagents and conditions: (a) 1N TBAF/THF, DMF, 100 °C, 100 min, 19%.

Scheme 6. Radiofluorination was performed using different reaction conditions as outlined in the table later.

1N TBAF/THF, DMF, 100 °C, 100 min, 19%
Firstly, cold fluorination reactions were carried out to assess fluoride incorporation for precursor (23). Amongst fluorination conditions tested, the best incorporation yield was found at 100 °C for 100 min in DMF using freshly prepared 1N TBAF/THF-solution (Scheme 5). The fluorinated intermediate (24) was purified by column chromatography on SiO$_2$ and its identity and purity (≥97%) confirmed by $^{1}$H/$^{19}$F-NMR, HR-MS and analytical HPLC (Supporting information). Notably, the $^{19}$F-spectrum showed a set of four signals due to the four diastereomers that the THP protection yields.

The obtained intermediate 24 was used as a cold reference standard for co-elution on HPLC in order to identify the hot intermediate.

Radiosynthesis of $^{18}$F-FAC

The $^{18}$F-incorporation to form intermediate (25) was carried out using an Eckert & Ziegler modular system and appropriate software for radiosynthesis. The reaction parameters of solvent, temperature and reaction time were varied in order to determine optimal $^{18}$F-incorporation conditions, using KHCO$_3$ as base (Scheme 6). [$^{18}$F]-Fluoride was eluted from the QMA cartridge using a K$_{222}$/KHCO$_3$ solution. Table 2 shows the calculated incorporation data, and Figure 5 shows $^{18}$F-incorporation plotted against the reaction temperature run in dry DMF as solvent. The highest incorporation values were obtained using DMF at 110 °C for 20 min. Although MeCN or tBuOH as solvent failed to show any conversion, the use of an alternative base such as TBA–OH or TBA–CO$_3$ in MeCN or tBuOH might provide an alternative strategy. However, higher temperatures and longer reaction times led to decreased $^{18}$F-incorporation in the arabinom position because of the formation of the anhydro intermediate. $^{18}$F-incorporation was calculated by peak integration of the crude reaction mixture after radiofluorination using radio-TLC (Supporting information).

The $^{18}$F-fluorinated intermediate (25) was confirmed by radio-HPLC (Figure 6) and co-injection of the $^{19}$F-containing reference standard (24) (Supporting information) after trapping the majority of the unreacted $^{18}$F-fluoride on a Sep-Pak alumina cartridge. The intermediate was eluted with 2.5 mL EtOAc into a new reaction vial. The solvent was then removed at 90 °C under a stream of nitrogen. 2N HCl solution was added to the dried residue (25), which was then stirred at 95 °C for 20 min (Scheme 7). A shorter deprotection time was found to lead to decreased final yields of the product $^{18}$F-FAC.

After neutralization with 2N NaOH solution, the mixture was loaded onto a semi-preparative HPLC column via a 10-mL injection loop. [$^{18}$F]-FAC was eluted after 22.5 min at a flow rate of 3.5 mL/min using 3% MeCN in H$_2$O as the mobile phase. Following concentration on the module (100 °C Table 2. [$^{18}$F]-incorporation results obtained under different conditions

<table>
<thead>
<tr>
<th>No.</th>
<th>Solvent</th>
<th>T [°C]</th>
<th>t [min]</th>
<th>[$^{18}$F]$^a$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MeCN</td>
<td>90</td>
<td>20</td>
<td>n/a</td>
</tr>
<tr>
<td>2</td>
<td>tBuOH</td>
<td>95</td>
<td>20</td>
<td>n/a</td>
</tr>
<tr>
<td>3</td>
<td>DMF</td>
<td>90</td>
<td>20</td>
<td>1.3 ± 0.3</td>
</tr>
<tr>
<td>4</td>
<td>DMF</td>
<td>95</td>
<td>20</td>
<td>2.6 ± 0.2</td>
</tr>
<tr>
<td>5</td>
<td>DMF</td>
<td>100</td>
<td>20</td>
<td>4.9 ± 0.6</td>
</tr>
<tr>
<td>6</td>
<td>DMF</td>
<td>105</td>
<td>20</td>
<td>7.7 ± 0.8</td>
</tr>
<tr>
<td>7</td>
<td>DMF</td>
<td>110</td>
<td>20</td>
<td>9.4 ± 0.8</td>
</tr>
<tr>
<td>8</td>
<td>DMF</td>
<td>110</td>
<td>30</td>
<td>7.9 ± 0.6</td>
</tr>
<tr>
<td>9</td>
<td>DMF</td>
<td>115</td>
<td>20</td>
<td>6.9 ± 0.5</td>
</tr>
<tr>
<td>10</td>
<td>DMF</td>
<td>120</td>
<td>20</td>
<td>4.1 ± 0.4</td>
</tr>
<tr>
<td>11</td>
<td>DMF</td>
<td>125</td>
<td>20</td>
<td>2.3 ± 0.2</td>
</tr>
<tr>
<td>12</td>
<td>DMF</td>
<td>130</td>
<td>20</td>
<td>1.8 ± 0.2</td>
</tr>
</tbody>
</table>

$^a$[$^{18}$F]-incorporation was calculated using radio-TLC. Each reaction was repeated three times under identical conditions.

Figure 5. [$^{18}$F]-incorporation plotted against the reaction temperature run in dry DMF as solvent.

Figure 6. Radio-HPLC was performed in order to confirm the formation of 25. The fluoridation product was co-eluted with reference compound 24 after 17.2 min (MeCN/H$_2$O 10:1, flow rate of 1 mL/min). The profile of the radio-HPLC diagram is consistent with the radio-TLC diagram shown earlier.

Scheme 7. Subsequent deprotection of the hot intermediate 25 using (a) 2N HCl at 95 °C for 20 min. A shorter deprotection time results in decreased final yields of the product $^{18}$F-FAC (1).
under a stream of nitrogen for 10 min), the sample was taken up in sterile saline, and an aliquot was co-eluted on an analytical HPLC in combination with both possible \(2',18^F\)-isomers FC (26) and FAC (27) as references (Figures 7 and 8). \(18^F\)-FAC was confirmed, and its purity calculated to \(\geq 98\%\) with specific activity of \(\geq 63 \text{ GBq/\mu mol}\). The total synthesis time was 168 min, including the concentration and saline formulation time. Finally, 0.75–0.86 GBq of \(18^F\)-FAC was obtained, resulting in radiochemical yields of 4.3–5.5\% (\(n=8\), decay-corrected from end of bombardment).

Additionally, a temperature study was performed to reveal why the \(18^F\)-incorporation decreases when the reaction temperature is increased. According to the mechanism, intramolecular anhydro formation is expected at higher temperatures leading to a lower yield of the correct stereoisomer. Figure 9 shows radio-HPLC traces that were recorded after radiofluorination at

![Figure 7](image7.png)

**Figure 7.** Radio-HPLC chromatogram of the purified fraction. Co-elution with both cold \(2',19^F\)-isomers 26 and 27 confirms that the correct stereoisomer was obtained.

![Figure 8](image8.png)

**Figure 8.** Chemical structures of the two different ‘cold’ \(2',19^F\)-cytidines FC (26) and FAC (27).

![Figure 9](image9.png)

**Figure 9.** Radio-HPLC traces recorded after radiofluorinations performed at different reaction temperatures.
110, 120 and 130 °C (20 and 40 min, bottom to top panel respectively) and co-elution with both 2-fluoro stereoisomers FC and FAC (bottom panel). The results show that higher temperatures do lead to anhydro formation and radiofluorination of the anhydro intermediate results in 18F-FC. The 18F-FAC/FC ratio decreases as the reaction temperature and time increase.

Conclusions
A novel radiochemical approach to 18F-FAC is described. For the first time, 18F-FAC was synthesized via late-stage radiofluorination of the intact pyrimidine nucleoside. Prior to synthesis, a novel fluorination precursor was designed using in silico methods and stability studies monitored by NMR. Precursor 23 was found to be a promising candidate for further testing out of the family of nucleosides that were investigated. 18F-incorporation studies led to optimal reaction conditions (110 °C, 20 min) that furnished 18F-FAC in 4.3–5.5% decay-corrected radiochemical yield. Considering that this new process involves late-stage fluorination and labelling of a cytidine derivative (more challenging than the equivalent uridine derivative), our radiochemical yield is good. Our overall synthesis time (168 min) is similar to that reported previously for 18F-FAC.11 Even though early 18F-introduction gives a higher radiochemical yield, its automation is more difficult and the reliability and reproducibility is less easy to maintain. In addition, the radiochemical yield and overall radiosynthesis time of our methods might be further improvable using other synthesizers; alternatively, one could obtain the same amount of labelled nucleoside product by starting with a higher 18F-activity. The reaction time for the radiolabelling step (20 min) is comparable with that used in the related 18F-FMAU radiosynthesis,12 reflective of the similar steric hindrance around the 2′-position in these nucleoside precursors. Shorter reaction times for this radiolabelling step were found to lead to lower yields of product through incomplete reaction.

Furthermore, it was shown that long reaction times and high reaction temperatures led to intramolecular anhydro formation identified by the detection of 18F-FC. This study shows that rational precursor design can at least minimize the issue of side-product formation and hence can offer synthetic access of important PET radiotracer as a single stereoisomer. This method should be transferable to GMP environments as it offers a quick and reproducible two-step route towards 18F-FAC comparable with the radiosynthesis of 18F-FDG. However, further optimization is required in order to improve the radiochemical yield and prepare for transfer to production of clinical grade radiotracer.

Acknowledgements
This work was supported by Cancer Research Wales through PhD studentship funding (to J.-P.M.). We thank Dr Andrea Brancale (Cardiff University, School of Pharmacy and Pharmaceutical Sciences) for the advice and support with DFT calculations. We acknowledge the EPSRC National Mass Spectrometry Facility (Swansea, Wales, UK) for the provision of HR-MS determinations.

Conflict of Interest
The authors did not report any conflict of interest.

References

Supporting Information
Additional supporting information may be found in the online version of this article at the publisher’s web-site.