LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Tubulin-Targeting Antiparasitic Herbicides: New Insights Into Putative Ligand-Binding Pockets and Computational Design of Molecular Equivalents

Presented by

Christopher T. Flood

A thesis submitted to the
University of Dublin, Trinity College
for the degree of
Doctor of Philosophy

School of Biochemistry and Immunology
Trinity College Dublin
Ireland

March 2013

TRINITY COLLEGE
19 JUN 2013
LIBRARY DUBLIN
Thesis 10128

Declaration

I hereby declare that this thesis is entirely my own work except where duly acknowledged and that it has not been submitted as an exercise for the award of a degree at this or any other university. I agree that the library may lend or copy this on request.

"If you wanna make turtle soup, first you gotta find some goddamn turtles"

- H.P. Voorheis

Abstract

Neglected diseases are a group of tropical infections which are especially endemic in low-income populations in developing regions of Africa, Asia, and the Americas. Collectively these diseases affect more more than a billion people worldwide. Malaria affects 300 million people and kills 1-1.5 million people every year. One of the main causes for the comeback of malaria is that the most widely used drug against malaria, chloroquine, has been rendered useless by drug resistance in much of the world. New anti-malarial drugs are presently available but the potential emergence of resistance and their cost make it of utmost importance to continue searching for new drugs. Similarly, Trypanosoma brucei affects the people of 36 sub-Saharan countries with 300,000 new cases reported every year. The few drugs used for the treatment of the disease are either toxic, cause severe side effects or suffer from parasite resistance. Recently, there has been increased focus on two classes of herbicide compounds, the dinitroanilines and the phosphorothioamidates, as potential selective anti-parasitic drugs. These compounds disrupt the microtubules of plants and also disrupt the microtubules of protozoa, including parasites such as Trypanosoma spp., Leishmania spp., Entamoeba spp., Plasmodium falciparum, Cryptosporidium parvum, and Toxoplasma gondii although the binding site of these compounds on the tubulin protein remains unconfirmed. The activity of these compounds appears to be restricted to plants and protozoa; these compounds are ineffective against vertebrate or fungal microtubules.

In the first section of this thesis, Chapters 2 and 3, homology modeling, computational ligand-docking and molecular dynamics simulations were employed to probe the surface of α-tubulin for potential herbicide binding sites. There are as yet no experimentally resolved structures of malarial or trypanosomal tubulin and so homology modeling based on the related mammalian tubulin protein was used to create models for the parasitic tubulins. Subsequent docking, molecular dynamics and MM-PBSA studies allowed the investigation of the interaction of the herbicides with the microtubule α-subunit, identifying a previously undescribed putative herbicide-binding site as the most energetically favourable pocket. In the second section of this thesis, Chapters 4 and 5, a ligand-based virtual screening platform was developed to identify new antiparasitic drugs based on the chemical features of the herbicide compounds. A tiered ligand-based protocol composed of a pharmacophore filter followed by a re-ranking of molecules with ROCS identified 15 compounds to be brought forward for biological characterization using *in vitro* parasite viability assays. Of these compounds, one displayed IC_{50}s of 5.45μM and 4.6μM against P. *falciparum* and T. *brucei* parasites respectively. Five additional compounds were shown to significant ability to inhibit the survival of T. *brucei* parasites in vitro with the most active compound having an IC_{50} of 3nM. None of the compounds identified have been previously described as antiparasitic compounds. More importantly, none of these compounds displayed any activity towards mammalian cell lines and none were shown to bind to mammalian tubulin, providing the basis for the

discovery of new potential parasite-selective treatments for Malaria and Human African Trypanosomiasis.

Acknowledgements

First and foremost I would like to thank my supervisor, Prof David Lloyd, for giving me the amazing opportunity to undertake this PhD. I can't thank you enough for the encouragement and seemingly endless patience you showed me.

To Andy, I would've gone mad without all the advice you've given me and all the wisdom you've imparted (for wisdom read gossip sitting outside BG's). Thank you so much for all the laughs over the years, if even 1% of your attitude rubbed off on me in all this time I'd consider myself a success. Hopefully my nagging and insecurities and more nagging have given you a taste of your future with all those PhD students that are going to come your way.

To Fernando, Spain's best export since David Villa, thank you so much for all the help you've given me over this past year. I really, really appreciate everything you've done to get me here, the encouragement you've given me over the last few months especially has made this possible. Thank you again.

A massive thank you to Giorgio and Martin for all the help early on, the ultimate computational tag-team. Thanks to Darren for keeping everything ticking over in the lab! Thanks also to all the former MDGers, Gemma and Tim especially, who took me under their wings from day one and always made me feel welcome.

A big thanks to Gus, Enda and Finian all the way over in Microbiology, I couldn't have done this without all your help.

To the lads, Daniel 'Dan' Nevin, Trevor 'Trev' Price, Nial 'Neil' Harte and Paul 'Paul' McKay. You often hear people talking about how tough doing a PhD is (didn't notice it m'self) but working with you guys was an absolute pleasure. I laughed every day and it never felt like work (which you'll also notice if you continue to read past this page). A huge thanks to Sandra, not only for showing me how to hold a pipette properly, but also for being a sound person who always wants to help. I'll never forget when you made fun of me for not knowing where the bin is though. And how can I forget the rest of my followers: Rob, Kate, Nick & Sinead, Niki, Clare, Emma, Kim, Jenni and Laura. You guys have been so much fun to work with, I don't know how I managed to end up with such an amazing, funny, intelligent (for the most part), attractive (Rob) group of friends. I've been very, very lucky. You lot have been lucky too, knowing me and all, but whatever.

To my non-science friends, Saar, Joe and Yams, thanks for keeping me sane and getting me drunk when I had no money. Thanks for never asking me about my PhD and thanks

for never talking about science. All the shite-talk (especially when Saar went through his awkward 'determinism' phase) and jokes made everything seem easier. I love you guys and I owe you so much.

Finally, and most of all, my family. To my mother, Jean, and my sisters Natalie and Sarah, thanks for putting up with me. You guys showed me so much support during the hardest times and always showed me how proud you were. I'll never be able to pay you back (especially not financially, no matter what I may have said) for all the patience you guys have shown with me, and I'll never be able to thank you enough. My only wish is that my Da, Tommy, was here to see this. When times were hardest I always thought of you and it helped me to keep going. This is dedicated to you.

Contents

Abstract iii

Acknowledgements v

List of Figures xi

List of Tables xv

Abbreviations xvi

1 Introduction 1

1.1 Neglected Diseases . . . 1
1.2 Malaria . . . 2
1.2.1 Life Cycle of the *Plasmodium falciparum* Parasite . . . 4
1.2.2 Current Chemotherapy . . . 6
1.2.3 Drug Resistance . . . 13
1.3 Human African Trypanosomiasis . . . 14
1.3.1 Life Cycle of *Trypanosoma brucei* . . . 16
1.3.2 Current Chemotherapy . . . 19
1.3.3 Problem of Drug Resistance . . . 23
1.4 Tubulin as an Anti-Parasitic Drug Target . . . 23
1.4.1 Tubulin & Microtubules . . . 24
1.4.2 Tubulin Structure . . . 25
1.4.3 Tubulin-Targeting Compounds . . . 27
1.5 Organophosphorus Herbicides: The Dinitroanilines & Phosprorothioamidates . . . 35
1.5.1 Putative Binding Site . . . 38
1.6 Chemistry and Toxicology of Organophosphorus Compounds . . . 43
1.7 Aims of the current study . . . 44

2 Homology Modeling of the Malarial and Trypanosomal Tubulin Heterodimer 45

2.1 Introduction . . . 45
2.2 Homology Modeling . . . 48
2.2.1 Template Selection . . . 48

2.2.2 Aligning Target and Template Sequences 48
2.2.3 Model Construction with MODELLER 49
2.3 Homology Model Validation with PROCHECK 50
2.3.1 Ramachandran Plot 51
2.3.2 Main-chain Parameters 54
2.3.3 Side-chain Parameters 56
2.4 Previous Modeling of Parasitic Tubulin: "Morrissette" Site Study 58
2.5 α-β Tubulin Heterodimer Homology Modeling Method 60
2.6 Results 63
2.6.1 Homology modeling of the *Plasmodium falciparum* tubulin heterodimer 63
2.6.2 Homology modeling of *Trypanosoma brucei* tubulin heterodimer . 69
2.7 Discussion and Conclusions 76

3 Molecular Dynamic Studies of Parasitic Tubulin Models 79
3.1 Introduction 79
3.2 Molecular Dynamics Simulations 79
3.2.1 Amber Molecular Dynamics Software 81
3.3 Ligand-Binding Site Elucidation 82
3.3.1 Geometry-Based Methods 83
3.3.2 Energy-Based Methods 83
3.4 MDpocket 84
3.5 Docking and Scoring 85
3.5.1 Systematic Search Methods 86
3.5.2 Stochastic Search Methods 86
3.6 Scoring Functions 87
3.6.1 Knowledge-Based Scoring Functions 87
3.6.2 Force Field Scoring Functions 88
3.6.3 Empirical Scoring Functions 88
3.6.4 Limitations of Scoring Functions 89
3.7 Combining Docking and Molecular Dynamics Simulations 90
3.8 Molecular Dynamics and MM-PBSA 90
3.9 Methods 93
3.9.1 Molecular Dynamics Simulaitons 93
3.9.2 Binding Pocket Analysis 94
3.9.3 ROCS shape matching 94
3.9.4 Ligand Conformer Generation 95
3.9.5 Docking 96
3.9.6 Molecular Dynamics Protocol 96
3.9.7 Analysing Molecular Dynamics Trajectories 97
3.9.8 MM-PB/GBSA 98
3.10 Results 100
3.11 Molecular Dynamics Simulations 100
3.12 Analysis of Transient Binding Pockets with MDpocket 102
3.13 Analysis of Amino acid differences across species 103
3.14 Using ROCS to identify potential binding pockets 104
3.15 Induced fit docking 111

3.16 Molecular Dynamics Study of potential binding pockets 111
3.16.1 Analysis of binding site 1 in parasitic tubulin 112
3.16.2 Analysis of binding site 4 in parasitic tubulin 120
3.16.3 Analysis of The Proposed Morrissette site in parasitic tubulin . . . 128
3.16.4 Analysis of the mammalian equivalent of binding site 1 128
3.17 Discussion and Future Work 131
3.18 Appendix A - Post-Molecular Dynamics Binding Pose Descriptions 135
3.18.1 Analysis of binding site 2 in parasitic tubulin 135
3.18.2 Analysis of binding site 3 in parasitic tubulin 142
3.18.3 Analysis of binding site 5 in parasitic tubulin 149
3.18.4 Analysis of binding site 6 in parasitic tubulin 156

4 Molecular Equivalents to Established Herbicides 163
4.1 Introduction 163
4.2 Virtual Screening and The Drug Discovery Process 164
4.3 Medicinal Chemistry Space 167
4.3.1 Description of Principal Component Analysis 168
4.4 The Concept of Molecular Similarity 170
4.5 Scaffold Hopping 171
4.6 Introduction to Pharmacophore Design 171
4.6.1 Generating a Pharmacophore Hypothesis 172
4.7 Pharmacophore Generation in MOE 178
4.8 Rapid Overlay of Chemical Structures 179
4.9 Successful Ligand-Based Virtual Screening Campaigns Against Parasitic Diseases 181
4.10 Methods 182
4.10.1 Mapping Anti-parasitic Medicinal Chemistry Space 182
4.10.2 Pharmacophore Design 184
4.10.3 ROCS Query Generation 187
4.10.4 Preparation of Validation Database 187
4.10.5 Validation of Ligand-Based Approaches 187
4.11 Results and Discussion 189
4.12 Navigating Parasitic Chemotherapy Space 189
4.13 Pharmacophore Design and Validation 191
4.14 ROCS Query Generation and Validation 193
4.15 Vendor Database Screening 193
4.16 PCA Analysis of Compounds Identified Through VS 198
4.17 Conclusion 200

5 Biological Characterisation of Computationally Designed Compounds 203
5.1 Introduction 203
5.2 Methods 205
5.2.1 Chemicals & Reagents 205
5.2.2 Culture of *Plasmodium falciparum* 205
5.2.3 *Plasmodium falciparum* viability assay 205
5.2.4 Culture of *Trypanosoma brucei* 206
5.2.5 *Trypansoma brucei* Cell Viability Assay 206

5.2.6 HeLa cell culture 207
5.2.7 HEK293 cell culture 207
5.2.8 Cell Cryo-storage 207
5.2.9 Alamar Blue cell viability assay 208
5.2.10 Thermal melt/shift assay 209
5.3 Results 210
5.4 Parasitic Viability Assays 210
5.4.1 *Plasmodium falciparum* Cell Viability Assay 210
5.4.2 *Trypanosoma brucei Cell Viability Assay* 212
5.4.3 Thermal Melt Assay 218
5.5 Discussion and Conclusion 223

6 General Discussion **227**

Bibliography **231**

List of Figures

1.1 The imbalance of drug discovery for neglected diseases. 2
1.2 Malaria endemic countries. 3
1.3 Estimated percentage of malaria cases due to *P. falciparum* 4
1.4 The Plasmodium life cycle 6
1.5 Commonly used natimalarial drugs. 11
1.6 Distribution of human African Trypanosomiasis (*T. b. gambiense* 15
1.7 Distribution of human African Trypanosomiasis (*T. b. rhodesience* 16
1.8 Life cycle of *Trypanosoma brucei* 18
1.9 Commonly used *T. brucei* drugs. 22
1.10 Microtubule Structure 25
1.11 Dynamic Instability in Microtubules 26
1.12 Tubulin Dimer Ultrastructure 28
1.13 Location of the colchicine and vinblastine binding sites on the tubulin protein 29
1.14 Tubulin destabiising compounds. 31
1.15 THe taxl binding site on tubulin 33
1.16 Tubulin stabilising compounds. 34
1.17 Structures of the tubulin depolymerising herbicide compounds Trifluralin, Chloralin, Oryzalin, Amiprophosmethyl and Butamifos. 36
1.18 The electron density of trifluralin and butamifos 37
1.19 The proposed Morrissette site of herbide binding to tubulin 41
1.20 Mislabeling of the tubulin protein with respect to the Morrissette binding site 42

2.1 Typical stages of the homology modeling process. 47
2.2 The homology modeling process employed by MODELLER 51
2.3 Example of a Ramachandran Plot 53
2.4 Example of Main-chain Parameters output from PROCHECK 55
2.5 The gauche(-), (+) and trans conformations of amino acid side-chains . . 57
2.6 Amino acid sequence alignment of *P. falciparum* alpha tubulin with the alpha tubulin sequences from bovine and pig sources. 65
2.7 Ramachandran plot for the homology model of *P. falciparum* tubulin. . . 66
2.8 Main chain parameter analysis for the homology model of *P. falciparum* tubulinl. 67
2.9 Side chain analysis for the homology model of *P. falciparum* tubulin. . . . 68
2.10 Alignment of the *P. falciparum* homology model (blue) and the template structure 1TUB (green). 69

2.11 Amino acid sequence alignment of *T. brucei* alpha tubulin with the alpha tubulin sequences from bovine and pig sources. 71
2.12 Ramachandran plot for the homology model of *T. brucei* tubulin. 72
2.13 Main chain analysis for the homology model of *T. brucei* tubulin. 73
2.14 Side chain analysis for the homology model of *T. brucei* tubulin. 74
2.15 Alignment of the *T. brucei* homology model (blue) and the template structure 1TUB (red). 75
2.16 Orientation of H1-B2 loop in *P. falciparum* Homology Model 78
3.1 An example of an equation used to approximate the atomic forces that govern molecular movement 80
3.2 Amber program suite workflow 81
3.3 MM-GB(PBSA) thermodynamic cycle 91
3.4 RMSD of backbone carbon atoms of malarial tubulin across the molecular dynamics simulation. 101
3.5 RMSD of the radius of gyration of malarial tubulin across the molecular dynamics simulation of 20ns. 101
3.6 RMSD of backbone carbon atoms of trypanosomal tubulin across the molecular dynamics simulation of 20ns. 102
3.7 RMSD of the radius of gyration of trypanosomal tubulin across the molecular dynamics simulation. 102
3.8 The positions of the amino acids on α-tubulin altered between herbicide susceptible species and herbicide resistant species (front view). 104
3.9 The positions of the amino acids on α-tubulin altered between herbicide susceptible species and herbicide resistant species (back view). 105
3.10 Illustration of the relative positions of putative binding sites 1-6 on the tubulin α-subunit 108
3.11 Close-up view of the loop region of α-tubulin comprising putative binding sites 1, 4, and the proposed 'Morrissette' site. 109
3.12 Results of ROCS shape-based searches on potential herbicide binding sites on the tubulin protein. 110
3.13 Post-molecular dynamics poses of APM and OZN in site 1 of Malarial and Trypanosomal Tubulin homology models 114
3.14 Post-molecular dynamics poses of APM and OZN in site 1 of Malarial and Trypanosomal Tubulin homology models 115
3.15 Ligand interaction diagram of APM in site 1 of malarial tubulin 116
3.16 Ligand interaction diagram of OZN in site 1 of malarial tubulin 117
3.17 Ligand interaction diagram of APM in site 1 of trypanosomal tubulin . . 118
3.18 Ligand interaction diagram of OZN in site 1 of trypanosomal tubulin . . . 119
3.19 Post-molecular dynamics poses of APM and OZN in site 4 of Malarial and Trypanosomal Tubulin homology models 122
3.20 Post-molecular dynamics poses of APM and OZN in site 4 of Malarial and Trypanosomal Tubulin homology models 123
3.21 Ligand interaction diagram of APM in site 4 of malarial tubulin 124
3.22 Ligand interaction diagram of OZN in site 4 of trypanosomal tubulin . . . 125
3.23 Ligand interaction diagram of OZN in site 4 of trypanosomal tubulin . . . 126
3.24 Ligand interaction diagram of OZN in site 4 of trypanosomal tubulin . . . 127

3.25 Graphs of binding MM-GBSA (blue line) and MM-PBSA (red line) free energies of APM and OZN in the malarial and trypanosomal tubulin homology models for the proposed "Morrissette" site. 129
3.26 Post-molecular dynamics poses of APM and OZN in the site 1 region of mammalian tubulin. 130
3.27 Post-molecular dynamics poses of APM and OZN in site 2 of Malarial and Trypanosomal Tubulin homology models 136
3.28 Post-molecular dynamics poses of APM and OZN in site 2 of Malarial and Trypanosomal Tubulin homology models 137
3.29 Ligand interaction diagram of APM in site 2 of malarial tubulin 138
3.30 Ligand interaction diagram of OZN in site 2 of malarial tubulin 139
3.31 Ligand interaction diagram of APM in site 2 of trypanosomal tubulin . . 140
3.32 Ligand interaction diagram of OZN in site 2 of trypanosomal tubulin . . . 141
3.33 Post-molecular dynamics poses of APM and OZN in site 3 of Malarial and Trypanosomal Tubulin homology models 143
3.34 Post-molecular dynamics poses of APM and OZN in site 3 of Malarial and Trypanosomal Tubulin homology models 144
3.35 Ligand interaction diagram of APM in site 3 of malarial tubulin 145
3.36 Ligand interaction diagram of OZN in site 3 of malarial tubulin 146
3.37 Ligand interaction diagram of APM in site 3 of trypanosomal tubulin . . 147
3.38 Ligand interaction diagram of OZN in site 3 of trypanosomal tubulin . . . 148
3.39 Post-molecular dynamics poses of APM and OZN in site 5 of Malarial and Trypanosomal Tubulin homology models 150
3.40 Post-molecular dynamics poses of APM and OZN in site 5 of Malarial and Trypanosomal Tubulin homology models 151
3.41 Ligand interaction diagram of APM in site 5 of malarial tubulin 152
3.42 Ligand interaction diagram of OZN in site 5 of malarial tubulin 153
3.43 Ligand interaction diagram of APM in site 5 of trypanosomal tubulin . . 154
3.44 Ligand interaction diagram of OZN in site 5 of trypanosomal tubulin . . . 155
3.45 Post-molecular dynamics poses of APM and OZN in site 6 of Malarial and Trypanosomal Tubulin homology models 157
3.46 Post-molecular dynamics poses of APM and OZN in site 6 of Malarial and Trypanosomal Tubulin homology models 158
3.47 Ligand interaction diagram of APM in site 6 of malarial tubulin 159
3.48 Ligand interaction diagram of OZN in site 6 of malarial tubulin 160
3.49 Ligand interaction diagram of APM in site 6 of trypanosomal tubulin . . 161
3.50 Ligand interaction diagram of OZN in site 6 of trypanosomal tubulin . . . 162

4.1 Hierarchical Virtual Screening 165
4.2 Virtual Screening Overview 167
4.3 Schematic of the Pharmacophore Design Process. 173
4.4 The pharmacophore design process. 175
4.5 The basis of the metrics used in virtual screening protocols. 176
4.6 Illustration of theoretical ROC curves. 177
4.7 Illustration of the ROCS shape-matching search. 180
4.8 Pharmacophore Training Set 185
4.9 Pharmacophore Test Set 186
4.10 Pipeline Pilot Protocol for Validation Database Preparation 187

4.11 PCA of anti-trypanosomal and anti-malarial medicinal chemistry space . . 189
4.12 PCA of anti-trypanosomal, anti-malarial and "druglike" medicinal chemistry space 190
4.13 Pharnacophore 3 192
4.14 ROC curve calculated after searching the decoy database with pharmacophore 3. 192
4.15 ROC curve calculated ater searching the decoy database with pharmacophore 3 and then re-ranking the pharmacophore output with ROCS searching. 193
4.16 Pipeline Pilot Protocol for Vendor Database Preparation 194
4.17 Compounds ordered from SPECS vendor database. 195
4.18 Compounds ordered from SPECS vendor database aligned with pharmacophore 3. 197
4.19 Compounds ordered from SPECS vendor database aligned with pharmacophore 3. 198
4.20 Chemistry Space of the APM herbicides (blue) versus the 15 compounds ordered from SPECS (red). 199
4.21 Chemistry Space of the APM herbicides (blue) versus the 15 compounds ordered from SPECS (red) and drug-like compounds in yellow. 199

5.1 24 h mean % survival data for the 15 compounds identified through virtual screening tested at 10μM versus bloodstream form of *T brucei*. 213
5.2 24h dose-response curves of compounds 5, 8, 10, 11, and 14 against bloodstream form of *T brucei*. 214
5.3 Analysis of the effects of compounds 1-15 on Human Embyonic Kidney Cell line. 216
5.4 Analysis of the effects of compounds 1-15 on HeLa Cell line. 217
5.5 Thermal melt profiles of bovine brain tubulin (1μM) in the presence of a range of concentrations of vinblastine (**A** and **B**), colchicine (**C** and **D**), vinblastine plus colchicine (**E** and **F**) and APM (**G** and **H**). Bovine brain tubulin in the absence of any compound is indicated in the black curves. First derivative plots of the raw data curves are on the right 220
5.6 (A) Thermal melt profiles of bovine brain tubulin (1 μM) in the presence of a range of 100 μM of compounds 1,4,5,6,7, and 8 or in the absence of any compound (black line). (B) First derivative plots of the raw data curves in (A). The Tm is defined as the temperature where the first derivative is maximal. The Tm values were: C1 not defined, C4 54.6 oC, C5 55.1 oC, C6 54.9 oC, C7 54.9 oC, C8 54.6 oC and bovine brain tubulin 54.9 oC 222
5.7 (A) Thermal melt profiles of bovine brain tubulin (1 μM) in the presence of 100 μM of compounds (C) 9, 10, 12, 13, 14 and 15 or in the absence of any compound (black line). (B) First derivative plots of the raw data curves in (A). The Tm values were: C9 54.7 oC, C10 61.7 oC, C12 55.0 oC, C13 54.9 oC, C14 54.8 oC, C15 54.8 oC and bovine brain tubulin 54.8 oC. 222
5.8 Molecular Surfaces of APM and Compound 11 225
5.9 Biological Activities and SAR of Compounds 1, 12 and 13 226

List of Tables

1.1 Table of some important neglected diseases 1
1.2 Summary of currently used drugs 12
1.3 Summary of potential new drug targets 13
1.4 Relative activities of anti-tubulin compounds between *Plasmodium falciparum and mammalian tubulin* 36

2.1 Tubulin Crystal Structures available from the Brookhaven Protein Database. 63

3.1 Table of Amino Acid differences between resistant (mammalian) and susceptible (*P. falciparum* and *T brucei*) alpha-tubulins. NH = Nonpolar Hydrophobic, PU = *Polar Uncharged*, Neg = *Negatively Charged* 106
3.2 Summary of Binding Sites used for ROCS shape-based searching. Residues in Italics are those that are present in herbicide-susceptible species and not present in herbicide-resistant species. 107
3.3 Summary of MM-GBSA binding energies for putative parasitic herbicide-binding sites 1-6 112

4.1 Lipinski's 'Rule of 5' descriptors 182
4.2 Descriptors used for PCA analysis 183
4.3 Descriptors used to filter the SPECS database prior to virtual screening . 194

5.1 Mean 50% inhibitory concentration data for the novel scaffolds identified through ligand-based virtual screening. The assay measures the lactate dehydrogenase activity of asynchronous *P. falciparum* 3D7 strain parasites as a measure for parasite growth and survival. The data presented is for 72-h cultures. 211
5.2 The biological effects of the most interesting compounds identified through ligand-based virtual screening. Included are the 72h effects of compounds on *P. falciparum*, the 24h effects on *T. brucei*, as well as 24h effects of compounds on HeLa and HEK293 mammalian cell lines (- no effect) . . . 215

Abbreviations

1D	**O**ne **D**imensional
2D	**T**wo **D**imensional
3D	**T**hree **D**imensional
ACHE	**A**cetyl**CH**olin**E**sterase
ADMET	**A**bsorption **D**istribution **M**etabolism **E**excretion **T**oxicity
AMBER	**A**ssisted **M**odel **B**uilding with **E**nergy **R**efinement
APM	**A**mi**P**rophos **M**ethyl
AQP2	**AQ**uaglycero**P**orin **2**
AUC	**A**rea **U**nder the **C**urve **2**
BBB	**B**lood **B**rain **B**arrier
BLAST	**B**asic **L**ocal **A**lignment **S**earch **T**ool
CADD	**C**omputer-**A**ided **D**rug **D**esign
CHARMM	**C**hemistry at **HAR**vard **M**olecular **M**echanics
CQ	**C**hloro**Q**uine
CSP	**C**onstraint **S**pace **P**rogramming
DFMO	**D**i**F**luoro**M**ethyl **O**rnithine
DNA	**D**i**N**itro**A**niline
DOPE	**D**iscrete **O**ptimised **P**rotein **E**nergy
FN	**F**alse **N**egative
FP	**F**alse **P**ositive
G3PDH	**G**lyceraldehyde **3**-**P**hosphate **DeH**ydrogenase
GA	**G**enetic **A**lgorithm
GAFF	**G**eneral **A**mber **F**orce **F**ield
GDP	**G**uanosine **DiP**hosphate
GTP	**G**uanosine **TriP**hosphate

HAT	**H**uman **A**frican **T**rypanosomiasis
HAPT	**H**igh **A**ffinity **P**entamidine **T**ransporter
HTS	**H**igh-**T**hroughput **S**creening
LBVS	**L**igand-**B**ased **V**irtual **S**creening
LO	**L**ead **O**ptimisation
MAP	**M**icrotubule **A**ssociated **P**rotein
MBP	**M**altose **B**inding **P**rotein
MD	**M**olecular **D**ynamics
MOE	**M**olecular **O**perating **E**nvironment
MM-GBSA	**M**olecular **M**echanics-**G**eneralized **B**orn **S**urface **A**rea
MM-PBSA	**M**olecular **M**echanics-**P**oisson **B**olzmann **S**urface **A**rea
MMP	**M**itochondrial **M**embrane **P**otential
MPXR	**M**elarsoprol/**P**entamidine cross **R**esistance
MRPA	**M**ultidrug **R**esistance **P**rotein **A**
NMR	**N**uclear **M**agnetic **R**esonance
ODC	**O**rnithine **D**e**C**arboxylase
OZN	**O**ry**Z**ali**N**
PCA	**P**rincipal **C**omponent **A**nalysis
PDB	**P**rotein **D**ata**B**ase
PFCRT	**P**lasmodium **F**alciparum **C**hloroquine **R**esistance **T**ransporter
PFMDR1	**P**lasmodium **F**alciparum **M**ulti**D**rug-**R**esistant gene **1**
PLDH	**P**arasite **L**actate **D**e**H**ydrogenase
PMF	**P**otential of **M**ean **F**orce
PPP	**P**otential **P**harmacophore **P**oint
PSA	**P**olar **S**urface **A**rea
QSAR	**Q**uantitative **S**tructure-**A**ctivity **R**elationship
RBC	**R**ed **B**lood **C**ell
RMSD	**R**oot **M**ean **S**quare **D**eviation
RMSF	**R**oot **M**ean **S**quare **F**luctuation
ROC	**R**eceiver **O**perating **C**urve
ROCS	**R**apid **O**verlay of **C**hemical **S**tructures
SANDER	**S**imulated **A**nnealing with **N**MR-**D**erived **E**nergy **R**estraints
SAR	**S**tructure-**A**ctivity **R**elationship

SBVS	**S**tructure-**B**ased **V**irtual **S**creening
SIF	**S**tumpy **I**nduction **F**actor
Tbg	**T**rypanasoma **B**rucei **G**ambiense
Tbr	**T**rypanasoma **B**rucei **R**hodesiense
TN	**T**rue **N**egative
TP	**T**rue **P**ositive **B**arrier
TPSA	**T**opological **P**olar **S**urface **A**rea
VSG	**V**ariable **S**urface **G**lycoprotein

Dedicated to my Ma and to the memory of my Da

Chapter 1
Introduction

1.1 Neglected Diseases

Neglected Tropical Diseases (NTDs) are a group of infectious diseases widespread in developing countries. Collectively, these diseases affect more than a billion people worldwide. These diseases are endemic in low-income populations and are also referred to as 'poverty promoting chronic infectious disease' as they thrive in and deepen already existing poverty. The World Health Organisation (WHO) has defined 17 of these conditions as core NTDs. The list includes African Trypanosomiasis, Malaria, Leishmaniasis and Schistosomiasis (Table 1.1).

TABLE 1.1: Table of some important neglected diseases

Disease	Organism	Scope
Malaria	*Plasmodium spp.*	500 million infections annually
Leishmaniasis	*Leishmania spp.*	2 million infections annually[1]
Trypanosomiasis	*T. brucei, T. cruzei.*	300,000 cases annually[2]
Schistosomiasis	*Schistosoma spp.*	more than 200 million existing infections[3]
Leprosy	*Mycobacterium leprae*	roughly 2 million infected[4]
Hookworm	*Ancylostoma duodenale* and *Necator americanus*	740 million infected[5]
Trachoma	*Chlamydia trachomatis*	90 million new cases/year[6]

As can be seen in Figure 1.1, drugs against neglected diseases represent an underwhelming proportion of the total drugs market. There exists a lack of effective, safe and affordable pharmaceuticals for neglected diseases and this represents a massive failure in pharmaceutical research. Between the years of 1975 and 1999, 1393 new chemical entities were marketed and only 16 of these were for the treatment of tropical diseases (Figure 1.1)[7]. This data suggests there is a 13-fold greater chance of a drug being brought to market for central nervous system (CNS) disorders or cancer than for a neglected disease. This massive problem has been brought on by the prevailing market-based intellectual property incentive model of global pharmaceutical innovation which focusses on health priorities of industrialised countries. In order to change this trend it will be necessary to develop new strategies in the development of drugs for NTDs, including new private sector obligations, public sector not-for-profit research, or a change to the current intellectual property-based incentive mechanism[8].

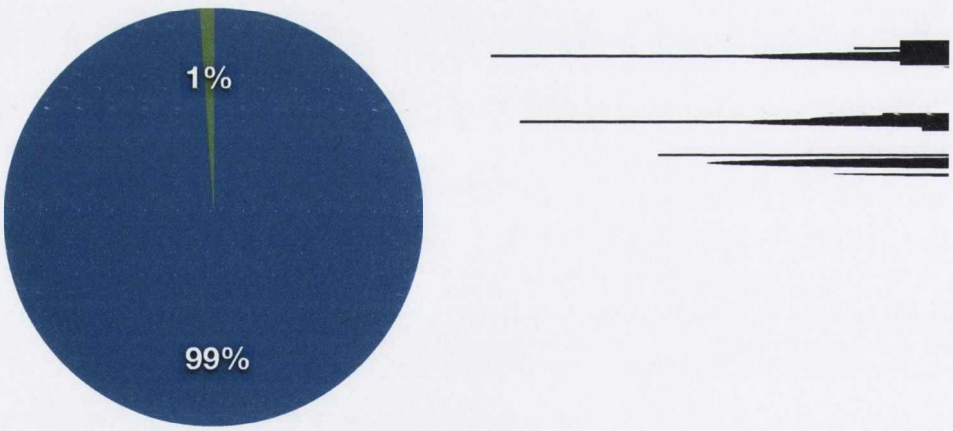


FIGURE 1.1: Pie chart illustrating the imbalance in drug production for neglected diseases.[7]

1.2 Malaria

Malaria is the most significant vector-borne disease in the world today. Despite more than a century of control efforts there are still approximately 3.2 billion people living in

107 countries at risk of infection as of 2004(Figures 1.2 and 1.3). The disease itself represents a serious burden to the socio-economic development of the countries affected[9,10]. There are approximately 350-500 million people infected annually worldwide, with sub-Saharan Africa containing 60% of the cases and 80% of the deaths, most being children under the age of five.

The malaria parasite belongs to the phylum Apicomplexa which comprises *Plasmodia*, *Cryptosporidium sarcocystis*, and *Toxoplasma gondii*. There are five species of malarial parasites known to cause disease in humans: *P. falciparum*, *P.vivax*, *P. knowlesi*, *P. malariae* and *P. ovale*. The natural vectors of the malarial parasites are female mosquitoes of the Anpholes genus, of which roughly 60 are responsible for the trasnsmission of malaria[11].

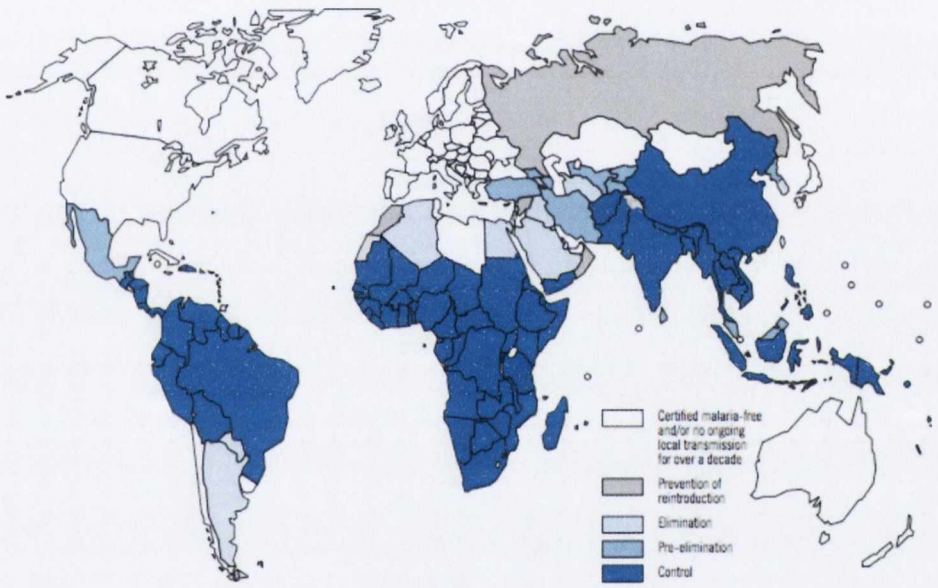


FIGURE 1.2: Malaria-free countries and malaria-endemic countries in phases of control, pre-elimination, elimination, and prevention of reintroduction, end 2007.[12]

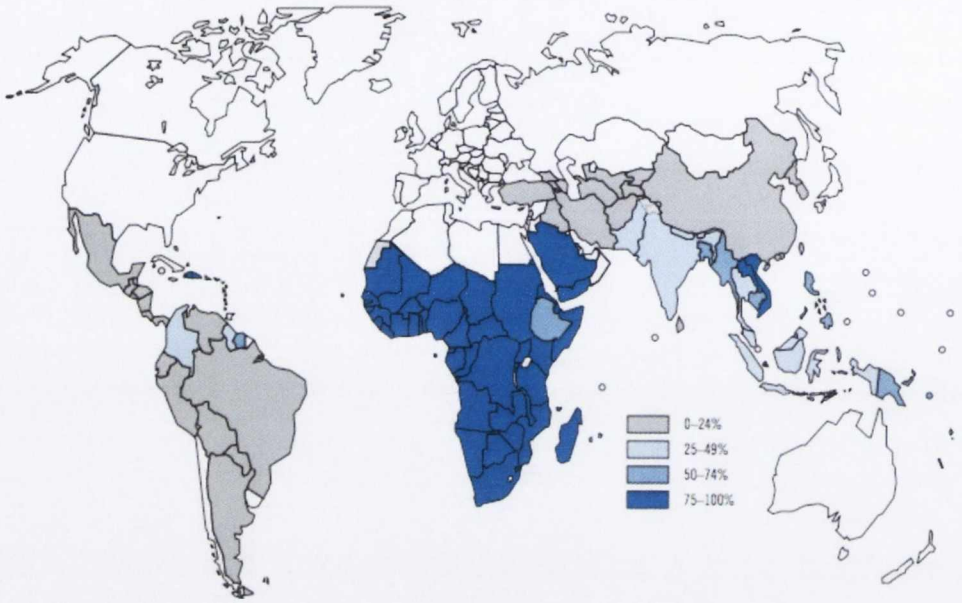


FIGURE 1.3: Estimated percentage of malaria cases due to *P. falciparum*[12]

1.2.1 Life Cycle of the *Plasmodium falciparum* Parasite

Malarial parasites have a complex bi-phasic life-cycle with an extrinsic sexual phase occuring in Anopholes and an intrinsic asexual phase taking place in humans[13]. The asexual stage can be further divided into a pre-erythrocytic schizogony stage existing in the parenchyma of the liver and an erythrocytic schizogony stage where the mature merozoites invade red blood cells (RBCs)[14–16].

Infection in humans begins with the bite of an infected female Anopheline mosquito (Figure 1.4). Sporozoites from the salivary gland of the mosquito enter the blood-stream then travel to the host liver sinusoids before invading the hepatocytes. The sporozoites burrow through several hepatocytes before invading a final one and beginning asexual reproduction. The exo-erythrocytic reproduction produces thousands of erythrocytic-infective merozoites inside the hepatocyte with each one capable of invading RBCs upon release from the liver[17]. These merozoites form the smallest cells in the parasitic life-cycle and are covered in a thick, bristly coat which is involved in adherence to erythrocytes[18]. Upon adherence the RBC surface is deformed, after which the parasite induces the construction of a vesicle inside the erythrocyte called a parasitophorous vacuole (PV), marking the beginning of intraerythrocytic cycle epitomised

by multiple rounds of asexual replication[19,20]. This cycle spans 48 hours in *P. falciparum* and is the main source of the clinical manifestations of the disease, causing both fever and anaemia. There are four well-defined intraerythrocytic parasite stages that can be seen: rings (0-16h), early trophozoites (16h-28h), late trophozoites (28h-36h) and schizonts (36h-48h)[21,22]. Within the RBCs the parasites grow by digesting the cell interior, including haemoglobin. The trophozoites undergo mitosis without cytokinesis, developing into schizonts. Each mature schizont contains about 20 merozoites which escape upon the degradation of the erythrocytic and PV membranes, going on to invade more RBCs[23]. The entire erythrocytic asexual cycle takes about two to three days. After several cycles some of the merozoites differentiate into the sexual forms, macro- and microgametocytes, which travel to peripheral capillaries just beneath the surface of the skin where they are taken up in the blood meal of feeding mosquitos, beginning the invertebrate phase of development and restarting the process[24].

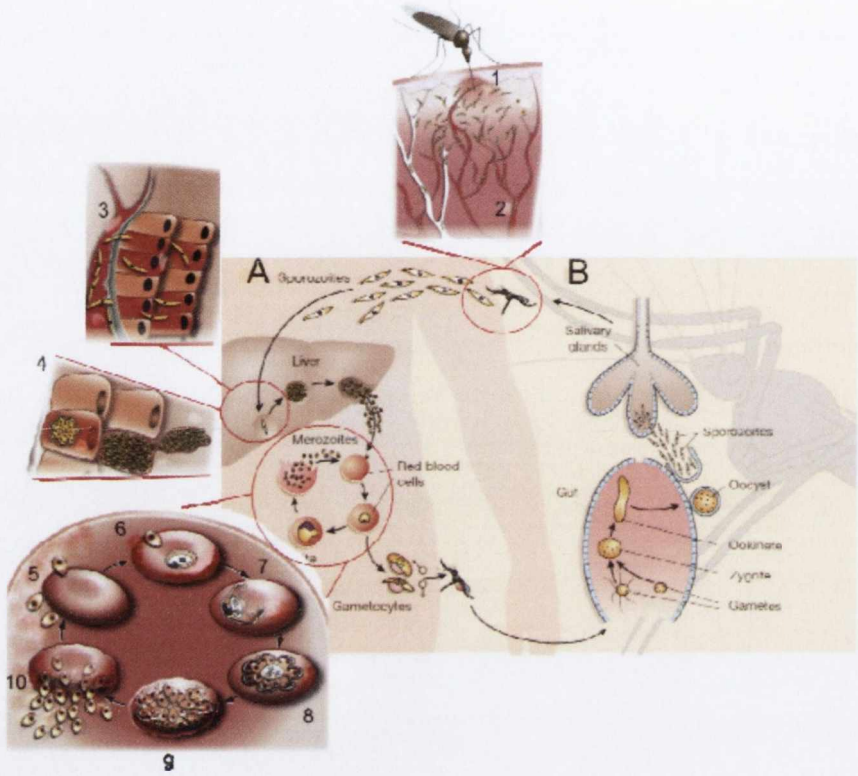


FIGURE 1.4: The Plasmodium life cycle. ***A***: The asexual life cycle in the human host. ***B***: The sexual life cycle in the mosquito. ***1***: Intradermal sporozoite injection when the female mosquito takes a blood meal. ***2***: Sporozoites migrate to the blood vessels to be distributed through the blood circulation. ***3***: Sporozoites invade the hepatocytes in the liver. ***4***: Parasites mature and multiply in the liver to ultimately release merozoites. ***5***: Start of the intraerythrocytic development cycle by the invasion of an erythrocyte by a merozoite. ***6***: Ring stage. ***7***: Trophozoite stage. ***8***: Schizont stage. ***9***: Preparation to release merozoites from the erythrocyte. ***10***: Released merozoites will then re-invade erythrocytes to start intraerythrocytic development cycle once more. *Adapted from Wirth 2002* [25]

1.2.2 Current Chemotherapy

The rapid evolution of malarial parasites has long been associated with the high death rate it creates. The parasites have evolved a strong resistance to the majority of conventional anti-malarial therapeutics available, as well as developing methods of avoiding the immunological responses of humans [26]. There are four major classes of anti-malarial drug

therapies: the blood schizonticides, tissue schizonticides, gametocytocides and sporontocides. Blood schizonticide therapies such as chloroquine (Figure 1.5) and quinine act on the blood-stream forms of the parasite as seen in the erythrocytic stage. Tissue schizonticides are used to fight the infection when the parasite begins attacking the liver. These drugs act on the primary tissue forms of *Plasmodia* within the liver, woth primaquine being the typical tissue schizonticide drug. Gametocytocide drugs act on the sexual forms of the parasite in the blood, the gametocytes, preventing transmission of the infection to the mosquito. Artemisinin, chloroquine and quinine display gametocytocidal activity against *P. vivax* and *P. malariae* and primaquine has gametocytocidal activity against all *Plasmodia*. The final class of anti-malarials, the sporontocides, have no direct effect on the gametocytes in the human host but instead interfere with the reproduction of the zygote within the mosquito, preventing infection of subsequent hosts.

Quinine

Quinine is a natural compound that has been used as an anti-malarial treatment for almost three centuries. The compound contains two major fused ring systems; the aromatic quinoline and the bicyclic quinuclidine. The accepted mode of action of quinine is based on the inhibition of hemozoin biocrystallisation. This inhibition facilitates the aggregation of cytotoxic heme which accumulates in the parasite leading to their death[27]. Due to the side effects associated with its use quinine is now only used to treat chloroquine-resistant *P. falciparum*.

Chloroquine

Chloroquine (CQ) is a 4-aminoquinoline class of drug that is used for the preventive treatment of malaria. It was first synthesised in 1934 by Andersag and colleagues at Bayer Laboratories as a cheaper alternative to quinine and has been the most successful single drug for treatment of malaria[28]. Despite the emergence in the 1960s of malarial strains resistant to CQ it is still considered the drug of choice for malarial treatment in humans because of its low cost and effectiveness. CQ is used in many disease states but its mechanism of action in the prevention of malaria is still a matter of debate. The common consensus is that CQ interferes with the ability of the parasite to digest

haemoglobin by increasing the intravesicular pH in parasitised cells, hindering nucleoprotein synthesis[29–31]. Another hypothesis is that it exerts its effect, at least in part, by reducing the activity of lysosomal enzyme arylsulphatase B, a protein involved in removal of 4-sulphate groups from chondroitin-4-sulphate, thus increasing the attachment of the *P. falciparum*-infected erythrocytes to endothelial cells of the vasculature[32].

Mefloquine

Mefloquine is a 4-quinoline methanol derivative that was initially synthesised among nearly 300 quinoline methanol compounds during World War II. It was first put to therapeutic use in 1985 where it was marketed as Lariam®. Mefloquine possesses high activity when tested against CQ-resistant strains of *P. falciparum*. Its use has been limited recently over the fear of the emergence of strains of *P. falciparum* displaying mefloquine resistance due to increased copy numbers of multidrug-resistant gene 1 (*pfmdr1*)[33,34]. In some parts of Thailand 60-70% of parasitic infections are mefloquine-resistant. Another drawback of mefloquine use is its association with serious adverse neuropsychiatric side-effects, thought to be caused in part by its long half-life. A BBC Watchdog television programme in the 1990s caused intensified public apprehension with the drug after they discovered that it caused more serious effects than had been reported by medical information available at the time. Side-effects reported included sleep disruption, panic attacks and depression[35–37]. The precise mechanism of action of mefloquine is also unclear, although it is postulated that it acts on a membrane target, inhibiting the release of Ca^{2+} ions and preventing the fusion of haemoglobin-shuttling vesicles with the digestive vacuole.

Halofantrine

Halofantrine is a phenanthrenemethanol anti-malarial that was, similar to mefloquine, developed during World War II by the Walter Reed Army Institute of Research as an alternative treatment for CQ-resistant *P. falciparum* and introduced clinically in 1988[38,39]. It is associated with a high-risk of cardiac arrhythmias (a prolongation of QT-interval caused by an inhibition of the inward K^+ current)[40]. The first case of halofantrine-associated death was reported in 1993[41], and since then 35 cases of fatal cardiotoxicity have been recorded, including the deaths of five children[42].

Lumefantrine

Lumefantrine (also known as Benflumetol) is a synthetic molecule structurally similar to halofantrine that was developed in the 1970s by the Academy of Military Sciences in China. It has been shown to be an effective anti-malarial agent when tested against CQ-resistant parasites and has displayed *in vitro* synergism with artemether[43,44], a combination that is being promoted by the WHO for treating uncomplicated malaria and which has shown high efficacy in Phase III clinical trials conducted in several countries[45]. The rationale behind this synergistic relationship is that artemether rapidly reduces parasitemia while lumafantrine eliminates residual parasites[46].

Primaquine

Primaquine, an 8-aminoquinaline derivative, is an analogue of one of the first synthetic anti-malarial agents developed by Bayer in the 1920s[47]. Its precursor, pamaquine, wasn't introduced clinically due to toxicity issues. However, by replacing the terminal diethylamino moiety of pamaquine with an unsubstituted primary amine, as is seen in primaquine, toxicity was greatly reduced. Primaquine sets itself apart from other antimalarials in that it is extremely effective against hypnozoites, a intrahepatic form of the malarial parasite, as well as being effective against sexual blood stages of different Plasmodia parasites[48]. The mechanism of action of primaquine has as yet no clear, solid understanding. It has been suggested that the active form of primaquine is a quinone metabolite which would lead to inhibition of the electron transport in the respiratory chain and could also contribute to oxidative stress[47], an effect seen in the primaquine analogue tafenoquine which targets Leishmania mitochondria, leading to an apoptosis-like death process[49].

One drawback of primaquine use is its relatively short half life of 4-6h, meaning it requires daily administration for 14 days to work. Tafenoquine is a long-acting analogue of primaquine and has a half-life of 2-3 weeks while displaying reduced gastrointestinal disturbances compared to primaquine[50]. The primary structural difference between the two is the addition of a triflouromethylphenoxy substituent, resulting in a 10-fold higher activity than primaquine against blood and liver stages of the parasite. One study

involving a daily dose of tafenoquine given to non-immune soldiers in Thailand reported a 95% protective efficacy against *P. Vivax* infections[51].

Artemisinin

Artemisinin is a naturally occurring antimalarial agent isolated from the herb sweet wormwood (*Artemesia annua*) and is one of the most potent antimalarial drugs currently available. It has been traditionally used in China for over 2000 years as a herbal remedy for febrile illnesses[52]. In 1971, the chemical structure of a sesquiterpene peroxide that displayed powerful antimalarial properties was isolated. X-ray crystallographers in the Research Institute of Earth Physics in Shanghai resolved its structure, showing the compound's key structural feature was an endoperoxide bridge in a seven-membered ring system present in all artemisinins[53]. The exact site of action of these compounds is still uncertain, although recently it has been shown that they act by inhibiting an endoplasmic reticulum-based Ca^{2+}-dependent ATPase (*Pf*ATPase6) that is essential for the parasite survival[54]. This mechanism is thought to involve the formation of radicals from the cleavage of the peroxide bridge within the compounds which react with and inactivate the ATPases. These compounds mainly act against late ring stages of the parasite and are highly active with an ability to reduce the parasite biomass 10,000-fold in a single asexual cycle. One disadvantage associated with artemisinin and its analogues is their rapid absorption and elimination in the body, the compounds having a half-life of less than 60 minutes. Any antimalarial drug with hopes of being effective on a large scale would need to remain at parasiticidal concentrations in the circulation for a week, ensuring activity for at least four asexual cycles of the parasite. It is for this reason that artemisinins are often used in combination with other antimalarial therapeutics. The use of combination chemotherapy offers several advantages over the use of single drugs to treat malaria. For example, combining two different drugs that have different mechanisms or sites of action may lead to a simultaneous attack on two parasitic systems. Furthermore, using two drugs will reduce the chance that a parasite will be resistant to both drugs as this would require a simultaneous change in at least two parasitic loci.

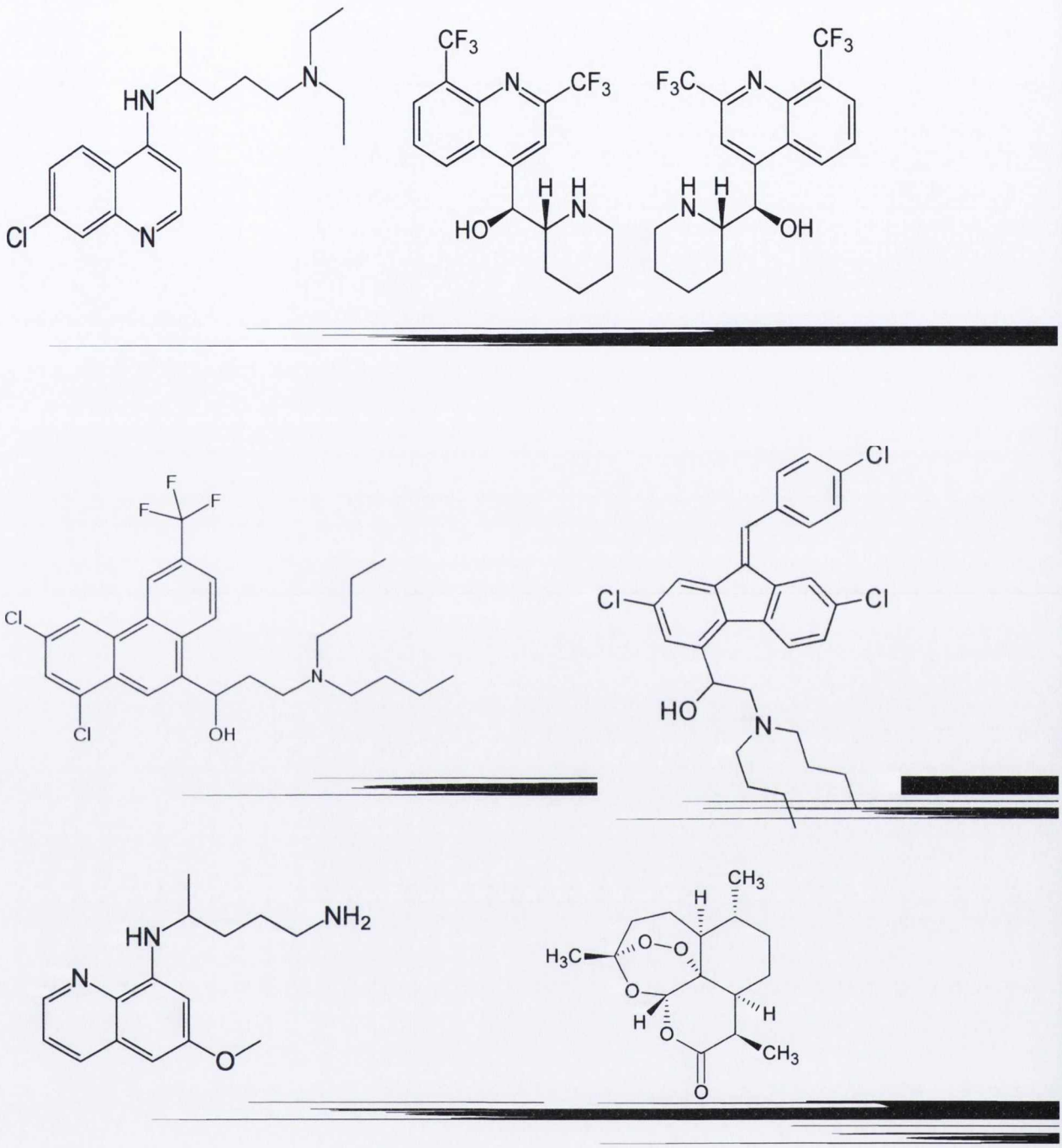


FIGURE 1.5: Commonly used natimalarial drugs.

TABLE 1.2: Summary of currently used drugs

Drug	Pharmaceutical name	Discovered	Cellular Target	Advantages	Disadvantages
Chloroquine	Daramal	1934	Heme metabolism	Cheap, fast-acting, bioavailability	Resistance, retinopathy
Quinine	Quininmax	1633	Heme met.	Fast acting, good bioavailablity	neurotoxicity
Mefloquine	Lariam	1963	Heme met.	Potent in RBC stage	Neuropsychiatric reactions
Halofantrine	Halfan	1960s	Withdrawn	Withdrawn	Cardiac arrhythmia
Artemisinin			*Pf*ATP6		slow acting

TABLE 1.3: Summary of potential new drug targets

Target/Pathway	Enzymes	Inhibitor
Polyamine biosynthesis	ODC, Adenosine deaminase	DFMO, Coformycin
Vit B synthesis	Pyridoxal Kinase	Aminphyline
Apicoplast	Fab H, Fab I	Triclosan, Thiolactomycin
Haemoglobin metabolism	Plasmepsin, falcipains	Pepstatin, chal-cones, leupeptin
Pyrimidine synthesis	Thymidate synthase	Pyrazofurin
Purine salvage, DNA/RNA	HGPRT, DNA Topoisomerase	Irinotecan, Levofloxacin
Glycolysis	Hexokinase	Brefeldin A
Transporters	Hexose transporter	O-3-hexose derivatives
Mitochondria	Cytochrome c oxidoreductase	Atovaquone
Membrane biosynthesis	Phospholipid biosynthesis	G25
Protein Kinases	Various	Xestoquinone

1.2.3 Drug Resistance

Many drugs which were once effective against malaria have lost their effectiveness through inappropriate use driving rapid evolution of the malaria parasite. In some cases it is more difficult to understand the basis of drug resistance because exact mechanisms of drug action are still not clearly understood.

Studies have shown that CQ-resistant parasites accumulate less of the drug than susceptible parasites[55]. Genetic studies looking for determinants of CQ resistance mapped CQ resistance to a 36 kb region that contained 8 putative genes. One of these genes, the *P. falciparum* CQ resistance transporter gene (*Pf*crt), was identified as the most important determinant of CQ resistance[56,57]. This gene encodes the chloroquine resistance transporter protein, a protein located in the parasite's digestive vacuole membrane. Along with P-glycoprotein homologue 1 protein[58] which is also embedded in this membrane, these two proteins are currently thought to be the major players in CQ resistance.

Similar studies have shown that *P. falciparum* parasites with mutations in their *Pf*ATPase6 have elevated IC_{50} values for artemisinins[59,60]. Mutations to the *Pf*mdr1 protein are also markers for resistance to artemisinins as well as being linked to decreased parasite sensitivity to mefloquine[55].

The failure of current anti-malarial drugs due to increasing drug resistance is fast becoming an all too common problem and there is a clear and urgent need to discover new drugs and new drug targets with novel mechanisms of action.

1.3 Human African Trypanosomiasis

Human African Trypanosomiasis (HAT), also known as *sleeping sickness*, is a fatal NTD caused by a single species of protozoan parasite - *Trypanosoma brucei* and transmitted by tsetse fly bites[61]. There are two forms of HAT depending on the parasite involved: Trypanasoma brucei gambiense (Tbg) which results in chronic infection and accounts for over 95% of reported cases, found primarily in west and central Africa[62] (Figure 1.6); and Trypanasoma brucei rhodesiense (Tbr), a form that causes acute infections and accounts for less than 5% of reported cases, found in eastern and southern Africa[63] (Figure 1.7).

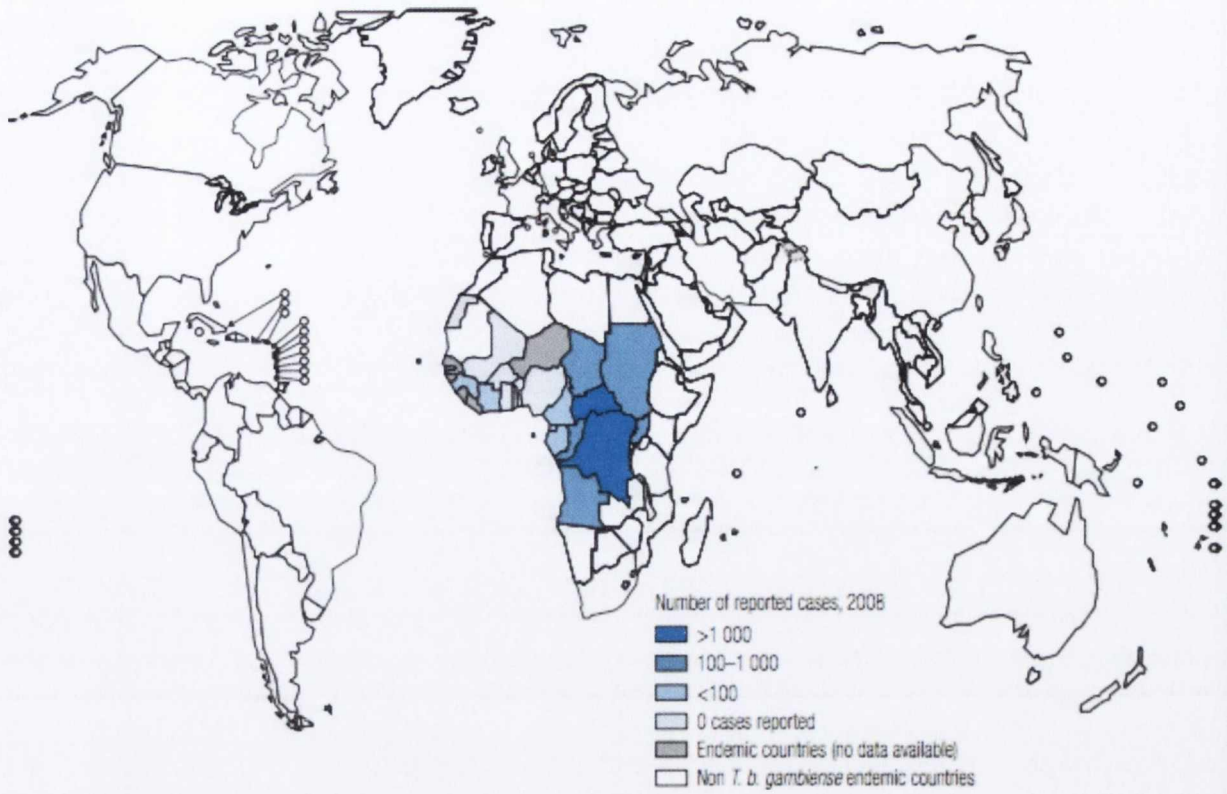


FIGURE 1.6: Distribution of human African Trypanosomiasis (*T. b. gambiense*, worldwide, 2008[64].

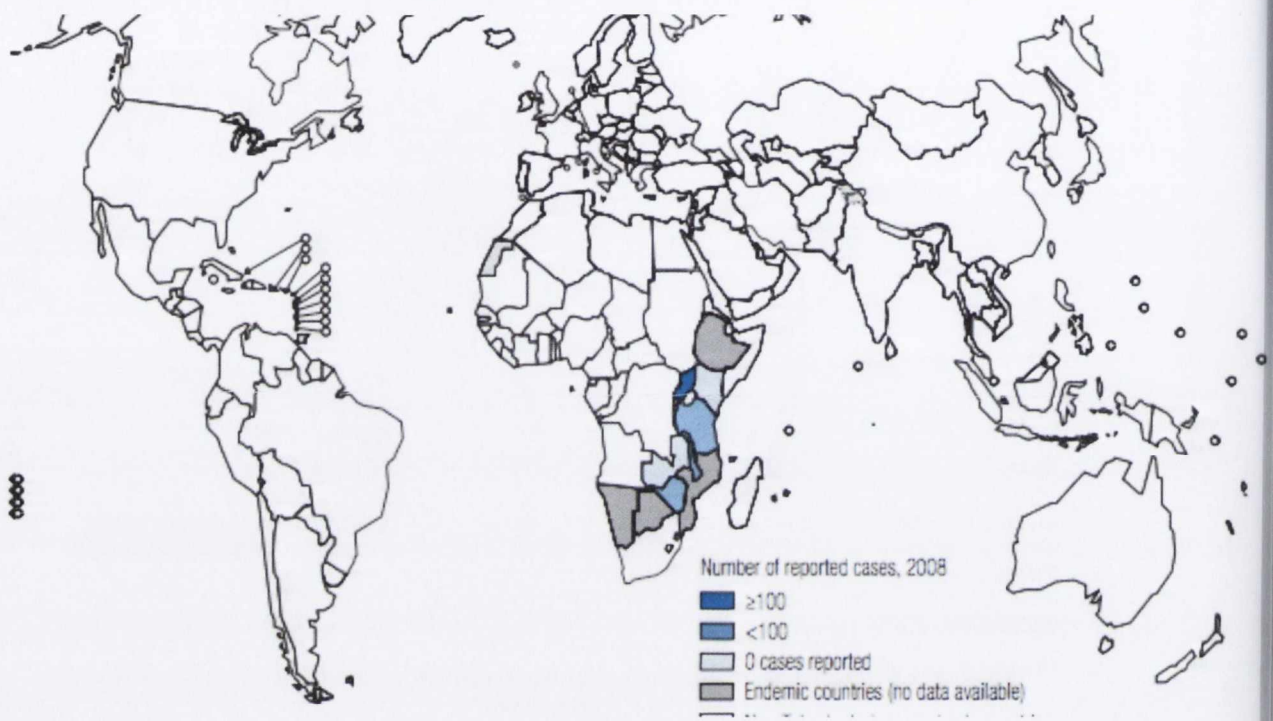


FIGURE 1.7: Distribution of human African Trypanosomiasis (*T. b. rhodesience*, worldwide, 2008[64].

1.3.1 Life Cycle of *Trypanosoma brucei*

The complex events of the *T. brucei* life cycle takes place in the tsetse salivary gland, the mammalian bloodstream and the tsetse midgut and parasite survival is dependent on the successful transmission between these radically different environments. Infected tsetse flies deposit metacyclic parasite forms into mammalian hosts during a blood meal (Figure 1.8). These metacyclic forms develop in the tsetse fly salivary glands and express a metacyclic variant surface glycoprotein (VSG) coat which serves to protect the parasite against the immune system of the host[65]. The parasites then differentiate into long slender forms which, unlike the metacyclic forms, can divide and proliferate in the bloodstream. In the bloodstream the long slender form begins to express the VSG gene from a distinct genomic location. This antigenic variation allows the existing bloodstream population to further evade eradication by the host immune system. Long slender bloodstream forms can then differentiate into the short stumpy bloodstream form, the

process occurring in a density-dependent manner due to the accumulated secretion of Stumpy Induction Factor (SIF)[66]. These forms do not divide and are pre-adapted to their future environment, namely the tsetse midgut. After ingestion by the tsetse fly during a blood meal, the short stumpy forms differentiate into proliferative procyclic forms that can move into the alimentary tract. Here they differentiate into the mesocyclic form which can migrate into the salivary glands. Further differentiation occurs in the salivary gland, with mesocyclic forms undergoing asymmetric cell division, the smaller of the two daughter cells being the progenitor of the epimastigote form. The proliferative promastigotes pack together closely after attachment in the salivary glands and can then differentiate to form the detached non-dividing metacyclic forms which are infective to the mammalian host. Once in the mammalian host these forms can re-enter the cell cycle and readopt a long slender morphology.

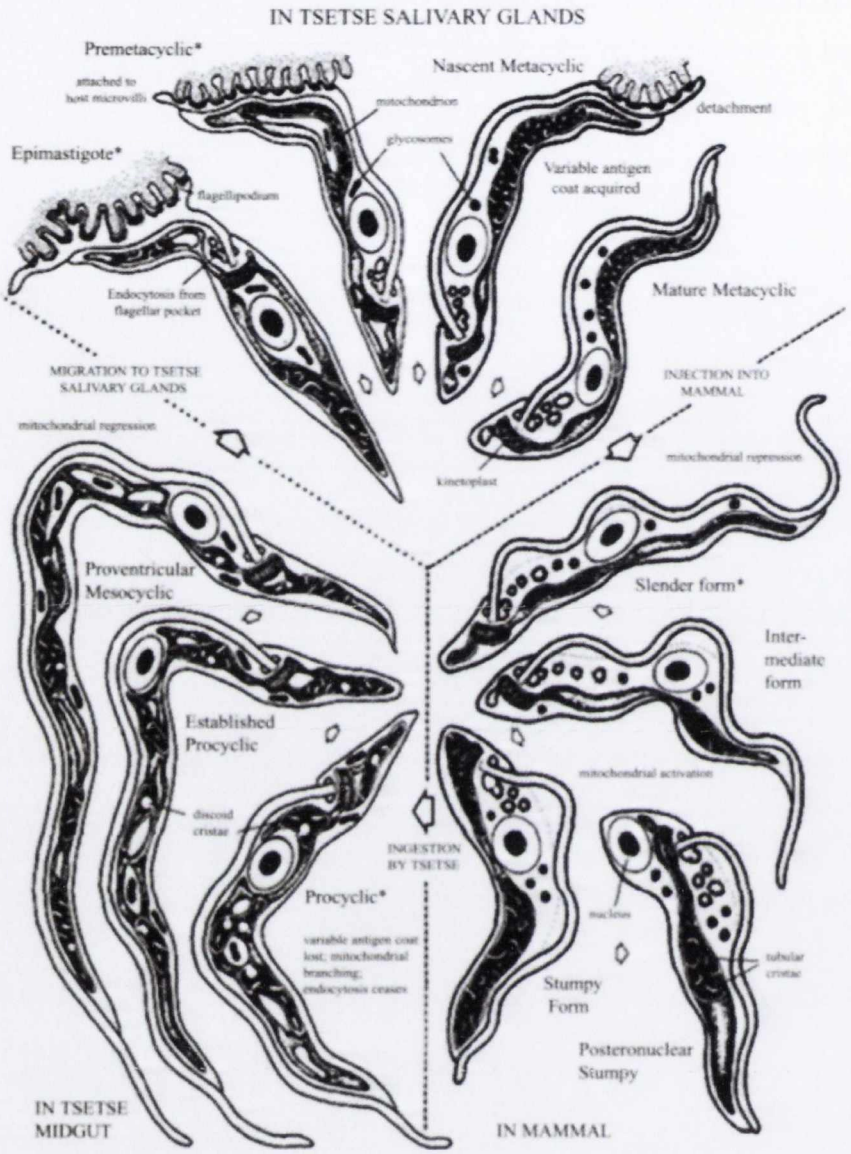


FIGURE 1.8: Life cycle of *Trypanosoma brucei*. The life cycle alternates between the mammalian host and the tsetse fly vector but involves adaption to three environments: the mammalian bloodstream and lymphatic system, the tsetse-fly midgut and the tsetse fly salivary gland[67].

1.3.2 Current Chemotherapy

There are currently only four drugs licenced for treating HAT: eflornithine, melarsoprol, pentamidine and suramin (Figure 1.9). Suramin and pentamidine are used in the first stages of the infection where the trypanosomes are multiplying in blood and lymph but have yet to cross the blood-brain-barrier (BBB) to infect the central nervous system. This stage is known as the second or neurological phase of the disease. Periodic switches in gene expression that cause the expression of antigenically distinct versions of the VSG coat has made the development of an effective vaccine too complex, meaning chemotherapy remains the best option.

Suramin

Suramin is a symmetrical, polysulfonated napthylamine polyanionic compound discovered by Oskar Dressel and Richard Kothe at Bayer in 1916. It was first used to treat sleeping sickness in 1922[68]. The development of suramin came about after observations that the related compounds trypan blue and trypan red had a marked trypanocidal activity. Like suramin, these compounds are precluded from passing through intact membranes due to the six negative charges they possess at physiological pH, making them inadequate for treatment of second stage sleeping sickness due to their inability to pass the BBB[69,70]. Due to its abundance of negative charges, suramin is thought to nonspecifically bind to many positively charged serum proteins making it difficult to conclude its definite mode of action. The binding to multiple proteins is also thought to be the reason for its slow excretion; suramin has been found in the blood of patients three months after injection, giving it one of the longest half lives ever documented for a drug given to humans[71]. Suramin binds to and inhibits many enzymes, including thymidine kinase and dihydrofolate reductase[72], but it is yet unclear whether its trypanocidal action is related to these electrostatic interactions.

Suramin has been found to deposit in the renal tubes and as a result it is advised against administering it to patients presenting with existing renal disease. Suramin is known to cause adverse reactions including proteinuria, reversible liver damage and nephrotoxicity and has also been observed to cause jaundice, hepatitis and haemolytic anaemia.

Pentamidine

Pentamidine is an aromatic diamidine effective against first stage HAT. Pentamidine was developed after it was observed that synthalin, a related compound and a potent hypoglycaemic agent, exhibited a marked trypanocidal activity[73]. Pentamidines are dicationic molecules, endowing them with very slow rates of diffusion across biological membranes. These compounds are reported to interact with cellular anions through tight binding with the network of circular DNA molecules that comprises the kinetoplastid and nucleus, thus inhibiting replication of nucleic acids[74]. However, this may not account for the trypanocidal activity as the generation of dikinetoplastic trypanosomes is not expected to cure the disease. Pentamidine is also involved in collapsing the mitochondrial membrane potential (MMP) while also acting as an uncoupler of oxidative phosphorylation in mammalian mitochondria[75].

Melarsoprol

The second stage of the disease is more difficult to treat due to the necessity for treatments to pass the BBB. Melarsoprol and eflornithine are two drugs that have both been shown to pass the BBB and reach sufficient drug levels in the central nervous system.

Melarsoprol, also known as mel b, is an organic trivalent arsenical, containing a markedly reactive arsenoxide group, that was introduced in the late 1940s and is still widely used as the first-line therapy for late stage HAT[76]. Due to its poor solubility, mel b is administered intravenously dissolved in propylene glycol, making the injections very painful.

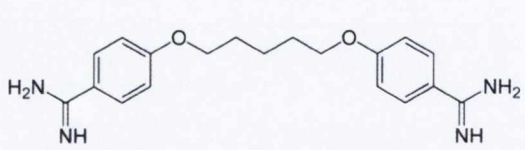
The mechanism of action of these compounds is still not entirely clear although research has shown that it interferes with enzymes of the glycolytic pathway, specifically thiol-containing enzymes like glycerol-3-phosphate dehydrogenase (G3PDH)[76]. It has also been shown to form a stable adduct with trypanothione[77] but whether this adduct is linked to toxicity is unclear. There are many adverse effects associated with mel b use. Up to 19% of patients suffer reactive encephalopathy (4-8% die[78]), making it necessary for patients undergoing treatment with mel b to be hospitalised for the duration of the treatment. Mel b appears to be increasingly less effective in some endemic areas of Uganda and Sudan, with up to 30% of treatment failing[79,80]. It is thought that

these cases involve patients accumulating sub-optimal levels of the drug in the brain. Resistance to mel b is thought to be down to a mutation in purine nucleoside transporters in the parasites[81]. These transporters are directly involved in the uptake of mel b and experiments on genetic variants of the TbAT1 adenosine transporter have been confirmed in relapse infections following mel b treatment. Experiments in Uganda involving 65 *T.b. gambiense* isolates with high melarsoprol treatment failure showed that 38 had a mutated TbAT1, with all of these containing the same nine mutations in their *Tb*AT1 gene[82]. More recently, a locus encoding two related aquaglyceroporins, family of pore proteins that facilitate the efficient and selective flux of small solutes across biological membranes, were linked to melarsoprol/pentamidine cross resistance (MPXR) through use of a high-throughput loss-of-function screen[83]. An AQP2-specific gene knockout generated MPXR trypanosomes but this resistance did not spread to all diamidines or arsenicals as a lipophilic arsenical, phenylarsine oxide, was still able to diffuse across the membrane.

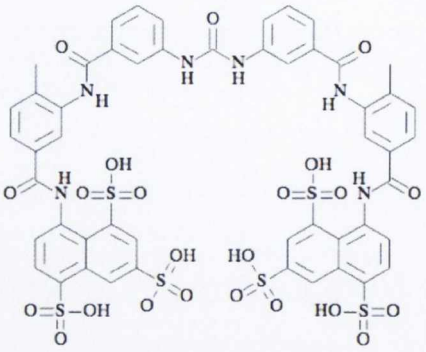
Nifurtimox

Nifurtimox is a 5-nitrofuran that has been in use for more than 40 years. It is currently recommended for late-stage west African sleeping sickness in combination with eflornithine, although there are existing worries regarding its toxicity, potential as a carcinogen and its limited efficacy[84].

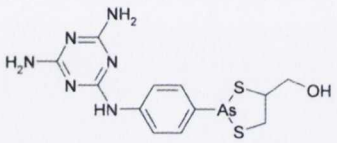
The mechanism of action of nifurtimox has been associated with the nitroreduction of the nitrofuran by NADH-dependent type 1 nitroreductases that are rare in eukaryotic cells[85], a chain of enzymatic reactions resulting in the generation of nitro anion radicals and then superoxide anions[86,87]. The reactive oxygen species in turn cause oxidative stress resulting in cell damage. In many cell types glutathione serves to protect against



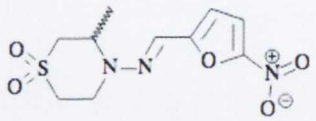
A. Pentamidine



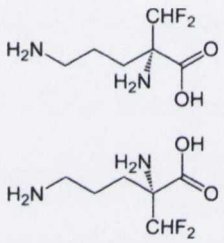
B. Suramin



C. Melarsoprol



D. Nifurtimox



E. Eflornithine

FIGURE 1.9: Commonly used *T. brucei* drugs.

oxidative stress and in trypanosomes, polyamines, the synthesis of which is interrupted by eflornithine, combine with glutathione to form trypanothione. This link between the two compounds was thought to be the cause of their synergy, however recent data suggests this to be untrue. Isobologram analyses have demonstrated that the action of nifurtimox and eflornithine did not synergise when nifurtimox was assayed in the presence of fixed concentrations of eflornithine[88].

Whether oxidative stress is caused by nifurtimox is still a matter of debate as there is little evidence of the production of superoxide anions *in vivo* compared to *in vitro*. More recent data has pointed to a mechanism of action whereby nifurtimox is activated upon metabolism to an open chain nitrile that is as toxic as nifurtimox itself and has been shown to inhibit growth of both parasitic and mammalian cell lines[84].

Eflornithine

Eflornithine, or Difluoromethyl ornithine (DFMO), was originally developed as an anti-cancer agent and introduced as a HAT therapeutic in the 1990s. This drug has the ability to cross the BBB and is currently licensed for the treatment of melarsoprol refractory cases, although the WHO recommend it as the first line treatment[89]. Eflornithine is one of the few anti-trypanosome drugs with a well understood mode of action. It irreversible inhibits ornithine decarboxylase (ODC), an enzyme essential in polyamine biosynthesis[90]. Eflornithine is active in T b gambiense but has only very limited activity in T b rhodesience. This is thought to be due to the T b rhodesience parasites having a naturally higher ODC turnover compared to the T b gambiense parasite[91]. The rapid rate at which the T b rhodesience ODC is replaced explains their natural refractoriness.

1.3.3 Problem of Drug Resistance

Drug resistance in HAT is usually associated with the loss or increased expression of transporter proteins. The loss of the aminopurine transporter 2 (P2) protein is associated with loss of parasite sensitivity to melarsoprol. This is especially worrying as melarsoprol is currently the only drug available against second stage T b rhodesiense infections[80]. Increased expression of the Multidrug Resistance Protein A (MRPA) has also lead to decreased sensitivity to melarsoprol treatment. Pentamidine resistance is linked to a loss in order of the P2 and High Affinity Pentamidine Transporter (HAPT), although this is at a cost to the fitness of the parasites[92,93].

Clearly, of the limited drugs available for HAT treatment, severe side-effects and increasing drug resistance are driving the need for new HAT drug targets and therapeutics.

1.4 Tubulin as an Anti-Parasitic Drug Target

During the ongoing search to identify new anti-parasitic drug targets tubulin has emerged as an attractive option for drug design. This protein has long been known as a target for anticancer agents but more recently it has been exploited in the search for species-selective anti-parasitic drugs[94].

1.4.1 Tubulin & Microtubules

Microtubules are long, hollow cytoskeletal protein polymers that are involved in a variety of cellular functions including intracellular transport, maintenance of cellular structural integrity and cell motility[95]. One of the most important roles of this protein is its task during cell division. Without a functioning tubulin network, genetic information, as duplicated chromosomes, would not be split between the two daughter cells, resulting in cell death. This fundamental function is what makes microtubules such an important target in the fight against cancer, a disease caused by unregulated cell growth.

There are three types of tubulin, α, β, and γ, which have similarities but are still relatively distinct. α-β tubulin heterodimers polymerise in a GTP-dependent manner to form linear protofilaments. 13 of these protofilaments then arrange side to side forming a hollow tube 25nm in diameter[96]. Longitudinally, tubulin heterodimers, and thus protofilments, are arranged end-to-end (Figure 1.10), resulting in a polar structure with α-tubulin oriented toward what is known as the 'minus-end' and β-tubulin toward the 'plus-end'. Growth and shortening rates are much more rapid at the plus-ends than the minus ends.

Dynamic Instability

Microtubules display a quality known as dynamic instability, the ability to switch between growth and shortening with non-equilibrium kinetics. This dynamic process is regulated by the hydrolysis and exchange of GTP within the β-subunit of each heterodimer, as well as due to presence or absence of small ions and various binding proteins termed microtubule-associated proteins (MAPs)[98,99]. A second type of dynamic instability is termed treadmilling[100]. During treadmilling, the plus end of the microtubule is

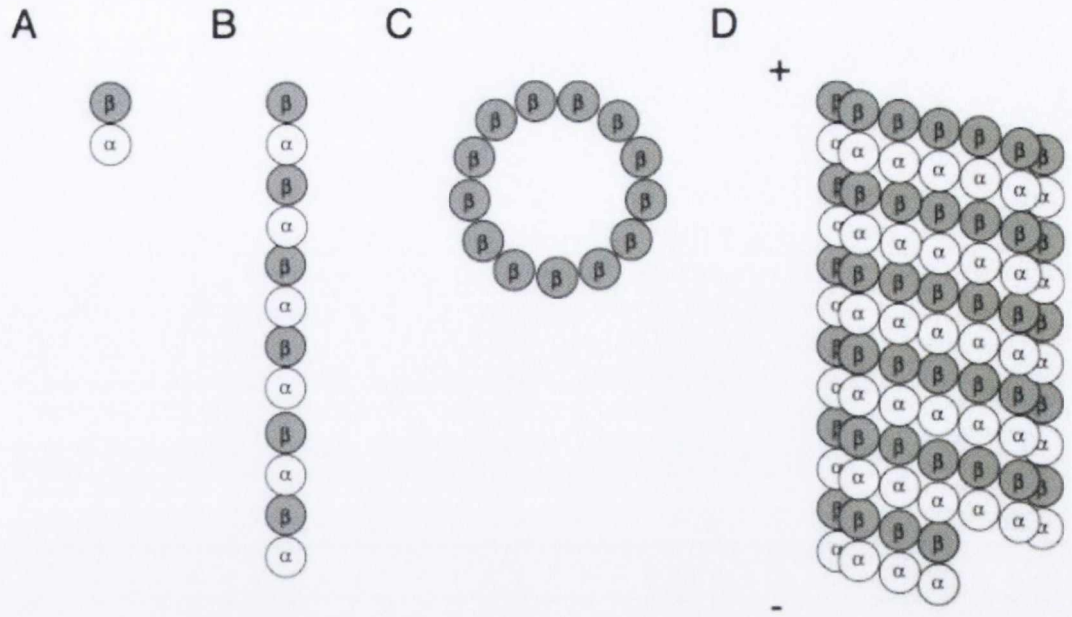


FIGURE 1.10: ***A***: Tubulin heterodimers are composed of α and β subunits that polymerise head to tail to form protofilaments ***B***. Thirteen protofilaments form lateral contacts to create the hollow cylindrical structure of the microtubule ***C*** and ***D*** with β-tubulin exposed at the plus end and α-tubulin exposed at the minus end[97].

undergoing a growth event while the minus end shortens at an equal pace, resulting in no net growth of the microtubule. At the ends of the microtubules there is a cap that stabilises the structure believed to be composed of a ring of tubulin subunits with GTP bound. When an α-β heterodimer is added to the end of the microtubule, the GTP present at the β-tubulin end of the microtubule is hydrolyzed, generating GDP-Pi along with a kink in the region between the new heterodimer and the previous heterodimer (Figure 1.11). This kink adds physical stress to the structure and it is thought that after a certain stress threshold the microtubule begins to depolymerise in a process termed catastrophe.

1.4.2 Tubulin Structure

Both α and β subunits have a size of 50 kDa and despite only sharing 43% amino acid identity both also have a similar protein structure with most of the variation occurs at the acidic C-terminus. Concerning the crystalised structure of bovine brain tubulin

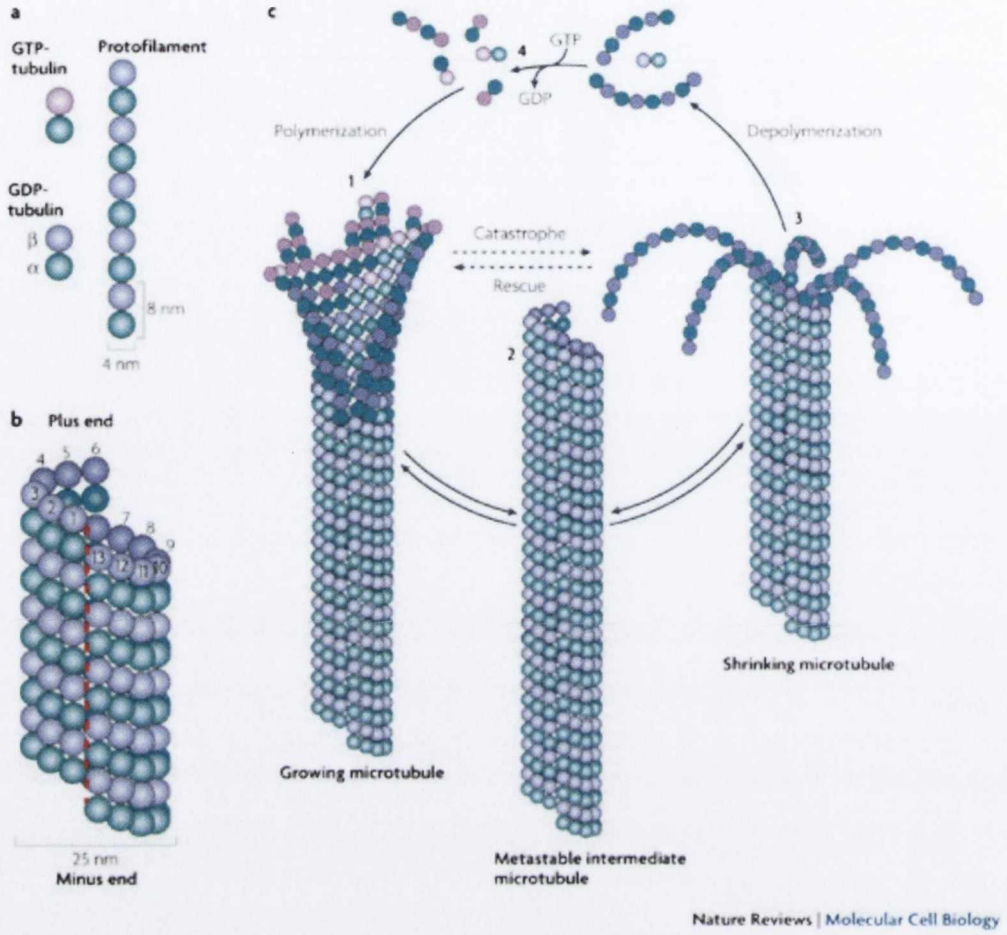


FIGURE 1.11: Illustration of the nucleotide-dependant switch in microtubules between polymerisation and depolymerisation[101].

(resolved at 3.58Å[102]), each momoner of the hetrodimer can be described through 3 well-defined regions: residues 1-206 represent the N-terminal nucleotide binding domain made up of 6 parallel β strands and alternating helices (Figure 1.12). In β-tubulin the GTP bound here is freely hydrolyzable and the site is called the E-site (exchangeable site), the GTP that binds in this site in α-tubulin is trapped by β-tubulin and is known as the non-exchangeable GTP site (N-site). Residues 207-384 are called the intermediate domain and it is composed of 5 α-helices and a mixed β-sheet. Contained in this region are the residues responsible for the E-site catalytic ability. The C terminal is made up of the remaining residues, forming two anti-parallel helices that cross over the previous two domains. When protofilaments align to form microtubules the major contacts are made laterally between homologous subunits. Primary lateral contacts occur through residues 279-287, also called the M-loop, and also through a loop between sheet-7 and

helix-9 that interacts with loop H1-S2 and H3.

1.4.3 Tubulin-Targeting Compounds

Tubulin-targeting compounds have been in use for decades in the treatment of illnesses including anthelmintic infections and cancer, two diseases that exhibit rapid cell division. Anti-mitotic drugs can be divided into two functional groups, microtubule-depolymerising (Figure 1.14) and microtubule-polymerising agents (Figure 1.16).

Colchicine

This group includes compounds such as colchicine, the vinca alkaloids, halichondrins and cembretastatins. Colchicine is a naturally occurring alkaloid originally isolated from the meadow saffron Colchicum autumnale. It exhibits anticancer properties and is also used in the treatment of Gout; its therapeutic uses are limited due high toxicity in humans. Colchicine binds to tubulin at the polymerising end and blocks further polymerisation (Figure 1.13). It binds in one high-affinity binding site located in the intermediate domain of β-tubulin, next to the GTP-binding site of α-tubulin. The crystal structure of tubulin complexed with colchicine displays a bent shape which lacks the ability to straighten, resulting in the obstruction of lateral contacts formed by the M-loop between subunits.

Several other compounds interfere with tubulin dynamics by binding at the colchicine site. Nocodazole is the lead compound of the benzimidazole family with antimitotic and anti tumoural potencies. It has been shown to bind tubulin through a colchicine competitive assay. It also depolymerises tubulin and binds rapidly, reversibly and in a two step mode and causes conformational changes to the protein[103]. Combrestatin[104,105] was first isolated from Combretum and it inhibits tubulin polymerisation by binding to the colchicine site. Their angiogenesis inhibition potency makes these compounds considerably attractive, in particular the best analogue A-4[106]. Its combination with a pro-drug permitted water solubility (A-4P) and enabled an impressive potency even in advanced clinical trials for solid tumour treatment[107].

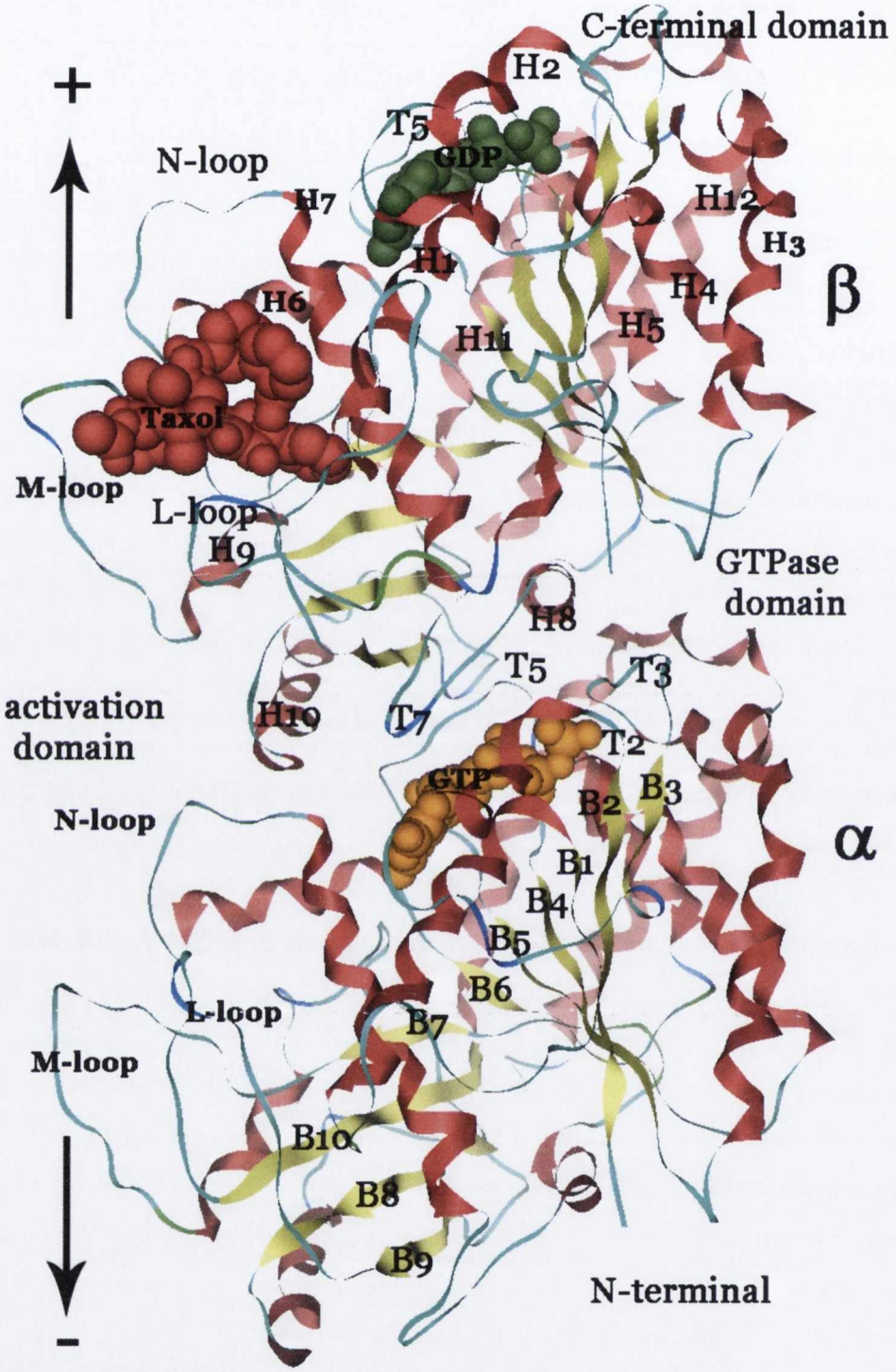


FIGURE 1.12: The tubulin dimer, showing the position and numbering of secondary structural elements.

Vinca Alkaloids

Vinblastine and vincristine, isolated from the plant Catharantus roseus, are two potent antimitotic compounds. They differ only in a single substitution of a formyl group

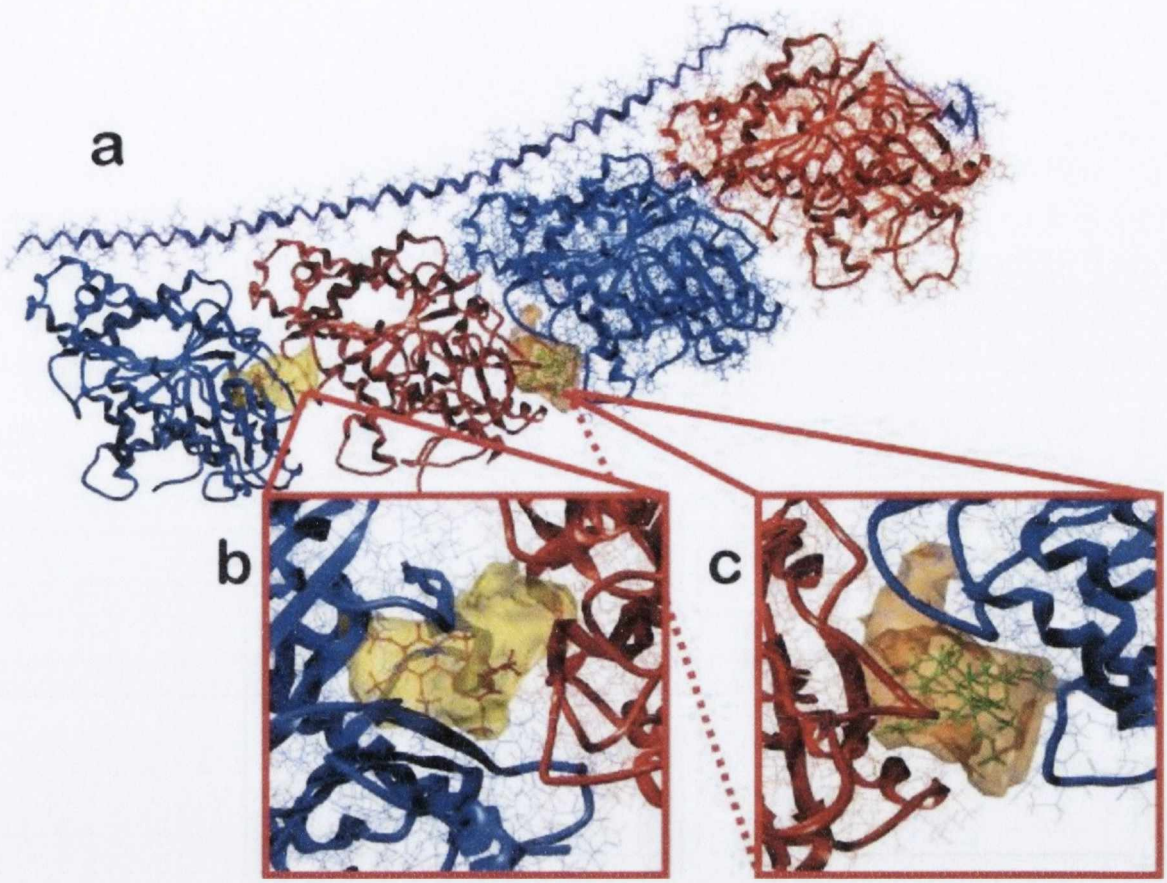


FIGURE 1.13: Tubulin protein (1Z2B) woth colchicine and vinblastine binding sites at interfaces between the α and β subunits. The tubulin polymer is rendered in ribbon and tube with the α subunits shown in red and β subunits shown in blue. (***(b)*** inset shows the colchicine binding pocket (yellow contour) at the intra-dimeric interface of the α-β-subunit; ***(c)*** inset shows the vinblastine binding site (orange contour) at the inter-dimeric interface between α-β-subunits[108].

for a methyl group. Vinblastine has been used to treat several cancer types including leukaemia, non-Hodgkins lymphoma and testicular germ cell tumours but has been linked to bone-marrow suppression and gastrointestinal toxicity.

Vinblastine binds to microtubules and disrupts their structure but its mechanism of action is not yet completely established. Vinblastine has been crystallised in complex with tubulin in the 1Z2B[108] structure deposited in the PDB. In this structure the protein-ligand interaction were found to occur between two heterodimers. Vinblastine interacts at the carboxy terminal turn of helix H6, the T5 loop and the N-loop of β-tubulin while the rest of the drug resides between the T7 loop, helix H10 and the B9 strand of α-tubulin (Figure 1.13).

The vinca-binding region of tubulin also accommodates a second overlapping site that interacts with toxic peptides such as dolastatin 10, isolated from the mollusk Dolabella auricularia. The area between two heterodimers is very large and can accommodate a wide range of compound scaffolds. X-ray studies identified that phomopsin A[109], a dolastatin 10 analogue, binds at this overlapping site in the tubulin heterodimer. These compounds exhibit nanomolar activity against many cancer cell lines but their toxicity profiles and poor *in vivo* activity preclude them from further clinical development.

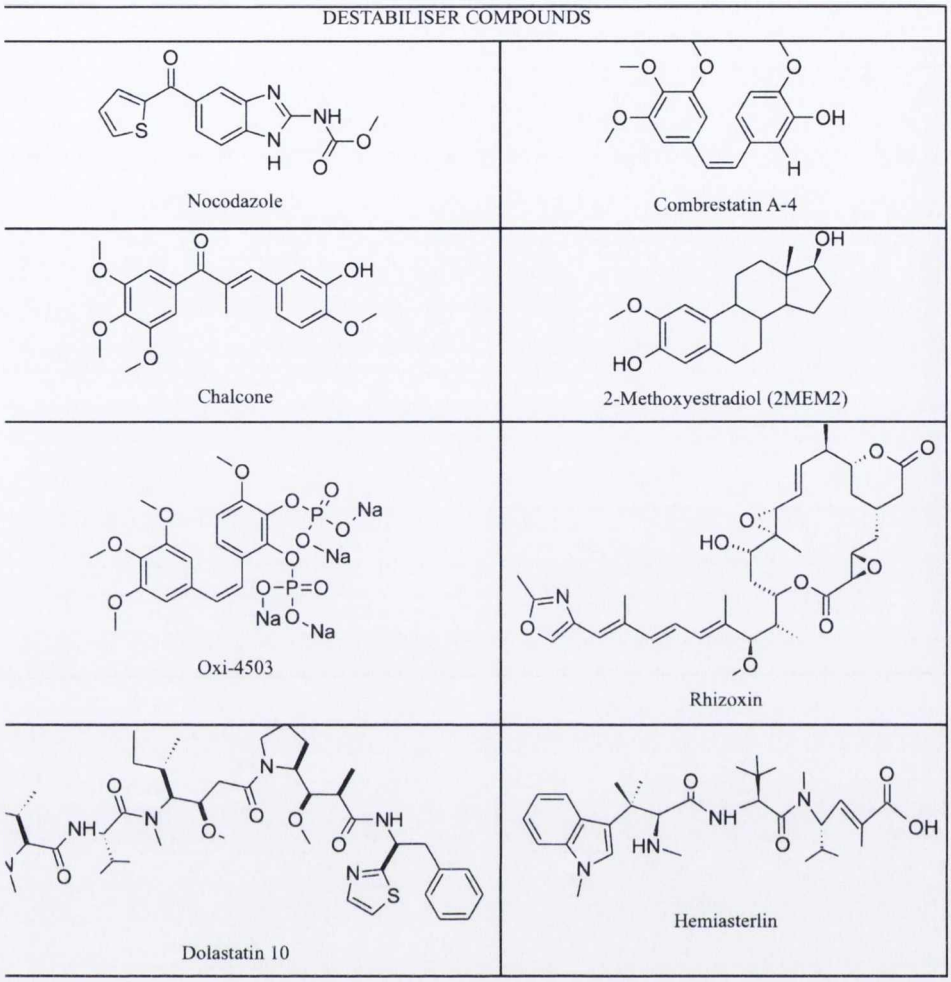


FIGURE 1.14: Tubulin destabiising compounds.

Taxanes

Taxol is a natural compound isolated from the bark of the pacific Yew tree. The taxane class of compounds stabilise microtubules and have been used as therapeutics for ovarian and breast cancer, although some, such as taxol, have severe dosage limitations due to high toxicity[110]. The taxol binding site is a hydrophobic pocket found on the inside face of the microtubule wall near the middle of the β tubulin subunit (Figure 1.15). Recent computational modeling has also put forward the existence of an intermediate site within microtubule nano-pores that are involved in aiding the drugÕs diffusion through the polymer[111,112]. Taxol binding causes the M-loop of one β tubulin monomer to move towards the H1-S2 loop of an adjacent β-tubulin monomer such that lateral contacts between the protofilaments are optimised.

The epithilones are a class of compounds that bind to the taxol site and stablise microtubules. The structure of tubulin complexed with the epithilone macrolide has been resolved, showing a similar interaction pattern as taxol, providing further data for the creation of a common taxol-site pharmacophore. Epithilones are associated with cytotoxic effects similar to taxol, although they are more soluble and easier to obtain.

FIGURE 1.15: The location of the Taxol and Paclitaxel binding site in β-tubulin[113].

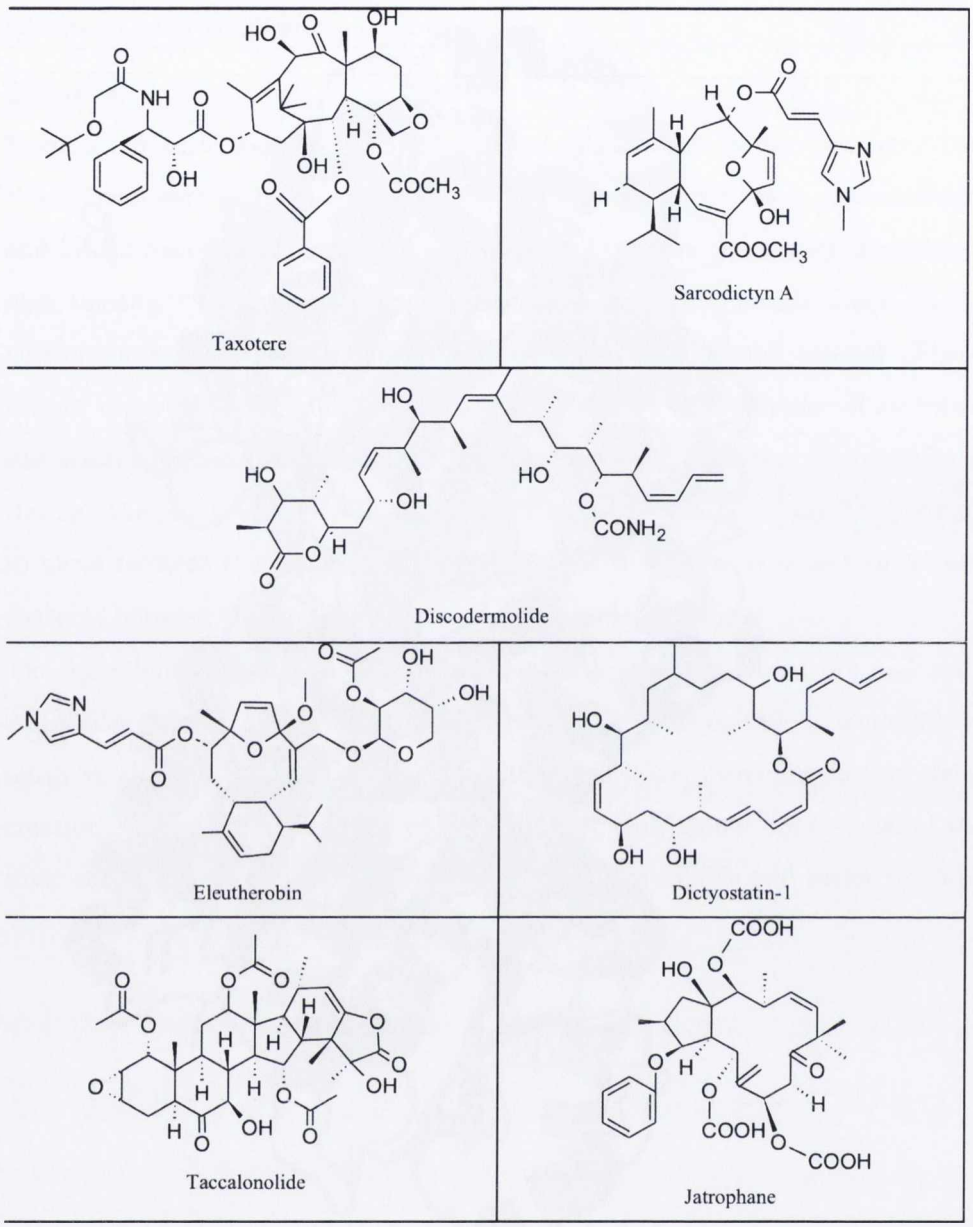


FIGURE 1.16: Tubulin stabilising compounds.

1.5 Organophosphorus Herbicides: The Dinitroanilines & Phosprorothioamidates

As of late there has been much effort put into identifying compounds that can effect parasitic tubulin dynamics without exhibiting the same effects on mammalian tubulin. Almost all compounds that have been shown to affect malarial tubulin *in vivo* are also highly toxic to mammalian cells. The only microtubule inhibitors thus far shown to be toxic to malarial tubulin and not mammalian tubulin are the dinitroanilines and phosphoro(thio)amidates, two classes of herbicides.

The dinitroanilines and phosphorothioamidates are two mechanistically related but chemically distinct classes of herbicide inhibitors that were historically marketed for crop protection. They allegedly act by disrupting tubulin polymerisation and thus inhibiting mitosis. The most well studied dinitroanilines are trifluralin, oryzalin (OZN) and chloralin, while the compound amiprophosmethyl (APM) is the best known phosphorothioamidate. The anti-protozoal activities of the dinitroanilines has been demonstrated against a number of protozoan parasites, including *Leishmania sp*[114], *T. brucei*[115,116] and Plasmodium species[117,118]. By contrast, these compounds are largely ineffective against mammalian tubulin, with IC_{50}s of over $30\mu M$[117,119]. Table 1.4 displays the interesting activity ratios between *P. falciparum* tubulin and mammalian tubulin compared to the classic mammalian tubulin binders. Overuse of these herbicides has resulted in the emergence of widespread resistance in certain crops. Tests revealed that this resistance may be due to a point mutation in α-tubulin that replaces threonine 239 with isoleucine[120]. This resistance extended to all dinitroanilines as well as the phosphorothioamidate herbicides, adding further weight to the claim that both classes of compound bind at the same site on tubulin.

TABLE 1.4: Relative activities of anti-tubulin compounds between *Plasmodium falciparum and mammalian tubulin*

Compound	IC_50 μM (*P. falciparum*)	IC_50 μM (*T. brucei*)	IC_50 μM (mammalian)
Colchicine	10.0	14.0	0.053
Vinblastine	0.028	0.410	0.0006
Vincristine	0.006	-	0.002
Paclitaxel	0.071	0.004	0.022
Docetaxel	0.003	-	0.039
cis-Tubuzole	3.0	-	<1.8
trans-Tubuzole	2.2	-	>18
Trifluralin	<1	6.0-7.0	>50
Oryzalin	6.1	11.0	32
Chloralin	16.0	-	64
Amiprophosmthyl	**3.5**	-	**>64**

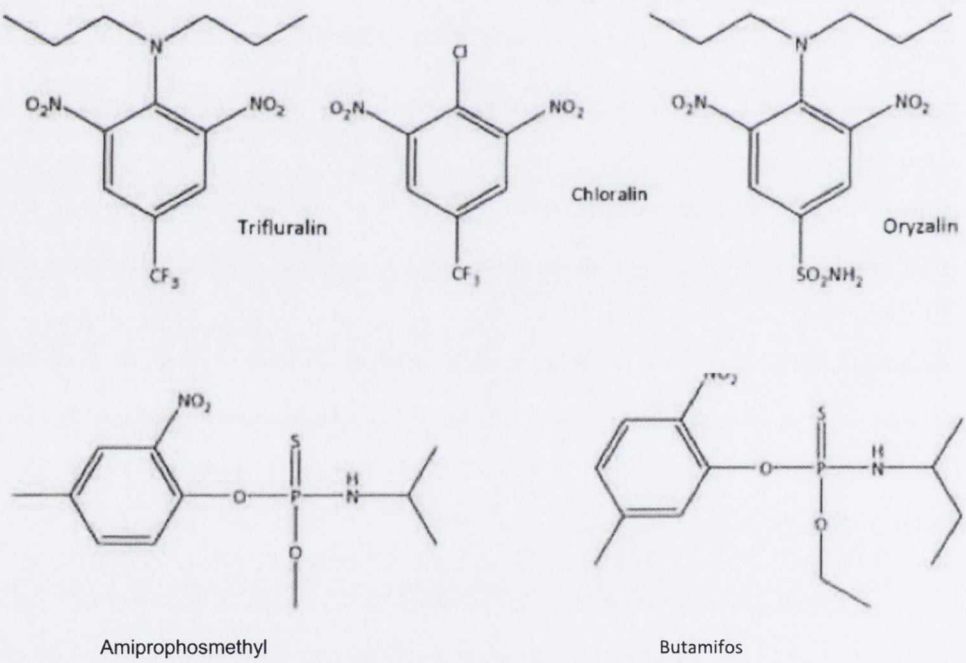


FIGURE 1.17: Structures of the tubulin depolymerising herbicide compounds Trifluralin, Chloralin, Oryzalin, Amiprophosmethyl and Butamifos.

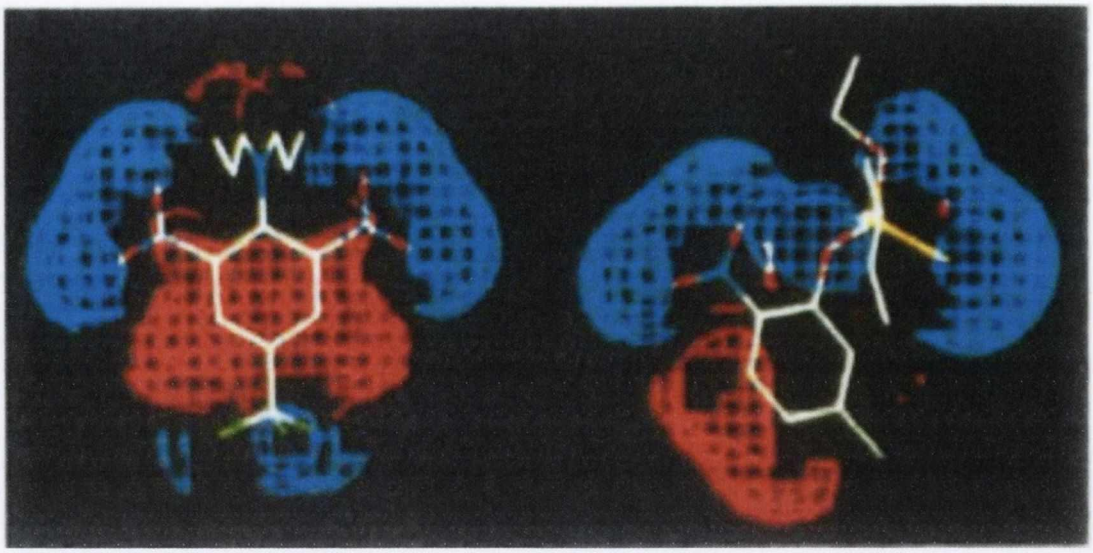


FIGURE 1.18: The electron density of trifluralin (left) and butamifos (right) (red areas denote negative regions and blue areas denote positive regions)[121]

Although Trypanosoma and Plasmodium are not closely related in evolutionary terms, the proposed target protein, tubulin, is highly conserved. The putative binding target of the dinitroanilines and phosphorothioamidates is α-tubulin[117,122,123] subunit of the heterodimer. They are reported to cause depolymerisation from the minus end resulting in shortening of the microtubule, eventually destroying the polymer. It has been argued that these two classes of compound occupy the same binding site on the tubulin heterodimer as the affinity of radiolabelled oryzalin for tubulin is competitively inhibited by APM[124]. This finding is supported by the similar electrostatic potential spacial arrangement profiles of the two herbicide classes, pictured in Figure 1.18[121]. In the diniatroaniline trifluralin, the areas of electron density are generated by the aryl nitro groups, whereas in the phosphorothioamidate butamifos they are centred on the aryl nitro group and the thiophosphate group. In the analysis of mutations that infer resistance to dinitroanilines in several plant species, it was found that the majority of mutations map to α-tubulin. However, the spatial distribution of mutations, as well as the fact that many mutations map to the core of the protein, implies that the mutations do not specifically define a binding site for the herbicide, but may be involved in protein ligand interactions during ligand binding.

The phosphorothioamidate APM was found to exhibit an IC50 of 3.5μM after 72-h versus asynchronous *P. falciparum*, compared with 2.9 μM for trifluralin and 6.1μM for oryzalin. The same work examined APM-treated cells for evidence of tubulin-disruption

using immunoflourescent microscopy. Treated parasites displayed a breakdown of normal hemispindles and microtubule organizing centres compared to untreated cells, with the effects of APM and trifluralin indistinguishable. To establish a definite link to the compounds binding to parasitic tubulin, the group attempted to purify recombinant tubulin and probe its interactions with radiolabelled trifluralin. Binding of C-14-trifluralin to maltose-binding protein(MBP)-tagged α-tubulin returned a value of 1437 dpm, while a value of 1580 dpm was calculated for binding to MBP-tagged β-tubulin[117]. However, it isn't clear whether or not this was a functional tubulin protein being examined, and the reported dpms do not indicate a preference to α over β tubulin.

In *T. cruzei*, there is evidence that dinitroanilines exert a greater depolymerising effect on microytubules involved in more dynamic processes. In one study, exposure of *T. cruzei* epimastigotes to trifluralin did not cause subpellicular or flagellar microtubule disruption. However, trifluralin treatment did have an effect on microtubule-dependent cellular delivery sytems[125]. This would imply that microtubules involved in cytokinesis or endocytosis may be the cellular targets of dinitroanilines.

1.5.1 Putative Binding Site

There are differing views concerning the location of the herbicide-binding pocket on the tubulin heterodimer. Three unrelated groups have carried out homology modeling in the hope of identifying the herbicide binding site. These are Blume *et al.*[121], using an *Eleusine indica* model, Delye *et al.*[120], (*Setaria viridis*) and Morrissette *et al.*[122], with *Toxoplasma gondii.*

Blume identified the binding site for dinitroaniline herbicides as being in the area of dimer-dimer contact and involving the following residues: Asp251, Asn253, Val252, Arg243, Leu136, Phe138, Cys4 and His8. These residues form a cavity with a positively charged electrostatic potential which is thought to be involved in an interaction with the partially negatively charged oxygen atom of the nitro groups on the dinitroaniline herbicides, specifically with the guanidino group of Arg2. This work was based on the observation that threonine 239 is conserved in all known plant and animal α-tubulins and as such implies a functional importance for this residue. In resistant forms of *Eleusine indica*, this threonine is mutated to an isoleucine, causing a change in the surface elecrostatics of the protein. However, this amino acid is located deep inside the

protein core and isn't immediately available to participate in ligand binding. This paper, instead, focuses on the nature of this electrostatic mutant and the nature of the spatial arrangements of the dinitroaniline's electronegative nitrobenzene rings and implies that the two are linked without properly testing the hypothesis.

The site which has received the most attention to date is known as the Morrissette site. This study was carried out on a homology model of Toxoplasma gondii and the site was predicted to exist under the N-loop of α tubulin, formed by the following residues: Arg2, Glu3, Val4, Tyr21, Phe24, His28, Ile42, Asp47, Arg64, Cys65, Thr239, Arg243, and Phe244 (Fig 1.19 A). The bovine crystal structure used to build this homology model did not have large parts of the α-tubulin N-loop resolved properly, a problem circumvented by using the analogous region in β-tubulin to properly model the loop. Morrissette *et al* used computational ligand docking to place oryzalin into this site before performing a very short (2.5ns) molecular dynamics simulation on the docked pose. They predicted that this was the correct binding site and that ligand binding in this cavity destabilised microtubules by disrupting lateral contacts through limiting the flexibility of the N-loop on α-tubulin. This work is limited by the detail presented in the molecular dynamics and docking studies. 2.5ns is short for a molecular dynamics simulation and it would be interesting to analyse the results of a longer trajectory.

Further mutations have been generated in T gondii to date that increase the parasite's resistance to oryzalin but that do not map to this binding site. At the present it is unknown whether these mutations in the proposed binding site of T gondii are interfering with binding of the herbicide compounds or instead are increasing the stability of tubulin. Another problem with this work is the mislabeling of the proposed binding site as the N-loop (Figure 1.20). In the tubulin protein, the N-loop is composed of the amino acids I276, S277, A278, E279, K280, A281, Y282, H283, E284, Q285, L286, S287 and V288. In this work, however, the N-loop is confused with a loop region in the N-terminus connecting helix 1 and β sheet 2. This region is not involved in coordinating lateral interactions between protofilaments in the growing microtubule. Figure 1.19 shows images from the publication in question. Panel A shows a view of α-tubulin with the N- and M-loops labelled incorrectly. This loop region appears to be especially disordered and it would be interesting to read a more detailed account of the homology

modeling protocol undertaken to arrive at this structure. Figure 1.20 shows the location of the N-loop and the Morrissette site on the same subunit.

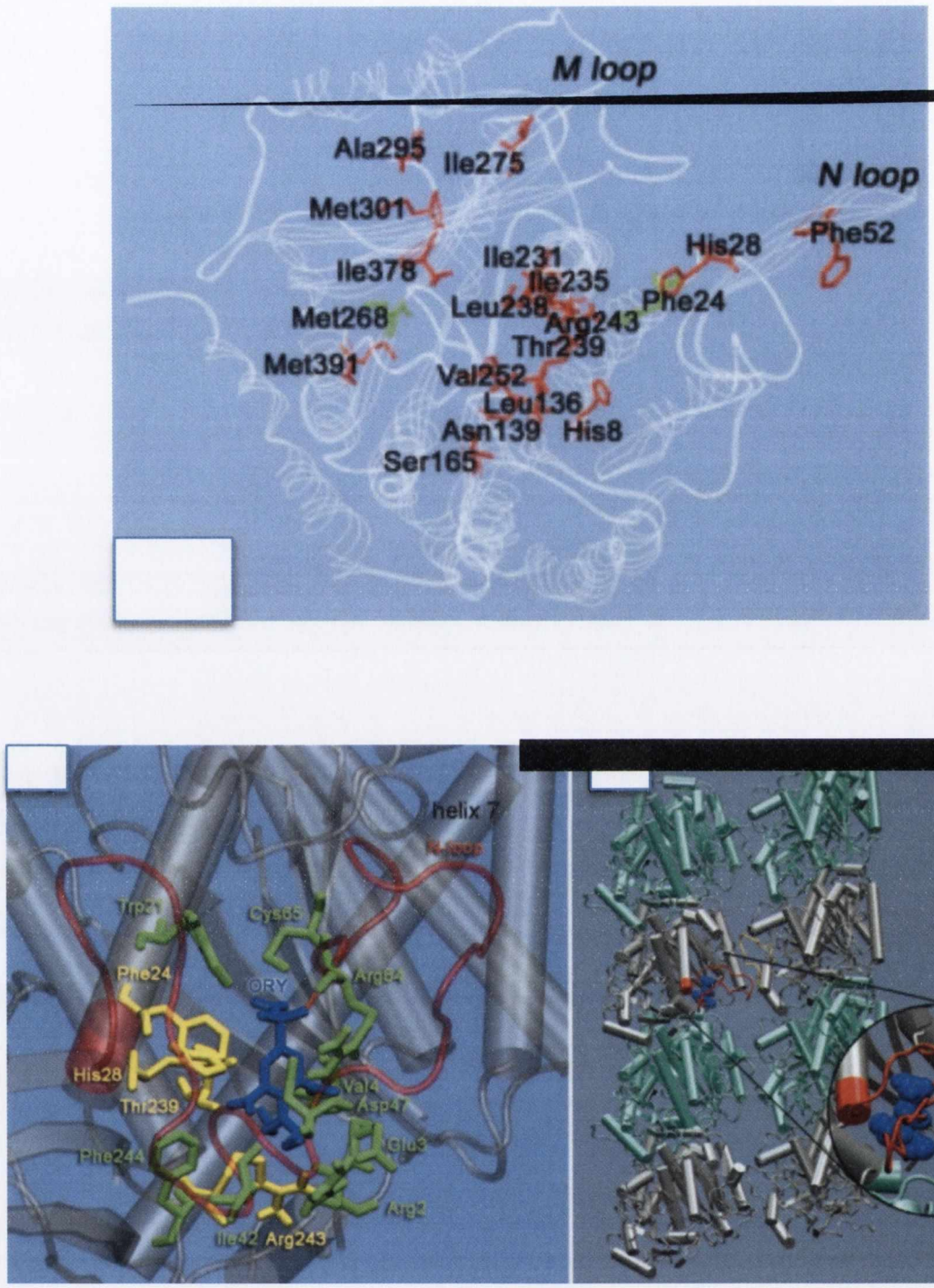


FIGURE 1.19: The proposed Morrissette site of herbide binding to the tubulin heterodimer. A shows a ribbon model of Toxoplasma α-tubulin (white). The residues that are mutated to confer oryzalin resistance are colored red. The plant mutations at residues 24 and 268 are colored green. B shows the structure of oryzalin bound to α-tubulin as predicted by docking simulations. C shows two protofilaments, each consisting of two α-β heterodimers, with bound oryzalin (blue)[122].

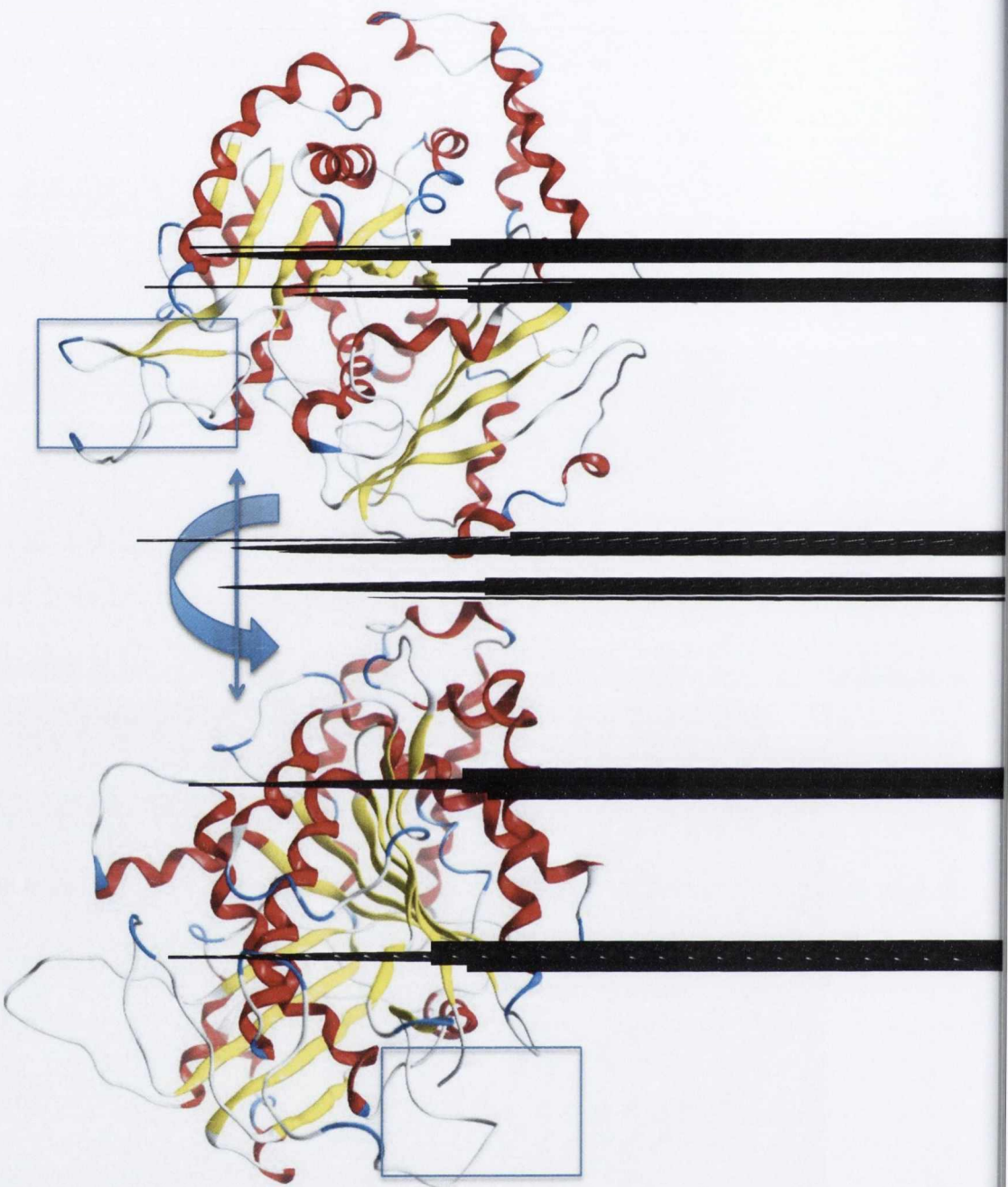


FIGURE 1.20: Mislabeling of the tubulin protein with respect to the Morrissette binding site. The Morrissette herbicide binding site was originally described as existing in the N-loop of tubulin. However, the N-loop exists 90 degrees to the right of the proposed Morrissette site.

1.6 Chemistry and Toxicology of Organophosphorus Compounds

Since their development in the 1950s, organophosphorus compounds have been shown to act as anticancer agents, neuromodulators, antibiotics, insecticides and herbicides. As the term "organophosphate" is generally used to refer to a subgroup of insecticides and neuromodulators, these compounds are sometimes thought of as having adverse effects on human health and thus negative connotations associated with them. Organophosphorus compounds have been shown to display varying degrees of toxicity and some have been used successfully to treat human disease. Cyclophosphamide has been in use as an alkylating agent to treat chronic myelocytic leukaemia and alendronate, a bisphosphonate, is used in the treatment of osteoporosis. Another phosphorylated compound in clinical use is foscarnet, a phosphonic acid derivative used as an antiviral treatment of herpes simplex virus. Organophosphorus compounds have also been shown to be effective prodrugs, one such example being fosamprenevir, a prodrug of the protease inhibitor and antiretroviral drug amprenavir.

However, it can be argued that the most recognised characteristic of the organophosphorus compounds is their effect on autonomic transmission. Some of the most notorious organophosphorus compounds are the ones that have been used in chemical warfare, with the nerve gas sarin perhaps being the most well known example. Sarin and other nerve agents act as reversible inhibitors of acetylcholinesterase, a serine protease that hydrolyzes the neurotransmitter acetylcholine and is found at neuromuscular junctions and cholinergic brain synapses. Inhibition of this enzyme leads to a build up of acetylcholine in the synaptic space, eventually resulting in paralysis and other irreversible neurological problems.

The phosphoroamidates have been widely used due to their moderate anti-cholinesterase activities. This moderate activity is thought to be due to the overlapping nitrogen p-orbital within the neighbouring phosphorus d-orbital, allowing resonance of the nitrogen lone pair, reducing the partial positive charge of the phosphorus atom responsible for enzyme phosphorylation. However, several compounds that possess this nitrogen-phosphorus orbital overlap have still been shown to possess anti-cholinesterase activity. Organophosphorus compounds such as fenamiphos, isophenphos, methamidophos and

crufomate all possess the overlapping orbital and have been shown to interfere with the cholinesterase enzyme. This data shows that if the phosphoro(thio)amidate compounds are ever to be considered as drug candidates then their potential anti-cholinesterase effects represent a stumbling block that must be overcome.

1.7 Aims of the current study

The main goal of this study is to generate homology models for *Plasmodium falciparum* and *Trypanosoma brucei* tubulin and to probe the surface of these structures for the most likely site involved in phosphorothioamidate and dinitroaniline binding. The second goal of this work is to take known ligand data and use it to develop a ligand-based virtual screening protocol. This protocol will then be used in an attempt to move away from the herbicide scaffold to a more 'drug-like' structure, while retaining the herbicide activity profile. This work would represent the first attempt at using virtual screening to identify new parasitic drugs based on the herbicide scaffolds as well as the first efforts at focussing on *T. brucei* and *P. falciparum* tubulin models to identify a potential herbicide-binding site.

Chapter 2
Homology Modeling of the Malarial and Trypanosomal Tubulin Heterodimer

2.1 Introduction

As discussed in Chapter 1 there are two related families of herbicides, the dinitroanilines and phosphorothioamidates, which have been demonstrated to interfere with microtubule polymerisation in several parasitic species but do not interfere with mammalian tubulin dynamics. As yet the definite binding site of these compounds on the parasitic tubulin heterodimer has not yet been elucidated. To date neither Malarial nor Trypanosomal tubulin structure had yet been resolved experimentally thus other methods of structure elucidation must to be utilised. In order to probe the surface of parasitic tubulin for the presence of a herbicide-selective binding pocket, it was necessary to work with structural models of these proteins.

Despite the number of proteins with a structure in the PDB[126] continuing to grow, it remains the case that the structure of the vast majority of proteins in existence remains unknown. As a result it has been necessary to develop methods to model the 3D structure of a protein from its amino acid similarity to structurally resolved proteins. Homology modeling is one such method. Homology modeling takes advantage of the fact that evolutionary related proteins usually have a similar tertiary structure. If the amino

acid sequence identity between the protein of interest and a related template protein is high enough, then homology modeling is regarded as a reliable technique for structure determination.

Briefly, the process of homology modeling is made up of 4 separate stages (Figure 2.1). The first stage concerns the identification of proteins that are related to the protein of interest and that have known tertiary structures. The next stage involves the arrangement of an optimal amino acid sequence alignment between the query protein and the template protein identified in stage one. The third stage calls for the construction of a structural model for the query protein given its alignment with the template sequence and the fourth stage is the evaluation of the resulting models. Software for homology modeling generally varies around the method used for 3D model calculation, of which there are three main methods. The first method is based on matching of protein segments, in which the query structure is divided into a series of segments which are then matched to regions of conserved atoms from a template fitted from the PDB. In a second method, known as fragment assembly, a model is built from core regions and loops and side-chains are added from related structures. The third group of methods is based on the satisfaction of spatial restraints. These restraints (a set of geometrical data used to create a probability density function) are calculated from the query protein and template protein alignment and satisfaction of these restraints is attempted.

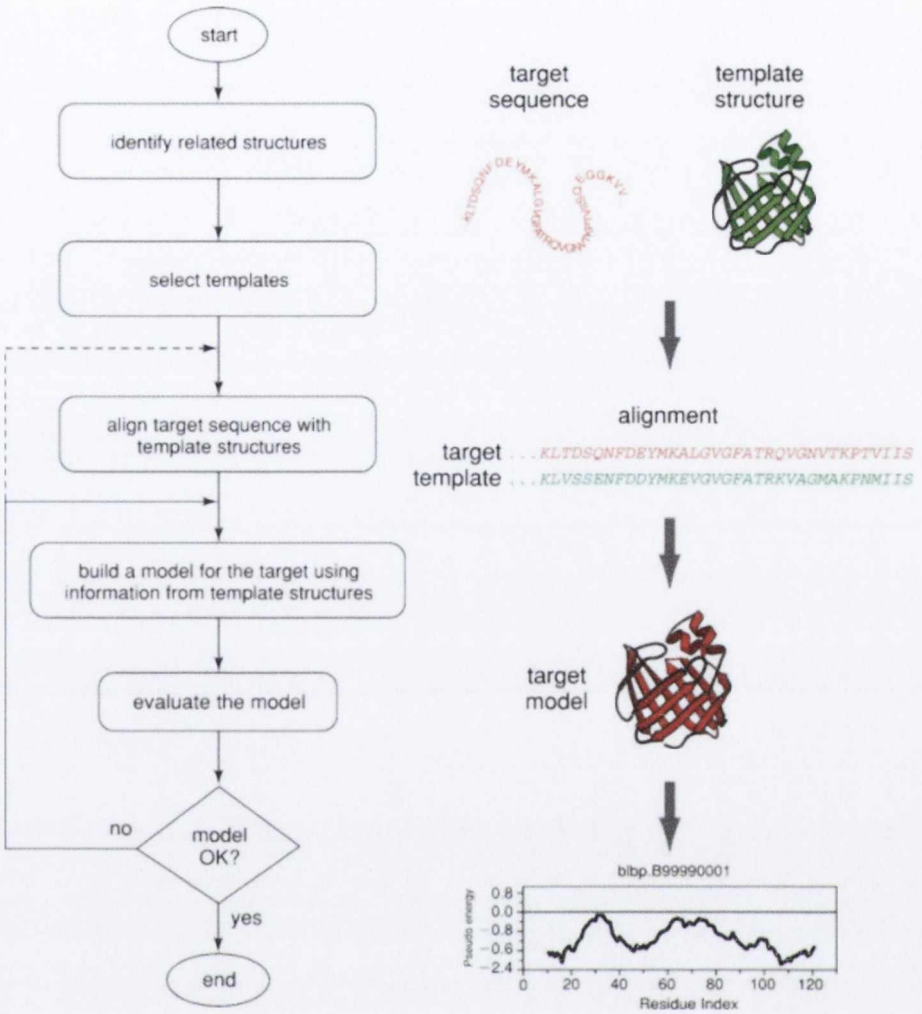


FIGURE 2.1: Typical stages of the homology modeling process[127].

2.2 Homology Modeling

2.2.1 Template Selection

The proper selection of template protein for homology modeling is a critical step as the correctness of the model depends largely on the quality of the templates. Generally speaking, the higher overall sequence similarity between the target and template sequences, the greater the chances that the structures predicted will be correct. Several other factors need to be taken into account when selecting the template. It is important to pay attention to protein families as selecting a template protein that is not the closest phylogenetic match of available templates may result in suboptimal model generation. Another important factor to consider when selecting a template is the similarity between the "environment" of the template and the environment in which the target needs to be modeled. In this sense "environment" may include everything that is not protein, including pH, ligands, etc. The quality of the experimental structure of the template is another important factor to be considered during template selection. This means taking into account details such as resolution and R-factors for X-ray structures as well as considering the number of restraints per residue for structures determined through NMR.

2.2.2 Aligning Target and Template Sequences

The purpose of sequence alignment is to position the amino acid sequences so that the matched stretches of amino acids correspond to common structural or functional features with gaps placed into regions where amino acid sequences are inserted or deleted. At high sequence similarity, >50%, a good alignment is fairly simple to obtain. However, at less than 30% sequence similarity a good alignment is more difficult to acquire. This will usually involve manual intervention to minimise the number of gaps and misaligned residues. At such low similarity the optimal alignment is critical to model quality.

The main methods of sequence alignment are FASTA[128] and BLAST[129], also known as heuristic search methods. These methods were originally developed to tackle the problem of increased search times for optimal alignments using the Needleman-Wunsch

algorithm[130] and the Smith-Waterman algorithm[131] methods of alignment. These methods, while excellent at finding optimal global and local sequence alignments, were too slow at searching the ever-expanding sequence databases FASTA and BLAST apply a scoring system for differences present between the target and template sequences, with the substitution of chemically similar residues incurring a small penalty and insertions, deletions and substitutions of non-similar residues having larger penalties. The sequences with the lowest penalty scores are identified as possible templates. If areas of the template structure are poorly defined in one template then it is sometimes possible to use multiple structures and to take the most well defined areas from each.

2.2.3 Model Construction with MODELLER

For the present study the program MODELLER[132] was chosen to perform all model construction. MODELLER belongs to the class of methods that build structures using the satisfaction of spatial restraints. In MODELLER, the core modeling procedure starts with an alignment of the target sequence with the template sequence(s) in an input file known as a PIR alignment file (this is the sequence file format read by MODELLER). The first step of the automatic modeling procedure is the calculation of many distance and dihedral angle restraints for the target sequence based on its alignment with the template sequence. These restraints are obtained from a statistical analysis of the relationships between many pairs of homologous structures which relied on a database of 105 family alignments that included 416 proteins with known crystal structures[133]. Tables quantifying various correlations were acquired by scanning these databases, such as correlations between two equivalent Cα - Cα distances, or between equivalent mainchain dihedral angles from related proteins. These correlations are expressed as probability density functions (*pdf's*) and are used directly as spatial restraints. The probabilities for different values of the mainchain dihedral angles are calculated from the type of residue considered, from the mainchain conformation of an equivalent residue, and from sequence similarity between the two proteins. MODELLER then combines the empirically-derived spatial restraints and CHARMM[134] force field energy terms enforcing proper stereochemistry into an objective function and the model is obtained by optimising this objective function in Cartesian space. Optimisation is performed employing methods of conjugate gradients and molecular dynamics with simulated annealing.

There are two approaches to modeling the positions of loops, *ab initio* loop prediction and the database approach to loop prediction. *Ab initio* loop prediction is based on a conformational search or enumeration of conformations in a given environment that is guided by a scoring or energy function[135]. The second approach consists of finding a portion of the mainchain that fits the two stem regions of a loop, usually involving a detailed search through a database of many protein structures. Often many alternative segments that fit the stem residues are found and these must be sorted according to either geometric data or sequence similarity between the template and target loops. Selected segments are then superposed and annealed on the stem regions, with initial models often optimised by energy functions. The main drawback of this approach is the limitations on the database search imposed by a relatively small number of nown protein structures. MODELLER implements an *ab initio* optimisation-based approach to loop modeling[136]. This approach makes it possible to simultaneously model several loop regions as well as loops interacting with ligands, a process that is not simple for database search approaches. Optimisation of loop regions in MODELLER is reliant on conjugate gradients and molecular dynamics with simulated annealing.

MODELLER contains an emphin-built energy minimisation program for the models being constructed. The first model evaluation criteria in MODELLER is the Molpdf (objective function) score mentioned previously. Briefly, a conjugate gradients algorithm and a molecular dynamics simulated annealing protocol add kinetic energy to the system so that molecules can overcome conformational barriers. The second model evaluation criteria is the Discrete Optimised Protein Energy (DOPE) assessment score. This is a statistical potential used to evaluate homology models in protein structure prediction. During the model generation stage, the models returning the minimum Molpdfs can be selected as best probable structures and are further evaluated with the DOPE score. This method is generally used to assess the quality of the strucuture as a whole but can also generate a residue-by-residue energy profile. DOPE scores will be used in this work to select the optimal models for further studies.

2.3 Homology Model Validation with PROCHECK

PROCHECK[138] is an online analysis software used to provide an idea of the stereochemical quality of all protein chains in a given PDB file. It can highlight regions of the

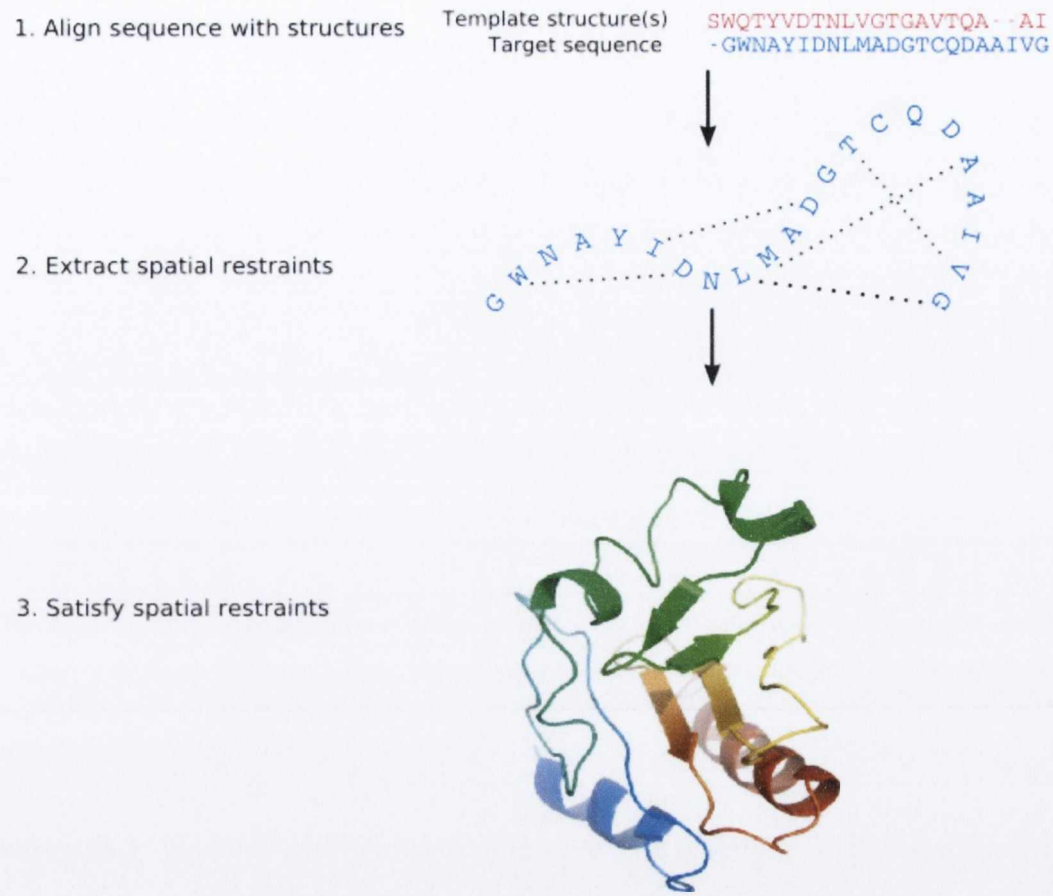


FIGURE 2.2: The homology modeling process employed by MODELLER[137].

proteins which appear to have unusual geometry and provide an assessment of the protein structure as a whole. PROCHECK is capable of generating various plots and text outputs describing the protein stereochemistry, including mainchain and all-residue Ramachandran plots, mainchain and sidechain parameter plots, mainchain and sidechain bond length and angle plots, as well as RMS planarity distance plots.

2.3.1 Ramachandran Plot

A Ramachandran plot (also known as a Ramachandran diagram or a ϕ-ψ plot), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan[139] is a way to visualize backbone dihedral angles ψ against ϕ of amino acid residues in protein structure. In a protein the backbone dihedral angles (angles between two planes) are called ϕ (involving backbone atoms C'-N-C^{α}-C'), ψ, (involving the backbone atoms N-C^{α}-C'-N) and ω (involving backbone atoms C^{α}-C'-N-C^{α}). A contour plot of the ϕ-ψ

angles of a protein is called a Ramachandran plot, named after G.N. Ramachandran who showed that amino acids in a protein structure were restricted to limited range of conformations. An examination of the X-ray structures of proteins shows that most amino acids occupy one of the low energy regions in the Ramachandran plot and it has become common practice to construct Ramachandran plots when assessing an X-ray or NMR structure.

Figure 2.3 shows a typical Ramachandran plot obtained for a protein containing both β-sheets and α-helices. The different regions of the plot are represented by different colourings, as described in Morris *et al.*[140]. The darkest areas (here shown in red) correspond to the "core" regions representing the most favourable combinations of ψ-ϕ values. Ideally, one would hope to have over 90% of the residues in these "core" regions. The percentage of residues in the "core" regions is one of the better guides to check the stereochemical quality of a model.

The different regions evident on the plot were obtained from the observed ψ-ϕ distribution for 121,870 residues from 463 known X-ray protein structures. The two most favoured regions are the "core" and "allowed" regions which relate to 10° x 10° pixels having more than 100 and 8 residues in them, respectively. The "generous" regions were defined by Morris *et al.* by extending out by 20° (two pixels) all round the "allowed" regions. The authors of the work found very few residues in these "generous" regions, so they can probably be treated much like the "disallowed" region and any residues in them investigated more closely.

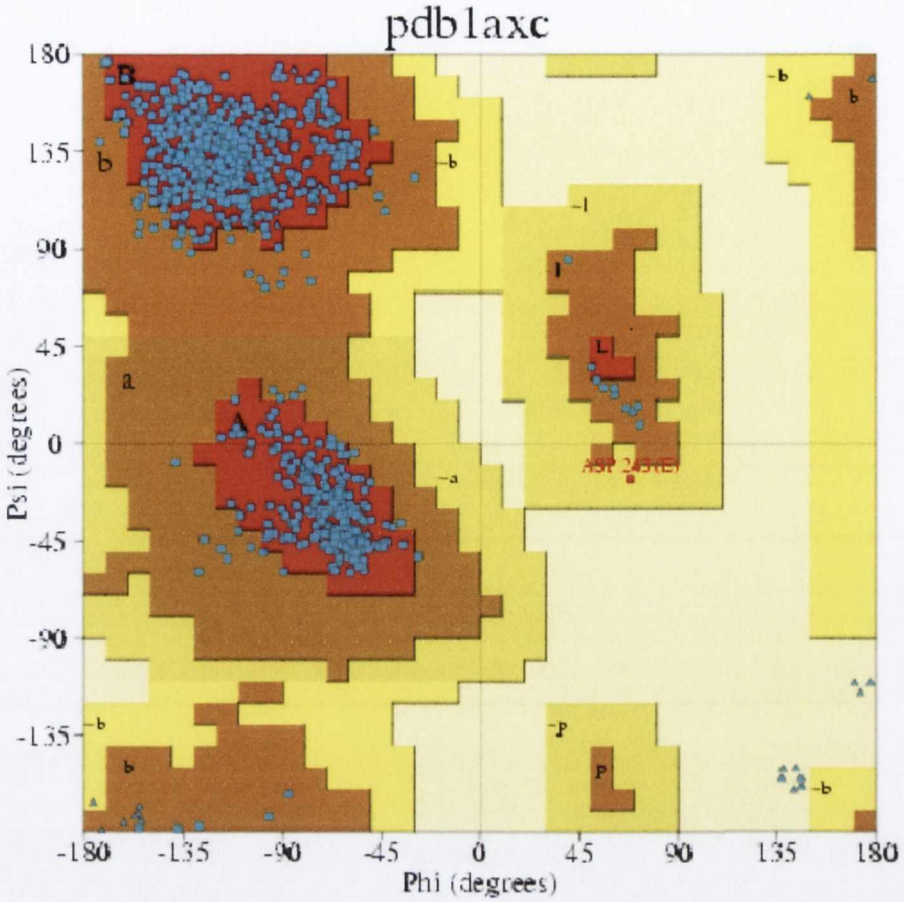


FIGURE 2.3: Ramachandran plot generated from human PCNA, a trimeric DNA clamp protein. Red, brown and yellow regions represent favoured, allowed and generously allowed regions as identified by PROCHECK.

2.3.2 Main-chain Parameters

The Main-chain parameters analysis[141], Figure. 2.4, generates six graphs that show how the main-chain residues of the structure (represented by a solid black square) compare with well-refined structures at a similar resolution. The purple band stretching across the six plots represents the results from the well-refined structures and the central line is a least-squares fit to the mean trend as a function of resolution. The width of the band on either side of the least-squares fit line corresponds to a variation of one standard deviation about the mean. The six properties plotted are:

1. Ramachandran Plot quality. Measured by the percentage of the protein's residues that are in the core region of the Ramachandran plot, expected to be over 90% for reliable models. As the resolution gets poorer, this value decreases.

2. Peptide bond planarity. Measured by calculating the standard deviation of the protein structure's ω tosion angles. Smaller values indicate a tighter clustering around the ideal value of 180°.

3. Bad non-bonded interactions. Measured by the number of bad contacts per 100 residues, defined as contacts where the distance of closest approach is less than 2.6Å.

4. Alpha-carbon tetrahedral distortion. Measured by calculating the standard deviation of the zeta torion angle, a notional torsion angle that is not defined about any actual bond in the structure but is defined by the C^{α}, N, C, and C^{β} atoms within a given residue.

5. Main-chain hydrogen bond energy. Measured by the standard deviation of the hydrogen-bond energies for main-chain hydrogen bonds, calculated using the method of Kabsch and Sander[142].

6. Overall G-factor. A measure of the overall normality of the structure, obtained from an average of all the different G-factors for each residue in the structure.

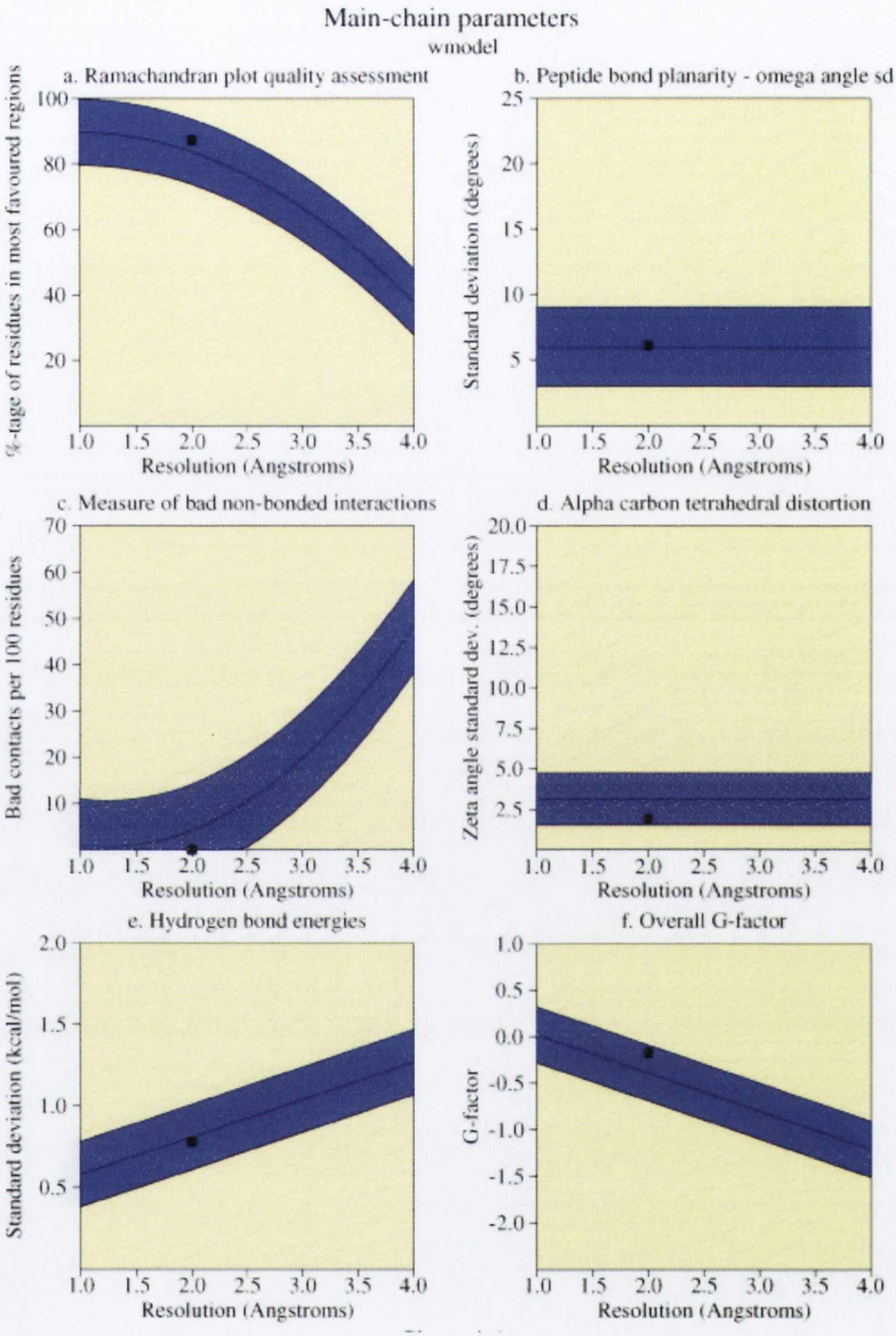


FIGURE 2.4: Example of Main-chain Parameters output from PROCHECK. This generates six graphs that show how the main-chain residues of the structure (represented by a solid black square) compare with well-reÞned structures at a similar resolution. The purple band stretching across the six plots represents the results from the well-reÞned structures and the central line is a least-squares Þt to the mean trend as a function of resolution. The width of the band on either side of the least-squares Þt line corresponds to a variation of one standard deviation about the mean

2.3.3 Side-chain Parameters

The Side-chain parameters analysis[143] generates five graphs that show how the structure (represented by a solid black square) compares with well-refined structures at a similar resolution. The purple band stretching across the six plots represents the results from the well-refined structures and the central line is a least-squares fit to the mean trend as a function of resolution. The width of the band on either side of the least-squares fit line corresponds to a variation of one standard deviation about the mean. Side-chain atoms of amino acids are named in the greek alphabet by the letter chi, or χ. Side chain torsion angles are thus numbered $\chi 1$, $\chi 2$, etc. For example, the amino acid Lysine has four χ angles, with $\chi 1$ occuring between C^{α} and the first atom in the side-chain. As such the $\chi 1$ angle is subject to certain restrictions which arise from steric hindrance between the side-chain atoms and the main-chain and the different conformations of the side chain as a function of $\chi 1$ are referred to as *gauche*(+), *trans*, and *gauche*(-) (fig 2.5).

The five properties plotted are:

1. Standard deviation of the $\chi 1$ gauche(-) torsion angles
2. Standard deviation of the $\chi 1$ trans torsion angles.
3. Standard deviation of the $\chi 1$ gauche(+) torsion angles.
4. Pooled standard deviation of all $\chi 1$ torsion angles.
5. Standard deviation of the $\chi 2$ trans torsion angles

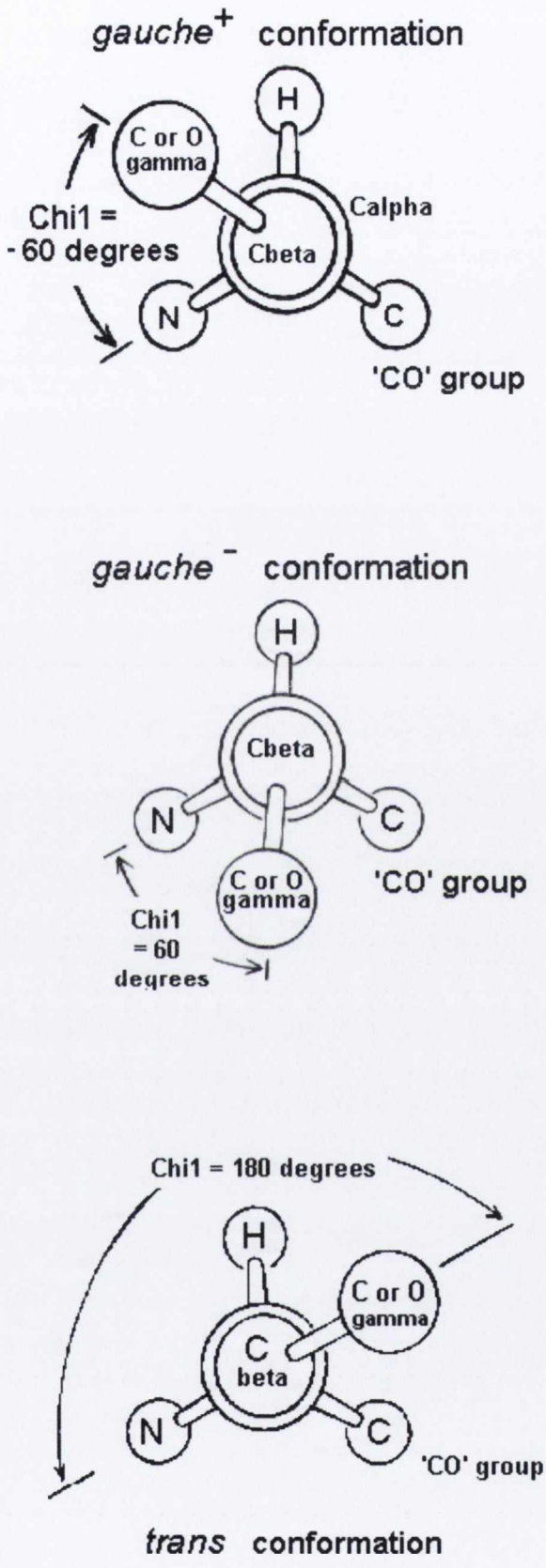


FIGURE 2.5: Illustration of the gauche(+), gauche(-), and trans conformations of amino acid sidechains viewed along the C^{β}-C^{α} bond. Adapted from [144].

2.4 Previous Modeling of Parasitic Tubulin: "Morrissette" Site Study

As stated in Chapter 1, the proposed site that has received the most attention on parasitic tubulin is the so called "Morrissette" site. In this work, homology modeling was carried out using the Swiss-Model Automated Comparative Protein Modeling Server[145] in the first approach mode, meaning an entirely automated approach where the user only supplies the protein sequence to be modeled. However, for the docking studies authors used Swiss-Model to build a complete model of bovine tubulin 1JFF[146] and this model in turn was used to create the *T. gondii* α-tubulin by mutating residues using WHAT IF software[147]. Briefly, the main idea behind the mutation option of WHAT IF is that the structure of the backbone, and the type and rotamer choice of the side-chain, influence each other. So if one wants to see the rotamer distribution for an alanine residue at position 17 in a sequence, WHAT IF takes the C^{α} positions of the residues 15-19 and looks at an internal database for all fragments of 5 residues that superpose well on this fragment and that have an alanine in the middle. This allows the user to pick and choose which residues in a protein structure are to be mutated, with the new residues having geometrically accepted orientations.

Swiss-Model constructs models using the same four-step procedure discussed earlier, namely template selection, sequence alignment, model building, then model evaluation. At the template selection stage, the Swiss-Model server template library ExPDB is extracted from the PDB. These entries are then split into individual protein chains and unreliable entries, e.g. low quality structures, are removed. Sequences in this template structure library are then searched with the target sequence. When template sequences have been identified, sequences are superposed using an iterative least squares algorithm. After incompatible templates are removed structural alignments are generated and a local pair-wise alignment of the target sequence to the template structure is calculated. Problems can arise here with respect to the presence of solitary residues in the final alignment, so-called "island" residues. These isolated residues are moved to the flanks in order to facilitate the process of loop construction. After an optimal alignment has been achieved, core modeling begins with an averaging of the backbone atom positions of the template structure. To generate the regions of insertions or deletions between the target-template alignment, an ensemble of fragments compatible with the neighbouring

stems is built using constraint space programming (CSP). In cases where CSP fails to offer a suitable solution, a loop library derived from experimental structures is used to find well suited loop fragments. Side-chains are then reconstructed based on the weighted positions of corresponding residues in the template structure, with conserved residues generated through iso-steric replacement of template side-chains with possible confirmations selected from a library of rotamers. Finally, a steepest descent energy minimisation is used to rid the structure of deviations in the protein geometry[148].

2.5 α-β Tubulin Heterodimer Homology Modeling Method

Amino acid sequences for alpha and beta-tubulin monomers from P flaciparum and T brucei were retrieved from the NCBI[149]protein sequence database in Fasta format (Accession numbers: *P. falciparum* α-tubulin CAA34101[150], *P. falciparum* β-tubulin AAA29780[151], *T. brucei* α-tubulin AAA30262[152], *T. brucei* β-tubulin AAA30261[152]). A sequence similarity search was performed using the PDB sequence search tool to identify tubulin proteins with resolved structures. BLAST was employed to align target and template sequences and each alignment was manually adjusted prior to model generation. Alignments were converted to modified PIR format for input into MODELLER. Automated homology model building was performed with Modeller9v5 operating on a Macbook pro 2.8 Ghz Intel Core 2 Duo running OSX v10.5.8. MODELLER was used to generate 100 models and these were evaluated using the program PROCHECK as well as MODELLER's own DOPE function. The discrete optimised protein energy (DOPE) score is an atomic distance-dependent statistical potential optimised for model assessment. Structure visualisation was carried out in MOE[153] and PyMOL.

MODELLER requires the alignment file to be in PIR format with the ".ali" exentsion. The body of this file is important for MODELLER to execute successfully. The first line of each sequence entry specifies the protein code after the >P1; identification line. The second line contains information necessary to extract the atomic coordinates of the sequence from the original PDB coordinate set. Each field in this line is separated by colon characters. The fields are:

1. **Field 1.** Specification of whether or not the 3D structure is available.
2. **Field 2.** The PDB code.
3. **Field 3-6.** The residue and chain identifiers for the first and last residue of the sequence given in the subsequent lines.
4. **Field 7.** Protein name (optional).
5. **Field 8.** Source of protein (optional).
6. **Field 9.** Resolution (optional).
7. **Field 10.** R-factor (optional).

The lines following this contain the aligned sequence. Chain breaks for multi-subunit proteins are denoted with a "/".

Once the targettemplate alignment is ready, MODELLER calculates a 3D model of the target completely automatically, using the automodel class. The model.py script will generate 100 similar models of your target based on the template.

Standard MODELLER Input File Model.py

```
# Homology modeling by the automodel class
from modeller import *
from modeller.automodel import *    # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

a = loopmodel(env,
              alnfile  = 'alignment.ali',     # alignment filename
              knowns   = '1TUB',              # codes of the templates
              sequence = '1Plfr')              # code of the target
a.starting_model= 1                 # index of the first model
a.ending_model  = 100                # index of the last model
                                    # (determines how many models to calculate)
a.md_level = None                   # No refinement of model

a.loop.starting_model = 1           # First loop model
a.loop.ending_model   = 100           # Last loop model
a.loop.md_level       = refine.fast # Loop model refinement level

a.make()                            # do homology modeling
```

In the model.py file the "knowns" tag specifies the name, as written in the .ali file, of the template structure. The "sequence" tag is the name as specified in the .ali file of the target sequence. The "assess_methods" tag ensures that, in this case, the DOPE and GA341 methods are used to evaluate the models generated. The second last line of the code dictates the number of models to generate. The loopmodel class specified in line 4 ensures that loop regions are automatically refined after building the standard models.

2.6 Results

2.6.1 Homology modeling of the *Plasmodium falciparum* tubulin heterodimer

P. falciparum tubulin subunits show a high level of amino acid similarity to mammalian tubulin. The alpha subunit displays 83% overall sequence similarity while the beta subunit shows 87% overall residue identity. Table 2.1 below lists tubulin crystal structures that have a high amino acid sequence similarity to Malarial alpha tubulin.

TABLE 2.1: Tubulin Crystal Structures available from the Brookhaven Protein Database.

PDB Code	Ligands	Resolution Å	Protein Source	Year
1Z2B[108]	GTP, GDP, Vinblastine, colchicine	4.10	Bovine	2005
1SA0[102]	GTP, GDP, colchicine	3.58	Bovine	2004
1SA1[102]	GTP, GDP, podophyllotoxin	4.20	Bovine	2004
1JFF[146]	GTP, GDP, taxol	3.50	Pig	2001
1FFX[154]	GTP, GDP	3.95	Bovine	2000
1TVK[155]	GTP, GDP, epitholine A	2.89	Bovine	2004
1TUB[156]	GTP, GDP, taxotere	3.70	Pig	1998

Using Clustal Omega[157], amino acid sequence alignments were prepared for *P. falciparum* α-tubulin with the α-tubulin sequences from 1JFF, 1SA0, 1SA1 (Bovine sources) and 1TUB (Pig source) (Figure 2.6). As the α-tubulin sequences had such a high level of homology, the alignment offered no problematic regions. The deciding factor on which structure to use as template for modeling came down to which crystal structure was the most complete. Several of the crystal structures available for template selection contain missing or weakly defined residues. Neither 1JFF or 1SA0 contain the initial methionine residue and also lack one α-tubulin loop with residues Q35-L60 missing. However, the

structure 1TUB contained a more complete protein and so was chosen as the template for modeling.

After model construction with MODELLER, the output structure was anaysed with PROCHECK package as well as through 3D structure analysis using MOE. The Ramachandran plot (Figure 2.7) shows 84.1% of residues to be in favoured regions (red), and 11.0% to be in allowed regions (yellow). Of the remaining residues, 2.5% were in generously allowed regions (beige) and 2.3% in disallowed regions (white). Based on structures with a resolution of at least 2.0Å, a model of very high quality should show up approximately 90% of residues in the favoured region.

Further analysis of values from the Ramachandran plot showed the main-chain residues to be of generally good quality (Figure 2.8). A bad contact score of 6.0 per 100 residues was calculated and an overall G-factor, a log-odds score based on the distributions of stereochemical parameters providing a measure of how unusual a given property is, of -0.4 was returned. Sidechain parameters were several deviations away from expected means and were of lesser quality. Side chain parameters were of lesser quality (Figure 2.9).

```
Malaria     MREVISIHVGQAGIQVGNACWELFCLEHGIQPDGQMPSDKASRANDDAFNTFFSETGAGK
1JFF        MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1sa0        MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1sa1        MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1TUB        MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
            *** *********:*:*******:*****************:  . **:************

Malaria     HVPRCVFVDLEPTVVDEVRTGTYRQLFHPEQLISGKEDAANNFARGHYTIGKEVIDVCLD
1JFF        HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1sa0        HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1sa1        HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1TUB        HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
            ****.*********:******************:********:**********:**: **

Malaria     RIRKLADNCTGLQGFLMFSAVGGGTGSGFGCLMLERLSVDYGKKSKLNFCCWPSPQVSTA
1JFF        RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1sa0        RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1sa1        RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1TUB        RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
            *******:******* :* :.******** .*::*************:*. :*:******

Malaria     VVEPYNSVLSTHSLLEHTDVAIMLDNEAIYDICRRNLDIERPTYTNLNRLIAQVISSLTA
1JFF        VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1sa0        VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1sa1        VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1TUB        VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
            *******:*:***: ***:* *:*:*******************************.*::**:**

Malaria     SLRFDGALNVDVTEFQTNLVPYPRIHFMLSSYAPVVSAEKAYHEQLSVSEITNSAFEPAN
1JFF        SLRFDGALNVDLTEFQTNLVPYPRGHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1sa0        SLRFDGALNVDLTEFQTNLVPYPRIHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1sa1        SLRFDGALNVDLTEFQTNLVPYPRIHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1TUB        SLRFDGALNVDLTEFQTNLVPYPRGHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
            **********:*********** ** *:*****:***********:****:.*****

Malaria     MMAKCDPRHGKYMACCLMYRGDVVPKDVNAAVATIKTKRTIQFVDWCPTGFKCGINYQPP
1JFF        QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1sa0        QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1sa1        QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1TUB        QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
             *.**************:************:******************** ****:**

Malaria     TVVPGGDLAKVMRAVCMISNSTAIAEVFSRMDQKFDLMYAKRAFVHWYVGEGMEEGEFSE
1JFF        TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1sa0        TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1sa1        TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1TUB        TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
            *********** *****:**:*****.::*:*:*******************************

Malaria     AREDLAALEKDYEEVGIESNEAEGEDEGYEADY
1JFF        AREDMAALEKDYEEVGVDSVEGEGEEEGEEY--
1sa0        AREDMAALEKDYEEVGVDSVEGEGEEEGEEY--
1sa1        AREDMAALEKDYEEVGVDSVEGEGEEEGEEY--
1TUB        AREDMAALEKDYEEVGVDSV-------------
            ****:***********::*
```

FIGURE 2.6: Amino acid sequence alignment of *P. falciparum* alpha tubulin with the alpha tubulin sequences from bovine and pig sources.

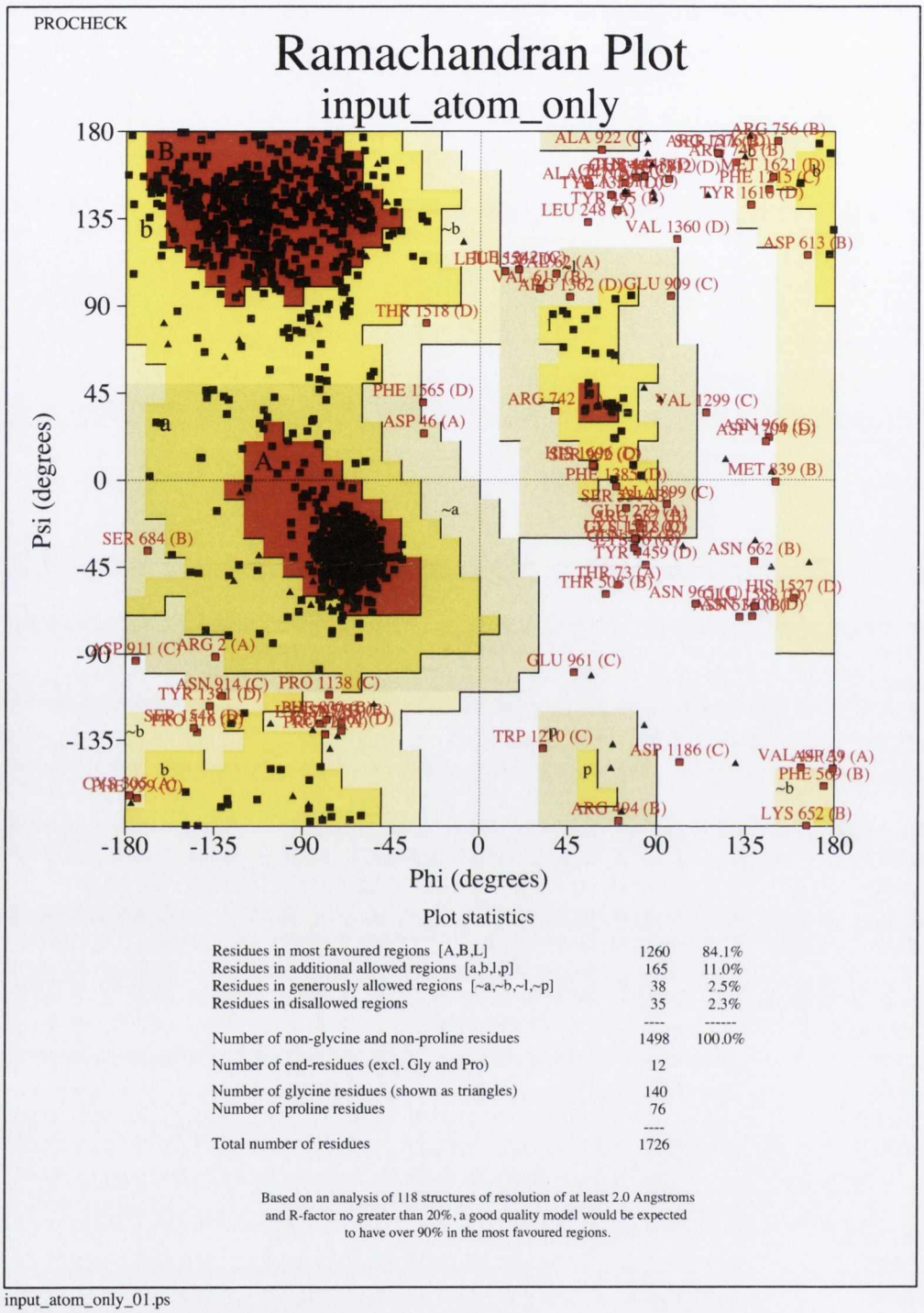


FIGURE 2.7: Ramachandran plot for the homology model of *P. falciparum* tubulin.

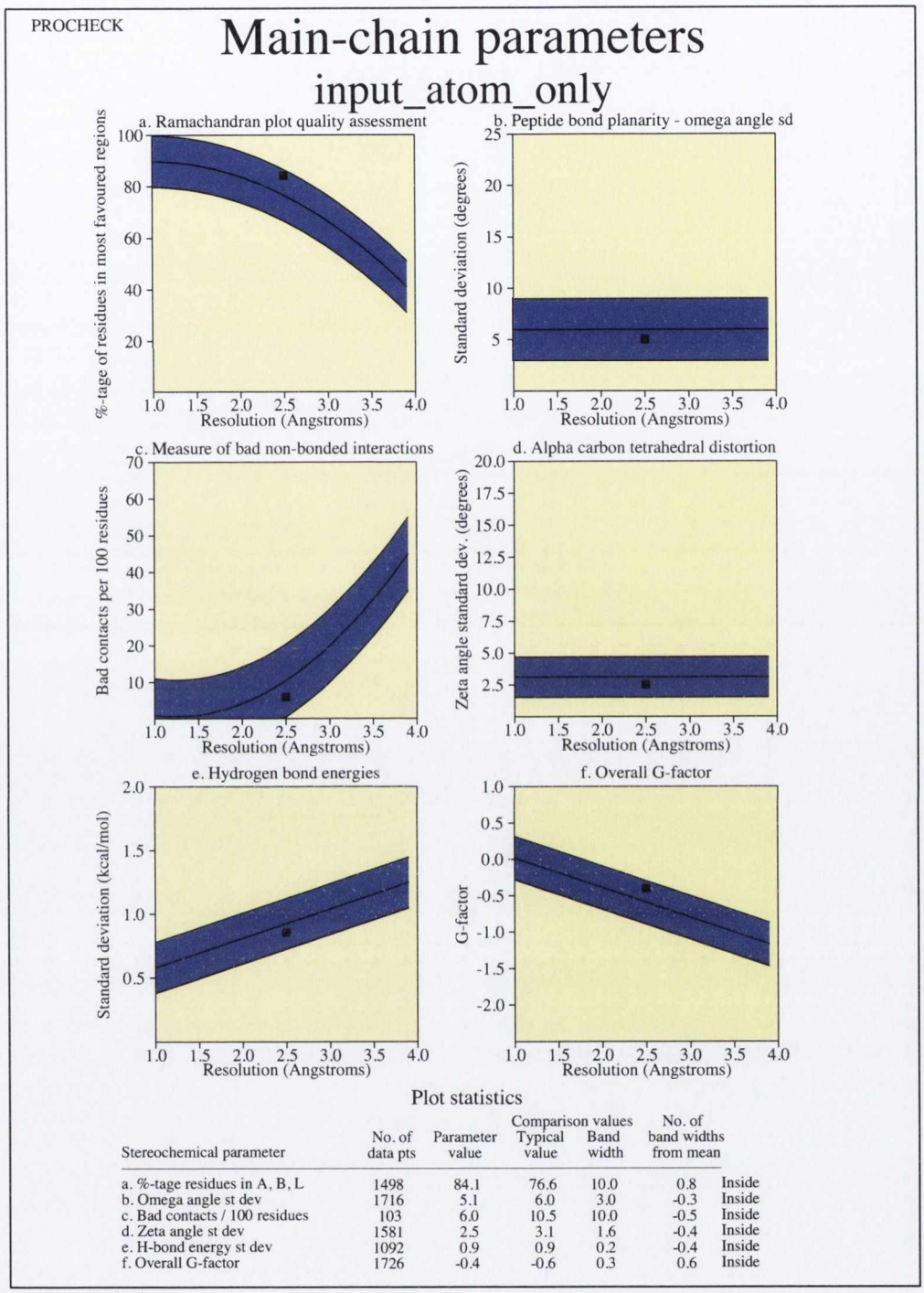


FIGURE 2.8: Main chain parameter analysis for the homology model of *P. falciparum* tubulinl.

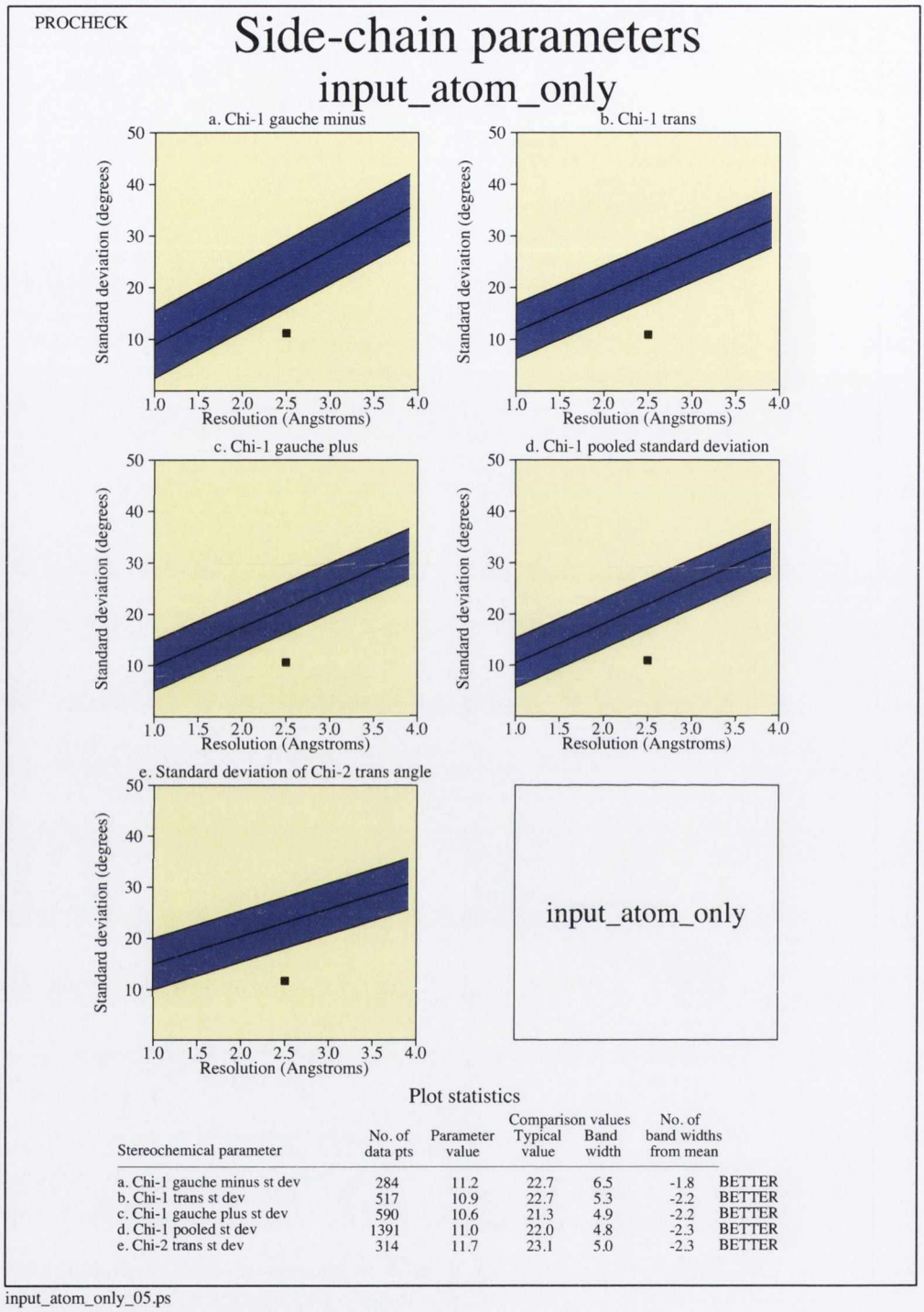


FIGURE 2.9: Side chain analysis for the homology model of *P. falciparum* tubulin.

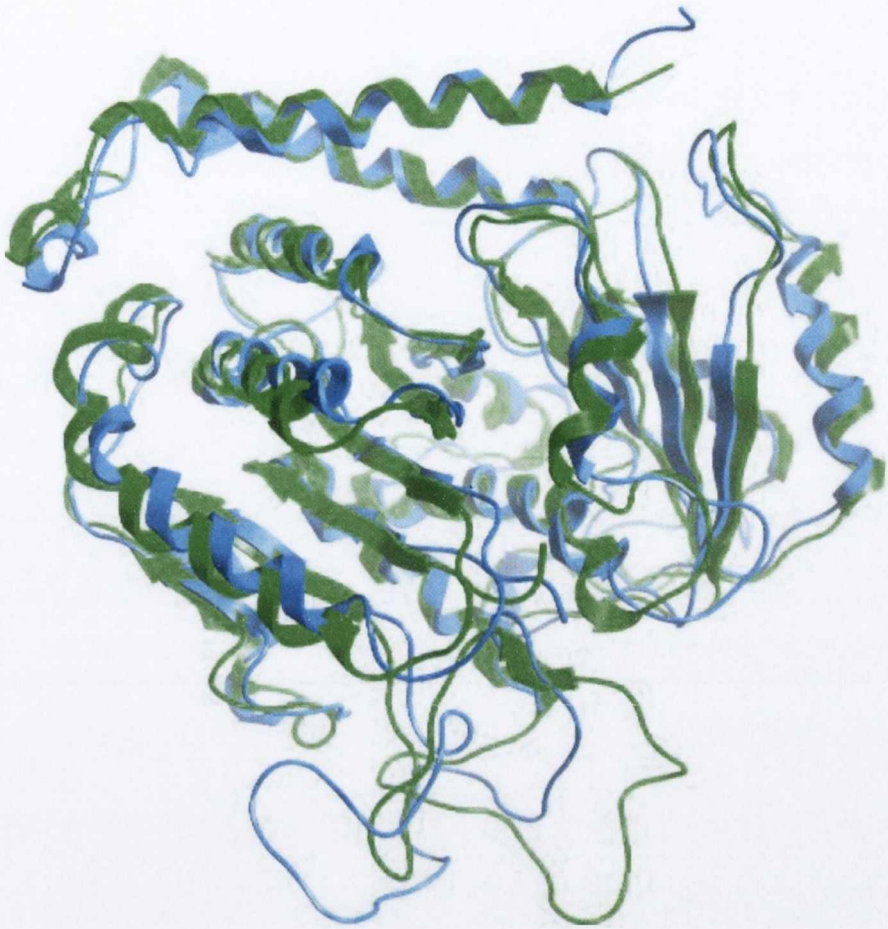


FIGURE 2.10: Alignment of the *P. falciparum* homology model (blue) and the template structure 1TUB (green).

2.6.2 Homology modeling of *Trypanosoma brucei* tubulin heterodimer

Trypanosoma brucei alpha and beta subunits have a sequence similarity with their porcine counterparts of 84% and 84% respectively. As in the previous case, the alignment put forward no problems due to the highly conserved nature of the protein. After model building, the structure was analsed with PROCHECK and results are presented in Figures 2.12 - 2.14. The Ramachandran plot shows 86.8% of residues to be in favoured regions (red), and 9.5% to be in additionally allowed regions (yellow). Of the remaining residues, 2.6% were in generously allowed regions (beige) and 1.1% in disallowed regions (white). Further analysis of values from the Rmachandran plot displayed main-chain residues occuring within allowed deviations from mean values. A bad contact score of

6.1 per 100 residues was generated and an overall G-factor of -0.2. A three dimensional structure representation of 1TUB versus the trypanosomal model is shown in figure 2.10. An RMSD of 3.725Å was calculated, with the majority of the deviations occuring in the poorly resolved loop regions.

Overall the models produced have well defined structures (Figures 2.10 and 2.15), although areas in the template that were missing show more variation in the models. Each monomer of the modeled heterodimers retain the 3 well-defined regions described in Chapter 1. The N-terminal nucleotide binding domain's 6 parallel β-strands and alternating helices are present, as is a well defined intermediate domain composed of 5 α-helices and a mixed β-sheet. The C terminal, a region missing from residue 438 onwards in the template structure, forms two anti-parallel helices that cross over the previous two domains and displays some variation across the models as would be expected.

```
tryp      MREAICIHIGQAGCQVGNACWELFCLEHGIQPDGAMPSDKTIGVEDDAFNTFFSETGAGK
1JFF      MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1sa0      MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1sa1      MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
1TUB      MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK
          ***.*.**:**** *:********:********** ********  **:**********

tryp      HVPRAVFLDLEPTVVDEVRTGTYRQLFHPEQLISGKEDAANNYARGHYTIGKEIVDLCLD
1JFF      HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1sa0      HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1sa1      HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
1TUB      HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD
          *******:******:******************:********************:** **

tryp      RIRKLADNCTGLQGFLVYHAVGGGTGSGLGALLLERLSVDYGKKSKLGYTVYPSPQVSTA
1JFF      RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1sa0      RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1sa1      RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
1TUB      RIRKLADQCTGLQGFSVFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFSIYPAPQVSTA
          *******:******* *:*:.*******: :***:************** :::***:******

tryp      VVEPYNSVLSTHSLLEHTDVAAMLDNEAIYDLTRRNLDIERPTYTNLNRLIGQVVSSLTA
1JFF      VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1sa0      VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1sa1      VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
1TUB      VVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA
          *******:*:**: ***:* * *:********: **********************:***:**

tryp      SLRFDGALNVDLTEFQTNLVPYPRIHFVLTSYAPVISAEKAYHEQLSVSEISNAVFEPAS
1JFF      SLRFDGALNVDLTEFQTNLVPYPRGHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1sa0      SLRFDGALNVDLTEFQTNLVPYPRIHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1sa1      SLRFDGALNVDLTEFQTNLVPYPRIHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
1TUB      SLRFDGALNVDLTEFQTNLVPYPRGHFPLATYAPVISAEKAYHEQLSVAEITNACFEPAN
          ************************ ** *:::********************:**:** ****.

tryp      MMTKCDPRHGKYMACCLMYRGDVVPKDVNAAVATIKTKRTIQFVDWSPTGFKCGINYQPP
1JFF      QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1sa0      QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1sa1      QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
1TUB      QMVKCDPRHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEPP
           *.**************:************:**************.***** ****:**

tryp      TVVPGGDLAKVQRAVCMIANSTAIAEVFARIDHKFDLMYSKRAFVHWYVGEGMEEGEFSE
1JFF      TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1sa0      TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1sa1      TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
1TUB      TVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYAKRAFVHWYVGEGMEEGEFSE
          ******************:*:*****.:**:*********:**********************

tryp      AREDLAALEKDYEEVGAESADMDGEEDVEEY
1JFF      AREDMAALEKDYEEVGVDSVEGEGEEEGEEY
1sa0      AREDMAALEKDYEEVGVDSVEGEGEEEGEEY
1sa1      AREDMAALEKDYEEVGVDSVEGEGEEEGEEY
1TUB      AREDMAALEKDYEEVGVDSV----------
          ****:***********.:*.
```

FIGURE 2.11: Amino acid sequence alignment of *T. brucei* alpha tubulin with the alpha tubulin sequences from bovine and pig sources.

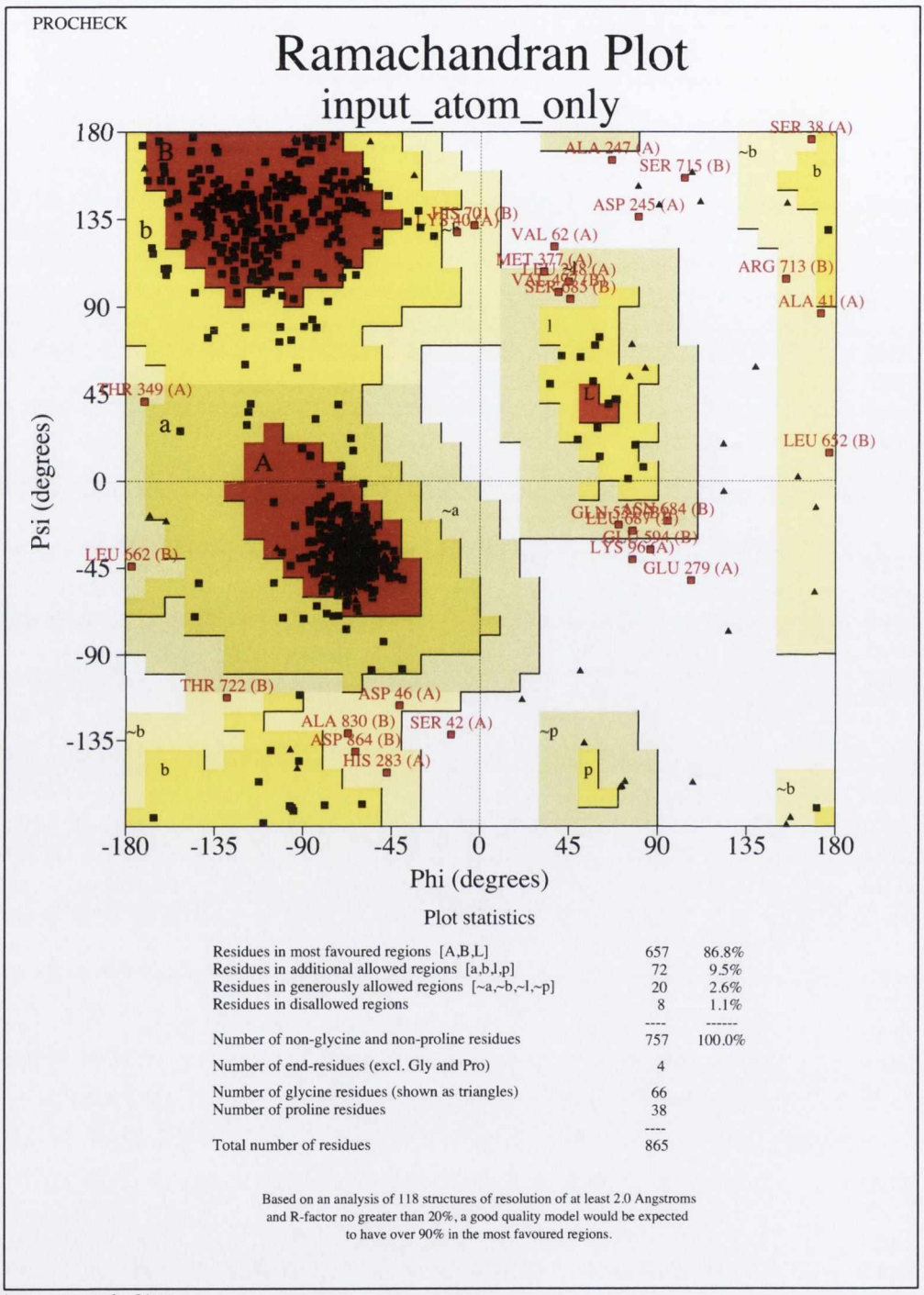
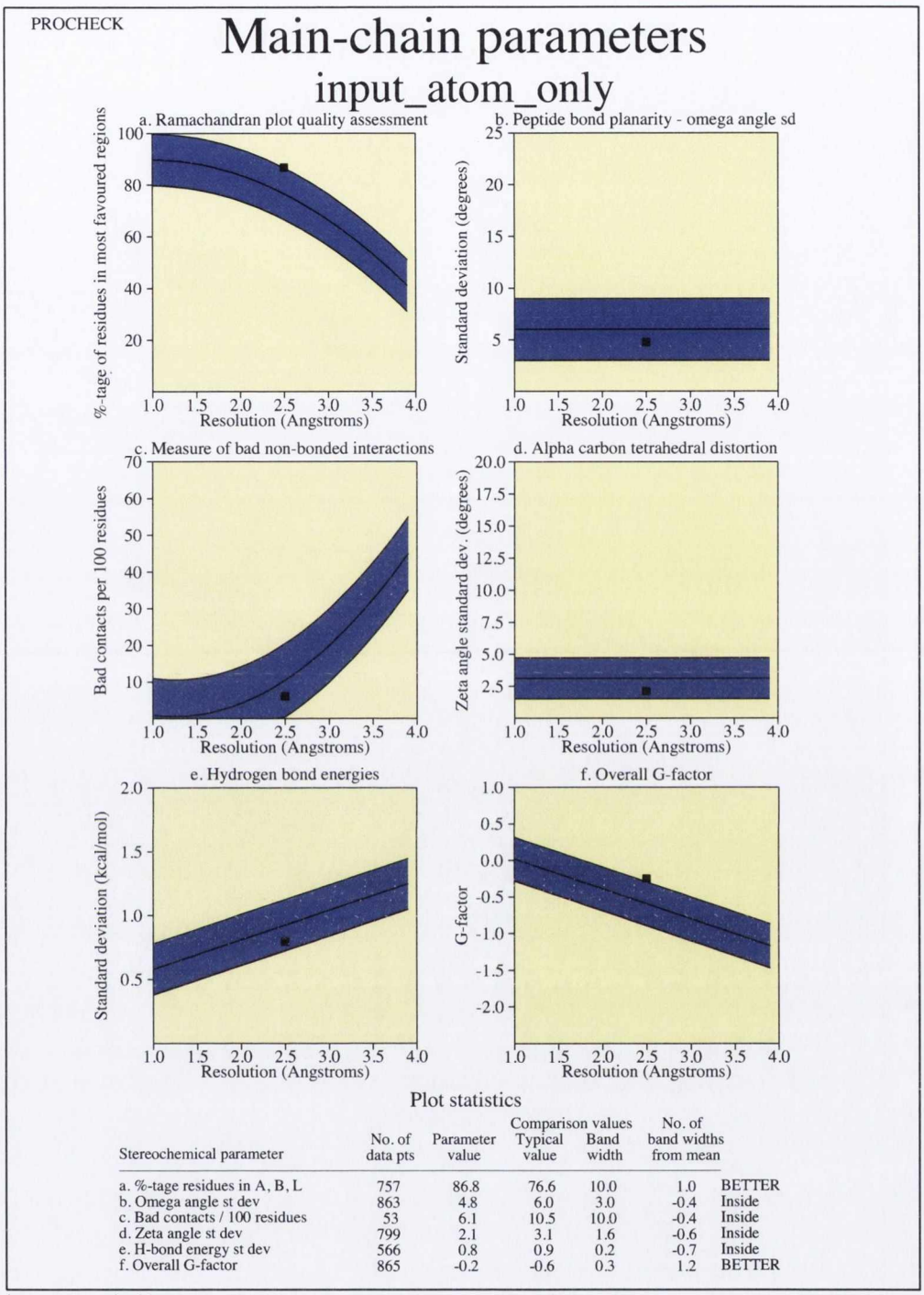


FIGURE 2.12: Ramachandran plot for the homology model of *T. brucei* tubulin.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	757	86.8	76.6	10.0	1.0	BETTER
b. Omega angle st dev	863	4.8	6.0	3.0	-0.4	Inside
c. Bad contacts / 100 residues	53	6.1	10.5	10.0	-0.4	Inside
d. Zeta angle st dev	799	2.1	3.1	1.6	-0.6	Inside
e. H-bond energy st dev	566	0.8	0.9	0.2	-0.7	Inside
f. Overall G-factor	865	-0.2	-0.6	0.3	1.2	BETTER

input_atom_only_04.ps

FIGURE 2.13: Main chain analysis for the homology model of *T. brucei* tubulin.

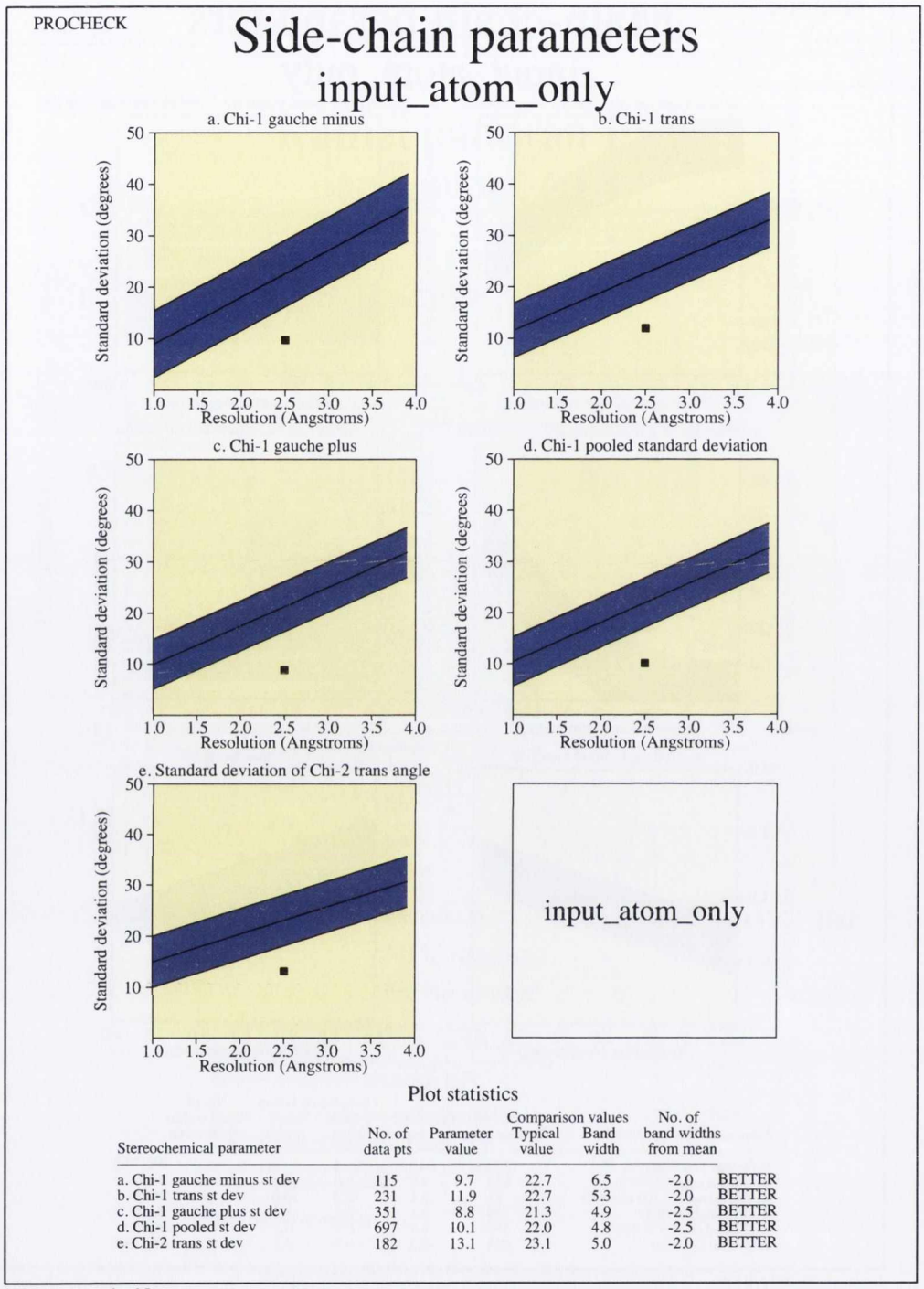


FIGURE 2.14: Side chain analysis for the homology model of *T. brucei* tubulin.

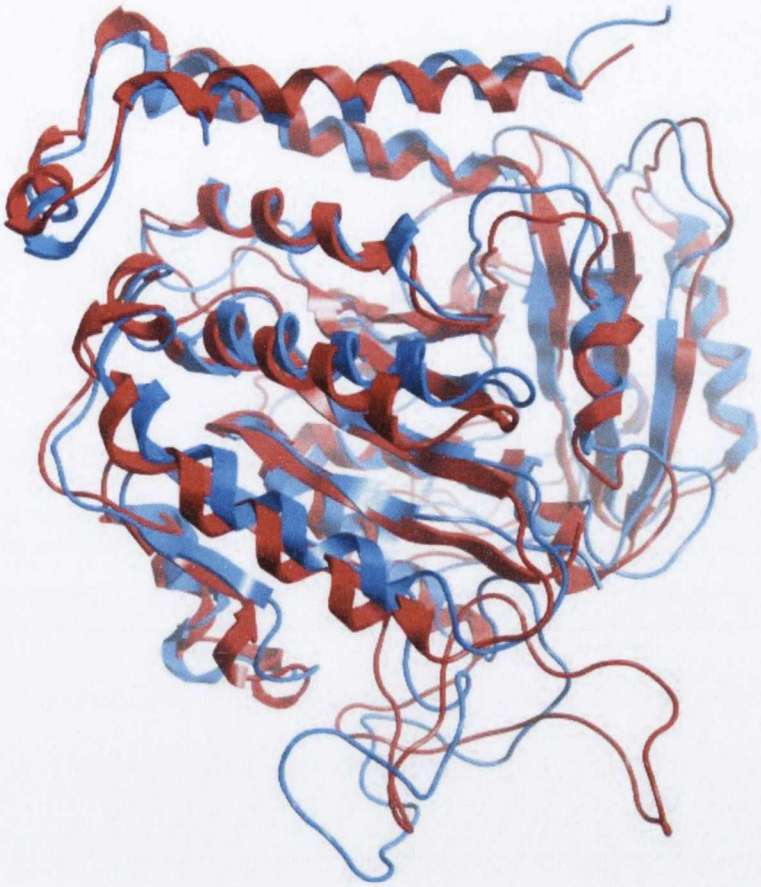


FIGURE 2.15: Alignment of the *T. brucei* homology model (blue) and the template structure 1TUB (red).

2.7 Discussion and Conclusions

Rational drug design techniques that require protein structural information are severely limited by the availability of experimentally resolved structures. Even though the Protein Data Bank expands daily, only a tiny amount of all known proteins have been experimentally resolved, leading to the increasing development and use of powerful software for sequence alignment, fold recognition, and protein modeling. Despite its limitations, homology modeling has been successfully used in several different ways. Homology models have been employed to design mutagenesis experiments[158,159] and, more commonly, to design new ligands[160,161].

Homology modeling of the parasitic tubulin heterodimer was straightforward in terms of sequence homology to experimentally derived structures, however, this was offset by the presence of missing and poorly resolved regions in the template structures which made modeling of small regions of the protein more problematic. As can be seen in the aligned structures of template and query, the majority of the variation observed was located in the loop region composed of residues 30-61. Previous efforts at modeling parasitic and plant tubulin structures have dealt with missing template regions in different ways. Morrissette *et al.*[122] managed this by using the corresponding loop in β-tubulin to model the H1-B2 loop in α-tubulin of *T. gondii*. In modeling the alpha subunit of Green Foxtail (*Setaria viridis* L. Beauv.), Delye *et al.*[120] ignored the missing residues as the herbicide binding site they proposed was not in the same region of the protein. Blume *et al.*[121] also created a homology model of plant tubulin, this time *Eleusine indica*, in an attempt to establish the binding site of the dinitroaniline and phosphoroamidate herbicides. However, nowhere in the paper do they detail the homology model construction, instead referring to a previous paper, written in Russian, for explanations. In this specific case we chose to let MODELLER select the best orientation for the loop region in question as at the time of model creation there were no complete tubulin crystal structures and the method used by Morrissette *et al.* introduced an unknown into an approach that comes with its own uncertanties. Figure 2.16 below shows the orientation of the H1-B2 loop (red) in the *P. falciparum* homology model. This loop adopts a different conformation compared to the equivalent loop region of the models made by Morrissette *et al*[122] (Figure 1.19, Chapter 1). This disparate structure of the models is a direct result of the different methods used to create them, but also a reflection of the

uncertainty of the final models generated by homology modeling. It is for these reasons that a molecular dynamics approach to putative binding site elucidation is preferable to simply taking the structures as is.

The structure 1TUB was chosen as the modeling template for both parasitic models due to its relative intactness compared to other available structures at the time and MODELLER was successfully used to build models for both malarial and trypanosomal tubulin. PROCHECK analysis revealed that the models generated were of good quality and could be used for further studies. In an ideal situation, experimentally derived structures would be the chosen route for structure and ligand-based design. However, in this specific case experimentally derived structures were unavailable and so homology modeling was the only recourse.

In conclusion, two homology models or malarial and trypanosomal tubulin heterodimers have been generated for the purpose of a computational approach to ascertain the putative binding site of two related classes of herbicide compounds on the tubulin protein.

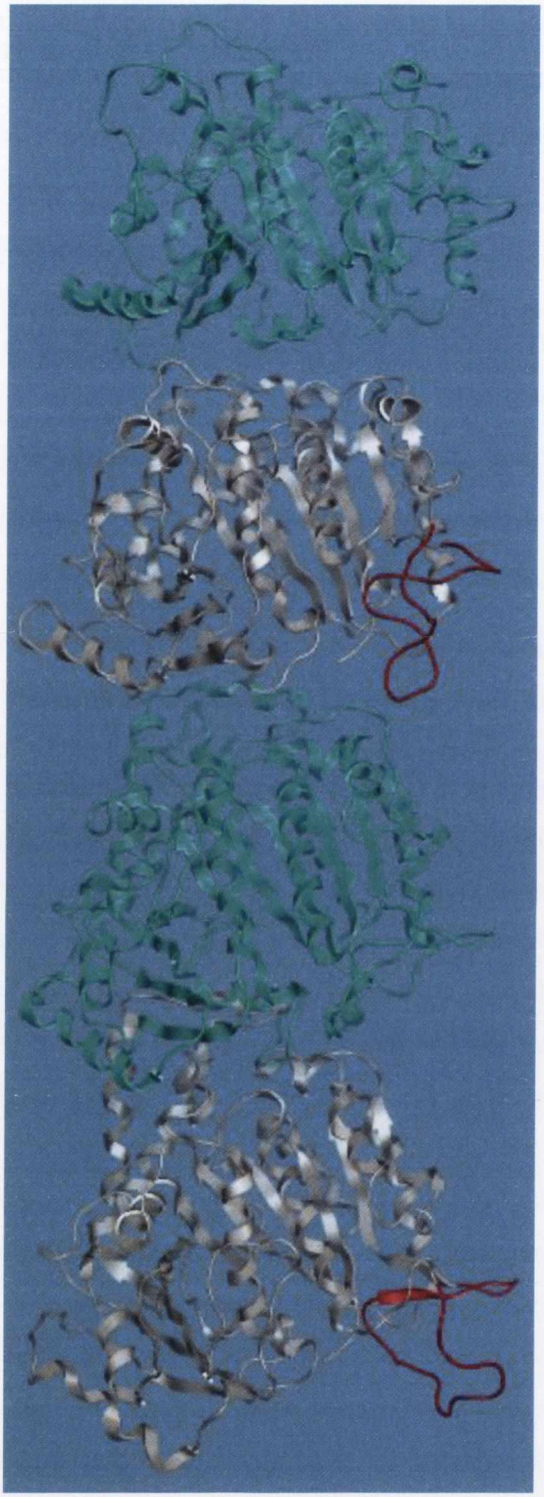


FIGURE 2.16: This image shows the orientation of the H1-B2 loop (red) in the *P. falciparum* homology model. This loop adopts a different conformation compared to the equivalent loop region of the models made by Morrissette *et al*[122] (Figure 1.19, Chapter 1)

Chapter 3
Molecular Dynamic Studies of Parasitic Tubulin Models

3.1 Introduction

3.2 Molecular Dynamics Simulations

Molecular dynamic (MD) simulations are employed to calculate movements of molecules, including the intramolecular atom motions, as a function of time. This is done by solving Newton's second law of motion in very small time steps (0.5 - 2.5 fs/step) that may be repeated billions of times, resulting in a continuous atom trajectory with sequential positions and velocities of the simulated system. During the MD simulation, force fields containing parameters describing how different types of atoms interact with each other are used to model the forces caused by atomic interactions that cause atoms to move.

The starting point for a MD simulation is a set of coordinates and initial velocities usually prepared from X-ray, NMR or homology modeling data. Using an equation similar to Figure 3.1 below, the forces acting on each of the atoms in the system is calculated, then each atom is moved according to those forces, then the simulation time is moved forward, and then the whole process is repeated many times.

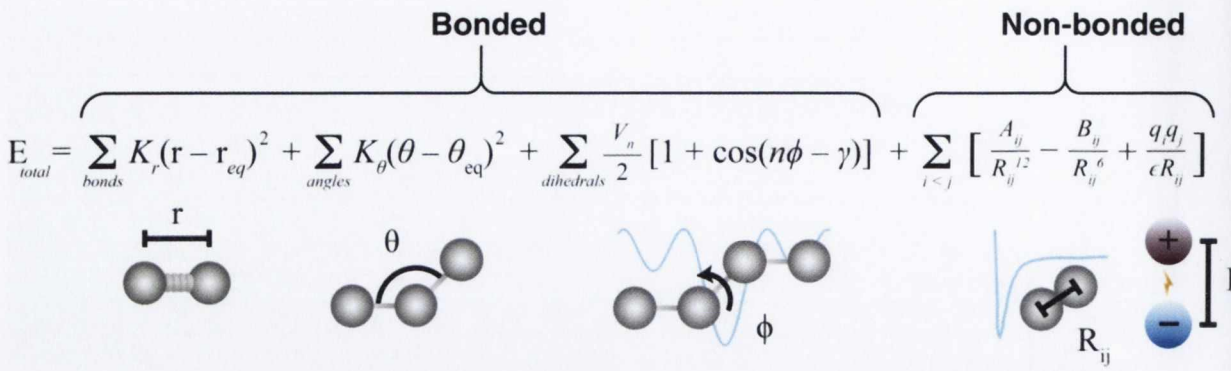


FIGURE 3.1: The atomic forces that govern molecular movement can be divided into those caused by interactions between atoms that are chemically bonded to one another and those caused by interactions between atoms that are not bonded. Chemical bonds and atomic angles are modeled using simple springs, and dihedral angles (that is, rotations about a bond) are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Non-bonded forces arise due to van der Waals interactions, modeled using the Lennard-Jones potential, and charged (electrostatic) interactions, modeled using Coulomb's law[162]

The contributions to the equation are made by forces derived from interactions between bonded and non-bonded atoms. The non-bonded forces result from van der Waals interactions and are modeled using the Lennard-Jones 6-12 potential with Coulomb's law employed to model electrostatic interactions. Dihedral angles are modeled using a sinusoidal function that approximates energy differences between eclipsed and staggered conformations, while chemical bonds and atomic angles are calculated using simple virtual springs.

The product of a MD simulation is a series of sequential snapshots of the protein, allowing detailed analysis of protein and binding site movements over time. These insights into protein motion can play an important role in new drug discovery. Structures solved by X-ray crystallography provide limited information regarding protein dynamics, but drug binding, in most cases, is a very dynamic process involving a flexible ligand and a receptor in constant motion. MD simulations allow us, for example, to look at how a binding site may change in volume over time, as well as showing up transient or cryptic binding sites and allosteric binding sites. The integration of protein motion in binding site prediction offers another avenue in the discovery of new drugs.

3.2.1 Amber Molecular Dynamics Software

Amber[163] (**A**ssisted **M**odel **B**uilding with **E**nergy **R**efinement) is a suite of programs used for executing molecular dynamics simulations and is also a collection of mechanical force fields. In the context of molecular dynamics simulations, the term "force field" refers to the combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates. A typical molecular dynamics simulation approach includes an initial structure generation step, a minimization step, an equilibration step, and production dynamics run with capture of structure conformations at regular intervals, and a trajectory analysis (Figure 3.2).

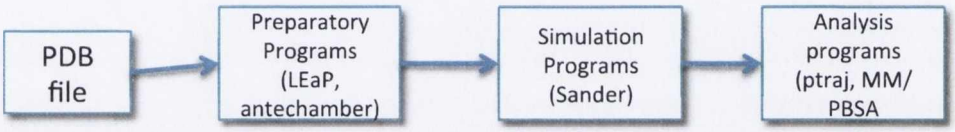


FIGURE 3.2: Amber program suite workflow. The PDB file is prepared using LEaP and if ligands are included they are prepared with antechamber. Sander is the module that performs md simulations and trajectories are analysed using ptraj.

LEaP is used in the protein preparation steps in order to generate a set of topology (containing atom types, charges and connectivity information) and coordinate (includes the force constants necessary to describe the system in terms of bond, angle torsion energies and non-bonded interaction energies) files from the input PDB structure, using the force field parameters from ff99sb. This step also includes the placement of counter-ions and the generation of a water box around the protein. For ligand preparation in Amber a separate module is used, known as antechamber. In generating topology files, antechamber uses information from the generalized amber force field (GAFF)[164]which contains parameters suitable to cover pharmaceutical small molecule chemical space, consisting of 33 basic atom types and 22 special atom types. Charge is calculated using HF/6-31G[165] or AM1-BCC[166] models.

Running an MD simulation with *sander* (**S**imulated **A**nnealing with **N**MR -**D**erived **E**nergy **R**estraints) requires three file types:

* *prmtp* file containing a description of molecular topology and the necessaty force field parameters
* *inpcrd* file, containing a description of atomic coordinates and current periodic box dimensions
* *mdin* file, the sander input file consisting of a series of namelists and control variables that contain the simulation run options.

3.3 Ligand-Binding Site Elucidation

There is a distinct lack of definition of what constitutes a ligand-binding site. Studies have shown that pocket volume is of importance for predicting binding sites; the average volume of a drug-binding cavity was found to be roughly 930Å using a geometry-based method. On the ligand side the average volume of drug-like compounds was gathered to be above 200Å and usually around 439Å mark, although the size of a ligand is not necessarily linked to the size of its binding site. Taken together, it is easy to see why coming up with a universal method for binding site prediction is such a difficult task when one considers the variation in shape, size and composition of binding pockets. This problem is exemplified by the fact that very similar ligands can bind to unrelated protein partners and highly related proteins can bind to many different ligands.
Access to the thousands of protein-ligand complexes in the PDB has enabled detailed studies concerning ligand-interacting residues in binding sites. One such study investigated the amino acid composition of ligand-binding sites and compared these to protein-protein binding interfaces and the protein surface in general. The research concluded that while residue composition of the general protein surface was comparable to protein-protein interfaces, there were considerable differences in residue propensities between these and ligand-binding sites with Cys, His, Met, Phe, Trp, Val, and Ile residues appearing most frequently.

The detection, comparison and analyses of potential ligand-binding pockets is of particular interest to structure-based drug design efforts. Knowledge of the location and

nature of ligand-protein binding events lends itself to computational ligand docking, virtual screening and novel compound design. The main tools for predicting the presence of binding pockets can be grouped into two classes: geometry-based methods and energy-based methods.

3.3.1 Geometry-Based Methods

Geometry-based methods of pocket prediction scan the protein along grid lines using a geometric probe to detect regions where grid points lay outside protein atoms. The program POCKET[167] was one of the first algorithms to employ this method of identifying cavities on protein surfaces. The protein is mapped onto a 3D grid, points of which are part of the protein if it is within 3Å of an atom coordinate, otherwise it is solvent. Next, x, y, and z-axes are scanned for pockets, which are distinguished by a sequence (known as a protein-solvent-protein event) of grid points which start and end with the protein label and a period of solvent grid points in between. Grid points that exceed a minimum threshold of protein-solvent-protein events are retained in the final pocket prediction. LIGSITE[168] is an enhancement of the POCKET program. With POCKET, the definition of a binding pocket is dependent on the angle of rotation of the protein relative to the grid axes. LIGSITE extends this by scanning along the four cubic diagonals in addition to the x, y and z directions.

PASS[169] is geometry-based pocket prediction tool that uses a probe-packing algorithm to detect protein pockets. Probe spheres are packed on the protein surface so that each probe touches a triplet of adjacent protein atoms. Probes that clash with protein atoms are removed and only probes with a burial count (the number of atoms within $8AA$ distance) above a threshold limit are retained. These steps are iterated until no new spheres can be placed. Probes are assigned a weight proportional to the number of probe spheres surrounding and then probes are clustered into active site points that define the pockets.

3.3.2 Energy-Based Methods

GRID[170] was one of the first pocket detection methods that took advantage of the interaction energy between the protein and chemical probes to evaluate the presence of

binding pockets. This method creates a grid around the protein with the interaction energy, composed of Lennard-Jones, Coulombic, and H-bonding terms, calculated between the protein and probes placed on the grid points. This method does not strictly identify binding pockets, rather it picks out regions of interest on the protein surface. SiteMap[171], part of the Schrodinger software suite, is a more modern approach that is more geared towards detecting drug binding sites. SiteMap uses energy calculations on a grid and retains probe points that obtain a favourable van der Waals interaction energy with the protein and are located outside the protein but have a level of enclosure within the protein. Predicted sites are scored in relation to their size, solvent exposure, hydrophobicity, and hydrogen-bonding potential among others.

3.4 MDpocket

While the programs described perviously are designed to detect pockets in static structures, the inherent plasticity of binding pockets represents a limitation to these methods. MDpocket[172]is able to identify and characterise binding sites and channels that are transiently formed in the protein through the analysis of conformational ensembles generated through molecular dynamic simulations. This program relies on 3 other programs: (i) sfpocket, a fast geometry-based cavity detection algorithm, (ii) dpocket which extracts pocket descriptors, and (iii) tpocket, a program that allows assessment of pocket scoring functions.

MDpocket takes as input a directory of MD trajectory snapshots in PDB format with the fpocket[173] program employed to detect binding pockets in the snapshots. Briefly, this platform relies on the concept of alpha spheres, a sphere that contacts four atoms on its boundary and that contains no internal atom. Having identified an ensemble of alpha spheres, fpocket then identifies clusters of spheres close together to locate pockets. The final step is to calculate properties from the pocket atoms in order to score the pocket. MDpocket provides information on the plasticity of detected pockets from normalised frequency and density maps, valuable output in the exploration of site opening and closing. Alpha spheres from different snapshots appear as two shades of grey. A 1Å grid is superimposed on to the alpha spheres and on each grid point the density of the surrounding alpha spheres and frequency are tracked. MDpocket allows the user to specify a value termed the Isovalue (expressed as the number of alpha sphere centers in a

8Å^3 cube around each grid point) which enables user to adjust the frequency with which a pocket must be open across the snapshots in order to appear as a binding site in the output. A smaller isovalue allows the detection of smaller pockets that may only be open for a short duration of the trajectory whereas a higher value will detect only conserved pockets. Additionally, this program also possesses a feature that allows the user to select specific binding pocket grid points and subject these to a second MDpocket analyses to measure volume, polar and non-polar surface area, and solvent accessible surface area of the pockets.

3.5 Docking and Scoring

As the number of protein structures available in the public domain has increased significantly in the last decade, structure-based virtual screening has become a tool that is widely used in the design and optimisation of novel drug candidates. Docking software simulates an *in silico* interaction of a molecule with a given target and can be applied to virtual screening of chemical databases and to lead optimisation. Most docking software makes several assumptions during the docking and scoring processes: (1) the docking site is usually user-defined, (2)water molecules are generally not considered during the calculations, and (3) the protein is frequently assumed as rigid even though protein-ligand interactions are dynamic and variable.

Calculating accurate protein-ligand interactions and binding energies is a key principle behind structure-based drug discovery. There are two main steps involved in molecular docking: (1) The precise pose prediction of the ligand inside the binding site (performed by the docking algorithm), and (2) the accurate prediction of binding free energy of the docked pose, calculated by a scoring function. The docking-challenge is twofold: the first challenge concerns the handling of conformational flexibility, most often regarding the ligand, but sometimes also receptor flexibility; the second challenge is the prediction of binding free energy. 'On the fly' conformational search methods may be classified into three categories: (1) Stochastic methods, (2) Systematic search methods, and (3) Deterministic methods.

Conformational sampling of small molecules is important in docking approaches as using a single ligand conformation does not sufficiently explore ligand flexibility. In addition

to this, the steps taken to process ligand data *in silico* prior to virtual screening has been shown to have an impact on results, as examined by Knox *et al.*[174], where the roles of SMILES representations, stereochemical information, protonation state information and ligand conformational ensembles were shown to be important in obtaining optimum enrichment rates in virtual screens. There are two general methods for taking ligand flexibility into account; 'on the fly' and 'off-line'. In the off-line method a set of ligand conformations are generated prior to docking and then rigidly docked into the binding site without any further consideration of ligand flexibility. In the 'on the fly' method ligand flexibility is explored as part of the docking program. 'On the fly' conformational search methods may be classified into three categories: (1) Stochastic methods, (2) Systematic search methods, and (3) Deterministic methods.

3.5.1 Systematic Search Methods

Systematic search approaches try to explore both rotational and translational degrees of freedom a molecule and this can occasionally lead to a combinatorial explosion with ligands containing many rotatable bonds. To avoid this scenario, systematic search methods employ fragmentation and construction algorithms. These algorithms generate conformers in the binding site by first splitting the ligand into rigid 'core' regions and flexible parts and then growing the ligand by docking the core region and incrementally adding the flexible parts with a systematic scanning of torsional angles. The program DOCK[175] was the first to implement an 'anchor and build' method. The core fragment is docked based on steric complementarity and the flexible sidechains are incrementally added one bond at a time. With each subsequent addition, dihedral angles are explored and energy minimisations are carried out.

3.5.2 Stochastic Search Methods

Stochastic search algorithms are also known as random search algorithms. They incorporate random translational and rotational alterations to the ligand to generate different docked poses. The Monte Carlo and Genetic Algorithm (GA) methods are the two most popular stochastic methods. Each newly formed conformation is evaluated by an energy function and is accepted if the energy is lower than the value derived from the previous

step or if it is within a range defined by a Boltzman probability. GAs involve the encoding of candidate solutions as genes (conformations) in a chromosome. These individual conformations are then evaluated and less fit solutions are removed from the gene pool. Pairs of surviving conformations form new 'children' conformations whose fitness is then assessed. GOLD[176] and AutoDock[177] employ GA search strategies.

3.6 Scoring Functions

Once a number of docked ligand poses have been generated it is necessary to employ scoring functions to rank the poses. The three main types of scoring functions are (1) Knowledge-based, (2) Force-field, and (3) Empirical scoring functions.

3.6.1 Knowledge-Based Scoring Functions

Knowledge-based scoring functions are also known as potentials of mean force (PMF) and are derived from structural information of protein-ligand complexes stored in databases[178,179]. These scoring functions depend on the assumption that the frequent occurrence of close intermolecular interactions between certain types of atom or functional groups are more energetically favourable than randomly occurring events. Knowledge-based scoring functions are generally quite robust as potentials involved in calculating binding energies are extracted from structures in training databases that are usually large and diverse. Three main examples are the PMF[178] score, DrugScore[179], and the London dG[153] scoring function, which was applied in this chapter. The London dG scoring function estimates the free energy of binding of the ligand from a given pose. The functional form is a sum of terms:

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{HB} f_{HB} + \sum_{m-lig} c_M f_M + \sum_{atoms} \Delta D_i \tag{3.1}$$

where c represents the average gain/loss of rotational and translational entropy; E_{flex} is the energy due to the loss of flexibility of the ligand (calculated from ligand topology only); f_{HB} measures geometric imperfections of hydrogen bonds and takes a value in [0,1]; c_{HB} is the energy of an ideal hydrogen bond; f_M measures geometric imperfections of metal ligations and takes a value in [0,1]; c_M is the energy of an ideal metal ligation; and D_i is the desolvation energy of atom i. The coefficients c, c_{HB} and c_M were fitted from around 400 X-ray crystal structures of protein-ligand complexes with available experimental pKi data. In this method the distribution of distances between ligand-protein atom type pairs are collated and inverted to generate a potential of mean force if a particular ligand atom is found at a certain distance from a particular protein atom very frequently, this is likely to be a favourable interaction.

3.6.2 Force Field Scoring Functions

These functions are developed based on physical atomic interactions and are derived from experimental data as well as *ab initio* quantum mechanical calculations. The physical atomic interactions involved include van der Waals interactions, electrostatic interactions, and bond forces such as stretching, bending, and torsional effects described using molecular mechanics force fields such as Amber[180] and CHARMM[134]. The van der Waals and electrostatic energy terms set out the internal energy of the ligand and its interactions with the protein, with the van der Waals term described by the Lennard Jones potential and electrostatic term described by a Coulombic formula with a distance dependent dielectric constant for charge separation. The introduction of the distance dependent dielectric constant in the Coulombic term solves the problem of how to treat the solvent effects by considering them implicitly, an advantage force-field methods have over other scoring functions.

The programs DOCK[175] and AutoDock[177,181] implement this kind of scoring function, with DOCK taking its energy parameters from Amber force fields.

3.6.3 Empirical Scoring Functions

Empirical scoring functions estimate the binding affinity of a protein-ligand complex based on a set of weighted energy terms

$$\Delta G = \sum Wi - \Delta Gi \tag{3.2}$$

Where ΔGi describes different enrgy terms such as van der Waals forces, hydrogen bond, desolvation, electrostatics, entropy, hydrophobicity, etc and the corresponding coefficients Wi are established by fitting binding affinity data of a training set of protein-ligand complexes with resolved 3D structures. A major drawback of empirical scoring functions is their trouble dealing with more flexible ligands although, compared to force field scoring functions, they are much faster at calculating binding energies due to their relatively simple energy terms. An example of an empirical based scoring function is the ChemScore[182] function where binding energy is evaluated as a function of the hydrogen bonding geometry and lipophilic contact area between ligand-receptor with the addition of a penalty term representing the loss of conformational entropy upon binding

3.6.4 Limitations of Scoring Functions

There exist many examples in the literature exploring the shortcomings of scoring functions[183]. However, scoring functions do not as such deliver accurate calculations of binding affinities, but rather rough estimates which usually suffice for virtual screening experiments. Typically, in a virtual screening experiment, a scoring function is prioritized on the basis of efficient separation of the active molecules from the known inactives, with very little overlapping or 'contamination'. Most of the scoring functions implemented in docking packages can accurately reproduce the X-ray crystallographic pose in 70-80% of cases with 2Å root mean square deviation. A well documented example of scoring function limitations is their bias towards high molecular weight molecules[184] due to the additive nature of parameters for empirical and force field-based scoring functions. Standard scoring functions also fail to take into account desolvation energy or entropy, important contributors to ligand-protein binding interactions.

3.7 Combining Docking and Molecular Dynamics Simulations

Docking and molecular dynamics simulations are complementary approaches when one considers their relative strengths and weaknesses. Docking, while being poor at handling crucial aspects such as receptor flexibility, binding entropy and solvent effects, has the advantage of being able to screen massive compound libraries quite rapidly. The weaknesses of docking lie in areas that molecular dynamics simulations are strong while MD simulations come with a computational cost much higher than that of docking.

MD simulations and docking strategies can come together in two main ways: (1) In accounting for receptor flexibility prior to docking and (2) for the refinement of docked ligands and in the calculation of binding free energies. The first option aids the docking process by providing an ensemble of protein structures extracted from an MD trajectory for docking allowing multiple conformational modes of the protein to be considered.

3.8 Molecular Dynamics and MM-PBSA

While scoring functions are useful for estimating binding affinities for very large datasets, due to their low computational costs they come with many deficiencies. Thermodynamics studies into the formation of protein-ligand complexes have demonstrated that these are equilibrium processes and thus assessing binding energies based on a single rigid structure (as in the case of most docking and scoring approaches) is not a sufficient analysis of the process. Many scoring functions also disregard terms such as solvation and entropic effects, forces that make important contributions to the free energy of binding. Conformational energy of the protein is rarely included in energy contribution calculations and it is reasonable to assume that as different binding pocket conformations will possess different free energies, their contributions to the free energy of binding would also be different.

Among the methods that have been introduced to address these problems, a list including thermodynamic integration, ligand interaction energy and free energy perturbation, MM-PBSA [163] approaches (Molecular Mechanics/Poisson-Boltzman/Surface Area) have probably been the subject of most attention in recent years.

MM-PBSA methods implemented in Amber combine force-field derived molecular mechanics (MM) energies, a continuum solvent Poisson-Boltzman (PB) model for modeling polar solvation effects, and a solvent accessible surface area (SA) dependent non-polar solvation term to calculate the absolute free energy of a system between two states, representing the bound and unbound state of a solvated molecule[185].

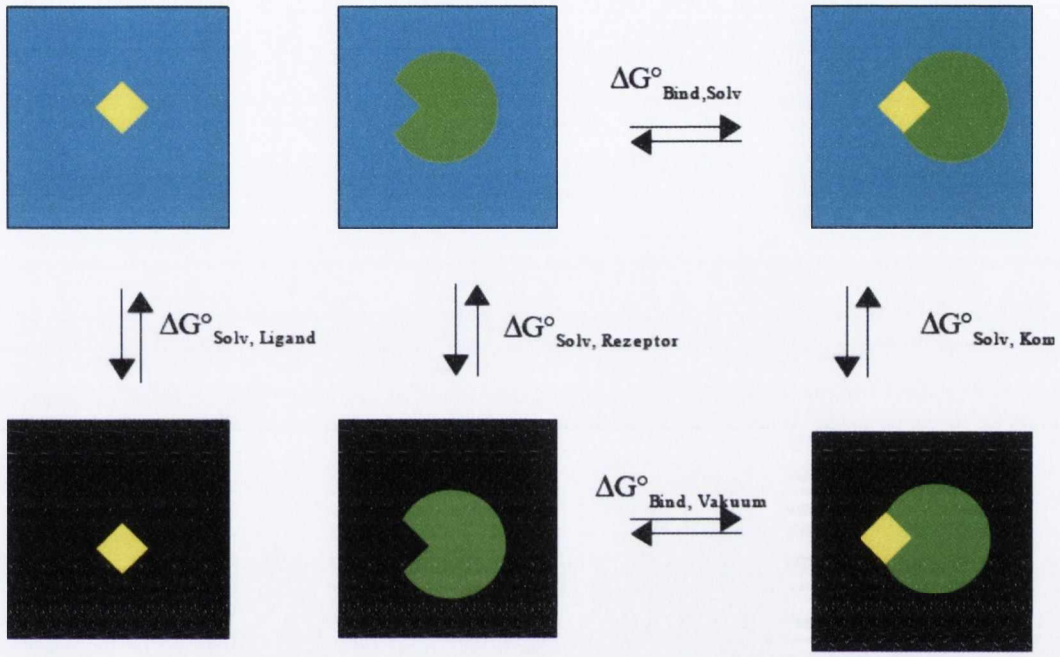


FIGURE 3.3: MM-GB(PBSA) thermodynamic cycle[186].

The approach of MM-PBSA divides the calculation up in accordance with the thermodynamic cycle (Figure 3.3) and can get multiple energy averages by using multiple structures from an MD trajectory. The binding free energy can be calculated as:

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta G_{bind,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,receptor}) \quad (3.3)$$

Explicit water molecules and ions are removed from the MD trajectory to increase the efficiency of calculations; implicit-solvent calculations are performed on MD trajectory snapshots by either using the generalised Born approach or by numerically solving the

Poisson-Boltzman equation. This can be solved for each of the three states by accounting for electrostatic and hydrophobic contributions to the solvation energy. Free energy *in vacuo* is calculated from the average interaction between ligand and receptor and takes entropy into account if necessary. Entropy contributions are usually neglected as they are computationally expensive and are associated with a large margin of error that may introduce substantial uncertainty into the results[186]. Average interaction energies are then collected by performing calculations on snapshots collected from an equilibrated MD simulation[187].

3.9 Methods

3.9.1 Molecular Dynamics Simulaitons

The MD simulations were performed in Amber 11.0 software package. Starting structures were pre-processed in MOE by removal of hydrogen atoms and further prepared by the LEaP module of Amber 11.0. Briefly, missing hydrogen atoms were added to the structure before overall charge was inspected and neautralised with counter-ions. The neutralised system was fully solvated in a rectangular water box using TIP3 model water molecules with a distance of 12Å between the wall of the periodic box and the closest atom in the solute.

The next step involved a seven step energy minimisation. The first step involved 500 steps of minimisation of the system's hydrogen atoms with 250 steps using the steepest descent method followed by 250 steps using the conjugate gradient method, followed by the second minimisation step of 5000 cycles of steepest descent on the system's water molecules. Steps 3-6 implemented 5,000 cycles each of steepest descent acting on the water and protein backbone atoms. During these steps the harmonic restraints were progressively reduced until an elastic constant force equal to 0kcal/mol was reached. The final step involved 10,000 steps of steepest descent minimisation on the system with no restraints on atom movement.

After minimisation the systems were gradually heated in the NVT ensemble (constant volume and temperature conditions) from 0 to 300K in 200ps. Heating was followed by equilibration for 50ps at constant temperature of 300k by coupling the system to a thermal bath with Berendsen algorithm with a time coupling constant of 1ps and a pressure of 1atm. SHAKE algorithm was used to constrain the bonds involving hydrogen atoms to their equilibration values. Production simulations were carried out for 20ns using the same non-bonded interaction parameters as in previous steps at constant periodic boundary conditions. Trajectories were analysed with the ptraj module of Amber 11 and system convergence was evaluated through analysis of energy and backbone RMSD plots.

3.9.2 Binding Pocket Analysis

Binding pocket analysis was carried using MDpocket on a MacBook Pro 2.8 GHz Intel Core 2 Duo running OS X 10.5.8. Snapshots from the MD trajectories were aligned on to a reference structure using the ptraj module of Amber 11 and a text file specifying the path of the snapshots was generated with a python script within MDpocket. Grid frequencies and densities for binding pockets across the trajectory were calculated to an Isovalue of 0.5 (default value) using the mdpocket command of MDpocket. Grid density points representing the 6 largest binding pockets on the alpha-tubulin subunits were saved as separate PDB files within MacPyMOL and subjected to more detailed pocket analysis and descriptor calculation within MDpocket.

3.9.3 ROCS shape matching

ROCS (**R**apid **O**verlay of **C**hemical **S**tructures) is a fast shape comparison application, based on the idea that molecules have similar shape if their volumes overlay well and any volume mismatch is a measure of dissimilarity. It uses a smooth Gaussian function to represent the molecular volume[188], so it is possible to routinely minimize to the best global matching. ROCS requires a database of conformers that it compares to the query molecule. Each conformer of each database molecule is overlaid rigidly on the query molecule, and the overlap of molecular volume between the query and the database conformer is optimized. Then, a measure of shape similarity between the query and the database conformer (the shape Tanimoto coefficient) is calculated using equation 3.4 below:

$$Tanimoto_{f,g} = \mathrm{O}_{f,g} \;/\; (\mathrm{I}_f + \mathrm{I}_g - \mathrm{O}_{f,g}) \tag{3.4}$$

The I terms are the self volume overlaps of each entity while the O term is the overlap between the two Gaussian functions. The I terms are independent of orientation but not O. Finding the orientation that maximises O, and hence minimises the differences

between f and g, is equivalent to finding the best overlay between the two objects. Once all conformers of the database molecule have been overlaid and the shape Tanimoto calculated, the conformer with the highest shape Tanimoto (highest shape similarity) is saved, along with the overlay of that conformer with the query molecule[189].

ROCS was used in an attempt to locate the APM binding site through a pure shape-based matching method. This approach involved selecting multiple snapshots from the *P. falciparum* and *T. brucei* molecular dynamic trajectories, enumerating potential binding sites from these snapshots, and then using these site shapes as ROCS queries.

Snapshots were selected every 2ns from a 20ns molecular dynamics trajectory for both *P. falciparum* and *T. brucei* proteins. Each snapshot was opened in MOE and Site Finder was used for active site identification and enumeration. The Site Finder tool is a geometric method based upon the alpha spheres approach. Briefly, alpha sites are placed in the active site using a modified Delaunay triangulation of collected 3D points. For each resulting collection of four points there is a related sphere called an alpha sphere that has a specific radius. The program then removes inaccessible of over-exposed alpha spheres, retaining small alpha spheres corresponding to locations of tight atomic packing in the receptor. Alpha spheres are then clustered using a single-linkage clustering algorithm to produce a collection of sites with each site consisting of one or more alpha spheres. After spheres were selected to denote the presence of a binding site, they were changed to carbon atoms using the MOE Builder tool and used as query structures for ROCS. A multiconformer database of APM was screened with the binding site shape queries with the top conformer, as scored by ShapeTanimoto, saved for analysis.

3.9.4 Ligand Conformer Generation

Multiple conformers were generated for a single APM structure using LowModeMD[190] in the MOE package. An energy minimised structure of APM was used as a starting point and conformers were generated with a 7 kcal gap separation and an RMSD of 0.25 for a maximum of 1000 conformers generated.

3.9.5 Docking

APM conformers were docked into tubulin homology model structure using the *induced fit* protocol implemented in MOE2011.10. This protocol gives the binding site freedom to adapt to the ligand during the docking process and flexibility is allowed for both ligand and protein side chains. Molecules were docked using MOE Triangle Matcher to generate an initial pose and poses were scored using the London dG function. 30 poses were retained, subjected to a forcefield refinement, and rescored using GBVI/WSA dG function. The two best scoring induced fit poses for each binding site were saved as separate PDB files. Prior to MD simulations the hydrogens in each saved PDB were removed and the ligand protonated using MOE Protonate 3D.

3.9.6 Molecular Dynamics Protocol

The LEaP module of Amber 11 was used to convert each PDB structure into Amber structure files. For the protein, forcefield parameters and partial charges form the ff99sb forcefield were applied; the ligand poses were prepared using the antechamber module of Amber 11 with the general Amber forcefield (GAFF) parameter assignment. Each structure was prepared with a quick optimisation step within the Amber forcefield in vacuum, and then placed in a TIP3P water box. The minimum distance between protein and the walls of the box was kept to 12Å. Prior to MD calculations, LEaP was used to add Na^+/Cl^- counterions to neutralise the complex.

All MD simulations were performed in the Amber 11.0 software package. Solvated systems were minimised in a multistep procedure as described in section 3.9.1. Following minimisation the systems were gradually heated in the NVT ensemble (constant volume and temperature) from 0 to 300K in 200ps. Systems were then equilibrated for 50ps at a constant temperature of 300K by coupling the system to a thermal bath with Berendsen algorithm with a time coupling constant of 1ps and a pressure of 1 atm (NPT). The MMPBSA approach was applied to the last conformation of the seventh minimisation step for each of the 12 poses providing energy binding values for generalized Born (GB) and Poisson-Boltzman (PB). All poses were then brought forward for a 10ns MD production step run in the NVT ensemble at constant temperature of 300K by coupling the system to a thermal bath with Berendsen algorithm with a time-coupling

constant of 2ps. Trajectories were analysed with the ptraj module of Amber 11 and system convergence was evaluated through constant ΔG binding energies during the 10ns simulation time.

3.9.7 Analysing Molecular Dynamics Trajectories

There exist several generic techniques that can be utilised for the analysis of MD trajectories. The two used in this work are briefly explained here.

The Root-Mean-Square Deviation (RMSD) gives the average distance d between two position vectors p and q of N equivalent atoms in 3D space:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_i \mathrm{d}_i{}^2} = \sqrt{\frac{1}{N}\|pi - qi\|}^2 \tag{3.5}$$

RMSD is a standard measure of structural distance between sets of coordinates.

Another common analysis tool for MD trajectories is the Root-Mean-Square Fluctuation (RMSF) which describes the atomic positional fluctuations within a considered time period. This is calculated by averaging over atom (or residue) i's deviations to its time-averaged position:

$$\text{RMSF}(i) = \sqrt{\langle(r_i\text{-}\langle r_i\rangle)^2} = \sqrt{\langle r_i{}^2\rangle - \langle r_i\rangle^2} \tag{3.6}$$

The angle brackets $\langle\rangle$ denote a time average. This measure is similar to RMSD except that the averaging is over time instead of atoms. In order to remove system-wide

translational and rotational movements, the coordinates should be aligned on a common structure prior to RMSF calculations.

3.9.8 MM-PB/GBSA

Binding free energy calculations were performed using the MM-PB/GBSA approach implemented in Amber 11. An average 500 snapshots were extracted form the last 5ns trajectory for calculations. The binding free energy for each calculation was computed for each molecular species (complex, protein and ligand) according to the following equation:

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand})$$

(3.7)

Each term contribution can be estimated as follows:

$$\Delta G = \Delta G_{mm(gas)} + \Delta G_{sol} - T\Delta S$$

(3.8)

Where $\Delta G_{mm(gas)}$ is the molecular mechanics free energy between protein and ligand (sum of the non-bonded electrostatic, van der Waals, and internal energy contributions), ΔG_{sol} is the solvation free energy and the $T\Delta S$ represents the conformational change entropy term. ΔG_{sol} can be expressed as the sum of an electrostatic ($\Delta G_{ele,sol}$) component and a nonpolar (ΔG_{np}) component:

$$\Delta G_{sol} = \Delta G_{ele,sol} + \Delta G_{np} \tag{3.9}$$

$\Delta G_{ele,sol}$ values of internal and external dielectic constants were 1 and 80 respectively.

3.10 Results

3.11 Molecular Dynamics Simulations

Simulation equilibration, general stability and occurrence of dynamic events were tracked by computing the coordinates relative deviations from the starting homology models as a function of time. Initially a steep rise was observed in the backbone RMSD for each of the three proteins, roughly over the first 2-4ns period. Subsequently the RMSD values reached a plateau and remained relatively settled during the remainder of the simulations. A certain amount of RMSD fluctuation is not unexpected given the large, flexible loops present in the tubulin protein. Much lower flexibility was observed for the central core helix and its surrounding beta sheets. There was a small amount of fluctuation at the C-terminal helices, mainly centered on helix 12. Given these results the dynamics production was deemed successful and the structures generated could be employed to analyse the time-dependent emergence, if any, of potential herbicide binding pockets. Figures 3.4 - 3.7 illustrate the stabilities of the malarial and trypanosomal tubulin simulations respectively.

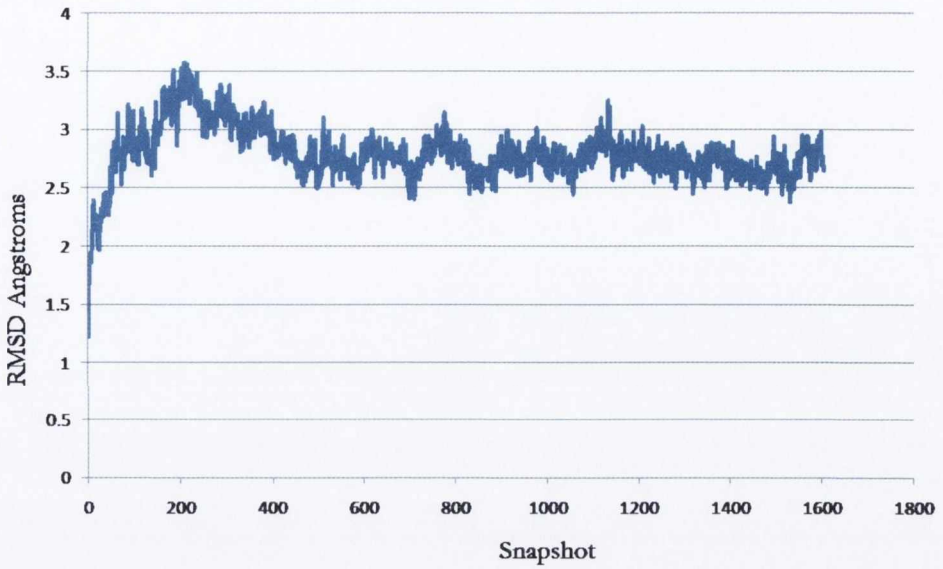


FIGURE 3.4: RMSD of backbone carbon atoms of malarial tubulin across the molecular dynamics simulation.

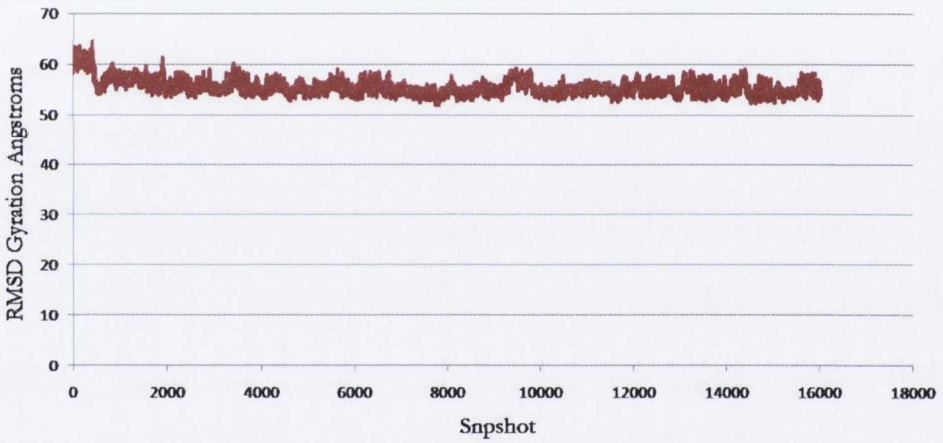


FIGURE 3.5: RMSD of the radius of gyration of malarial tubulin across the molecular dynamics simulation of 20ns.

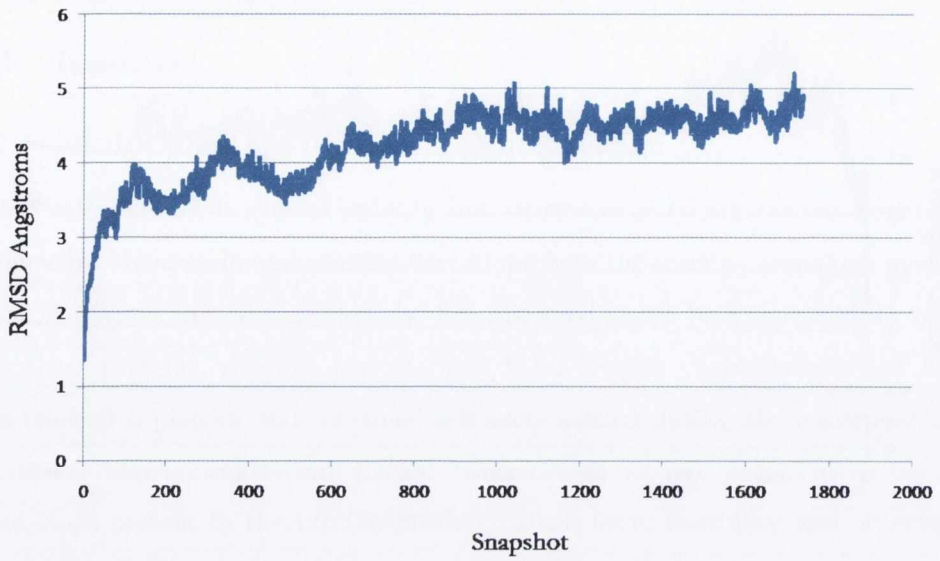


FIGURE 3.6: RMSD of backbone carbon atoms of trypanosomal tubulin across the molecular dynamics simulation of 20ns.

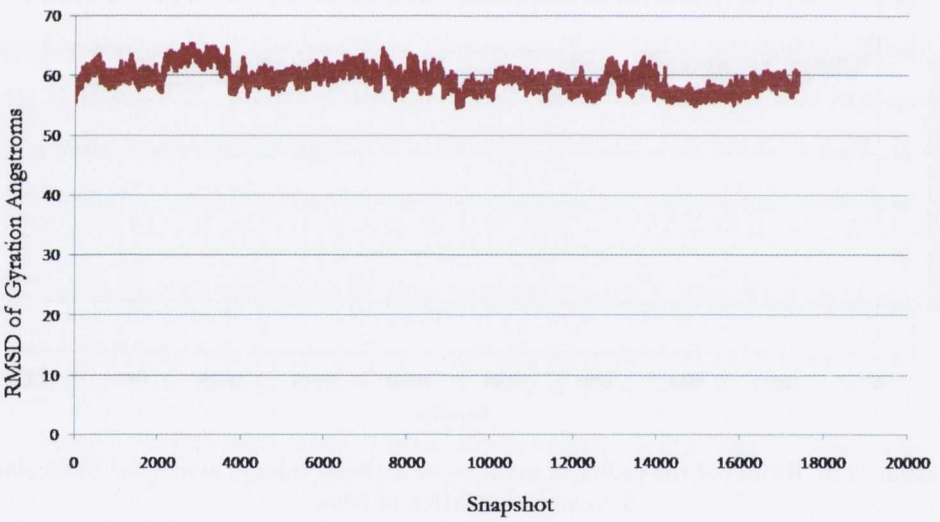


FIGURE 3.7: RMSD of the radius of gyration of trypanosomal tubulin across the molecular dynamics simulation.

3.12 Analysis of Transient Binding Pockets with MDpocket

MDpocket was used in an attempt to identify the presence of any transient binding pockets that appear on the tubulin surface during the molecular dynamics simulation. The results of mdpocket analysis using an isovalue of 0.3 failed to return any significant results. This result was perhaps not surprising given the overall fold similarity and amino acid conservation between the mammalian template protein and the parasitic models.

This also raises the problem of homology modeling in an effort to look for differences between two proteins. These results, for the time being, reject the hypothesis that there is a herbicide-selective binding pocket that opens up during protein fluctuations that is present on the surface of parasitic tubulin but not on the surface of mammalian tubulin. Instead, what is more likely is that the herbicide binding site is formed due to amino acid differences between the susceptible and resistant tubulins rather than an entirely new binding site.

3.13 Analysis of Amino acid differences across species

The alpha tubulin sequences from *P. falciparum* and *T. brucei* parasites were aligned with the alpha-tubulin sequence from the porcine source to facilitate the identification of amino acids potentially involved in the formation of the herbicide-binding site. There are 31 amino acid substitutions across the 451 amino acid sequences, representing a 6.8% difference between resistant and susceptible species (Table 3.1). Substitutions of note include four polar uncharged amino acids changing to non-polar hydrophobic residues occurring at C118, S174, S289, and C353 positions. In addition to this, there are another six residues changing from non-polar hydrophobic to polar uncharged amino acids at F24, A48, A140, L194, V200, and M301.

Figures 3.8 and 3.9 detail the relative positions of these amino acids in the tubulin alpha subunit.

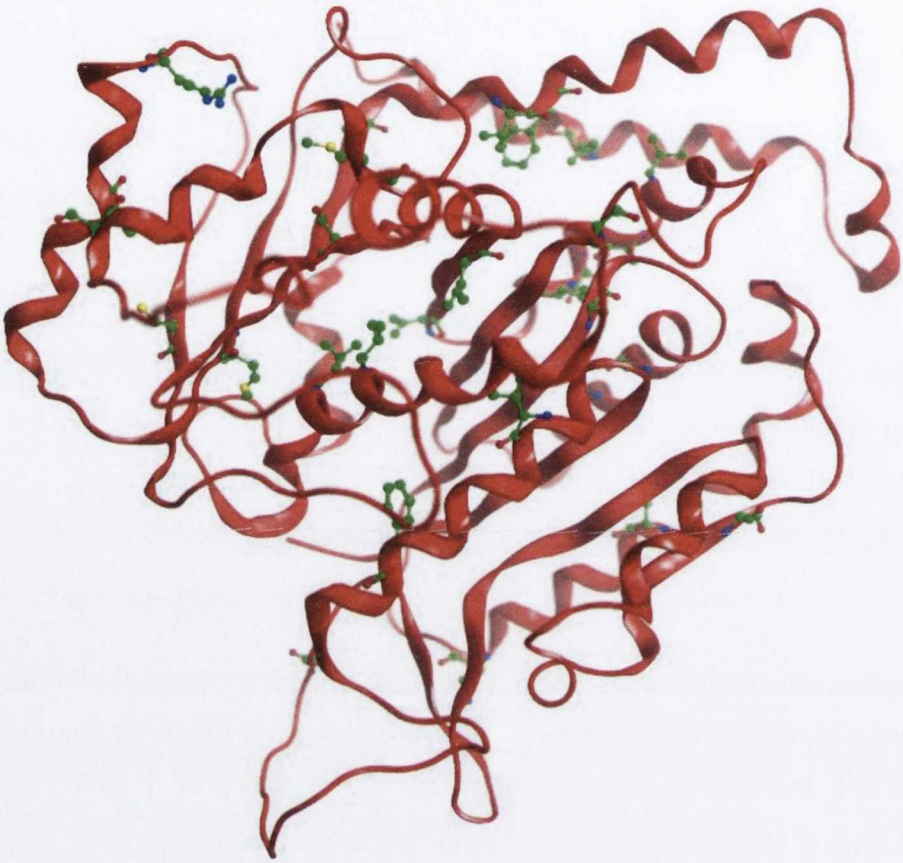


FIGURE 3.8: The positions of the amino acids on α-tubulin altered between herbicide susceptible species and herbicide resistant species (front view).

3.14 Using ROCS to identify potential binding pockets

In an attempt to locate a herbicide-selective binding pocket, ROCS was used to identify the best-fit pockets based purely on pocket shape similarity to APM. This methodology involved selecting binding site representations using sitefinder in MOE then replacing the dummy atoms with carbon atoms. These carbon atoms then serve as the shape query to screen a database containing multiple conformations of APM with receptor flexibility taken into account by using 10 snapshots from the MD trajectories. In total, six protein cavities were selected for shape analysis with ROCS based on their size, relative level of buriedness, and having in close proximity at least one amino acid that differs between susceptible and resistant tubulin proteins.

Binding site 1 is located in a hydrophobic pocket in the N terminal of alpha tubulin, residing in a highly disordered loop region connecting helix 1 and beta sheet 2 and also

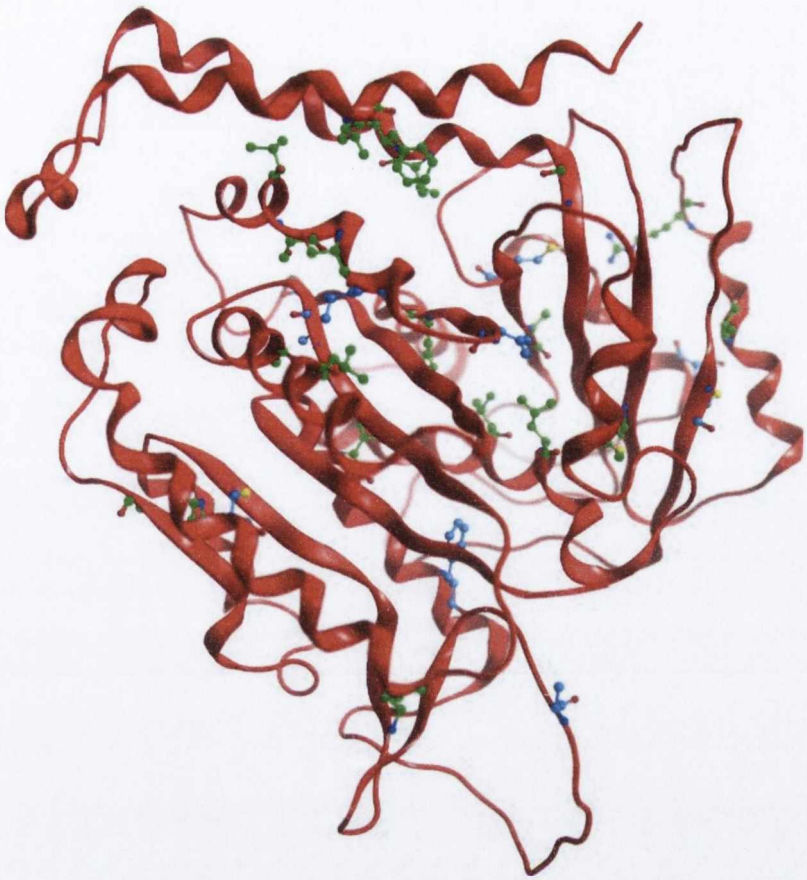


FIGURE 3.9: The positions of the amino acids on α-tubulin altered between herbicide susceptible species and herbicide resistant species (back view).

comprising amino acids of helix 1. Amino acids present in this binding site include M1, R2, E3, A4, F24, H28, I30, M36, P37, S38, D39, L40, A48, F49, N50, T51, F52, F53, H61, and R243. Binding site 2 contains amino acids L269, P298, M301, C305, D306, P307, R308, H309, G310, Y312, R339, T340, I341, Q342, S381, A383 and I384. This pocket is centered in the loop region of helix 9 and beta sheet 9 and takes in residues from beta sheet 7. Binding site 3 is located in a shallow hydrophobic channel between helix 5 and helix 12. Residues present in this pocket include H192, L195, E196, T198, Q256, Y262, P263, R264, I265, N424, L425, A427, L428 and N431. Binding site 4 is centered around residues T56, V62, P63, C65, V66, H88, E99, R121, K124, L125 and N128, located between helix 3 and beta sheet 2. Binding site 5 is surrounded by helix 11, and beta sheets 4 and 5. The main amino acids in this pocket are Val141, W172, P173, S174, P175, T179, A180, V181, V182, E183, P184, Y185, Ser187, T191, K304, R2390, M391, K394, M398, A403, and F404. Binding site 6 contains the amino acid

TABLE 3.1: Table of Amino Acid differences between resistant (mammalian) and susceptible (*P. falciparum* and *T brucei*) alpha-tubulins. NH = Nonpolar Hydrophobic, PU = *Polar Uncharged*, Neg = *Negatively Charged*

AA Position	AA type Susceptible	AA Charge	AA type Resistant	AA charge
16	Val	NH	Ile	NH
24	Phe	NH	Tyr	PU
48	Ala	NH	Ser	PU
75	Val	NH	Ile	NH
94	Ser	PU	Thr	PU
118	Cys	PU	Val	NH
128	Asn	PU	Gln	PU
140	Ala	NH	Ser	PU
150	Gly	PU	Thr	PU
154	Leu	NH	Met	NH
174	Ser	PU	Ala	NH
188	Val	NH	Ile	NH
190	Ser	PU	Thr	PU
193	Ser	PU	Thr	PU
194	Leu	NH	Thr	PU
200	Val	NH	Cys	PU
204	Leu	NH	Val	NH
234	Val	NH	Ile	NH
238	Leu	NH	Ile	NH
271	Ser	PU	Thr	PU
289	Ser	PU	Ala	NH
301	Met	NH	Gln	PU
318	Met	NH	Leu	NH
332	Val	NH	Ile	NH
339	Thr	PU	Ser	PU
353	Cys	PU	Val	NH
381	Ser	PU	Thr	PU
387	Val	NH	Ala	NH
388	Phe	NH	Trp	NH
425	Leu	NH	Met	NH
438	Glu	Neg	Asp	Neg

residues F103, H107, Y108, L189, H192, S193, M413, E414, E417, F418, A421 and D424 and is located between helix 12, helix 5, and helix 3. Also included was the Morrissette

site, composed of residues M1, R2, E3, R43, D46, D47, A48, T51, F53, S54, E55, R64, N128, C129, T130, and G131. This site is located in a highly variable loop region; loops between helix 1 and beta sheet 2 and between helix 3 and beta sheet 2 comprise this pocket, not the N-loop as previously stated by Morrissette *et al.* The binding sites are summarised in Table 3.2 and illustrated in Figures 3.10 and 3.11.

TABLE 3.2: Summary of Binding Sites used for ROCS shape-based searching. Residues in Italics are those that are present in herbicide-susceptible species and not present in herbicide-resistant species.

Site Number	AA involved	Secondary Structure Involved
1	M1, R2, E3, A4, *F24*, H28, I30, M36, P37, S38, D39, L40, *A48*, F49, N50, T51, F52, F53, H61, R243	H1,H1-B2 loop
2	L269, P298, *M301*, C305, D306, P307, R308, H309, G310, Y312, R339, *T340*, I341, Q342, *S381*, A383, I384	H9-B9 loop, B7
3	H192, L195, E196, T198, Q256, Y262, P263, R264, I265, N424, *L425*, A427, L428, N431	H5, H12
4	T56, V62, P63, C65, V66, H88, E99, R121, K124, L125, *N128*	H3, B2
5	Val141, W172, P173, *S174*, P175, T179, A180, V181, V182, E183, P184, Y185, Ser187, T191, K304, R390, M391, K394, M398, A403, F404	H11, B4, B5
6	F103, H107, Y108, L189, H192, *S193*, M413, E414, E417, F418, A421, D424	H3, H5, H12
Morrissette	M1, R2, E3, R43, D46, D47, *A48*, T51, F53, S54, E55, R64, *N128*, C129, T130, G131	H1-B2 loop, B32-H3 loop

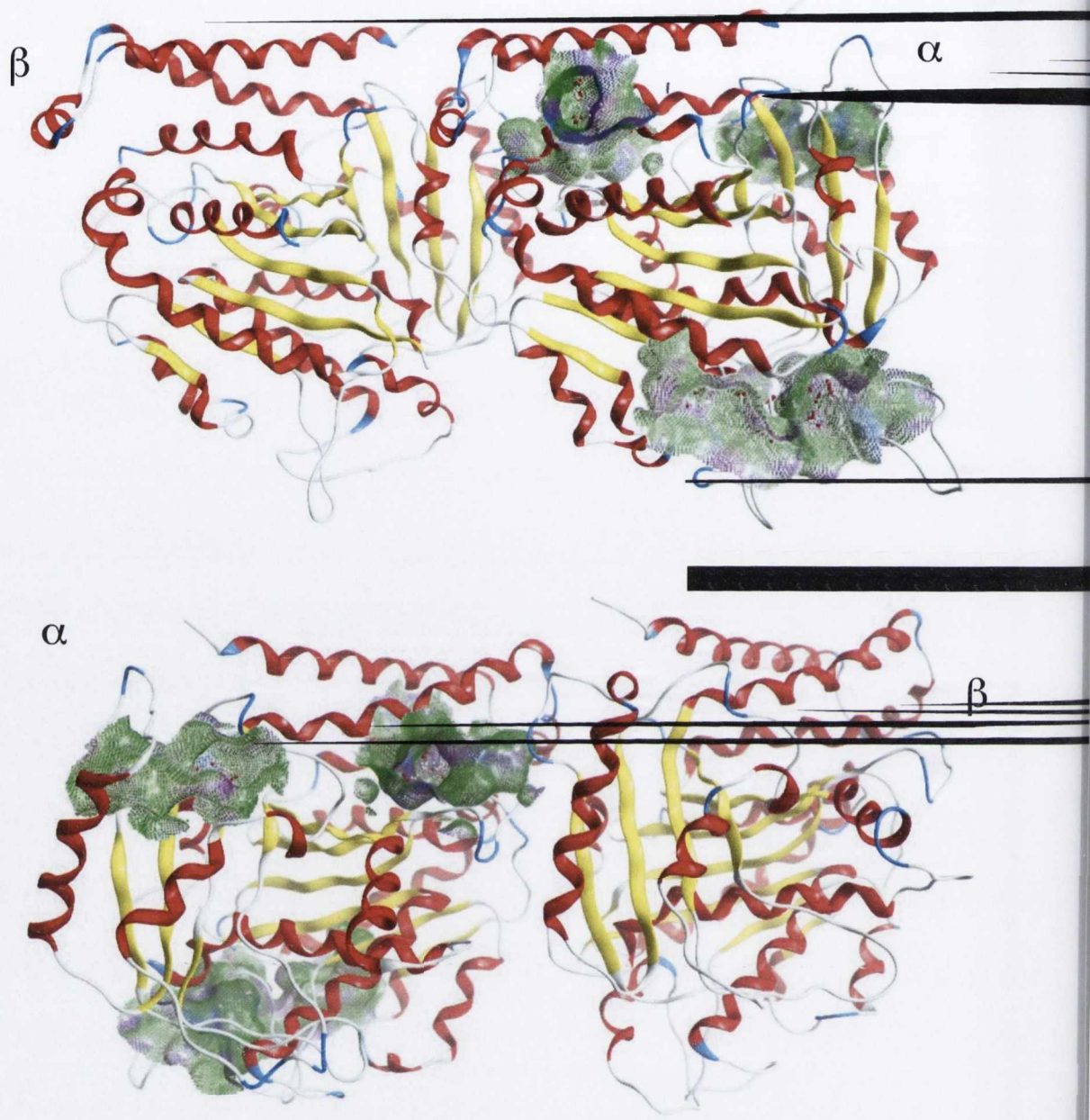


FIGURE 3.10: Illustration of the relative positions of putative binding sites 1-6 on the tubulin α-subunit. A is front view and B is rear view.

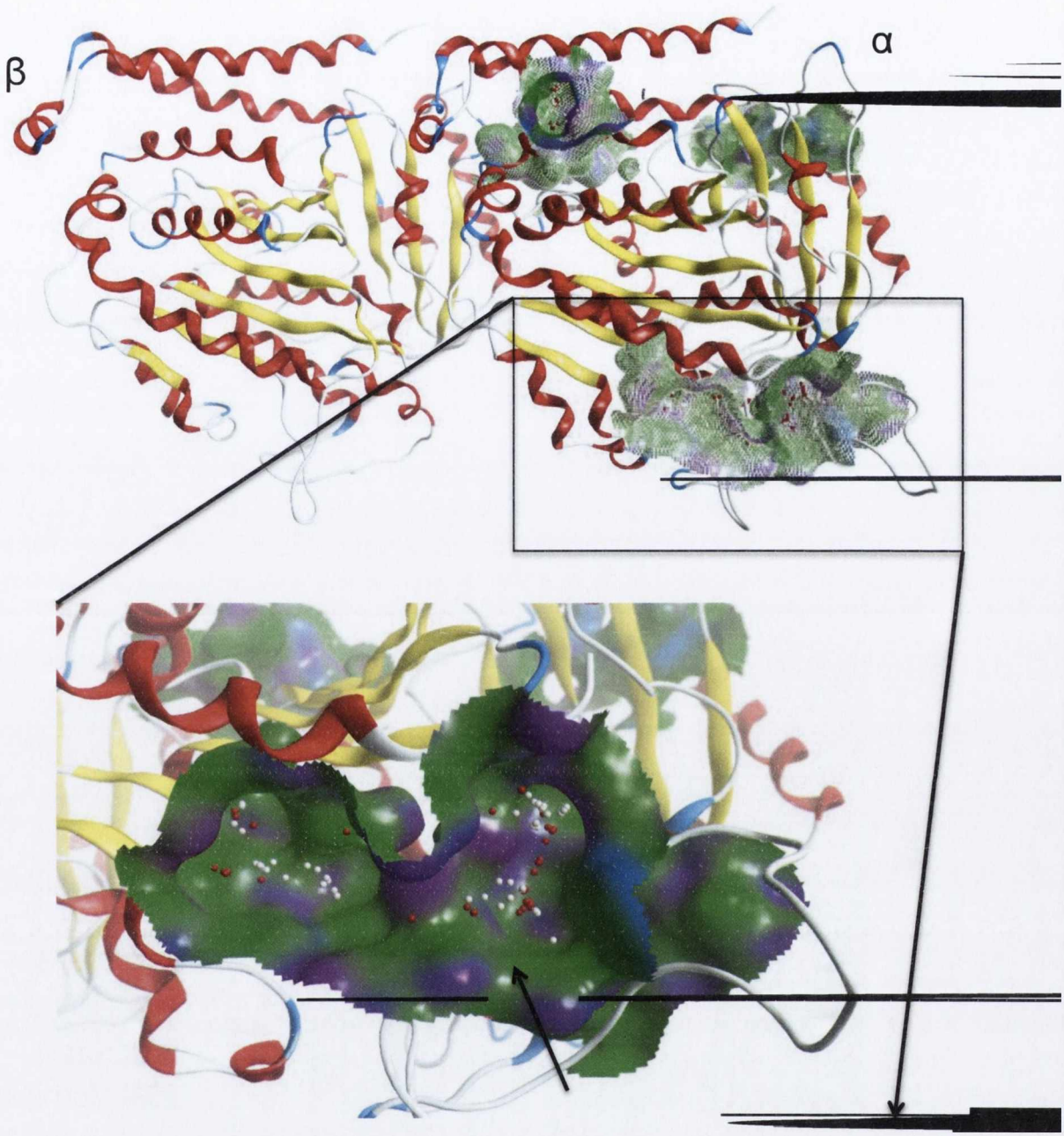


FIGURE 3.11: Close-up view of the loop region of α-tubulin comprising putative binding sites 1, 4, and the proposed 'Morrissette' site.

As can be seen in Fig. 3.12 below of the results from ROCs searching using binding site shapes as represented by alpha spheres, no one binding site stands out above the others as an obvious shape match for the APM herbicide. A p-value of 0.0904 was obtained using a one-way analysis of variance, indicating the means were not statistically significant. This was reflected in an identical run using the known binding site of colchicine and 3 non-colchicine binding pockets as ROCS queries in an attempt to establish the validity of the protocol. The inability of the ROCS shape-based searching method may be caused by the observation that smaller sites generally score better than the larger sites, indicating a penalization of areas of unoccupied space within the query-herbicide overlap. In the graph below, sites 1 and 7, the largest sites, scored the lowest average ShapeTanimoto values.

Following the inability of ROCS to identify a suitable binding pocket, a more accurate method was chosen in an attempt to establish the location of the APM binding site in parasitic tubulin.

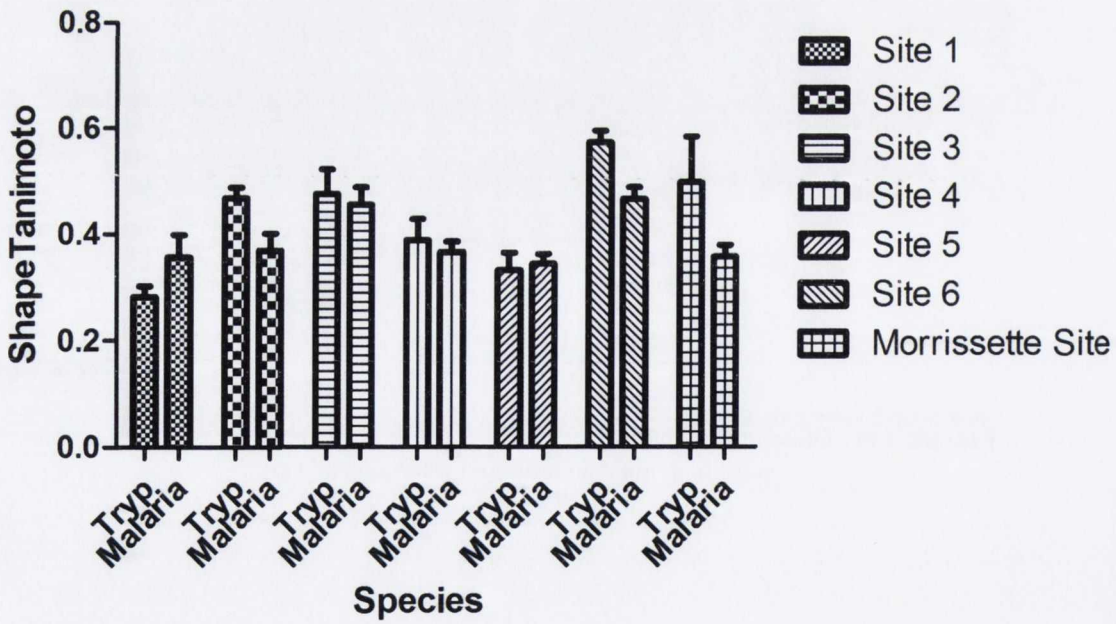


FIGURE 3.12: Results of ROCS shape-based searches on potential herbicide binding sites on the tubulin protein.

3.15 Induced fit docking

In order to probe the surface of the *P. falciparum* and *T. brucei* tubulin models for a potential common site of herbicide action, we used an induced fit docking protocol in MOE. This protocol was well suited to deal with the binding of a small molecule at a flexible and solvent exposed surfaced compared to traditional rigid docking methods. This protocol was used to provide an initial docked pose of APM and oryzalin, in *P. falciparum* and *T. brucei* tubulin binding sites, to be used as a starting point for a more exhaustive exploration using molecular dynamics studies in Amber 11.

Based on the evidence of amino acid substitutions between herbicide susceptible and resistant tubulin proteins (Table 3.1), 6 cavities were identified and selected for docking studies using Sitefinder in MOE. A database containing precomputed conformers of APM and oryzalin were docked to these sites with docked poses ranked using the London dG scoring function (E-score) and only the top ranked pose for each site and compound per parasite tubulin was saved and prepared for MD simulations.

3.16 Molecular Dynamics Study of potential binding pockets

Poses selected from the induced fit docking step were subjected to a molecular dynamics protocol comprised of an Amber protein-ligand preparation including explicit water solvation, minimization, heating and equilibration steps. Following this, all poses generated from the minimisation step were brought forward for a more extensive molecular dynamics study involving a 10ns production run followed by calculation of the mean ΔG binding free energy of the protein-ligand complexes.

When performing MD calculations it is important to establish the stability of the system under study when results are being analysed. This is done by inspecting the convergence of the ΔG of the protein-ligand complexes during the production dynamics run. Once convergence was reached, we focused on the last 5ns of the production run for the MM-GB/PBSA calculations. In all cases observed all of the protein-ligand complexes showed a convergent ΔG profile before the last 5ns.

TABLE 3.3: Summary of MM-GBSA binding energies for putative parasitic herbicide-binding sites 1-6

Site Number	*P. falciparum* APM (kcal/-mol)	*T. brucei* APM (kcal/-mol)	*P. falciparum* OZN (kcal/-mol)	*T. brucei* OZN (kcal/mol)
Site 1	**-41.4945**	**-37.0681**	**-46.4057**	**-37.5128**
Site 2	-30.3928	-10.372	-32.2916	-10.0183
Site 3	-16.7229	-22.1881	-7.9115	-5.1069
Site 4	**-28.3664**	**-24.3178**	**-27.8527**	**-30.2805**
Site 5	-32.7877	-17.1902	-39.1363	-10.9651
Site 6	-21.2734	-18.8363	-3.8089	-21.168

Table 3.3 summarises the results of the MM-GBSA biding energy calculations for sites 1-6. From these results it is obvious that sites 1 and 4 are the only sites that obtained a consistent ΔG of binding after the 10ns MD simulation. As such the following sections will discuss the binding poses obtained in sites 1 and 4 in detail, while discussions of sites 2, 3, 5 and 6 can be found in the appandix to this chapter.

3.16.1 Analysis of binding site 1 in parasitic tubulin

Analysis of the poses generated at the end of the production dynamics simulation for APM within Malaria and Trypanosome binding site 1 revealed that the final poses lie in two opposite orientations with no ligand overlap whatsoever (Figure 3.13 A). Closer examination of the overlap of binding site 1 with APM bound in malaria and trypanosome tubulin revealed that there is a large disparity in the region of residues (G34-A44) between helix 1 and beta sheet 2; this movement created two subpockets in the cavity. With respect to Tryp site 1, the hydrophobic nitrobenzene moiety of APM sits in a hydrophobic sub-pocket made up of amino acids M36, P37, F53, F54, and H61 (Figure 3.17). The methoxy group is slightly exposed to solvent while the complex is stabilised due to interactions between the sulfur atom of APM and F52 and F53, although these are weak interactions. In the malarial site 1 (Figure 3.15) the ligand appears to be overall more exposed to solvent than in trypanosomal site 1, although the interactions between protein and ligand contribute more to stabilising enthalpy. There is a weak hydrogen bond between the amine hydrogen and the centre of the benzene ring of F53, while N50 is also hydrogen bonded to APM through its terminal amine hydrogen and the amine group on APM. A hydrogen bonding network also appears

through T51 and an oxygen atom on the ring nitro group of APM. This is reflected in the total ΔG figures from Table 3.3 where the Malaria site1-APM figure is more negative than the figure for tryp site1-APM. Both have similar contributions to binding, with the malaria-APM complex having a higher contribution to binding through the ΔG ele term.

The poses of oryzalin in site 1 of malaria and tryp tubulin adopt similar orientations (Figure 3.13 B) . In the malaria site site, OZN makes 2 hydrogen bonds from the sulfonamide amine to H28 and and N50 (Figure 3.16). The residue R243 makes a hydrogen bond with an oxygen atom of the sulfonamide and T51 forms a hydrogen bond with one of the two ring nitro groups. In the trypanosomal site (Figure 3.18), there is a bigger interaction focus on the ring nitro groups than the sulfonamide group. R243 again forms a hydrogen bond with a sulfonamide oxygen and there are three hydrogen bonds to the ring nitro groups involving K40, T51 and F53. In terms of energy contributions to binding, there is a striking difference between OZN and APM in terms of the electrostatic component. With the OZN poses, a very large, positive ΔG ele was calculated, caused by the presence of unpaired polar groups in the ligand.

With reference to the malaria tubulin solely (Figure 3.14 A), the bound OZN and APM structures take up very similar binding orientations within the binding site and both interact with residues H28, F49 ad T51. In the trypanosome tubulin (Figure 3.14 B), the loop region between helix 1 and beta sheet 2 does not align well and the APM and OZN poses are also aligned poorly. However, the two poses share an interaction with F53 and similar hydrophobic interactions with D39.

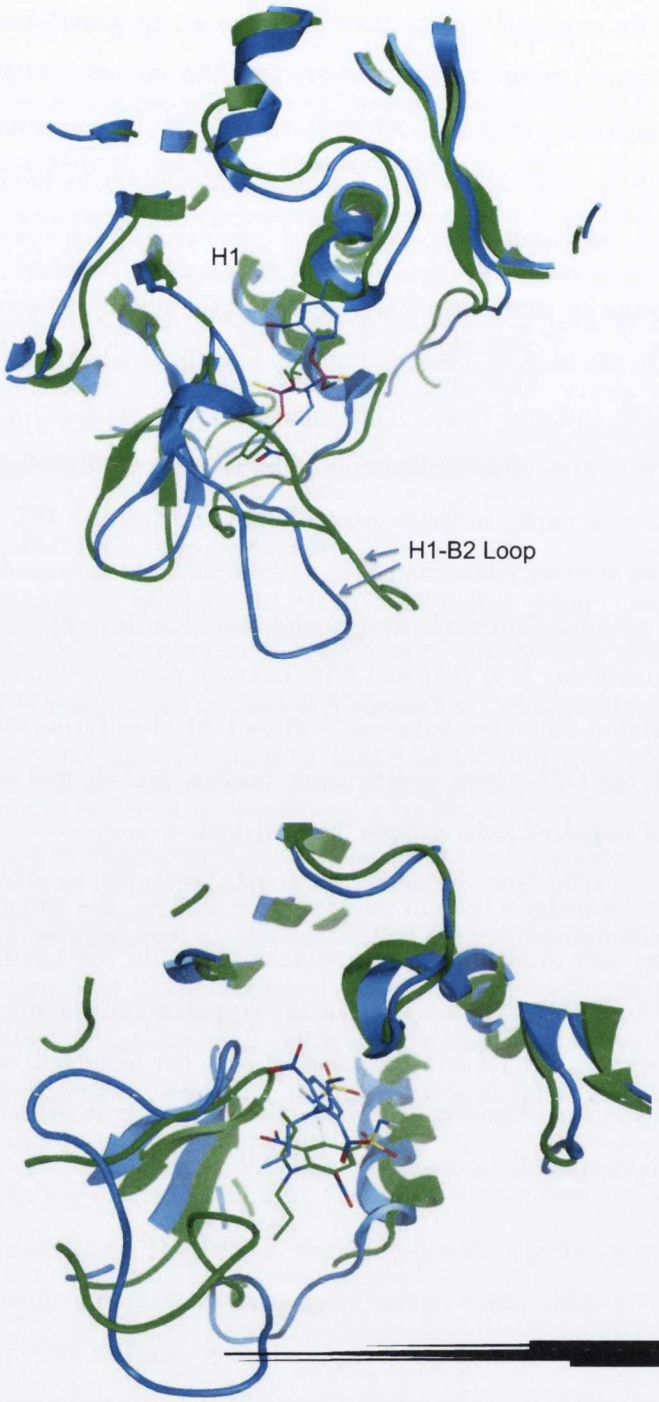


FIGURE 3.13: Post-molecular dynamics poses of APM and OZN in site 1 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

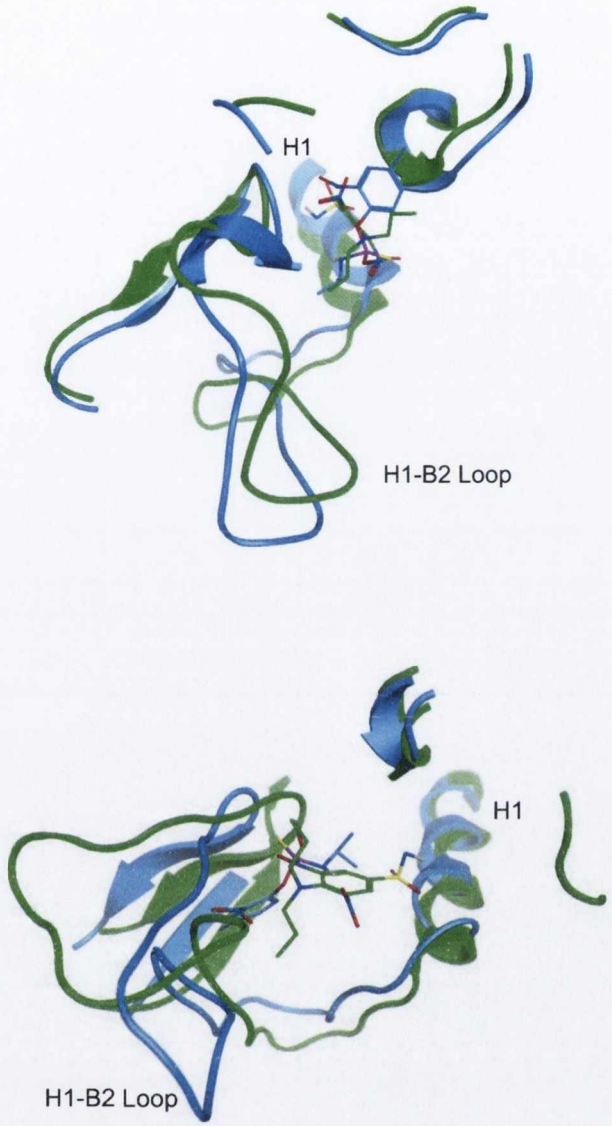


FIGURE 3.14: Post-molecular dynamics poses of APM and OZN in site 1 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 1 of malarial tubulin. (B) The orientations of APM and OZN in site 1 of trypanosomal tubulin

FIGURE 3.15: Ligand interaction diagram of APM in site 1 of malarial tubulin.

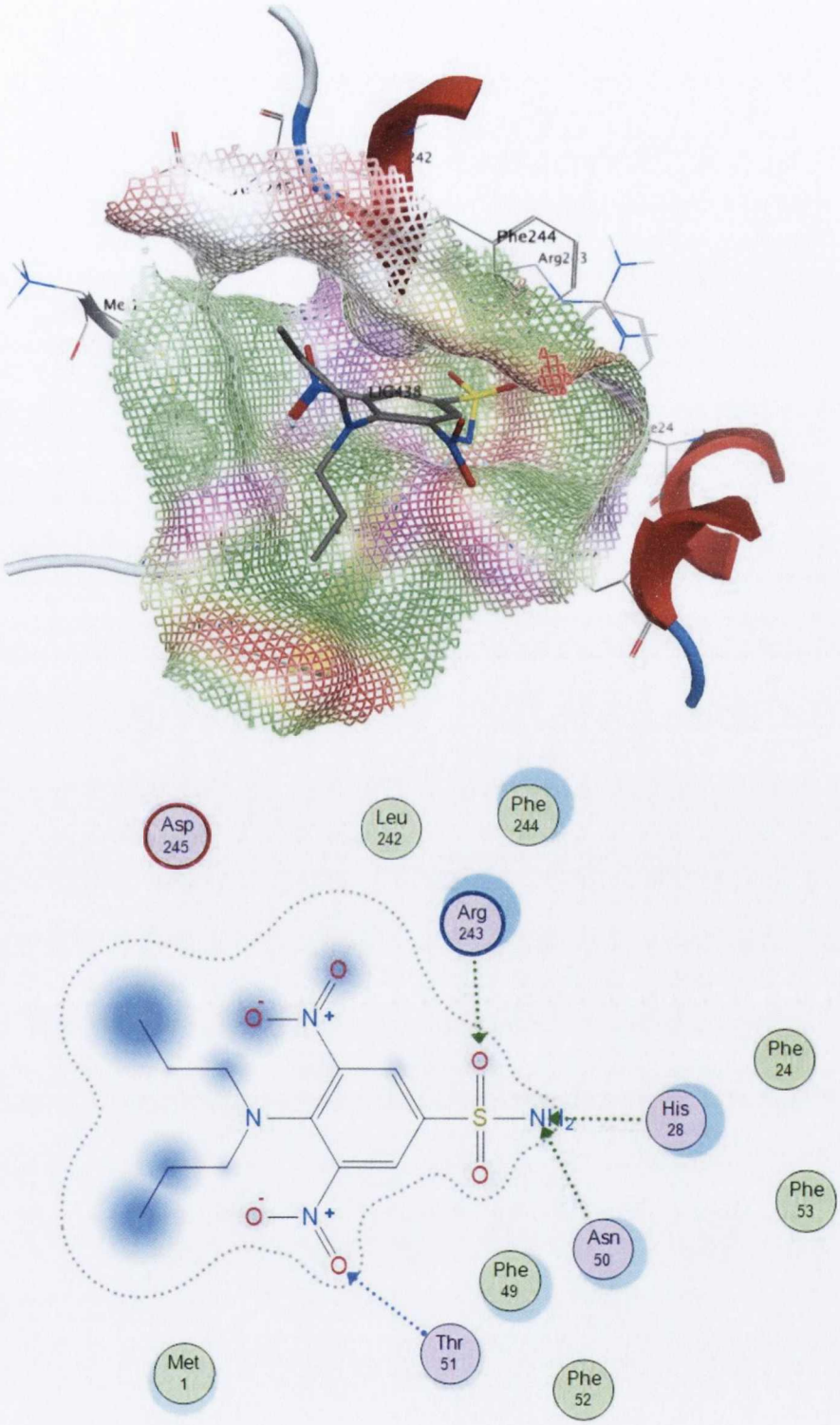


FIGURE 3.16: Ligand interaction diagram of OZN in site 1 of malarial tubulin.

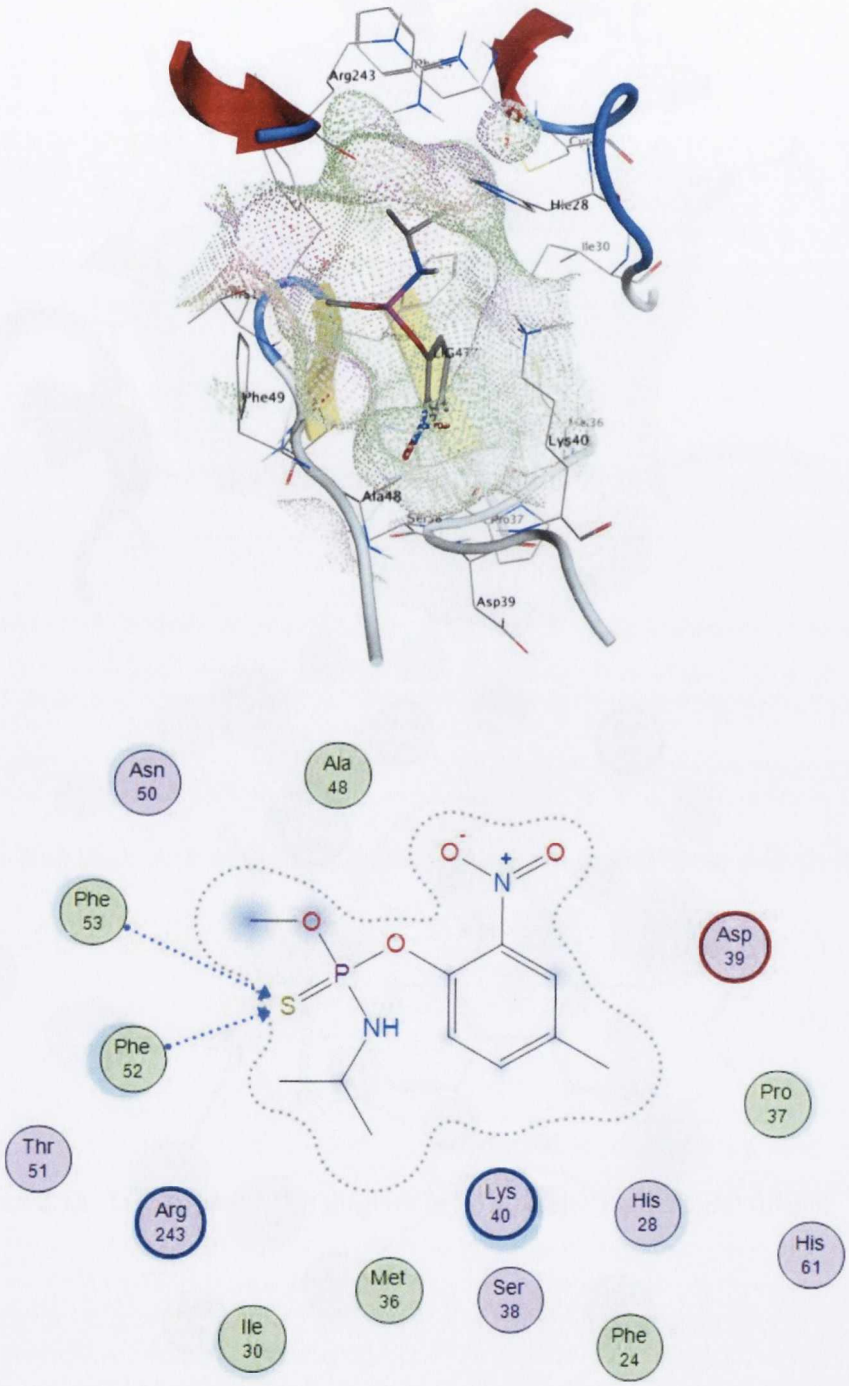


FIGURE 3.17: Ligand interaction diagram of APM in site 1 of malarial tubulin.

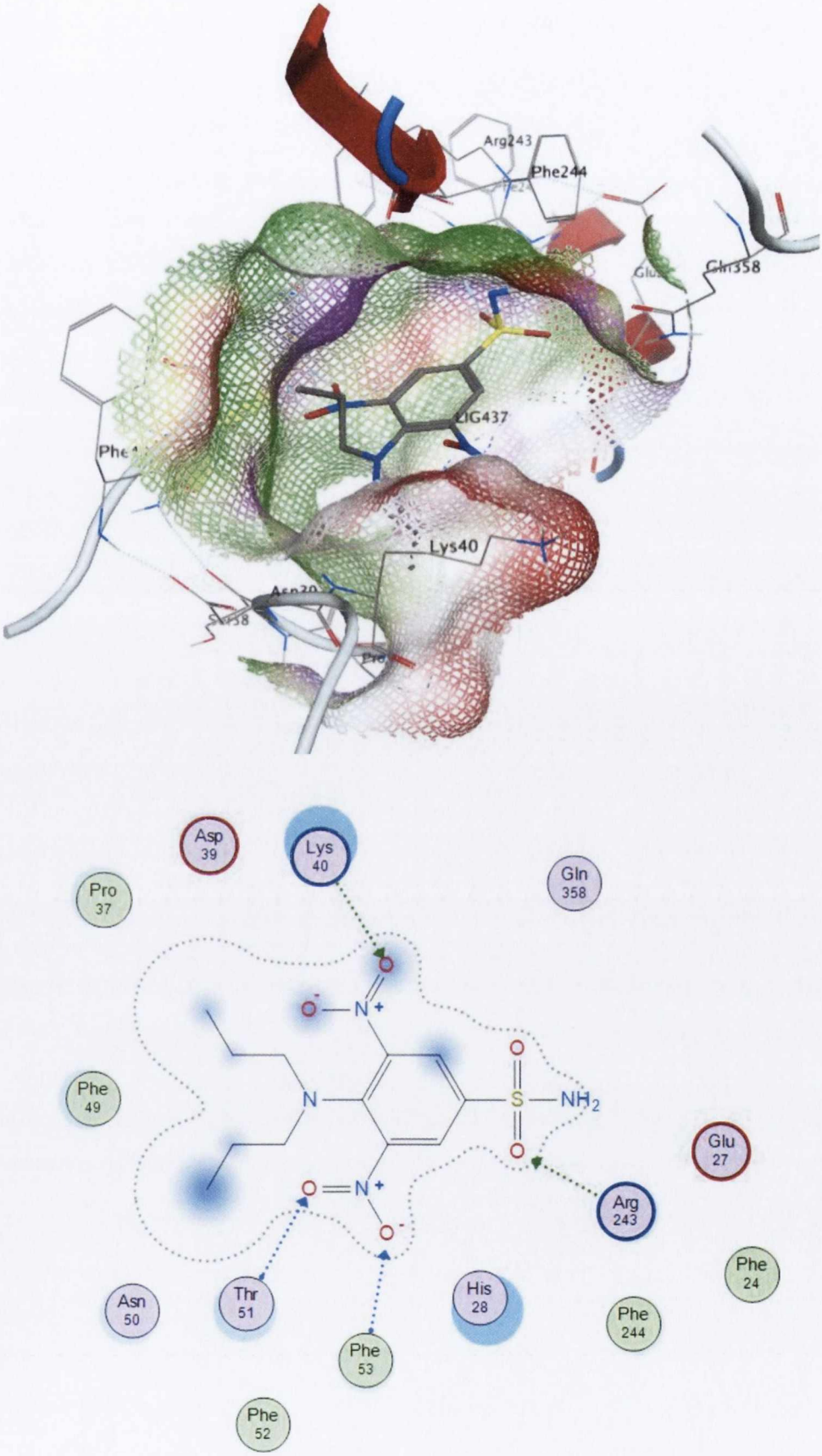


FIGURE 3.18: Ligand interaction diagram of OZN in site 1 of malarial tubulin.

3.16.2 Analysis of binding site 4 in parasitic tubulin

Examination of the post-MD poses of APM in site 4 of the Malaria and Trypanosomal models showed that both poses occupied roughly the same area of the pocket (Figure 3.19 A and B). According to the ΔG total values from Table 3.3, the malarial tubulin-APM complex had a more favourable binding free energy than the trypanosomal tubulin-APM complex. A more in depth look at the shape of the binding pocket revealed the reason for this; binding site 4 contained two smaller hydrophobic sub-pockets (in both models) and in the malaria tubulin-APM complex, the nitrophenyl group showed great complimentarity with subpocket 1 (Figure 3.21) while the hydrophobic isopropyl group pointed into subpocket 2. This interaction had greater hydrophobic contacts with the receptor compared to the trypanosomal tubulin-APM complex which only interacted with subpocket 2 through its nitrophenyl group (Figure 3.23). This left the isopropyl group exposed to solvent, reducing the number of hydrophobic contacts available for the ligand to make with the protein.

With reference to the OZN post-MD docked poses, both the malarial and trypanosomal poses occupy nearly identical positions within the binding pocket (Figure 3.19 B). In the malarial pocket one arm of the dipropylamide group of OZN points into subpocket 2 of site 4 while one of the ring nitro groups makes hydrophobic contacts with helix 3 (Figure 3.22). There is also a π(arene) hydrogen bond to the phenyl ring of OZN and a hydrogen bond between carbon 1 of the dipropylamide group and H88. In the trypanosomal tubulin-APM pose (Figure 3.24), the ΔG total is more favourable and was reflected in the interactions the pose makes in the binding site; one of the ring nitro groups pointed directly into subpocket 2, making hydrophobic contacts with V66, E90, and R121, while one arm of the dipropylamide group of OZN points into subpocket 1, interacting with helix 3. R121 also makes a hydrogen bond with an oxygen atom on the OZN sulfonamide group and a final hydrogen bond is established between the sulfonamide amine and E90.

Aligning the final MD poses of APM and OZN in malarial tubulin reveals a common interaction with H88. Both also have a hydrophobic group pointing into subpocket 2 (isopropyl group of APM and dipropylamide group of OZN). The ring groups of both compounds lie in opposite orientations after the final frame of MD; the sulfonamide group of OZN points out into the solvent (a position stabilised by the π(arene) hydrogen bond)

while the APM nitrophenyl group points down towards a hydrophobic pocket made up of residues 55-65. The hydrophobic contacts brought about through this placement probably result in the greater ΔG total, although the difference between the poses is not statistically significant.

In binding site 4 the difference between herbicide-susceptible species and herbicide-resistant species is the sole change of a glutamine in the resistant species to an asparagine in the susceptible species. It is possible that the glutamine in resistant species causes a higher level of steric hindrance towards the herbicide compounds. Instead, an asparagine residue at the beginning of helix 3 would offer less steric hindrance towards the phenyl rings of OZN and APM, thereby improving the likelihood of increased ligand stability.

FIGURE 3.19: Post-molecular dynamics poses of APM and OZN in site 4 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

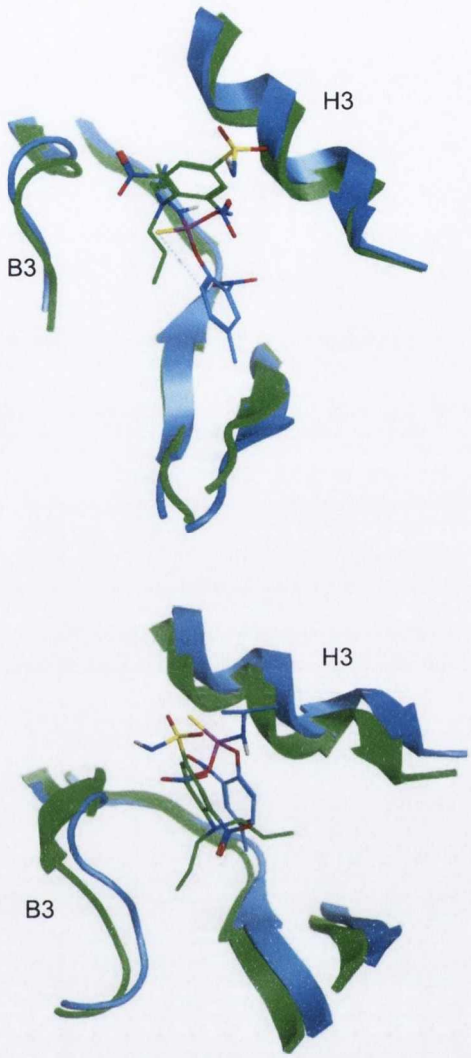


FIGURE 3.20: Post-molecular dynamics poses of APM and OZN in site 4 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 4 of malarial tubulin. (B) The orientations of APM and OZN in site 4 of trypanosomal tubulin

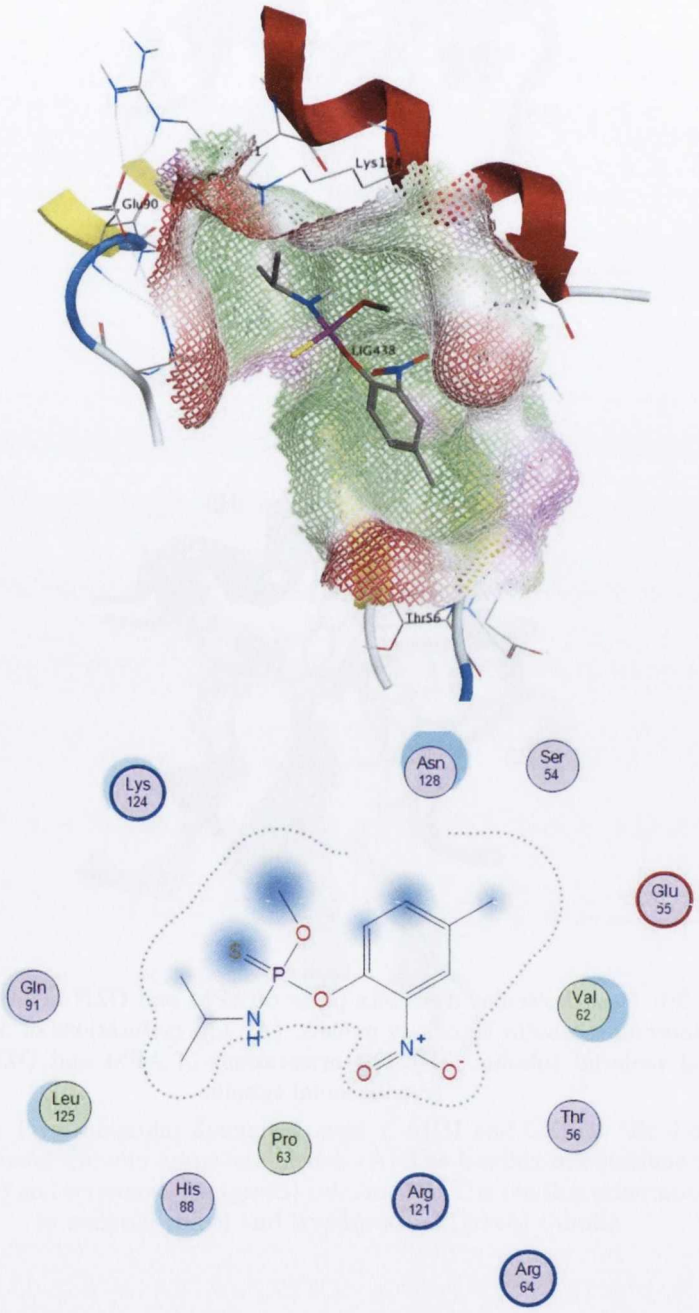


FIGURE 3.21: Ligand interaction diagram of APM in site 4 of malarial tubulin.

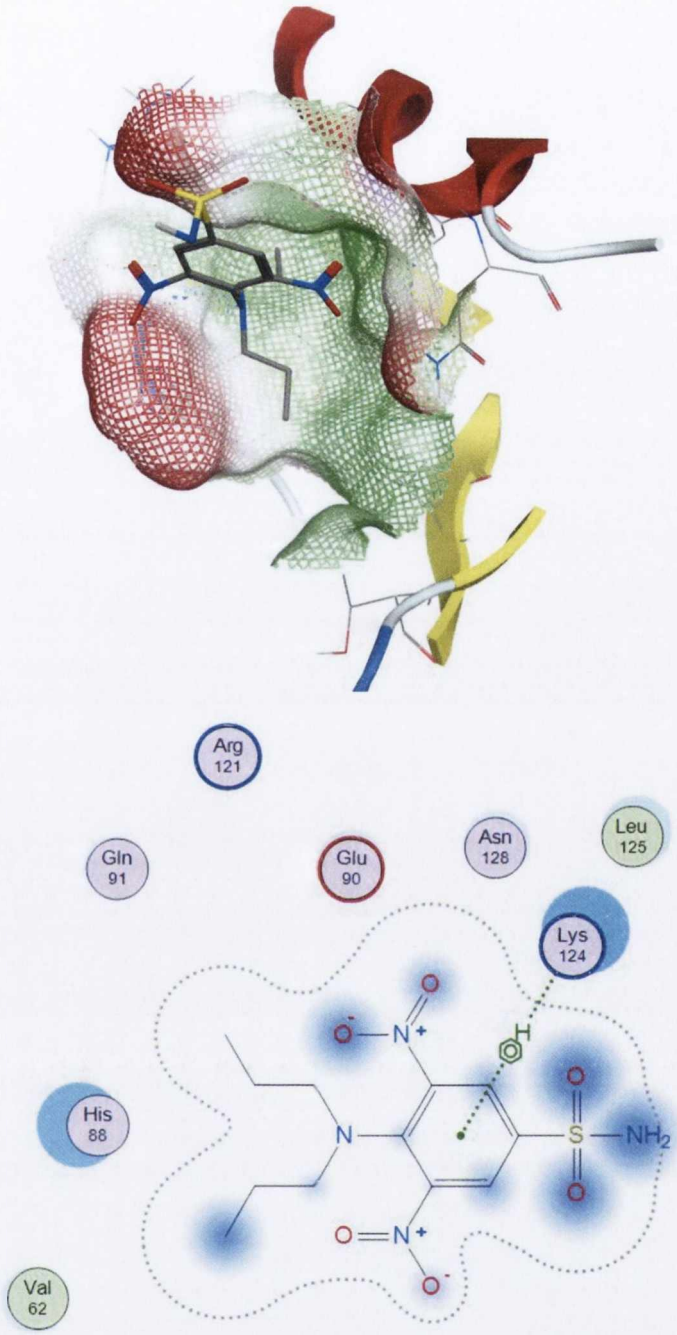


FIGURE 3.22: Ligand interaction diagram of OZN in site 4 of malarial tubulin.

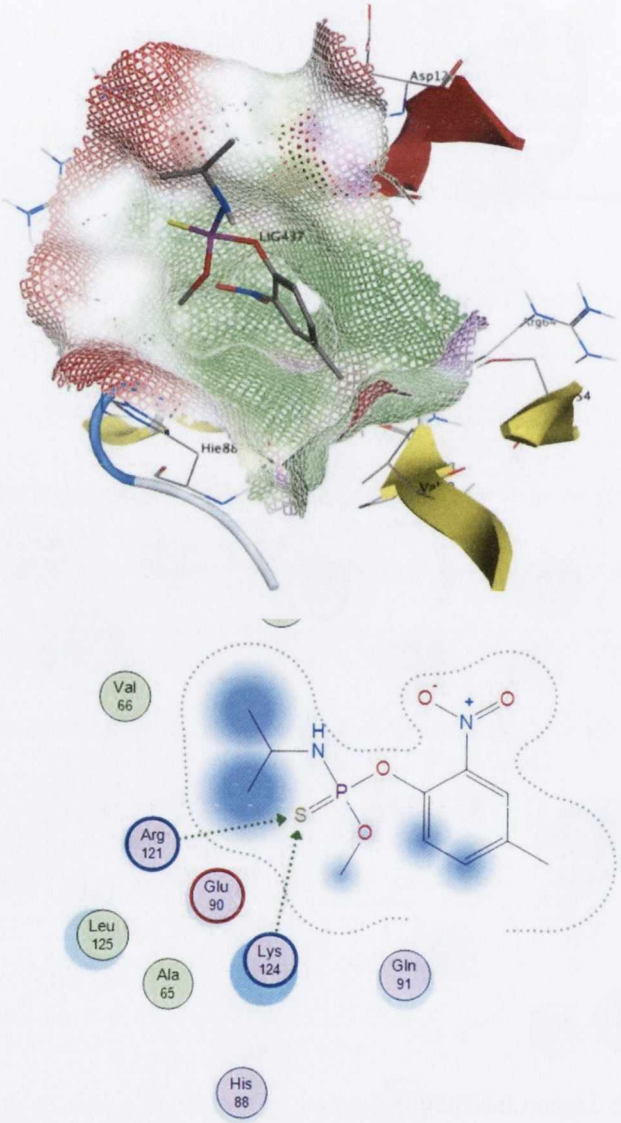


FIGURE 3.23: Ligand interaction diagram of APM in site 4 of trypanosomal tubulin.

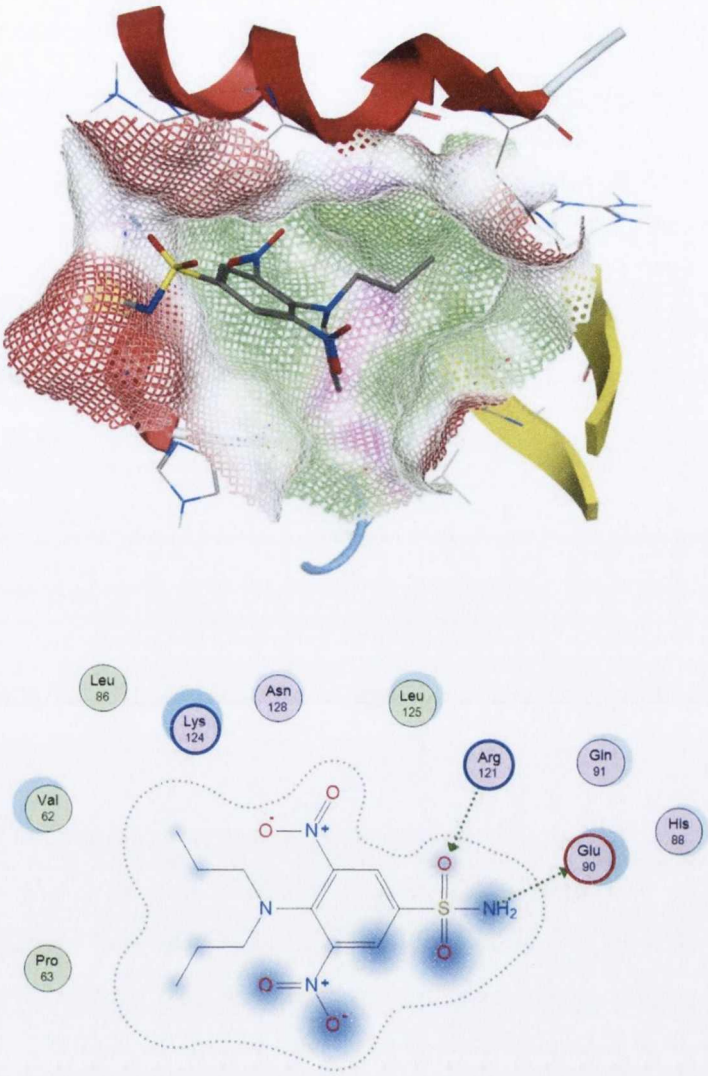


FIGURE 3.24: Ligand interaction diagram of OZN in site 4 of trypanosomal tubulin.

3.16.3 Analysis of The Proposed Morrissette site in parasitic tubulin

In addition to the 6 binding sites analysed, we decided to apply the same techniques to calculate the free binding energies of APM and OZN to malarial and trypanosomal tubulin. Figure 3.25 below displays the graphs of the free energies obtained from the 10ns produciton dynamics simulations of the APM- and OZN-bound tubulin models. In reference to the trypanomal graphs, the ΔG of binding never stabilises and by 8ns into the simulation OZN had completely left the binding site and APM hs also reached a near zero ΔG. For the malarial poses, both APM and OZN are stronger than the trypanosomal poses, but still almost -30kcal/mol weaker than the equivalent energies calculated for binding site 1. The malarial poses of APM and OZN also appear to be highly unstable and multiple times along the 10ns simulation the ΔG approaches zero.

3.16.4 Analysis of the mammalian equivalent of binding site 1

The most energetically favourable binding site from the parasitic MM-PBSA studies, binding site 1, was chosen for a direct comparison with the corresponding region on the 1TUB tubulin protein. Looking at the 1TUB protein it was clear that the H1-B2 loop was smaller than in the homology models (caused by the missing residues) and the entrance to what would be binding site 1 was blocked by H28, D47 and F49. A small binding pocket was detected in the region, overlapping slightly with binding site 1. Figure 3.26 shows the docked poses of APM and OZN post-MD in the 1TUB site 1 equivalent. However, this site was very shallow and this was reflected in the results of the MM-PBSA calculations. After the MD simulations and free energy calculations, OZN and APM had binding free energies of -3.2 and -18.6 kcal/mol respectively. As can be seen in fig 3.37, OZN migrated away from the binding site entirely. These results indicate that binding site 1 from parasitic tubulin is not present in this mammalian structure.

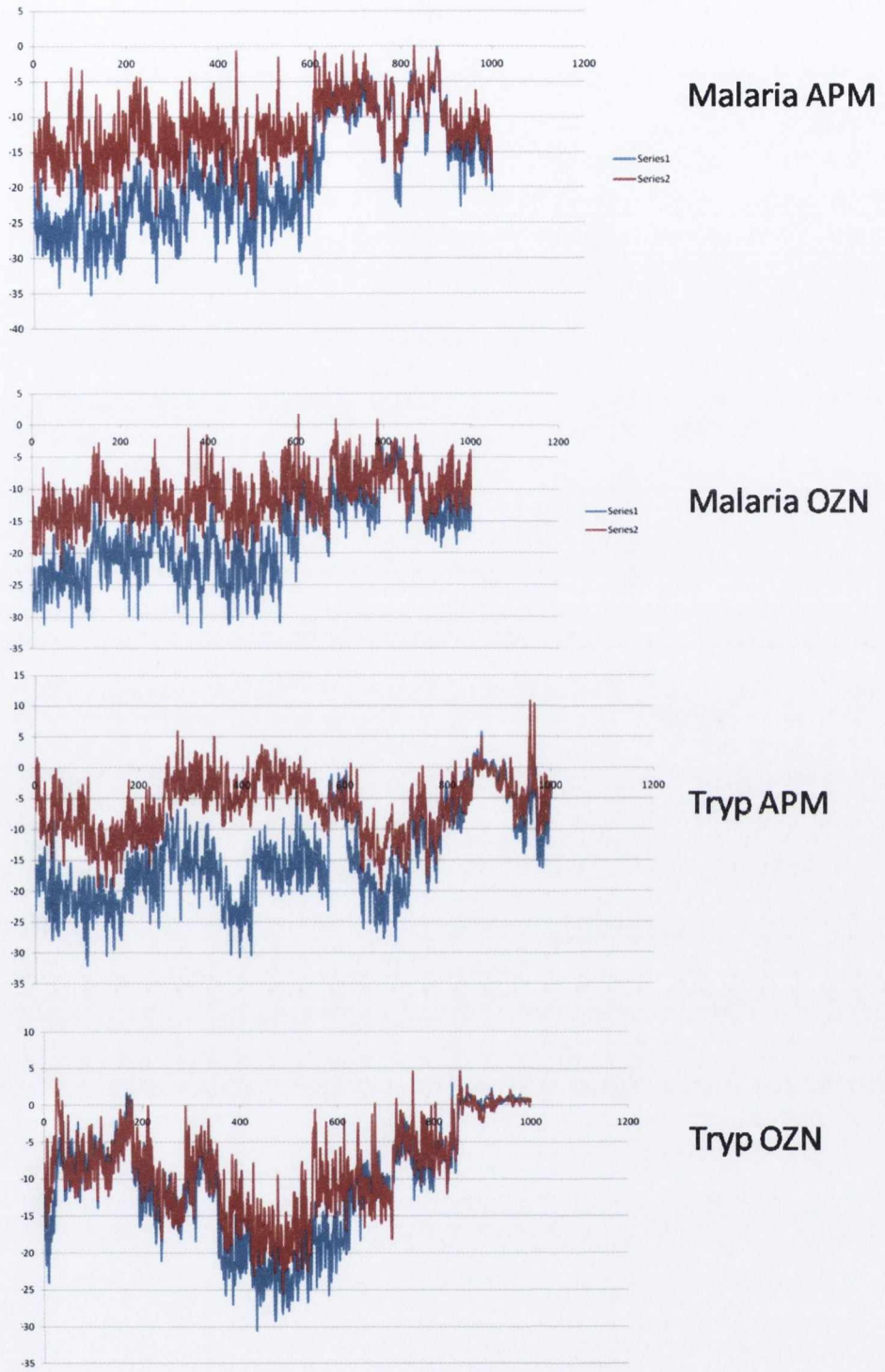


FIGURE 3.25: Graphs of binding MM-GBSA (blue line) and MM-PBSA (red line) free energies of APM and OZN in the malarial and trypanosomal tubulin homology models for the proposed "Morrissette" site.

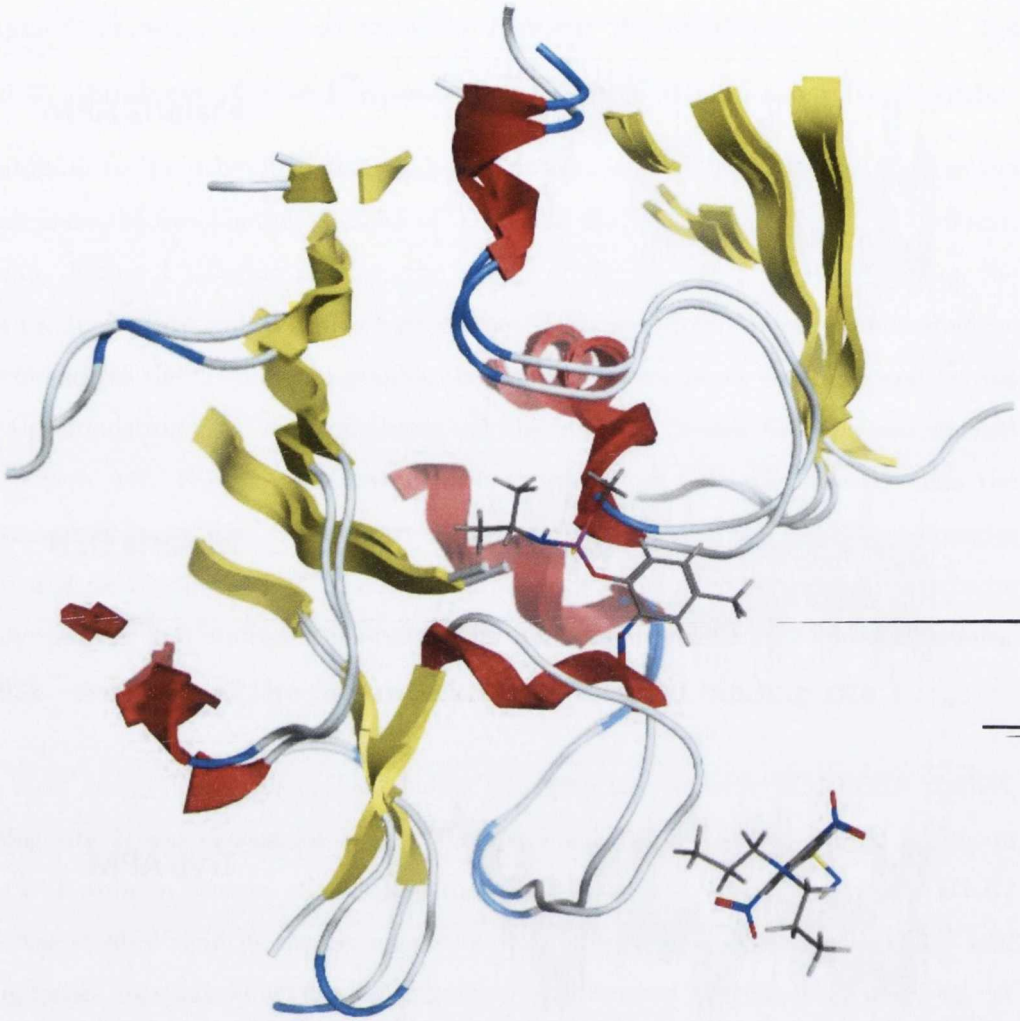


FIGURE 3.26: Post-molecular dynamics poses of APM and OZN in the site 1 region of mammalian tubulin.

3.17 Discussion and Future Work

In this chapter we discussed a computational approach to investigate the existence of a parasite-selective herbicide binding site on the tubulin protein. The final protocol involved homology modeling followed by induced fit docking and then a 10ns molecular dynamics simulation with a view to calculating binding free energies of the herbicide poses.

Six binding sites were chosen on the parasitic tubulin models based on amino acid substitution data between susceptible and resistant species. In an attempt to find a simpler approach to locating a potential binding site, ROCS shape-based searching was used to rank the six binding sites based on their shape similarity to the herbicides APM and OZN. This approach didn't offer any clear information on which binding site might be the herbicide site, and attempts at validating this approach using known binding sites and their bound ligands also failed due to an apparent punishment of negative space between the binding site and herbicide shapes during the ROCS shape matching procedure. Following this, a more accurate and sensitive approach was devised involving induced fit docking followed by a 10ns molecular dynamics simulation with binding free energy calculations. This approach offers many advantages over docking alone, such as a more detailed treatment of solvation effects and protein-ligand interactions. Typical scoring functions treat solvation effects with overly simplistic models such as the distance dependent dielectric. Many recent papers have detailed the superiority of MM-PBSA energy functions over typical scoring functions, showing them to be more robust and having less performance fluctuations.

With this in mind, both APM and OZN were docked into the 6 chosen binding sites using an induced fit protocol in MOE before being subjected to MM-PBSA calculations. In addition to this, the ligands were docked to the proposed 'Morrissette' site on parasitic tubulin to provide a comparison. Of the binding sites selected for this investigation MM-PBSA results of APM and OZN show a clear preference for binding site 1 in terms of predicted binding free energy. However, when the poses for APM and OZN in binding site 1 were investigated further it revealed that there was no apparent consensus binding pose between the two ligands. Binding site 4 was the only binding site that offered up an agreement in binding poses of APM and ORZ between and within the 2 parasitic models. However, in terms of binding free energies calculated, binding site 4 recorded a

value between 10-20 kcal/mol lower than binding site 1. These results could have been biased by the starting binding site shapes as binding site 1 is much larger than binding site 4 (and much more flexible) and so during the induced fit docking procedure and molecular dynamics simulations there is a larger chance of positional variation within the docked poses. This would perhaps account for the tight poses generated in site 4 and the variation in poses generated within site 1. A second source of bias introduced into the study was that obtaining the exact same binding site shape between malarial and trypanosomal tubulin models was problematic due to the inherent flexibility of the proteins and stochastic nature of homology modeling and molecular dynamics simulations. As it stood, acquiring matching binding site topologies between the two proteins was unfeasible. This problem could be overcome with the use of multiple protein snapshots at the induced fit docking stage to remove starting structure bias, as well as a cluster analysis of the most stable poses generated instead of selecting only the top ranked pose and selection of multiple poses for MM-PBSA calculations. However, in terms of storage space required for a single molecular dynamics trajectory of a single ligand in a single binding site, expanding this already sizable calculation to even 9 additional snapshots was impractical. With a single run requiring at least 10Gb of storage space, the space demands for 3 proteins (malarial, trypanosomal and mammalian), with 7 binding sites and 2 ligands over 10 snapshots would top out at nearly 4.5Tb.

Even with the limitations discussed, there is a clear trend towards site 1 as the most energetically favourable binding site for the herbicide inhibitors and binding site 4 in terms of a common ligand binding pose. . This goes directly against previously published data, specifically the proposed Morrissette binding site. In analysing the binding free energies of APM and OZN in the Morrissette site this work would infer that this is not the preferred binding site for these compounds. In these studies the Morrissette site appeared shallow and overly exposed to solvent and during the molecular dynamics simulations the trpanosomal poses of APM and OZN migrated out of the binding site after roughly 8ns, while the malarial poses started to approach zero after 6ns. In the work of Morrissette *et al.* describing the search for the OZN binding site on *Toxoplasma gondii* α-tubulin, they reported a binding affiinity of 23nm (after a 2.5ns MD trajectory) in their predicted 'Morrissette' site. If we in our work analyse only the first 2.5ns of the APM and OZN trajectories in the Morrissette site favourable binding energies are obtained. However, the binding free energies for APM and OZN in the 'Morrissette' site

at this stage are still undergoing large fluctuations, a clear indication that the protein-ligand complex has not reached an energetically stable state. Therefore, taking the binding free energy after 2ns in this work offers up highly misleading results.

Another apparent problem with the Morrissette work, and perhaps a reason why they failed to recognise binding site 1 in their work, is that during homology model creation they chose to model the missing residues in the α-subunit of the 1JFF template structure using the equivalent region in the β-subunit of 1JFF. Analysis of these two short regions of the α and β sequences, roughly 26 amino acids long, shows that homology between the two is only 20%. In the β subunit these amino acids form an alpha helix composed of residues 40-53. This same problem may also reflect a bias existing in the present work, namely that the H1-B2 loop is in the wrong conformation after homology modeling. Recently, several crystal structures have emerged that have the H1-B2 loop fully resolved. These structures show that the H1-B2 loop contains a small helix composed of F49, N50, and T51 (residues present in the parasitic tubulins) close to the B2 loop that is not present in the homology models. These results would imply that there is a bias present in the current study that makes site 1 appear more open in the models than it would perhaps be in their actual structure. It also illustrates the bias in the work of Morrissette *et al.* as the β-tubulin H1-B2 loop has a much larger helix in it, completely disfiguring the corresponding loop in their α-tubulin model. A short molecular dynamics simulation could reveal how flexible the 3 residues blocking entrance to site 1 are, detailing their fluctuations and potentially disclosing how often the binding site may be blocked. The B-factors for Asparagine 50 are particularly high and would indicate that the certainty of this amino acid sidechain position is in doubt.

Taking the parasitic MM-PBSA results into consideration, APM and OZN were docked into the equivalent region of binding site 1 from mammalian tubulin (1TUB) in order to assess the presence of selectivity in the resulting binding free energies. The results showed that although there is a small, shallow pocket present in the region, the parasitic site 1 was not present in mammalian tubulin. However, the cause of this is perhaps down to the reasons previously discussed; the missing residues in the 1TUB structure made it difficult ro predict the presence and orientation of a small helix at the end of the H1-B2 loop that blocks entry to the site observed in the parasitic models lacking this 3 residue helix. These problems arose out of a lack of sufficient structural data to correctly inform the modeling process.

The work detailed in this chapter captures an attempt at a rational, computational approach to identify a herbicide-selective binding site on parasitic tubulin models. However, in reality, efforts to locate ligand-binding sites are more generally the realm of complex biophysical techniques such as X-ray crystallography, NMR spectroscopy, epitope matching approaches in mass spectrometry, or genetics studies such as site-directed mutagenesis. However, for methods such as X-ray crystallography and NMR spectroscopy require large amounts of stable protein material and, as our collaborators found, purification of tubulin directly from *P. falciparum* is hindered by the scarcity of of the protein in the erythrocytic stage of the parasitic life cycle. To circumvent this recombinant maltose-binding protein(MBP)-tagged fusion tubulins were generated using the pMAL-c2x vector in *E. coli* expression systems, but attempts at cleaving the MBP tag using factor Xa were unsuccessful. Attempts to purify the malarial tubulin without the MBP tag resulted in a tendency of the tubulins to aggregate in the *E. coli* cells, most likely caused by the fact that the protein was expressed in isolation thus preventing the formation of a natural $\alpha\beta$ heterodimer or because the proteins had not folded properly. *T. brucei* tubulin has been purified previously to near heterogeneity in the laboratory of Prof Keith Gull in the early 1990s. This purified protein retained the ability to form structurally normal microtubules and didn't require the addition of taxol or nucleating microtubule fragments. In the course of this study the purification of *T. brucei* tubulin was not a goal and was not undertaken. However, following the failure of purification of *P. falciparum* tubulin, it is perhaps the most logical step forward in the continuation of this work to proceed with setting up a *T. brucei* functional tubulin purification protocol in an effort to provide a source of the protein for structural characterisation of herbicide binding and also for on-target ligand assays.

Future work will involve streamlining the molecular dynamics scripts for more efficient output. As it stands, the scripts generate many files that occupy much needed disc space but contain no valuable data. Current efforts are focussing on generating a list of unwanted files and updating the scripts so as to automatically remove them from the directories at the end of script executions. These more efficient scripts might allow us to pursue a more rigorous blind-docking and MM-PBSA approach that will remove some of the bias and errors that were detailed previously, specifically enabling the introduction of multiple receptor snapshots per binding site to dismiss any prejudice involved in using single receptor conformations.

3.18 Appendix A - Post-Molecular Dynamics Binding Pose Descriptions

3.18.1 Analysis of binding site 2 in parasitic tubulin

The poses of APM in malarial and trypanosomal tubulin also adopt dissimilar orientations, again associated with a disordered loop region in the trypanosome tubulin between beta sheet 8 and helix 9 (Figure 3.27 A). This resulted in APM coming out of site 2 of trypanasome tubulin and forming a weak interaction between R308 and the sulphur atom of APM (Figure 3.31). In the malarial tubulin, APM sits with the nitrophenyl group pointing into a hydrophobic pocket with G310 interacting through a π(arene) hydrogen bond to the phenyl ring (Figure 3.29). The interaction is further stabilised by interactions between the sulphur atom of APM and R339 and between T340 and the amino group of APM. The differences in binding pose is further emphasised through the anlysis of ΔG total in Table 3.3. The malaria APM tubulin complex has a ΔG total that is more than twice that (-30kcal/mol) of the trypanosome APM-tubulin complex (-10kcal/mol).

The OZN poses originally docked to site 2 also occupied different areas of the protein post-MD (Figure 3.27 B). In the malarial protein, OZN stayed bound to site 2 in between the H9-B8 loop and hydrogen bonds are made between R308 and the terminal amine group of OZN and also between the nitro groups of OZN and R308 and Q342, contributing to a ΔG total of -32 kcal/mol (Figure 3.30). In the trypanosomal tubulin, OZN has migrated away from site 2 and down towards a shallow pocket near helix 6. The complex is stabilised by hydrogen bonds between the terminal amine group of OZN and residue D211 and between a sulfonamide oxygen and R215 (Figure 3.32).

Analysing the poses of OZN and APM within malarial site 2 shows that the two compounds have similar poses and bind in the same pocket, although they don't make any overlapping protein-ligand interactions. Within the trypanosomal site 2, the two compounds occupy completely different areas of the protein, with only APM staying in the area of site 2.

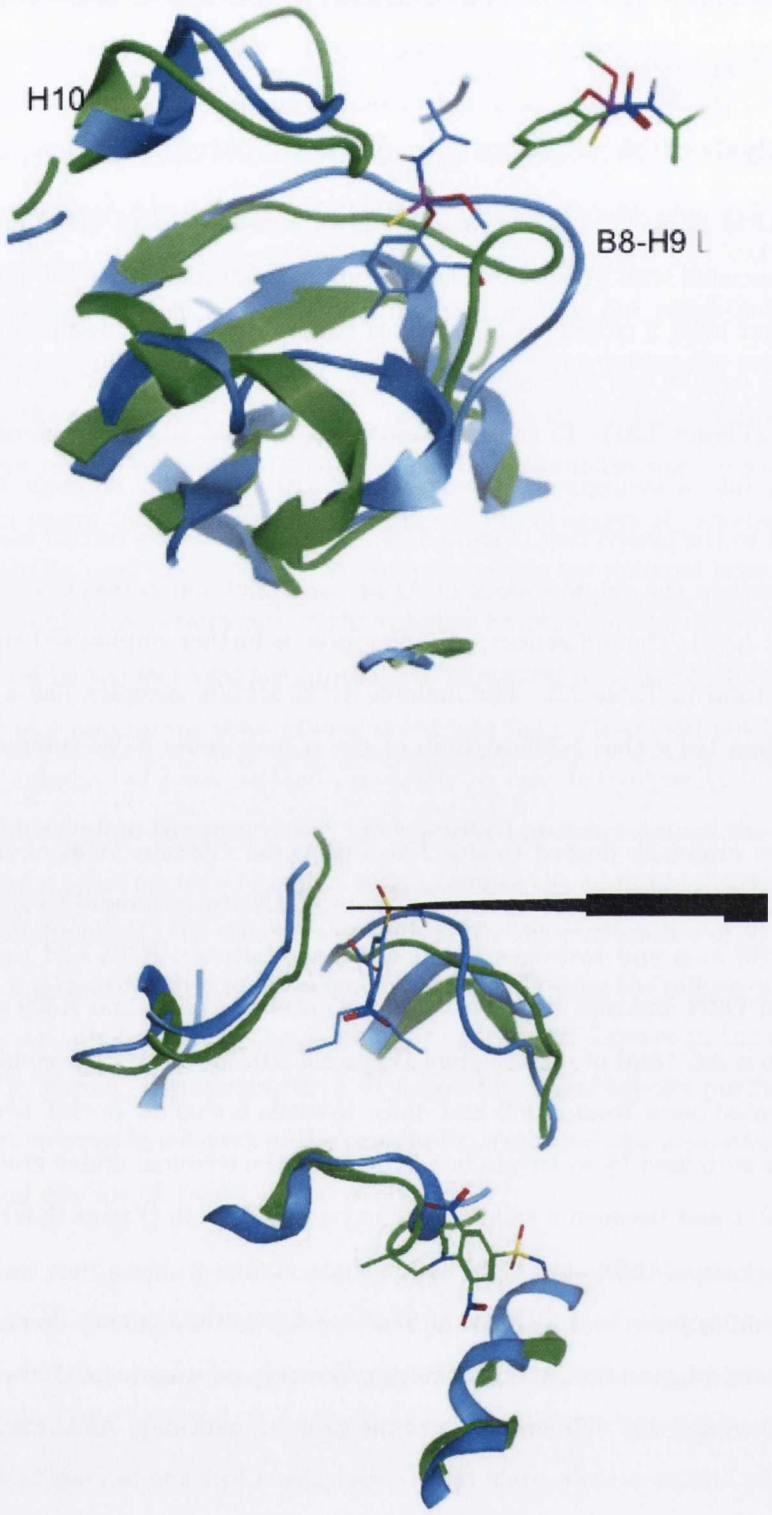


FIGURE 3.27: Post-molecular dynamics poses of APM and OZN in site 2 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

FIGURE 3.28: Post-molecular dynamics poses of APM and OZN in site 2 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 2 of malarial tubulin. (B) The orientations of APM and OZN in site 2 of trypanosomal tubulin

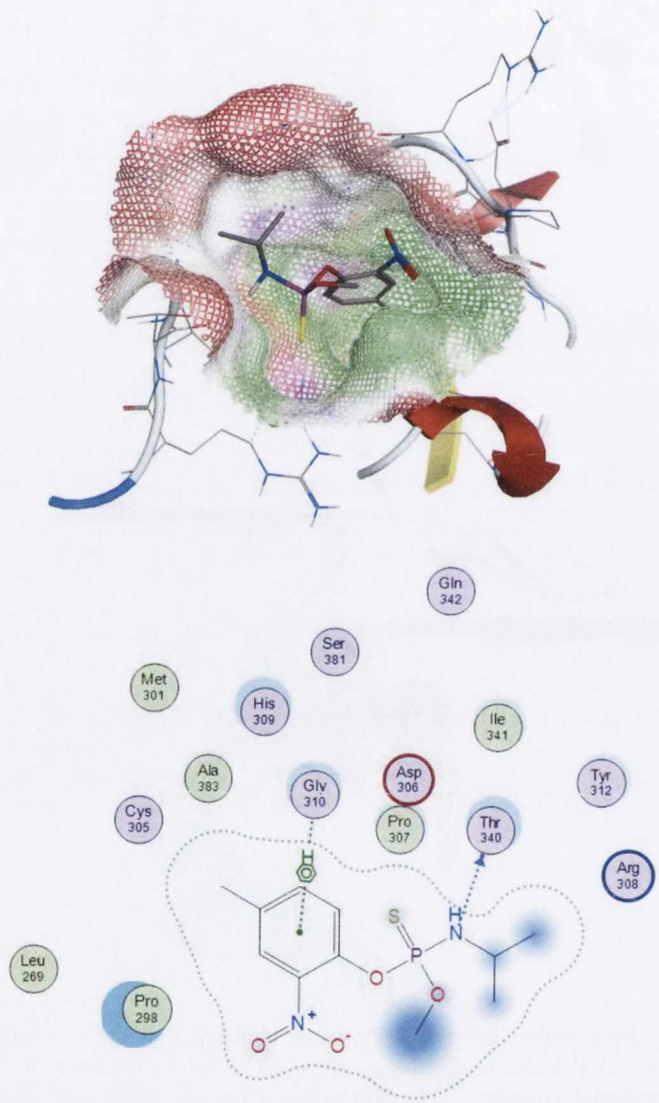


FIGURE 3.29: Ligand interaction diagram of APM in site 2 of malarial tubulin.

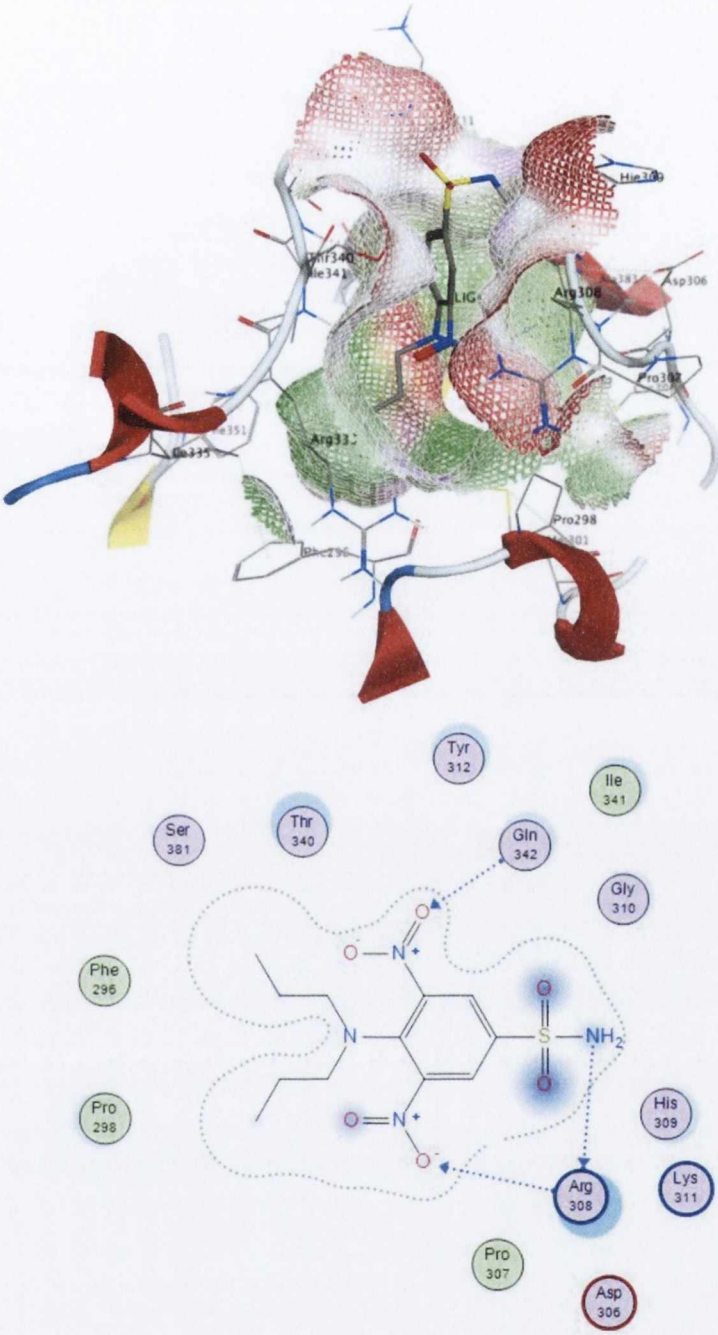


FIGURE 3.30: Ligand interaction diagram of OZN in site 2 of malarial tubulin.

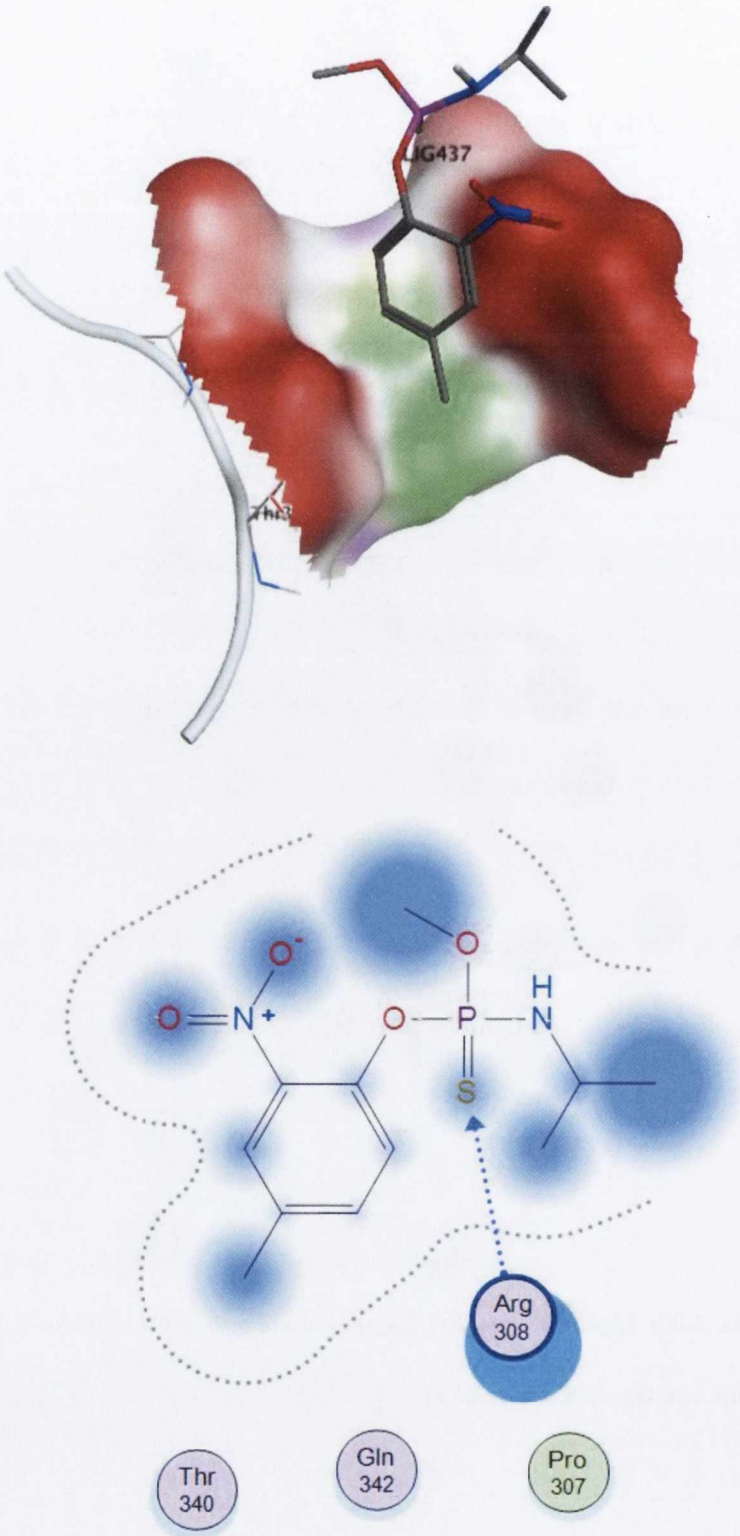


FIGURE 3.31: Ligand interaction diagram of APM in site 2 of trypanosomal tubulin.

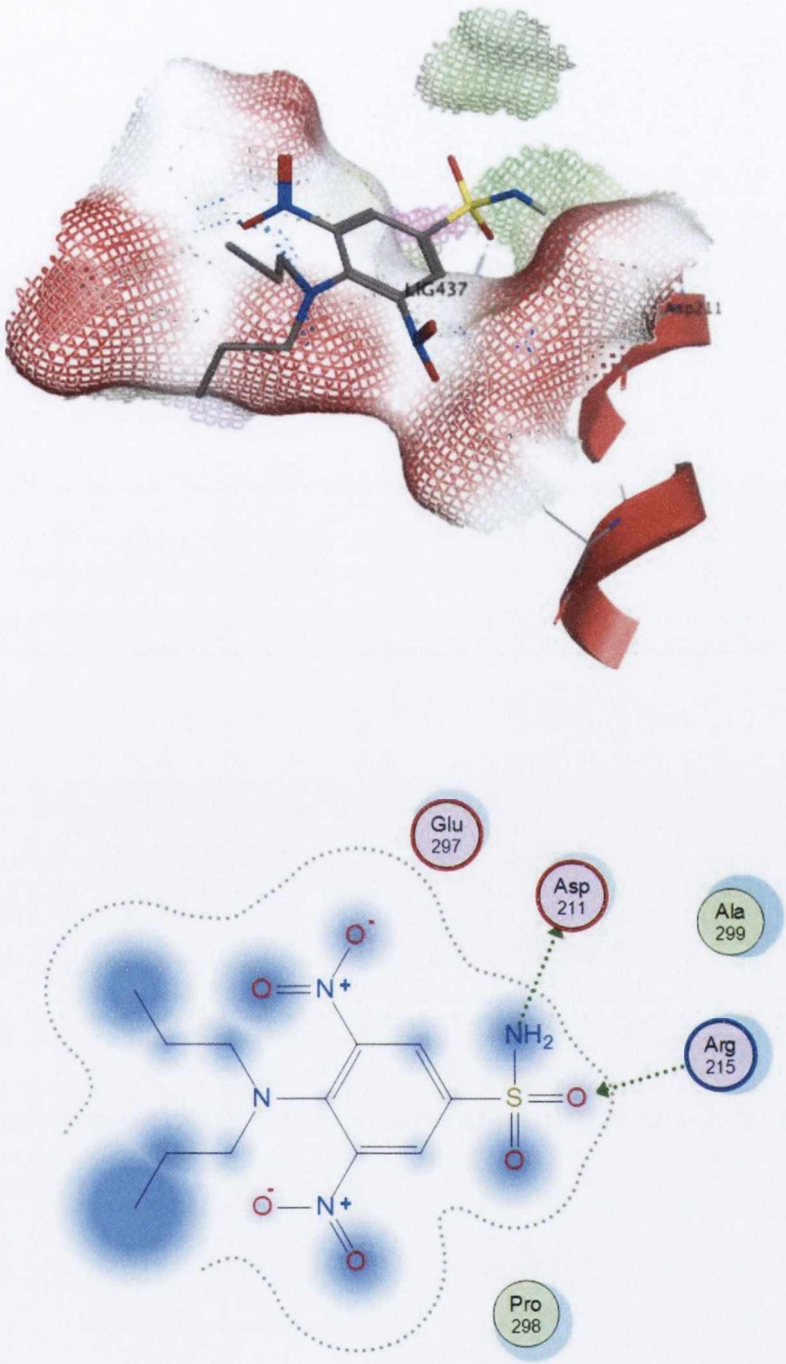


FIGURE 3.32: Ligand interaction diagram of OZN in site 2 of trypanosomal tubulin.

3.18.2 Analysis of binding site 3 in parasitic tubulin

The site 3 post-MD APM poses in malarial and trypanosomal models occupy the same area of the site, with the trypnasomal pose being slightly more exposed to solvent (Figure 3.33 A and 3.35). The malarial APM pose is more deeply buried in the site 3 pocket, with the position of loop region between helix 8 and beta sheet 7 being the main difference between the two poses. In the trypanosomal pose this loop encroaches on the binding site, hindering the position of the nitrophenyl group of APM through steric clashes with R264 and I265 (Figure 3.37); In the malarial model these residues take up positions that offer no resistance to ligand binding. Table 3.3 shows the ΔG total data for these poses. Although the ΔG total is more favourable for the trypanosomal APM pose, when the standard deviations of both poses are taken into account (data not shown) the differences are not significant.

Regarding the post-MD OZN poses in the models, both OZN poses left the binding pocket entirely and this is reflected in their near zero ΔG energies (Figure 3.33 B). In the malarial tubulin OZN has moved away from site 3 and in the final frame was found interacting weakly with helix 4, making a hydrogen bond to G162 while the rest of the ligand remained exposed to solvent effects(Figure 3.36). In the trypanosomal model (Figure 3.38), OZN is found next to helix 8 outside the binding pocket, and is also exposed to solvent. Due to the solvent exposed nature of the binding pocket, as well as an apparent lack of complementarity between the site and the ligands, both poses migrated away from the pocket and into the surrounding waters. This made any comparison between the APM and OZN parasite-specific MD poses unworkable.

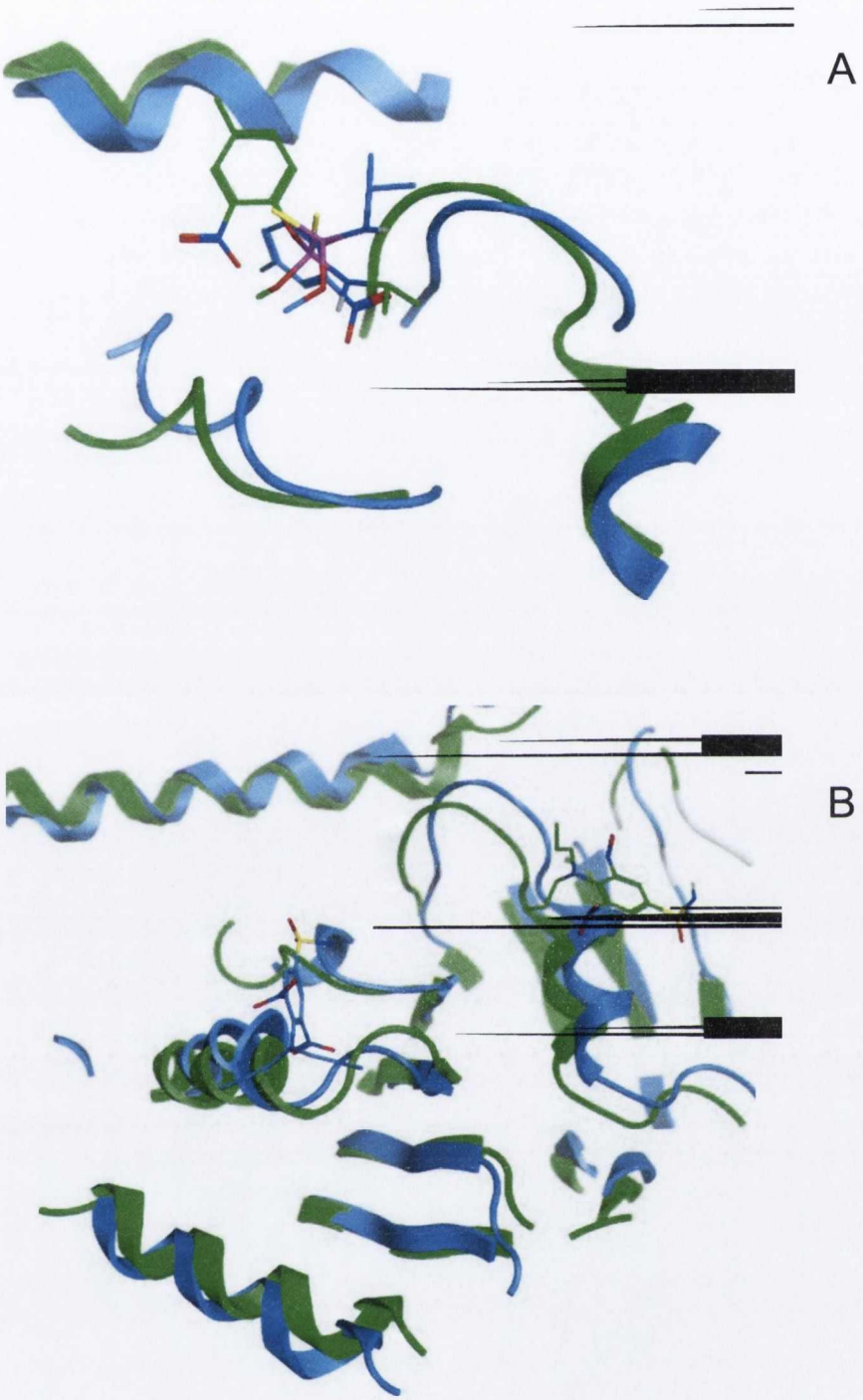


FIGURE 3.33: Post-molecular dynamics poses of APM and OZN in site 3 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

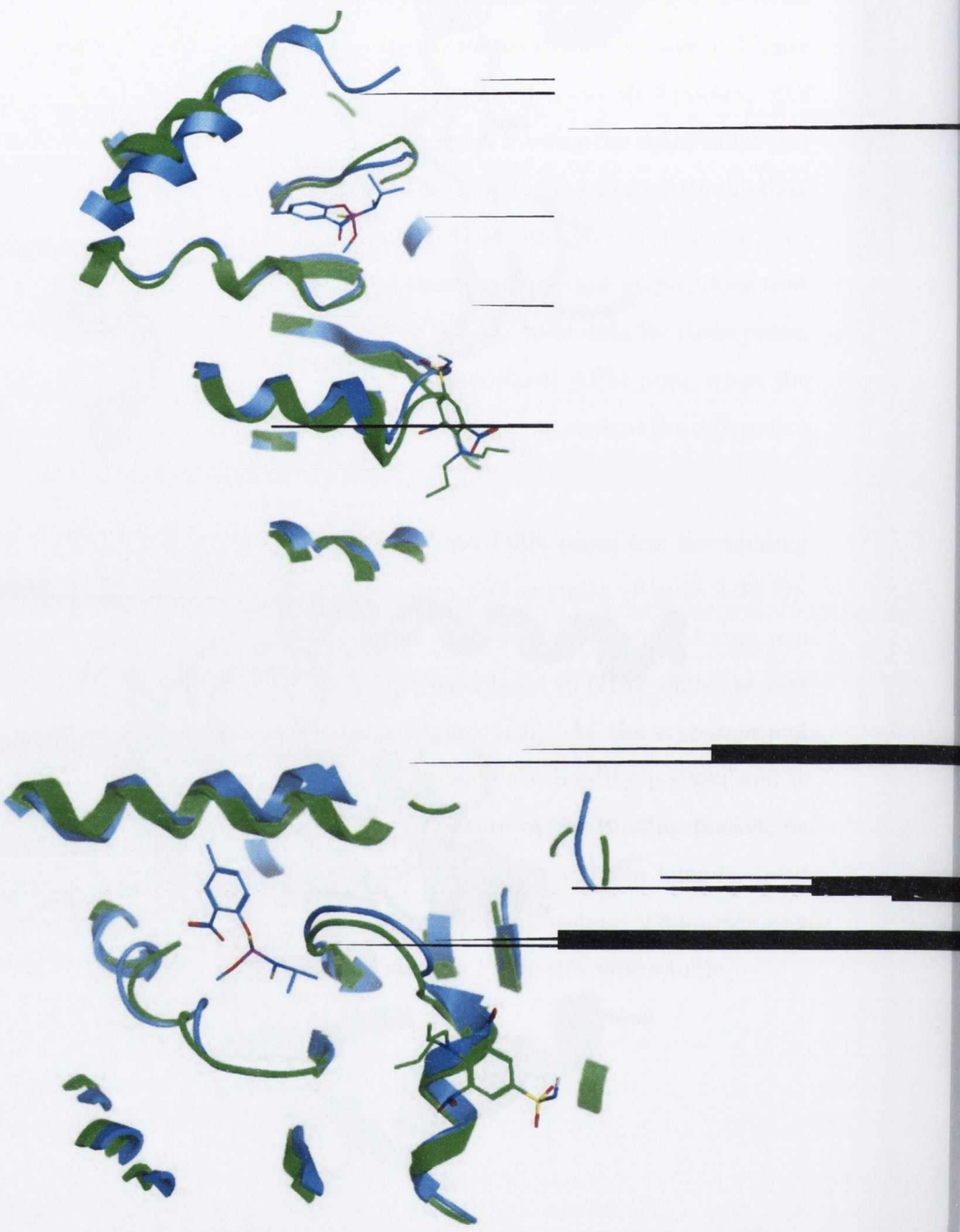


FIGURE 3.34: Post-molecular dynamics poses of APM and OZN in site 3 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 3 of malarial tubulin. (B) The orientations of APM and OZN in site 1 of trypanosomal tubulin

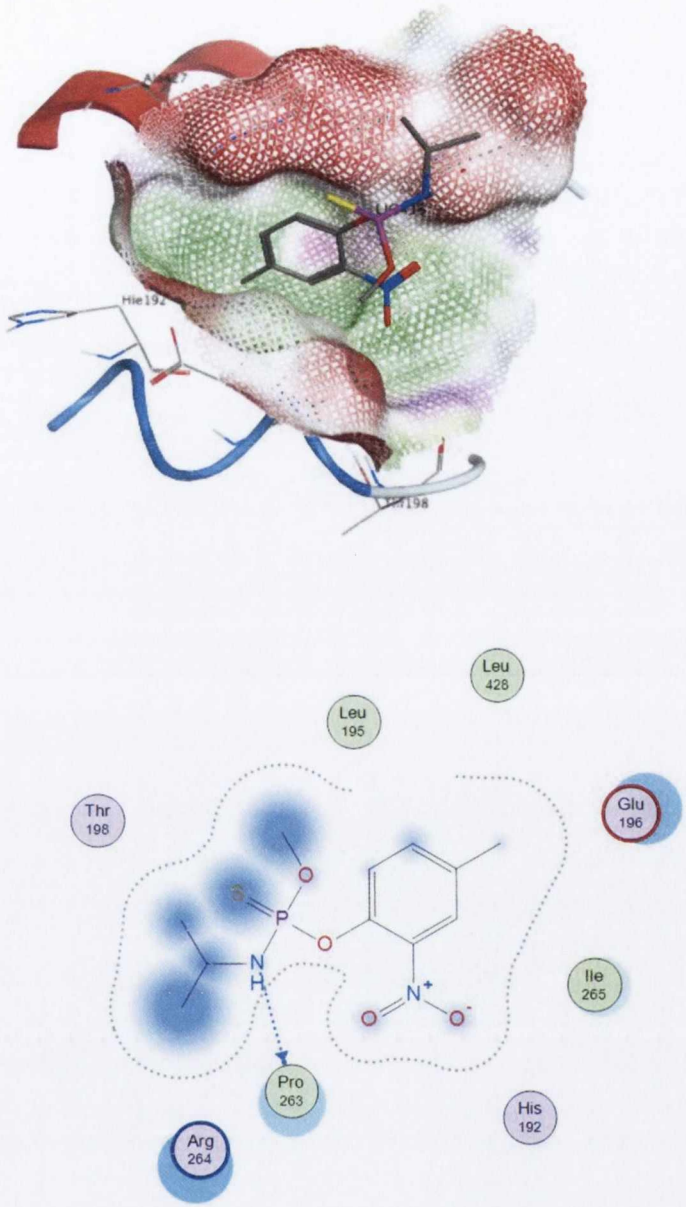


FIGURE 3.35: Ligand interaction diagram of APM in site 3 of malarial tubulin.

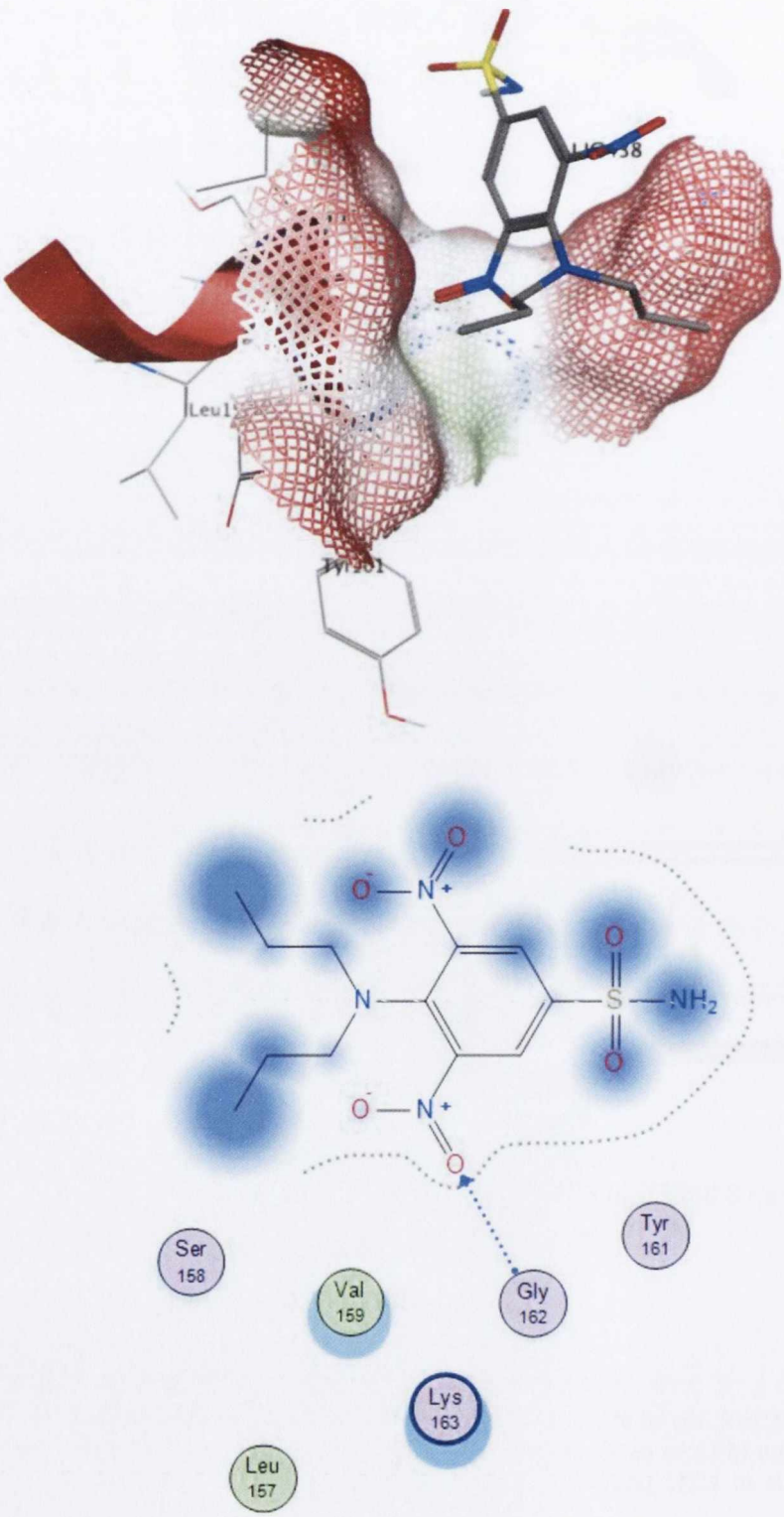


FIGURE 3.36: Ligand interaction diagram of OZN in site 3 of malarial tubulin.

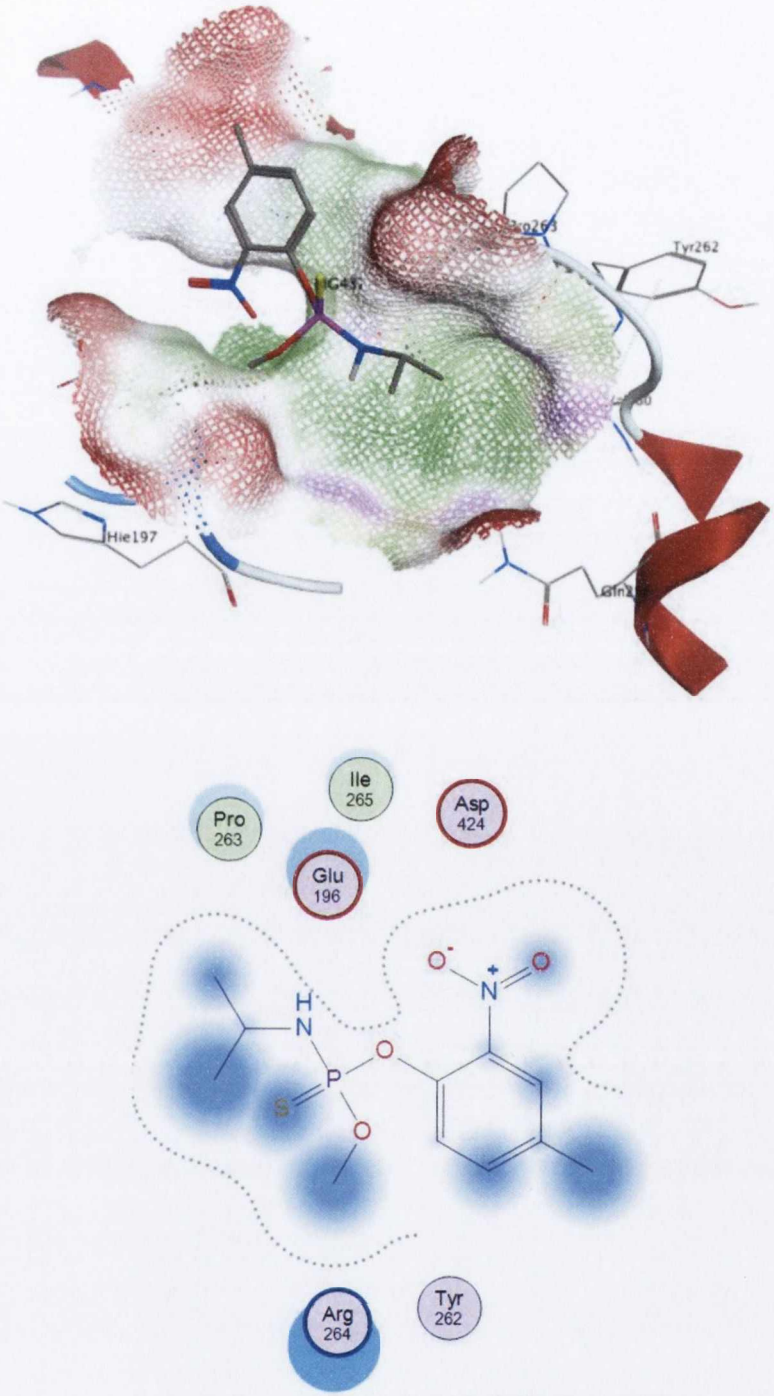


FIGURE 3.37: Ligand interaction diagram of APM in site 3 of trypanosomal tubulin.

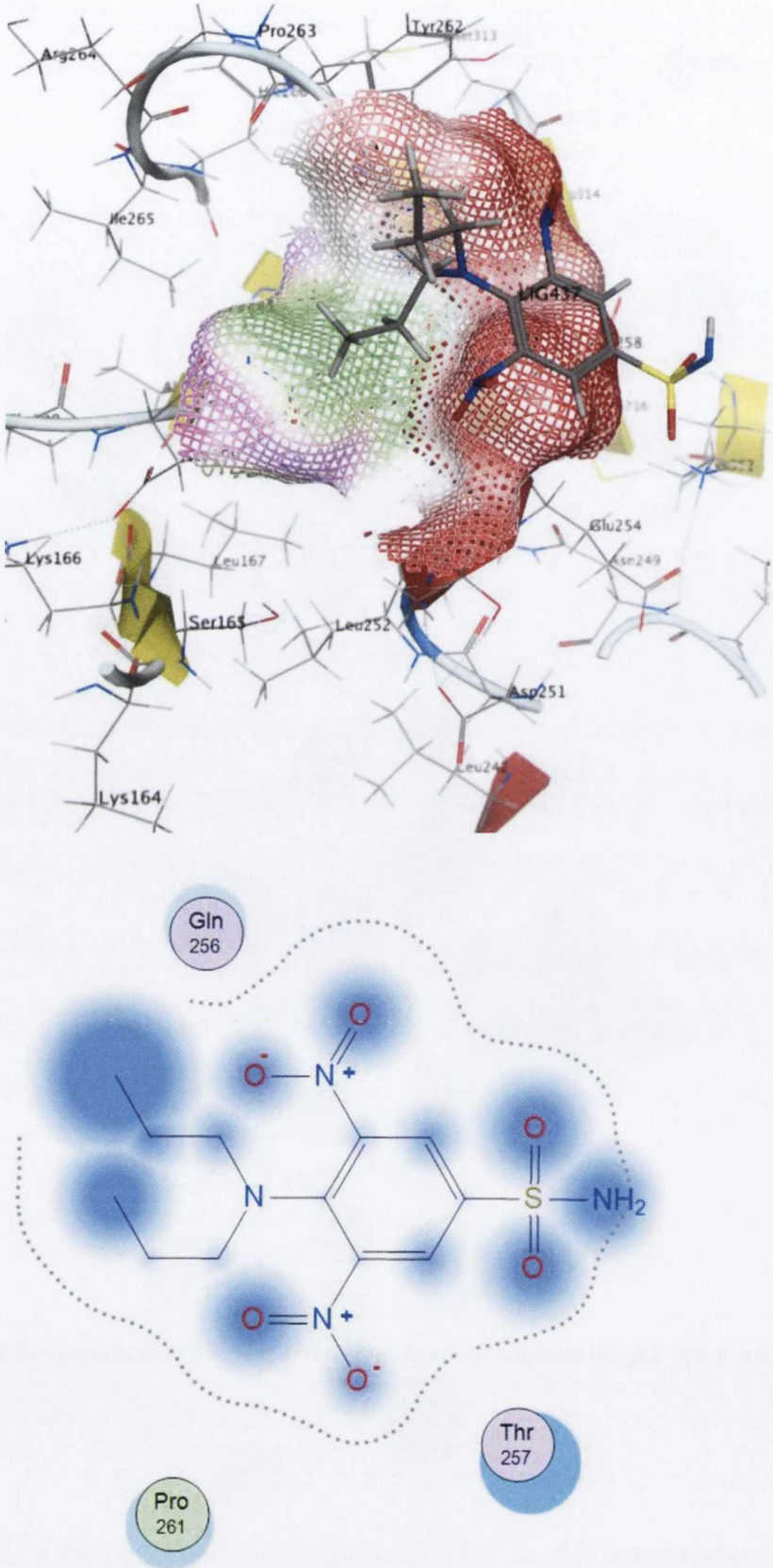


FIGURE 3.38: Ligand interaction diagram of OZN in site 3 of trypanosomal tubulin.

3.18.3 Analysis of binding site 5 in parasitic tubulin

Analysis of the poses generated at the end of the production dynamics simulation for APM in site 5 of malarial and trypanosomal tubulin showed no common binding orientation. In the malarial tubulin-APM complex, the hydrophobic nitrophenylgroup is buried inside the pocket forming hydrophobic interactions with helix 11 and the loop region of beta sheet 4 (Figure 3.41). Hydrogen bonds also stabilise the structure, occurring between P175 and the sulphur atom of APM and also between K394 and the APM amine. In the trypanosomal tubulin-APM complex, APM has left the pocket completely and is interacting with the protein through a very shallow, mildly polar pocket under the loop between helix 12 and helix 11 (Figure 4.43). A single hydrogen bond between F404 and the sulphur atom of APM is found in the complex. This unfavourable position taken up by APM in the trypanosomal site 5 is reflected in the ΔG total from Table 3.3 which shows that the pose in the malarial site had a binding free energy nearly twice that of the trypanosomal equivalent.

The poses of the post-MD OZN in malarial and trypanosomal tubulin follow the arrangements of the APM poses. In malarial tubulin, the OZN pose overlaps well with APM (Figure 3.40 A), with the dipropylamide group of OZN sitting in the same hydrophobic region as the nitrophenyl group of APM. The ring nitro groups of OZN also take part in hydrophobic interactions with P175 while the amine of the sulfonamide in OZN forms a hydrogen bond with E207 (Figure 3.42). In the trypanosomal tubulin, OZN and APM adopt divergent orientations within site 4 as a result of a highly disordered loop between helix 12 and helix 11 (Figure 3.40 B). The OZN pose within tryp tubulin occupied a region closer to the initial site 5 pocket and made a single hydrogen bond form its amine group to K401 (Figure 3.44). The APM pose had apparently migrated away from the initial pocket towards the helix11-12 loop, becoming exposed to solvent in the process.

The movement of the APM and OZN ligands away from the pocket of site 5 is clearly reflected in the ΔG table below. The trypanosomal poses, having significant solvent exposure, have much lower binding free energies than the malarial poses.

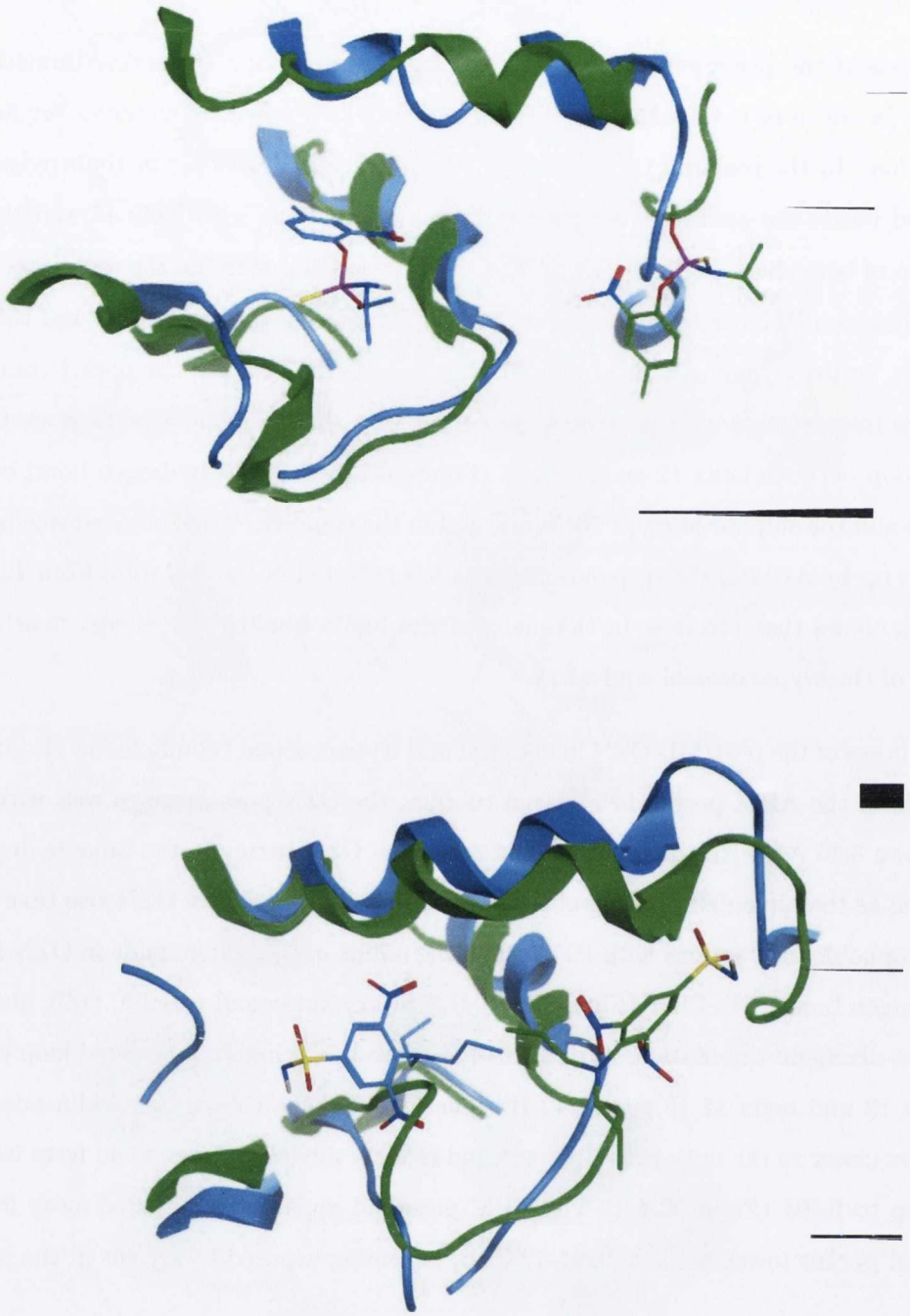


FIGURE 3.39: Post-molecular dynamics poses of APM and OZN in site 5 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

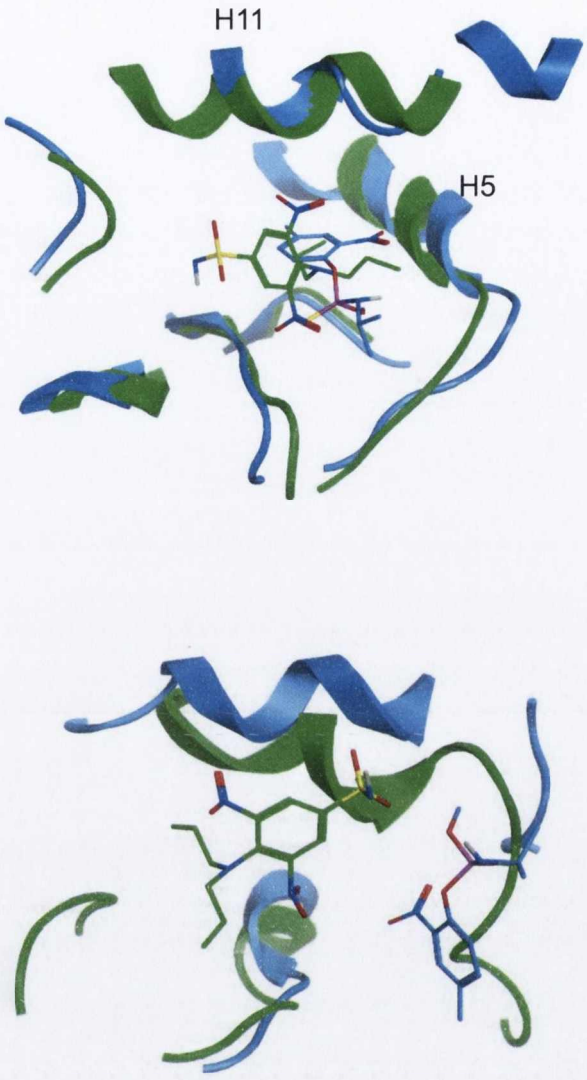


FIGURE 3.40: Post-molecular dynamics poses of APM and OZN in site 5 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 5 of malarial tubulin. (B) The orientations of APM and OZN in site 1 of trypanosomal tubulin

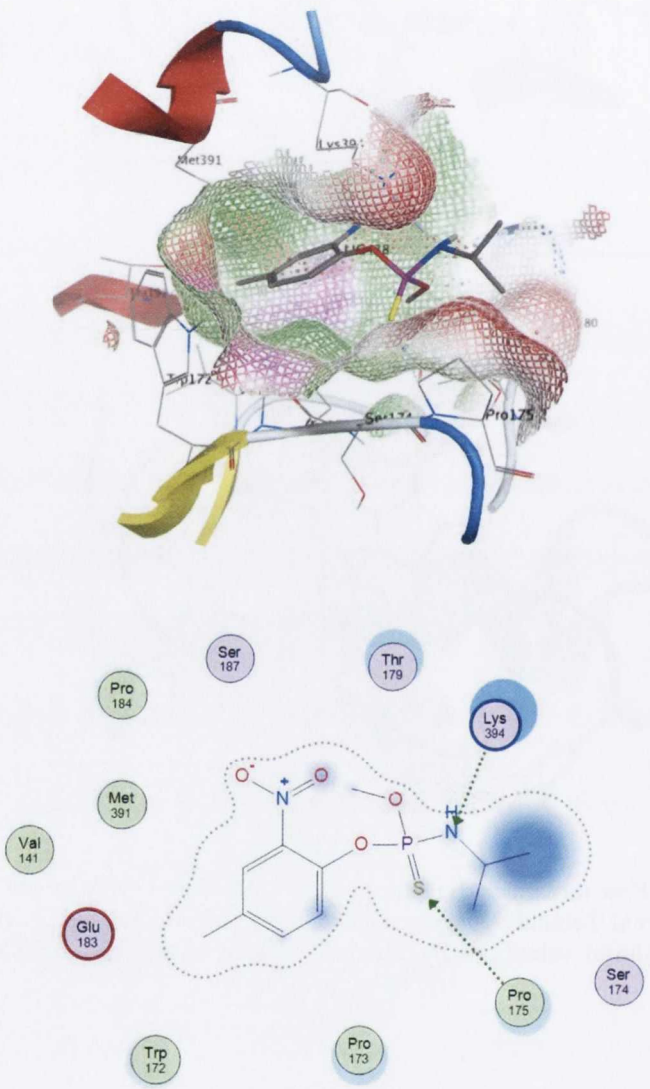


FIGURE 3.41: Ligand interaction diagram of APM in site 5 of malarial tubulin.

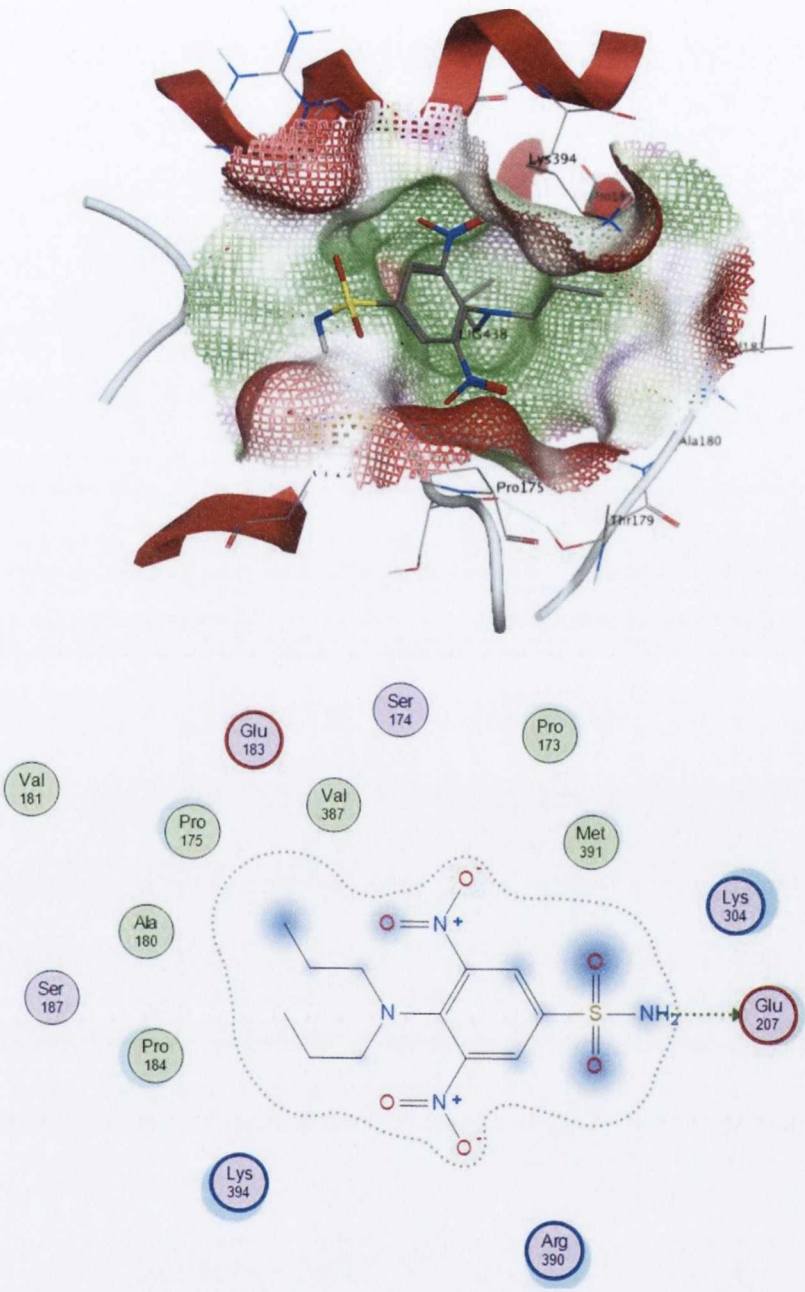


FIGURE 3.42: Ligand interaction diagram of OZN in site 5 of malarial tubulin.

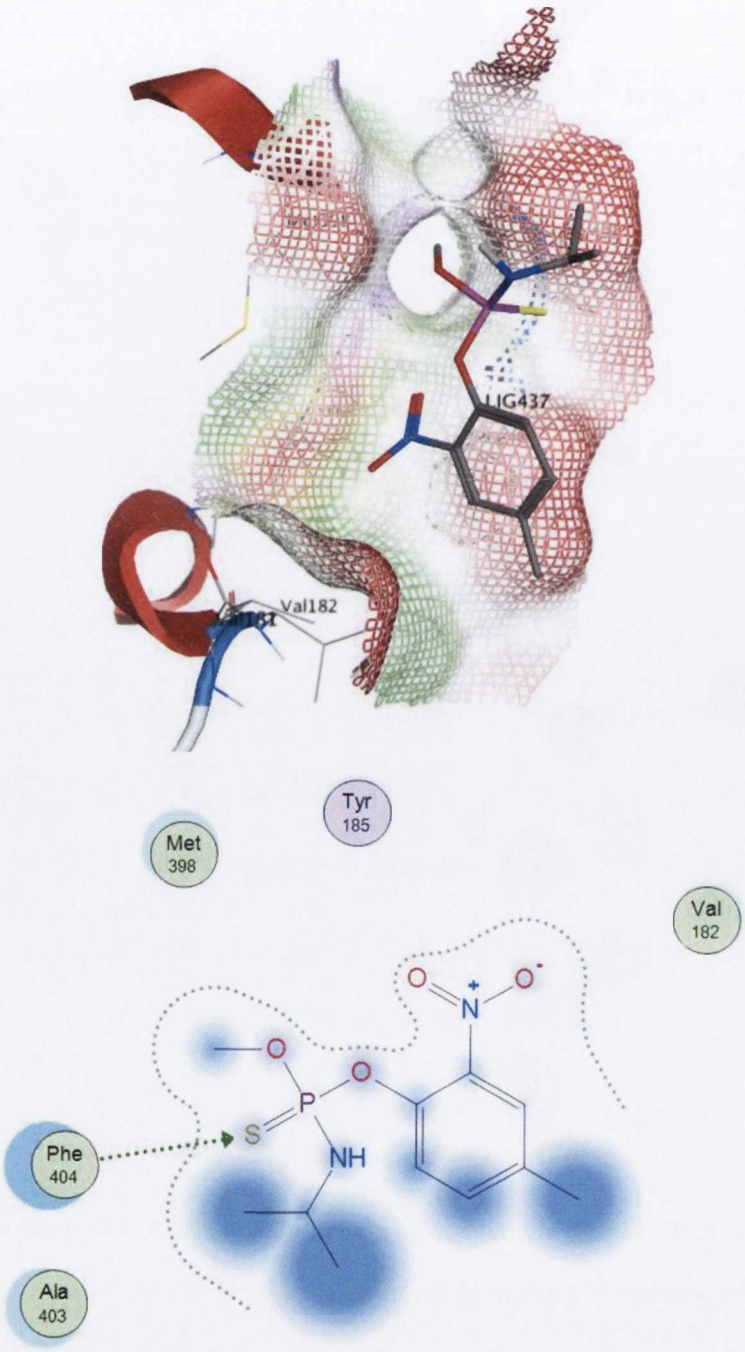


FIGURE 3.43: Ligand interaction diagram of APM in site 5 of trypanosomal tubulin.

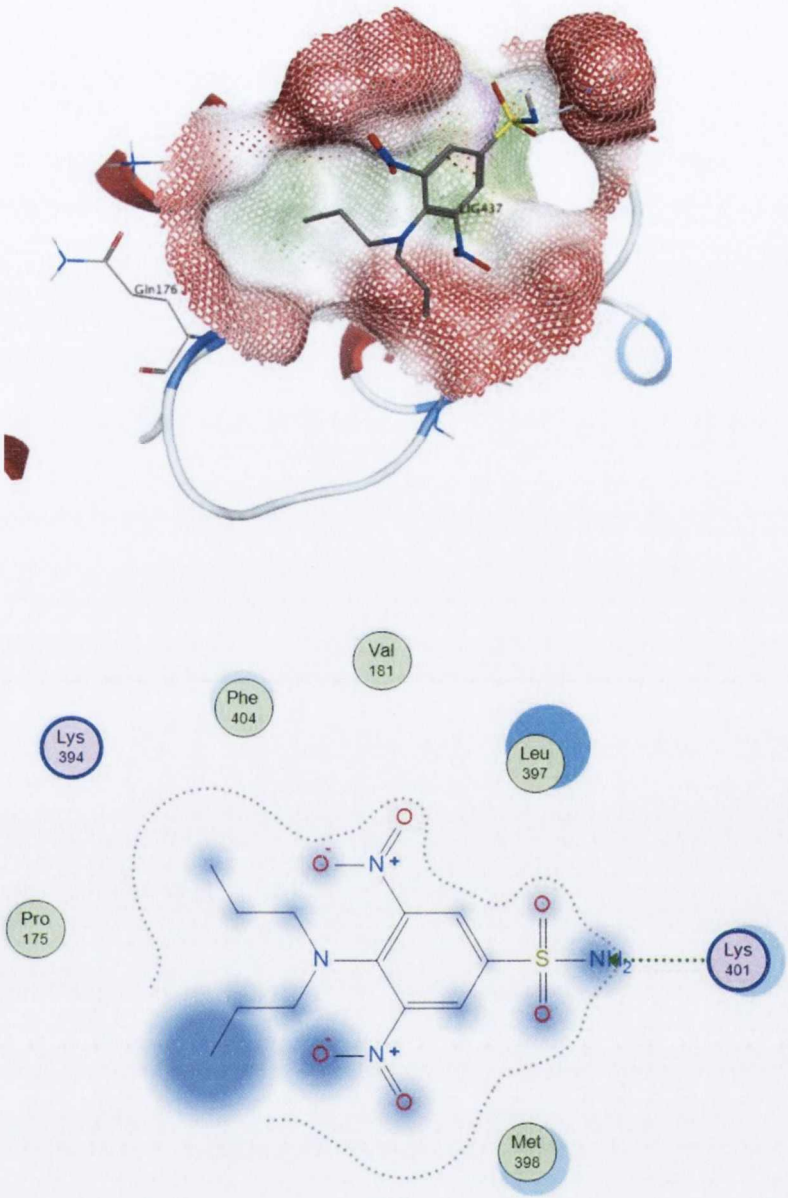


FIGURE 3.44: Ligand interaction diagram of OZN in site 5 of trypanosomal tubulin.

3.18.4 Analysis of binding site 6 in parasitic tubulin

Inspection of the APM-tubulin poses between the malarial and trypanosomal models revealed the ligands adopted a very similar pose in the binding site, the slight disparity in their relative positions being caused by a notable movement in the H11-12 loop (Figure 3.45 A). In the malarial tubulin-APM complex the nitrophenyl group of APM is buried in a small hydrophobic region under the H11-12 loop and hydrogen bonds are made between H192 and the sulphur of APM (Figure 3.47). In the trypanosomal tubulin-APM complex the ligand obtained approximately the same pose as in malarial tubulin; the nitrophenyl group again occupied a hydrophobic pocket under the H11-12 loop and the sulphur atom again formed a hydrogen bond, this time with F418, while a π(arene) hydrogen bond to the phenyl ring of APM formed with Y408 (Figure 3.49).

With respect to the OZN poses in the malarial and trypanosomal tubulin models, the malarial OZN pose migrated outside of the pocket, exposing the phenyl ring and sulfonamide groups to surrounding solvent (Figure 3.45 B). This unfavourable pose is indicated in table 3.3, where the malarial-OZN complex has a near zero ΔG total. In the trpypanosomal OZN-tubulin complex, one of the ring nitro groups and the dipropylamide group occupy a space under the hydrophobic H11-12 loop (Figure 3.50). Hydrogen bonds are formed between a sulfonamide oxygen atom and T109 and between the sulfonamide amine and A104. This favourable position of OZN is in direct opposition to the OZN pose in malarial tubulin.

In the malarial protein, the APM and OZN ligands adopted similar positions, albeit with the OZN ligand having migrated away from the more energetically favourable pose assumed by APM and becoming exposed to surrounding water. The only common interaction between the two ligands was a hydrogen bond to H192. The OZN and APM poses in the trypanosomal site were very similar, both adopting poses in the hydrophobic pocket of site 6 underneath the H11-12 loop. The shift in positions of OZN and APM in this site were caused by a small movement of this loop region.

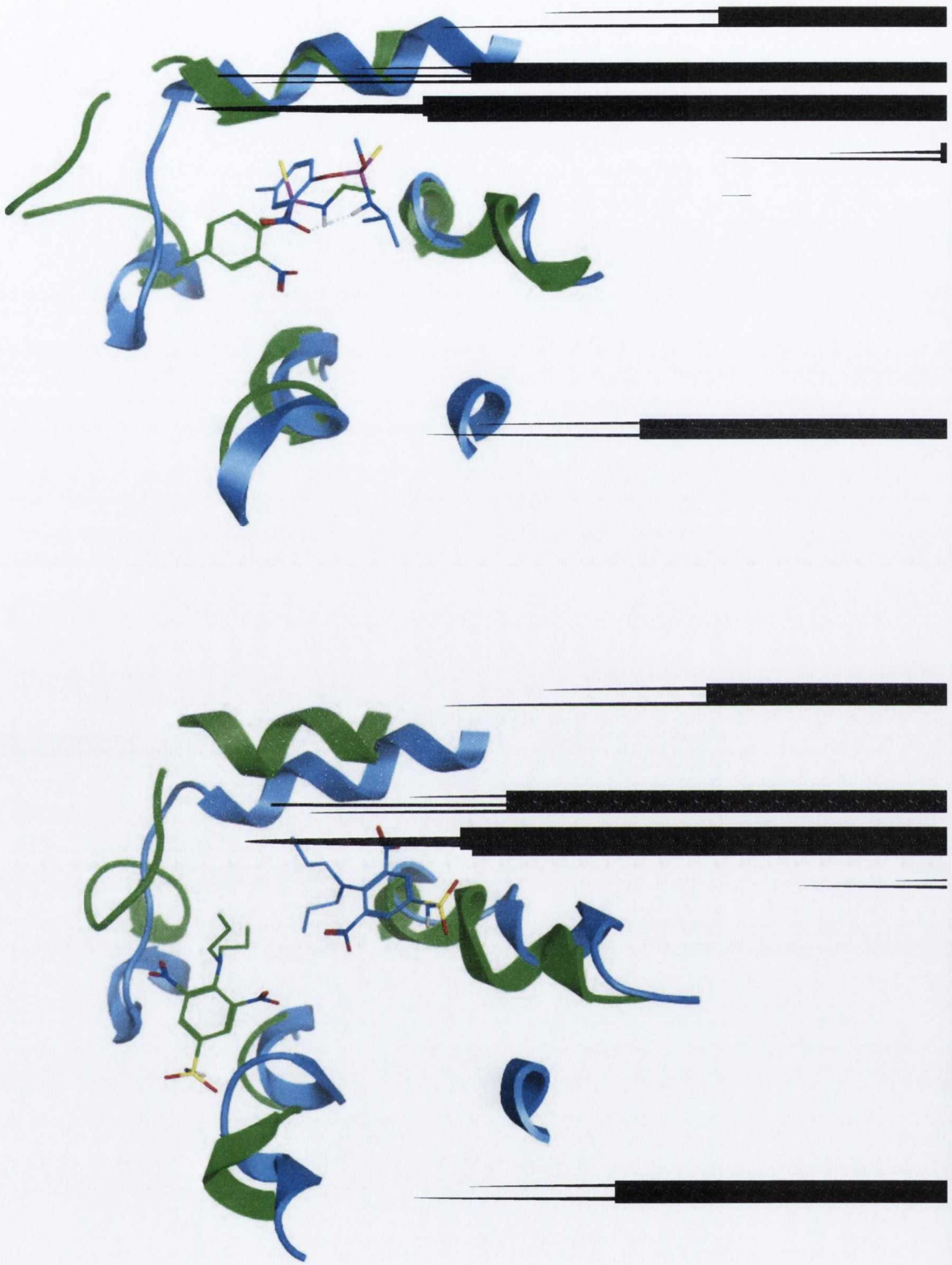


FIGURE 3.45: Post-molecular dynamics poses of APM and OZN in site 6 of Malarial and Trypanosomal Tubulin homology models.(A) The binding orientations of APM in malarial (blue) and trypanosomal (green) tubulin. (B) The binding orientations of OZN in malarial (blue) and trypanosomal (green) tubulin

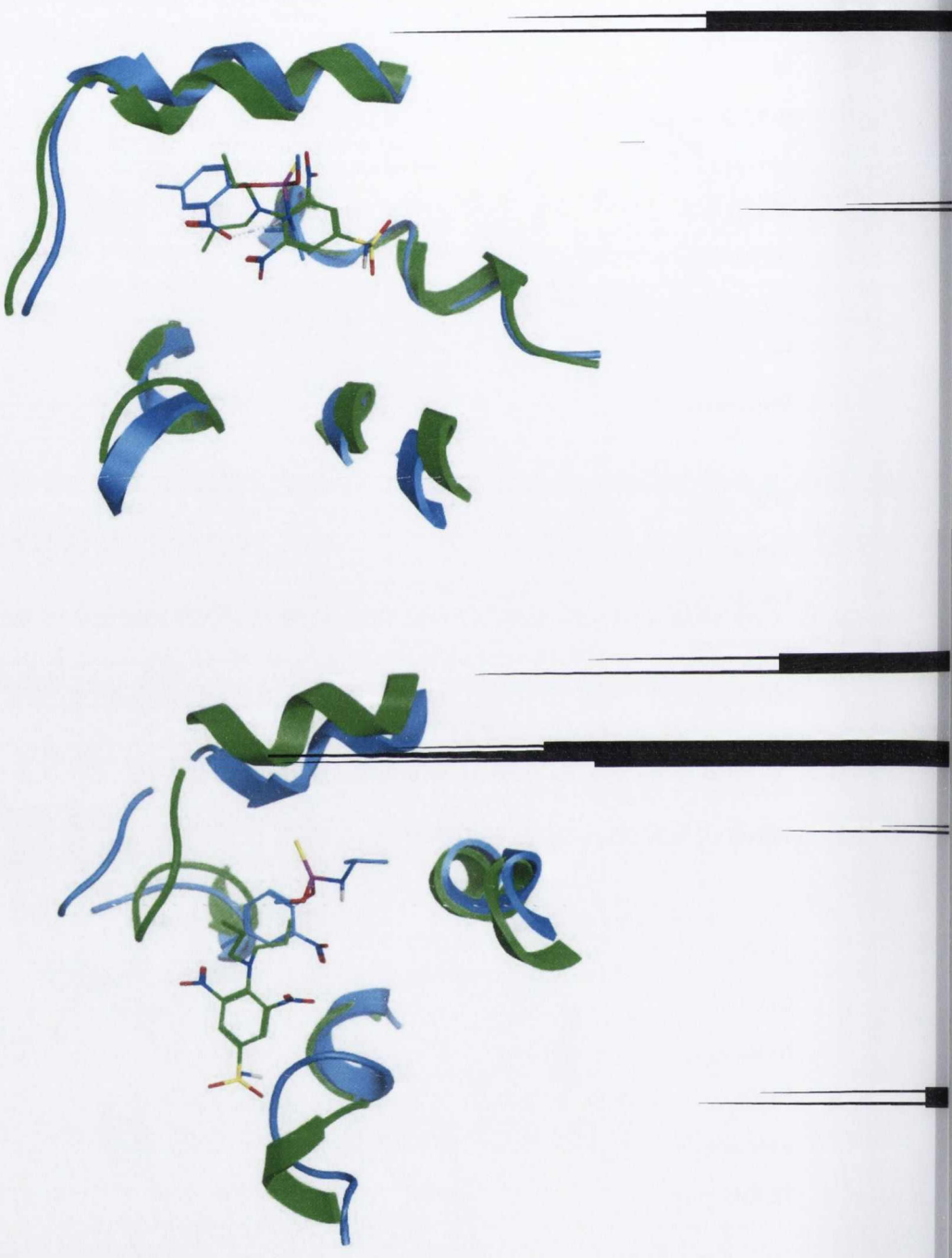


FIGURE 3.46: Post-molecular dynamics poses of APM and OZN in site 6 of Malarial and Trypanosomal Tubulin homology models. (A) The orientations of APM and OZN in site 6 of malarial tubulin. (B) The orientations of APM and OZN in site 1 of trypanosomal tubulin

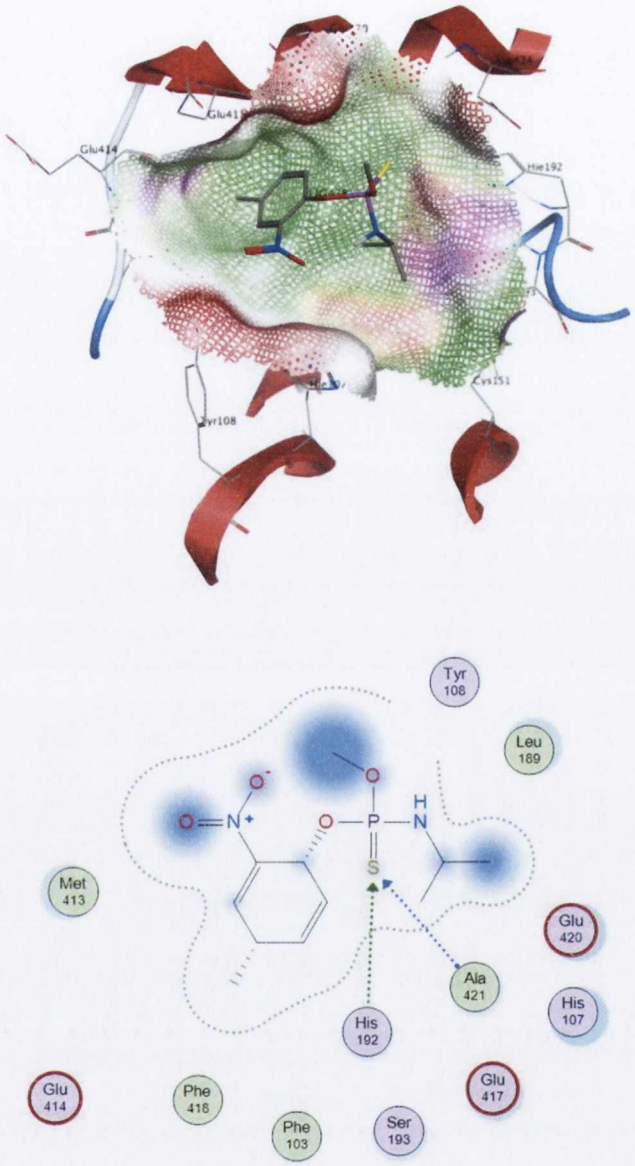


FIGURE 3.47: Ligand interaction diagram of APM in site 6 of malarial tubulin.

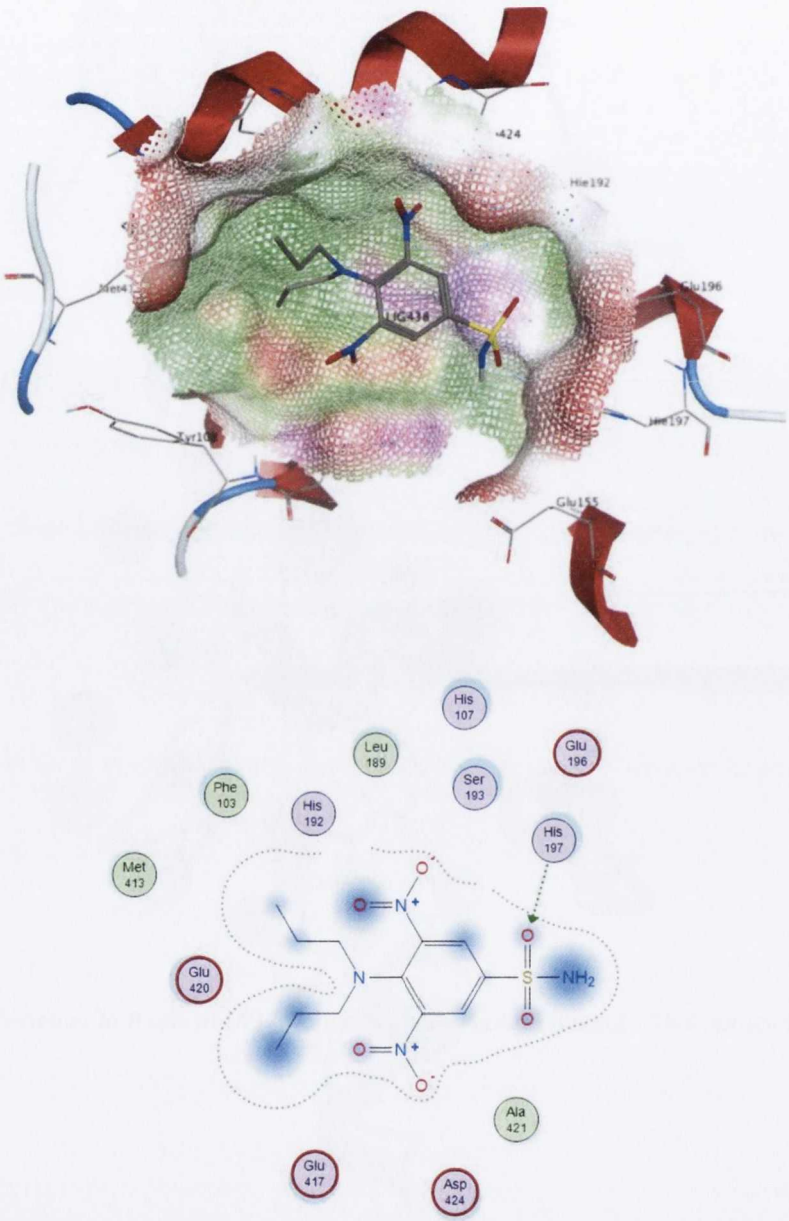


FIGURE 3.48: Ligand interaction diagram of OZN in site 6 of malarial tubulin.

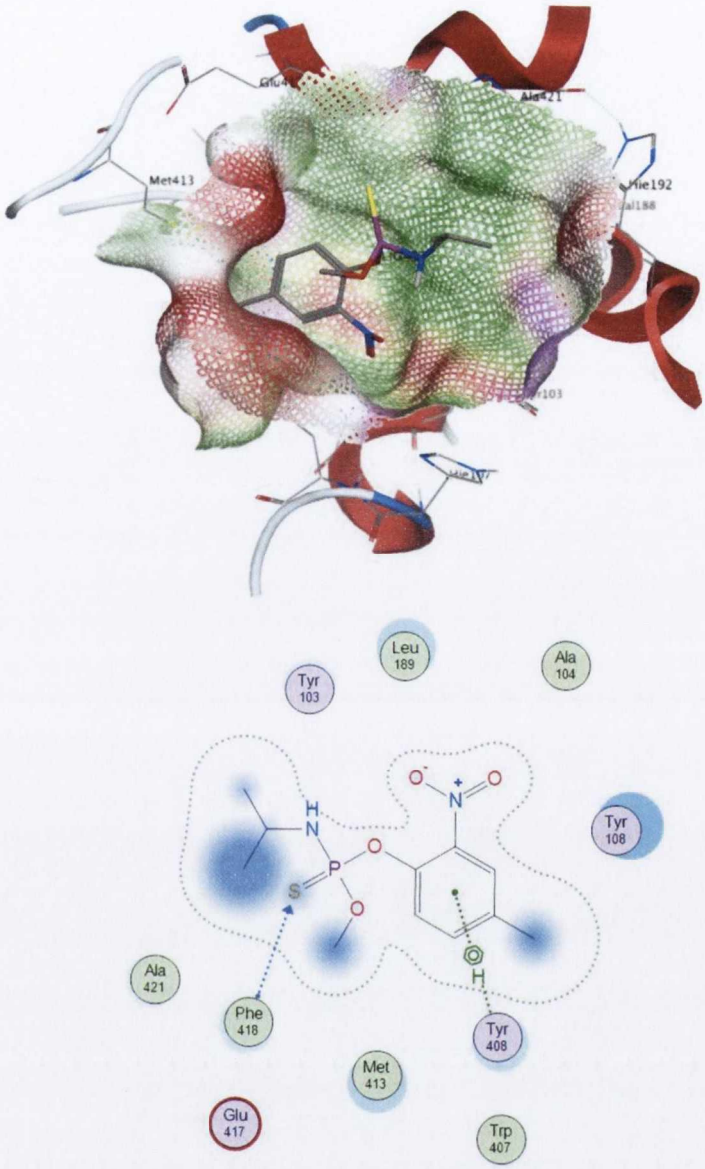


FIGURE 3.49: Ligand interaction diagram of APM in site 6 of trypanosomal tubulin.

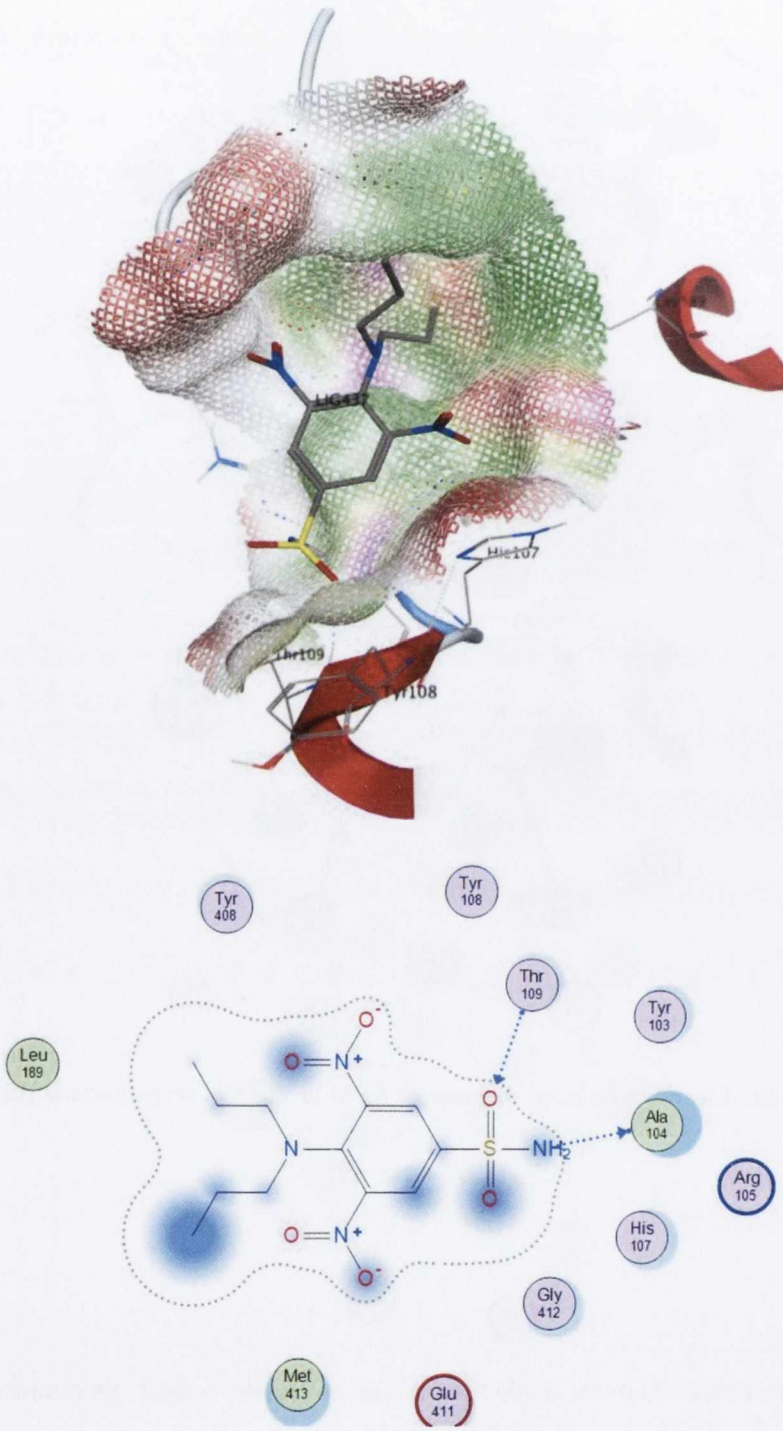


FIGURE 3.50: Ligand interaction diagram of OZN in site 6 of trypanosomal tubulin.

Chapter 4
Molecular Equivalents to Established Herbicides

4.1 Introduction

In Chapter 1 we introduced the dinitroaniline and phosphorothioamidate herbicides, two mechanistically related classes of compounds that have been shown to possess antiparsitic activity against *P. falciparum*, *Leishmania*, and Trypanosoma parasites. This inhibitory activity is seemingly achieved through the impact of the compounds on parasitic tubulin, an effect not displayed against mammalian tubulin. Despite the efforts at identifying their putative binding site in Chapters 2 and 3 the exact binding site remains undetermined. The apparent parasite-selectivity makes these compounds ideal candidates for rational design of second generation antiparasitic drugs. However, as stated in section 1.5.2 in Chapter 1 these compounds are limited as potential drugs due to their low solubility and potential anti-cholinesterase effects and traditional synthetic chemistry approaches to generate new phosphoro(thio)amidate compounds have thus far failed to address these concerns or drastically improve compound potency. In situations such as this, computational drug design techniques represent an ideal alternative approach to novel compound discovery.

This chapter details a computational drug design approach to identify new antiparasitic compounds based on the biological information available for the known herbicide compounds. Due to the lack of structural information regarding the binding site of these

compounds a ligand-based methodology was pursued. The aim of this study was to move away from the herbicide chemical class while retaining the selective tubulin-targeting nature of the compounds, with the focus ligand-based virtual screening methods for "scaffold hopping"[191]. The scaffold hopping approach attempts to retrieve molecules that have a different topology compared to known active molecules, in other words we were interested in methods that were able to retrieve the non-obvious hits from the vast chemical space.

4.2 Virtual Screening and The Drug Discovery Process

The development of a drug is an expensive and time-consuming process which can take on average up to 14 years and cost 800 million USD to bring a compound from hit identification through to approved drug. Within the industry at the moment there is increasing attempts at reducing this protracted hit-to-lead timeline, decreasing the number of drugs that fail to progress through clinical trials (the so called attrition rate which stands at roughly 90%) and improving the quality of candidate drugs making the transition from discovery to clinical development. This has resulted in the Pharmaceutical industry relying heavily on *in vitro* high-throughput screening (HTS) for the identification of novel hit compounds. HTS involves testing disease states with every available compound, often hundreds of thousands at a time. However, rather than decreasing the attrition rate, it appears HTS might have aggravated the situation with hit rates staying low as HTS emerged as a highly inefficient procedure[192]. A potential reason for failings of HTS might be based on the emphasis on large numbers of tested molecules instead of high quality experiments, i.e. testing the right molecules. Computer based methods might provide a means to rationalize these experiments incorporating the challenges provided by the high-throughput experiments.

Virtual Screening is an *in silico* approach to drug design that incorporates a variety of molecular modeling techniques that allow for the reduction of large virtual library databases through early assessment of a predicted activity of lead candidates, as well as removing potentially toxic or unwanted compounds from the datasets (Figure 4.1). The number of chemically feasible molecules which could be in principle used as drug candidates has been estimated to be 10^{100}[193] which is larger than the number of atoms in the universe. This number has two main consequences: first, it should be possible

to find a ligand with appropriate characteristics for each biological macromolecule and second, it is a practical impossibility to experimentally characterise all these ligands. Thus, this approach helps to converge on possible active molecules from large molecular libraries and shift the focus on eventual physical assaying of smaller subset of promising compounds. VS isn't limited to simply reducing screening database sizes for HTS, it can also be employed to search for new compounds that are more 'drug-like' or 'lead-like'.

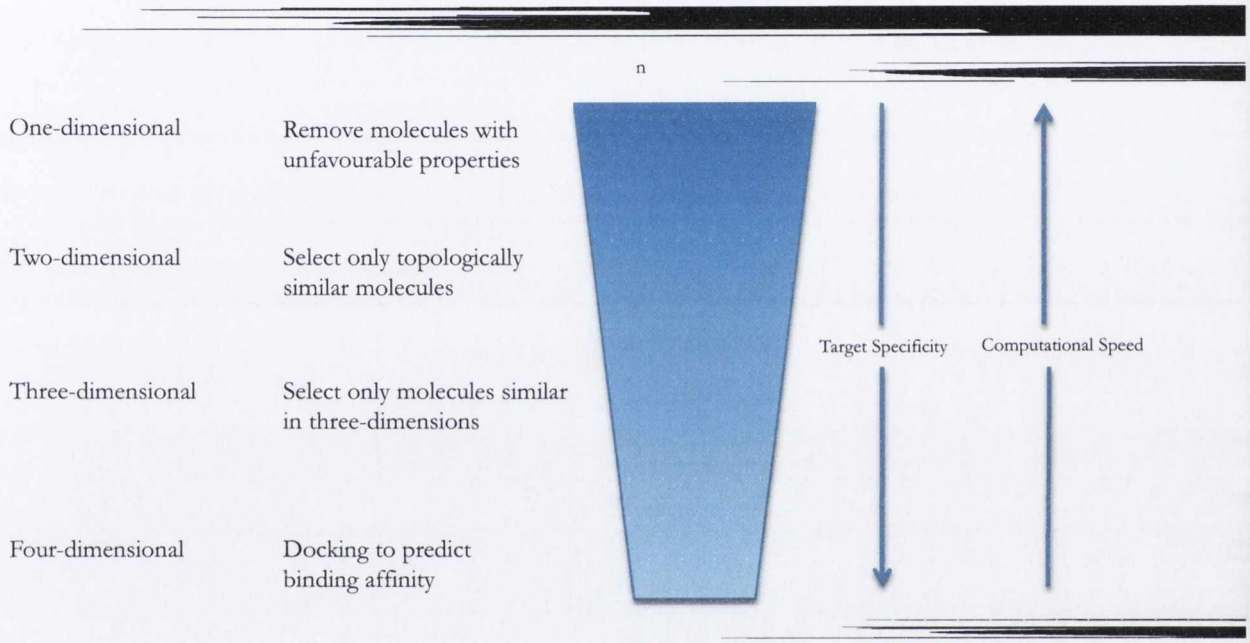


FIGURE 4.1: Virtual screening campaigns are often organised into hierarchical protocols involving simple filters such as one-dimensional screens based on filtering out potentially toxic groups to more complex and accurate methods such as binding affinity prediction.

Broadly speaking, in the process of VS, two main techniques can be applied, namely ligand-based or Structure-Based (receptor-based) virtual screening (Figure 4.2). For ligand-based virtual screening (LBVS) approaches, the typical strategy uses information provided by a compound or set of compounds with known activity against the biological target with a view to discovering other compounds in external databases with similar activities. This approach is applied in the absence of structural information on the target, as well as being implemented as a tool for lead compound modification and optimization. The main idea behind LBVS is that structurally similar compounds are

likely to have similar biological properties and activities. LBVS methods that are routinely used are substructure searching, pharmacophore searching, similarity searching and Quantitative Structure-Activity Relationships (QSAR) methods, all of which use structural information from an active compound. LBVS techniques are often used to scaffold hop, the aim of which is to use known active compounds to identify structurally novel compounds (that retain the desired biological activity of the starting compound) through modification of the central core of the compound. Many successful examples of scaffold hopping exist in the literature. Once a lead compound has been identified, the process of Lead Optimisation (LO) starts. LO is a term used to describe the direction of research after a lead compound has been identified through a VS campaign and biologically validated in the lab. The new direction stems from the desire to find compounds that share the lead compound's chemical class but have more desirable properties such as lower IC_{50}, better absorption, free of intellectual property (IP) claims, no observed cytotoxicity, and high selectivity to name a few. This may be achieved through methods similar to the ones used to identify the lead compounds, including pharmacophore, similarity and substructure searching.

Alternatively, when the structure of the biological target is known, structure-based virtual screening (SBVS) methods may be applied. These methods involve the explicit molecular docking of compound libraries into the binding site of the target, producing multiple binding poses for each compound, followed by a measurement of the quality of fit of the compound in the binding site. The structure used for docking may be one resolved by NMR, X-ray diffraction studies or a homology model, but typically a crystallographically-derived structure is best due to the high resolution that can be achieved.

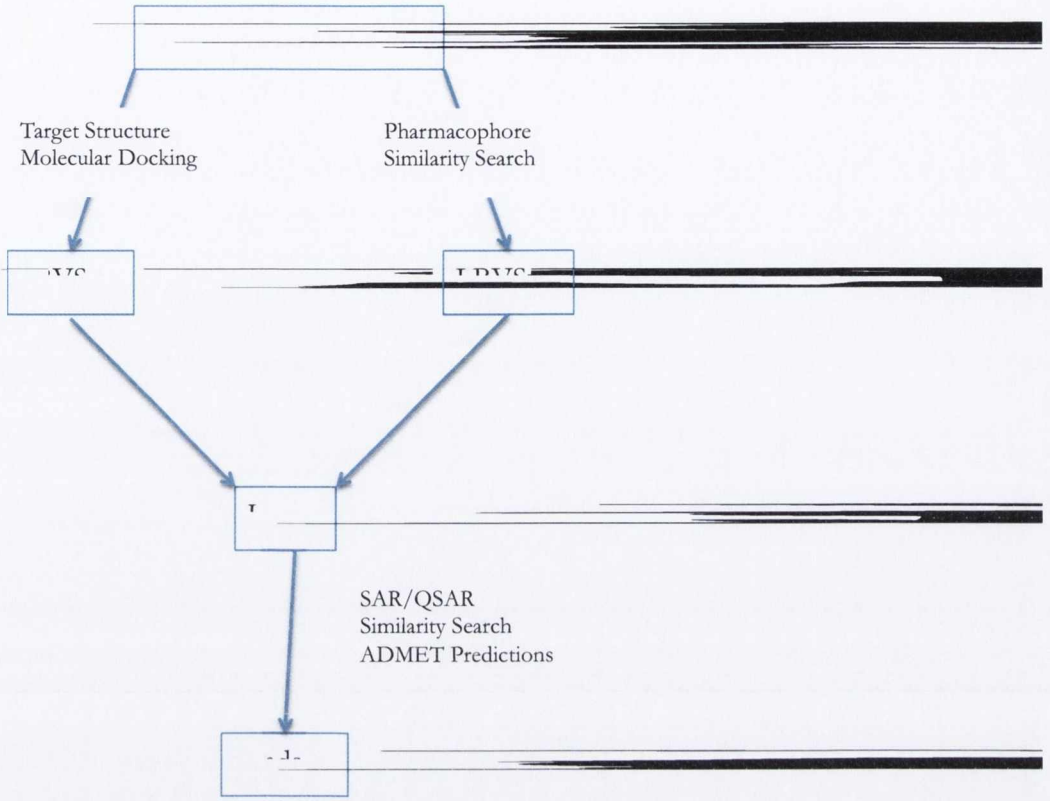


FIGURE 4.2: Virtual Screening overview. VS can be broadly divided into structure-based and ligand-based virtual screening. Both techniques have common output - the obtainment of 'hit' compounds.

4.3 Medicinal Chemistry Space

Chemistry space, much like the cosmic space referred to by Douglas Adams in *The Hitchhiker's Guide To The Galaxy*, is 'mind-bogglingly' big. This chemical space is so vast that it is advantageous to partition it into smaller, more easy to use sub-sections, such as 'drug-like' space or a specific drug target chemical space. Attempting to navigate chemistry space, where the number of unique chemical entities is thought to a number of 10^{100} *in potentia*, may be impossible in that respect, so to make this enormous space more manageable it is possible to describe it in terms of a coordinate system with multiple axes representing a number of physico-chemical properties. In this work, multi-dimensional chemistry space was represented by 14 physico-chemical descriptors and transformed into 3D space represented by 3 principal components using Principal Component Analysis (PCA), where each principal component is a combination of the 14 vectors.

4.3.1 Description of Principal Component Analysis

It is necessary to create a graphical distribution of these datasets in 3D space so as to present the relationship, if any, of anti-parasitic medicinal chemistry space in the context of wider chemical space. PCA is a fairly simple method of transforming an n-descriptor space into a more tractable 3D space.

Given m molecules, each of which described by an n-vector of real numbers xi=(x_{i1}, ... , x_{in}), consisting of the descriptors for molecule i. Each molecule has an associated importance weight, wi, a non-negative real number. These weights can be thought of as the relative probability that the associated molecule will be enqountered and are usually all equal to 1; however, in some applications unequal weights are used.

Let W denote the sum of all weights. Dimension reduction through principal component analysis can be interpreted in the following manner. Let X denote a random n-vector and let Z denote a random p-vector with mean 0 and covariance matrix equal to the identity matrix.

If we assume that X=RZ+x_0 for some n by p linear transform R and some n-vector x_0 then PCA is the estimation of the Z vectors from a sample of X vectors. IF we integrate both sides of the supposed affine transform we obtain:

$$E(X) = \int xPr(X = x) = \int (Rz + x_o$$

$$) \Pr(Z = z) = R0 + x_0 = x_0 \quad (4.1)$$

which shows that x_0 is the mean of the distribution of the X vectors. Turning to the covariance of the X vectors, we observe that:

$$Cov(X) = \int (x - x_0$$

$)(x\text{-}x_0)^T \Pr(X = x) = \int Rzz^T R^T \Pr(Z = z) = RR^T$ (4.2)

These equations suggest the following method for estimating the Z vectors. We use the sampled data to approximate both E(X) and Cov(X):

$$E(X) \sim \bar{x} = \frac{1}{W} \sum_{i=1}^{m} w_i$$

x_i, $\mathrm{Cov}(X) \sim S = \frac{1}{W} \sum_{i=1}^{m} w_i x_i x_i^T - \bar{x}\bar{x}^T$ (4.3)

The sample covariance matrix S is now symmetric semi-definite; hence, all of its eigenvalues are real and non-negative. We can therefore diagonalise S so that:

S =QT D D Q

where

- Q is orthogonal
- D is diagonal-sorted in descending order from top left to bottom right.

Let p be the number of non-zero diagonal values in D (the square roots of the eigenvalues of S). We can estimate R as the first columns of QTD and say that the X vectors have p principal components. In practice, we restrict the selection of p further by limiting the condition of the estimated RTR matrix; that is, we choose p so that the largest eigenvalue divided by the smallest eigenvalue is less than some specified threshold. The foregoing

considerations are integrated into the PCA computational method by an algorithmic procedure.

Algorithm Description

- Calculate the sample average vector x0 and covariance matrix S
- Diagonalise S
- Remove all components associated with zero eigenvalues
- Remove all components that exceed the matrix condition
- Remove all components above the minimum variance threshold
- Remove all components past the component limit
- From the PCA transform Q form the p remaining components so that Z=Q(x - x0) has identity covariance and zero mean
- For each i write the p-vector zi = $Q(x_i - x_0)$ to the output

4.4 The Concept of Molecular Similarity

Ligand-based virtual screening is based on the common assumption that two compounds that display similar chemical properties are likely to exhibit similar biological effects than compounds that have dissimilar structures[194]. Since it is necessary to have only one or more active compounds for these models they offer a valid alternative when no 3D target protein structure is available.

There are many different methods of calculating molecular similarity, roughly divided into 2D and 3D methods. 2D methods incorporate substructure searching, fingerprint searching and similarity metrics. Substructure searching is very much what it says on the tin; searching for the presence of a desired substructure in a database of molecules. The use of molecular fingerprints to compare two molecules is a more common method. Path fingerprints can be generated by the exhaustive systematic analysis of all the bonds connecting the atoms of a molecule. These methods then encode the molecular structure as a series of bits (binary digits) that illustrate the presence or absence of particular fragments in the molecule. These

bitstrings can then be used to compare two molecules to determine the similarity between them. The most routinely used metric to calculate similarity is the Tanimoto coefficient. This is described as the ratio of bits present in both molecules divided by the number of distinct bits in each molecule. Other metrics include the Cosine and Hamming methods.

4.5 Scaffold Hopping

While the similarity concept discussed previously has the advantages of being quick and simple, these have the distinct disadvantage in that the hits they return tend to have the same structural and chemical features as the query compounds. While this is an adequate method if one wishes to find analogues of a given compound, it is not sufficient when the aim is to identify compounds that retain the biological activity of the query compound but possess a different chemical scaffold. This is a challenging goal in VS and the ideal method would not only return a large number of compounds from a given screening approach, but also a highly diverse set of active compounds. Having diversity amongst hit structures is important in terms of giving synthetic chemists a choice with respect to chemical accessibility and prospects for lead optimization as well as facilitating the creation of IP.
Two popular methods for scaffold hopping are pharmacophore searching and Rapid Overlay of Chemical Structures (ROCS).

4.6 Introduction to Pharmacophore Design

The pharmacophore concept was first introduced by Paul Ehrlich in 1909 as 'a collection of essential features (phoros) in a molecular framework which is responsible for a drug's biological activity (pharmacon).' This definition has since evolved and is now considered as the 'ensemble of steric and electronic features that is necessary to enable the optimal supramolecular interactions with a specific biological target structure and to trigger (or block) its biological response'. The pharmacophores are composed of pharmacophore features that have a characteristic 3D geometry and may be considered the highest common denominator of a group of molecules exhibiting a similar pharmacological profile and which are recognised by the same

site of the target protein. Pharmacophore features, or potential pharmacophore points (PPP), are generally categorised by potential ligand-receptor interactions. The interactions can be clustered into three general groups: hydrophobic, polar positive and polar negative, which can be further broken down into hydrophobic-alkyl, aromatic, hydrogen-bond donor, cation, hydrogen-bond acceptor and anion. Pharmacophore queries can be derived based on the spatial arrangement of these features in a ligand.

Pharmacophores can be derived via analogue-based or receptor-based methods. Analogue-based methods use a ligand or ligands that have a desired biological activity against a target of unknown structure or where the ligand-binding site is unknown. Receptor-based pharmacophore models take advantage of the resolved protein-ligand complex to design pharmacophore hypotheses. It is important to note that description of ligand-receptor interactions by pharmacophores is a perhaps an oversimplification as they do not consider entropy or solvatation effects, while some groups like fluorine, which can interact like hydrophobic groups and as hydrogen-bond acceptors are hard to model correctly by pharmacophore types.

4.6.1 Generating a Pharmacophore Hypothesis

The steps in generating a pharmacophore hypotheses (Figures 4.3 and 4.4) generally include: (1) data selection, (2) alignment of active compounds, (3) feature extraction, and (4) pharmacophore validation.

Data selection refers to the creation of training (to build the pharmacophore hypothesis) and test (to test the pharmacophore hypothesis) sets and is probably the most important aspect of pharmacophore design. As with most computational approaches, the quality of the output data reflects the quality of the input data; if you put rubbish in you get rubbish out. The chief considerations to be made when selecting input data are usually (but not limited to) the types of ligands available,

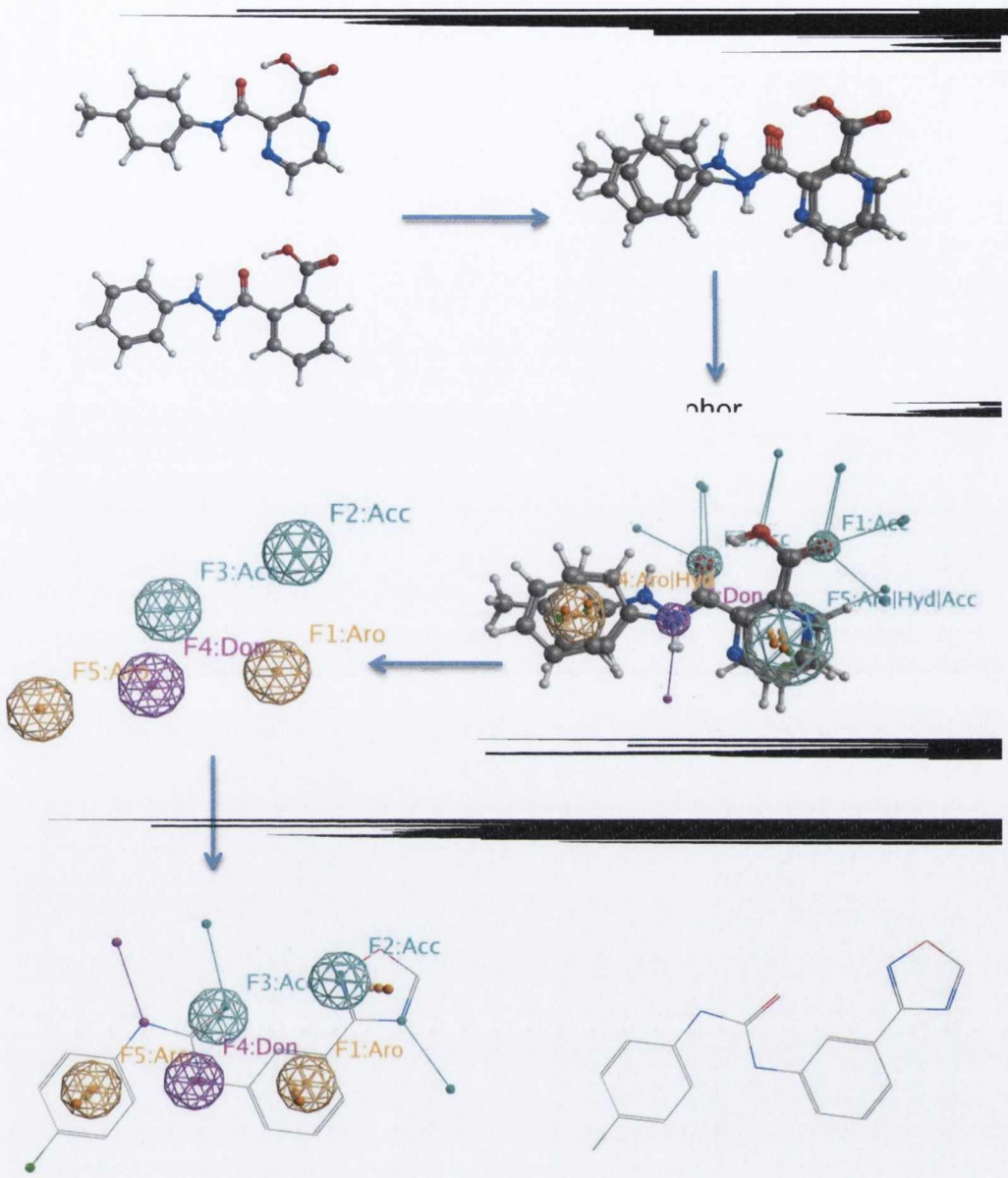


FIGURE 4.3: Schematic of the Pharmacophore Design Process.

the size of the dataset, and the chemical diversity of the compounds. In an ideal scenario, a training set would contain over 15 compounds with structural diversity, a known bioactive conformation, and possessing a range of biological affinity. However, as often happens, a bioactive conformation isn't always known and a pharmacophore may have to be generated using only one or two known active compounds. In this case it is necessary to use the lowest energy conformations as a starting point. Following data selection the training set must be aligned so that

common pharmacophore features can be identified and extracted. Following feature extraction, it is necessary to test the predictive ability of your pharmacophore hypotheses.

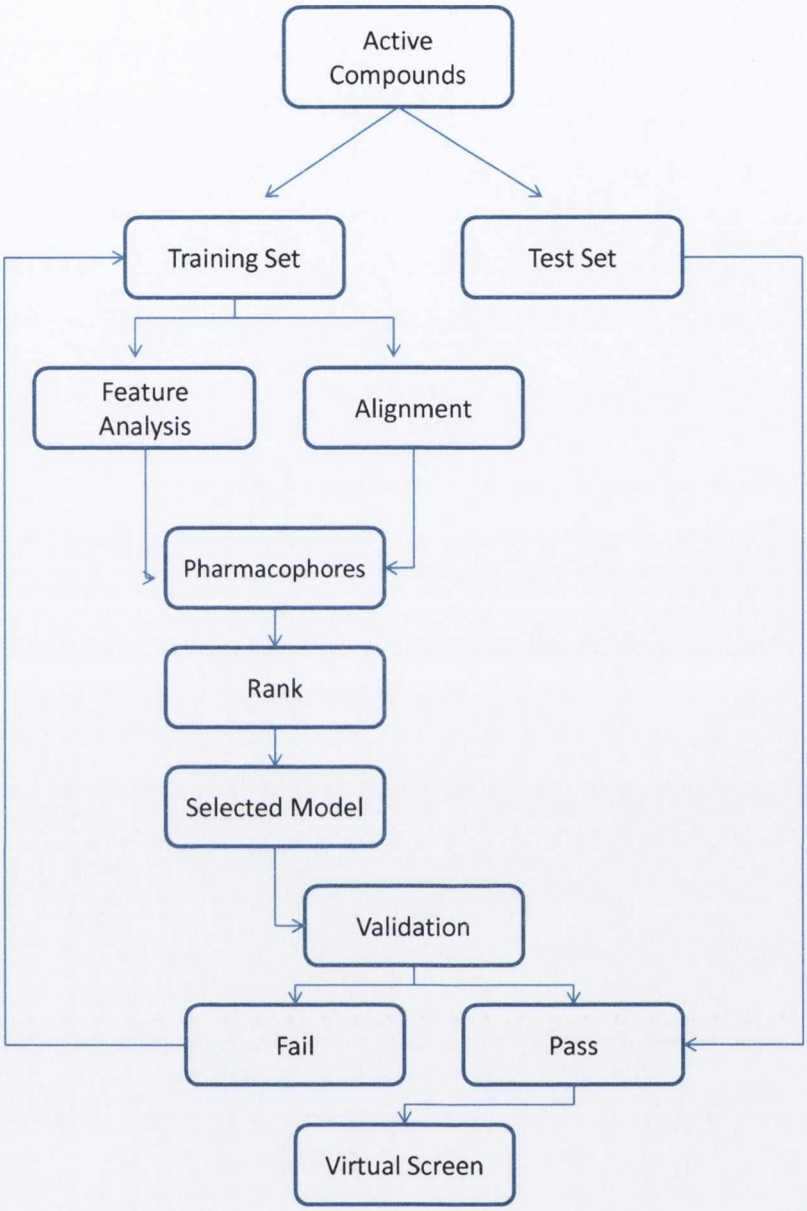


FIGURE 4.4: General workflow for pharmacophore development. The first step is collecting the active dataset and generating the conformers, following this it is necessary to define the training and test sets. The next step involves flexible alignment of the training set compounds and extraction of common pharmacophore features. The resulting pharmacophore hypotheses must be ranked to assess the model accuracy and finally, when the best model has been selected, a validation step using the test set or decoy database must be performed before executing a virtual screen of vendor libraries.

Validating a Pharmacophore Hypothesis

Independent of the computational method chosen, whether it is ligand- or structure-based virtual screening, it is essential to test and validate the model developed. Measurement of the model quality can be assessed with several methods discussed in the literature[195,196]. All evaluation methods aim to describe the aptitude of the model with respect to retrieving active compounds from a database containing both active and decoy molecules, known informally as a "haystack" database. The accuracy of the model tested is defined as the percentage of molecules that are correctly classified.

Proper selection of decoy molecules is an important aspect of validation dataset design. The dataset should contain not only a structurally diverse set of active compounds but also a set of decoy molecules that are similar to the actives, thus insuring that any bias towards selecting actives over decoys is removed.
Figure 4.5 below illustrates the basis of the evaluation metrics used in VS protocols.

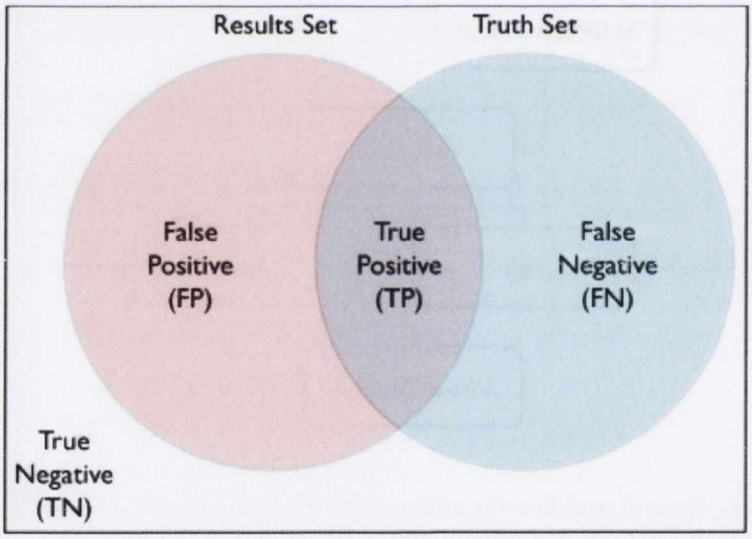


FIGURE 4.5: The basis of the metrics used in virtual screening protocols.

- True Positives (TP) = molecules correctly predicted to be active
- False Positives (FP) = molecules incorrectly predicted to be active
- True Negatives (TN) = molecules correctly predicted to be inactive
- False Negatives (FN) = molecules incorrectly predicted to be inactive

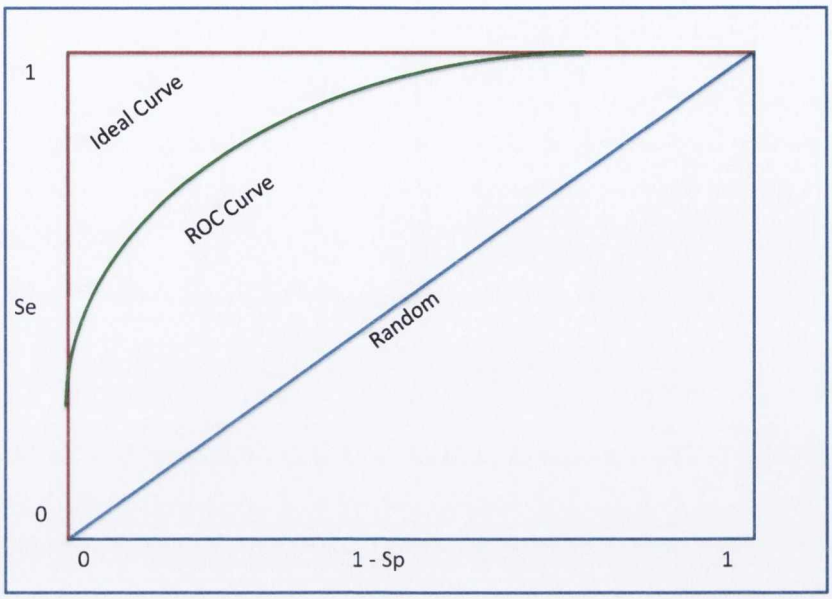


FIGURE 4.6: Illustration of theoretical ROC curves.

The Receiver Operating Characteristic (ROC) curve is used to assess the ability of a model to distinguish actives from inactives or decoy molecules within a virtual screening protocol (Figure 4.6). In a ROC curve the true positive rate is plotted as a function of the false positive rate. Each point on the ROC curve represents a sensitivity/specificity pair, with sensitivity defined as the number of TP divided by the sum of the TP and FN, and specificity defined as the number of TP divided by the sum of the TN and FP. An ideal test with perfect discrimination has a ROC plot that passes through the upper left corner of the graph (100% sensitivity, 100% specificity).

4.7 Pharmacophore Generation in MOE

The Pharmacophore Query Editor of MOE was used to generate the pharmacophore hypotheses with the aligned molecules. This tool allows the user to create a pharmacophore query composed of a set of constraints on the position and type of chemical properties (aromatic, hydrophobic, etc.) of non-hydrogen atoms of each molecule, referred to as ligand annotation points. These annotation points can be divided into three categories: 1. If an annotation is located directly on an atom, it is denoted an atom annotation. Such annotations include the H-Bond donor/acceptor and cation type features. 2. If an annotation is located at the geometric centre in a molecule (eg. a ring) it is called a centroid annotation. 3. Projected annotations are used to describe the location of possible hydrogen bond partners, metal ligation partners, or potential R-group atom locations.

There exist several pharmacophore annotation schemes in MOE which define how each ligand is to be annotated. Default PCH-type scheme was employed in pharmacophore generation. Existing structure-activity data for these compounds was used where possible to guide feature selection. Selected annotation points exist as pharmacophore features represented as a point in 3D space with a radius-like tolerance on spatial proximity. The pharmacophore features fall into four main categories: H-bond donors/acceptors and metal ligators, ionic groups, aromatic and π-rings, and hydrophobic features. H-bond donors/acceptors and metal ligators contains the H-bond donor/acceptor annotations and metal ligator features, including the projected types as well. The ionic groups contain the features for anionic and cationic heavy atoms and also they are bioisosteres of -CO_2 and -CN. The aromatic and π-rings feature group contains the annotations for defining the centroids of aromatic and pseudo-aromatic rings, other π-system rings and for the centroid projection. The hydrophobe annotations contain the planar and non-planar descriptions of hydrophobic centroids. Pharmacophore features were manually selected from the molecule alignment and saved as .ph4 files. MOE Pharmacophore Query Editor allows the user to add constraints, partial matches and exclusion volumes to pharmacophore queries. Constraint features allow the user to ensure that a specific feature is present in screened molecules while partial matches allow a degree of violations within the selected features (ie there exists four features in this query but a passing molecule need only match 3).

4.8 Rapid Overlay of Chemical Structures

Rapid Overlay of Chemical Structures (ROCS) is a method of shape-based virtual screening, the objective of which is to find and score an optimal overlap of two molecules in 3D space. in ROCS, the continuous functions that are used to calculate volume overlap is constructed from atom-centred Gaussians[197]; the advantage of this method, as opposed to a hard-sphere approach to molecular volume, is that they are easily integrable and therefore can be calculated faster. Once ROCS has found the optimal overlap between two molecules, it uses a shape Tanimoto function to score the result with a value of 1.0 for two identical shapes and 0.0 if they are completely different. Although the Tanimoto score in ROCS is indifferent of atom type, it offers a ColorScore function which uses SMARTS to describe the 'colour' of atoms or groups of atoms in an approach very similar to pharmacophore annotation points. ComboScore in ROCS uses a combination of shape and colour for ranking molecules and gives a score between 0 and 2. Figure 4.7 shows a simplified illustration of how a ROCS search works.

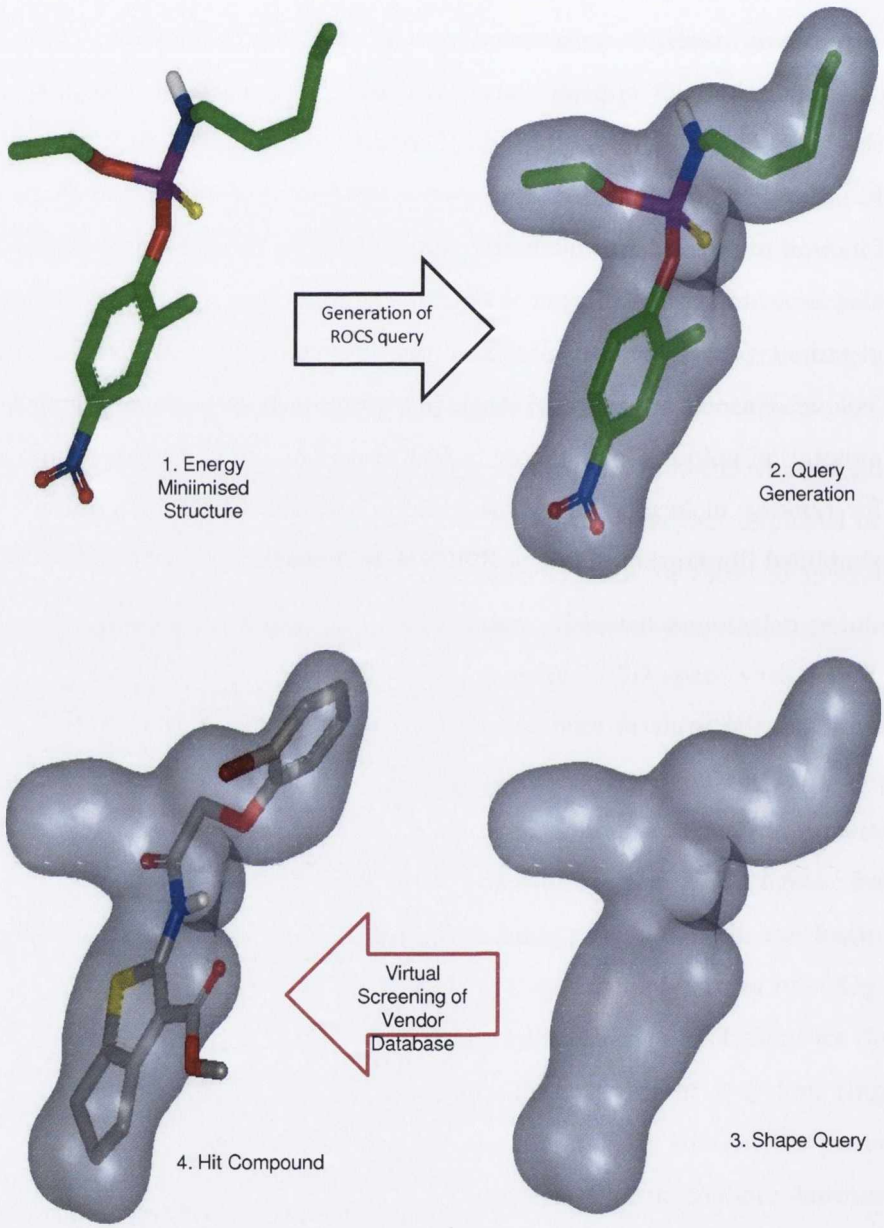


FIGURE 4.7: Illustration of the ROCS shape-matching search.

4.9 Successful Ligand-Based Virtual Screening Campaigns Against Parasitic Diseases

A search on PubMed reveals that while there exists many cases of VS being successfully utilised in the design of new anticancer and anti-HIV agents, there are far less examples to be seen in the application of ligand-based design methods against neglected diseases. Given the imbalance in the representation of drugs for NTDs in the pharmaceutical market, this is perhaps unsurprising. However, while there is an obvious imbalance in the application of computational tools for the design of new drugs for NTDs, there still exists some successful examples of their use.

A recent example combining both ligand- and structure-based virtual screening approached for the identification of novel anti-malarial compounds was carried out in 2011 by Rogrigues, T. *et al*[198]. This work was directed towards the discovery of new inhibitors of the malarial cytochrome bc_1 based on a pharmacophore modeling using a single active compound, GW844520 4(1H)-pyridone. In this paper, two pharmacophores, A and B, were generated based based on a possible bioactive pose of GW844520 obtained through docking at the cytochrome bc_1 Q_o site (quinol oxidation site). Following this, the ZINC database of commercially available compounds was screened with pharmacophore A and the MOE drug-like database was screened with the more restrictive pharmacophore B. Pharmacophre screening reduced a Lipinski-filtered ZINC database from 136,996 compounds down to 1,106 compounds while the MOE database of 600,000 compounds was reduced to 717 compounds after screening. The authors then employed a two-stage docking of the pharmacophore hits, the first stage involving a quick docking approach to remove poorly scoring compounds and the second stage involving a more exhaustive approach to docking the top-ranked 100 compounds. After a visual inspection of the top 200 docked compounds, 100 for each pharmacophore, 23 compounds were purchased and tested on a *P. falciparum* chloroquine-resistant cell line. Of the 23 compounds tested, 6 displayed EC_{50}s in the 2-30μM range.

4.10 Methods

4.10.1 Mapping Anti-parasitic Medicinal Chemistry Space

Compound Selection

For the PCA analysis, APM and 31 APM analogues were selected along with oryzalin and trifluralin to represent herbicide medicinal chemistry space. To represent the relationship of anti-trypanosomal medicinal chemistry space in the context of wider chemistry space, a compound search was performed in Integrity by Thomson Reuters to identify anti-trypanosomal compounds. An identical search was used to identify a set of anti-malarial compounds for PCA analysis. The final database contained 126 anti-trypanosomal compounds and 215 anti-malarial compounds that had either made it to preclinical phases, phases I-III, or had been launched. To establish drug-like medicinal chemistry space, a subset of the Maybridge vendor database was selected using Lipinski's 'Rule of 5' (RO5) (Table 4.1) method. Briefly, the RO5 is based on a study of the characteristics of orally bioavailable drugs in which Lipinski put forward an analysis of reasons why potential drugs may fail during the drug design process. Lipinski suggested that certain molecular properties are important for a drug's pharmacokinetics in the human body, and that breaking as few of these rules as possible should aid the drug design process. The Maybridge subset was generated using a Pipeline Pilot protocol that first filtered the Maybridge database using Lipinski's RO5 and then randomly selected 0.5% of the resulting drug-like compounds.

TABLE 4.1: Lipinski's 'Rule of 5' descriptors

Descriptor	Value
Molecular Weight	$\leq 500Da$
Octanol/water partition coefficent	$logP \leq 5$
Number of H-bond donors	≤ 5
Number of H-bond acceptors	≤ 10

Descriptor Calculation

Physico-chemical descriptors were calculated with MOE's Calculate Descriptors tool. 14 descriptors were chosen, encompassing 1D and 2D classes (Table 4.2). 1D descriptors are physical or chemical properties use only atom type information while 2D descriptors use atom information as well as connectivity or bond-type information. A total of 15 molecular descriptors were chosen for this study. The descriptors covered a range of drug-like properties relating to size, lipophilicity, polarizability, flexibility, rigidity, and hydrogen bond capacity. These descriptors included counts of oxygen, carbon, and nitrogen atoms, molecular weight, molar refractivity, h-bond donor and acceptor counts, topological polar surface area (TPSA), sLogP, and counts of rotatable and rigid bonds.

TABLE 4.2: Descriptors used for PCA analysis

Descriptor	Description
a_nC	Number of carbon atoms
a_nO	Number of oxygen atoms
a_nN	Number of nitrogen atoms
a_acc	Number of h-bond acceptors
a_don	Number of h-bond donors
rotb	Number of rotatable bonds
weinerPath	Weiner path number
MW	Molecular Weight
vdw_vol	Van der Waals volume
rings	Number of rings
opr_brigid	Oprea rigid bond count
balabanJ	Balaban averaged distance sum connectivity
TPSA	Topological polar surface area
mr	Molar refractivity

4.10.2 Pharmacophore Design

23 compounds in total were used in the present study including APM and 18 APM analogues synthesized by Dr Christine Mara of Royal College of Surgeons Ireland and 4 dinitroanilines taken from the literature (chloralin, oryzalin, trifluralin and gb-II-5[199]). These compounds were divided into a 6 compound training set for model construction (Figure 4.8) with the remaining 17 compounds used as a test set for model validation (Figure 4.9).

Pharmacophore generation was initiated by taking the training set molecules and flexibly aligning them in MOEv2011.10. 3D structures were generated with MOEÕs Builder tool and protonated as they were expected to be found in aqueous solution at physiological pH. Flexible alignment was carried out with an iteration limit of 200, an energy cutoff of 15 energy value plus the minimum generated value, failure limit of 20 configurations in a row, RMSD tolerance 0.5Å and maximum energy minimisation steps 500. Compounds 1 and 2 were aligned and the alignment with the lowest alignment score, S (a sum of the terms for average energy strain of the molecules and the similarity measurement of the configuration) was manually selected and had its potential fixed relative to each other for further alignments. Compound 3 was then superimposed onto this orientation and a low scoring overlay was saved and fixed for the subsequent alignment procedures until all six molecules were aligned. Model validation was carried out in two stages. The first stage involved assessing the ability of the pharmacophore to select all the molecules that constituted the test set. If a pharmacophore successfully passed all test set compounds it progressed to the second stage of validation involving screening a 'haystack' database whose construction is detailed in section 4.10.4 below.

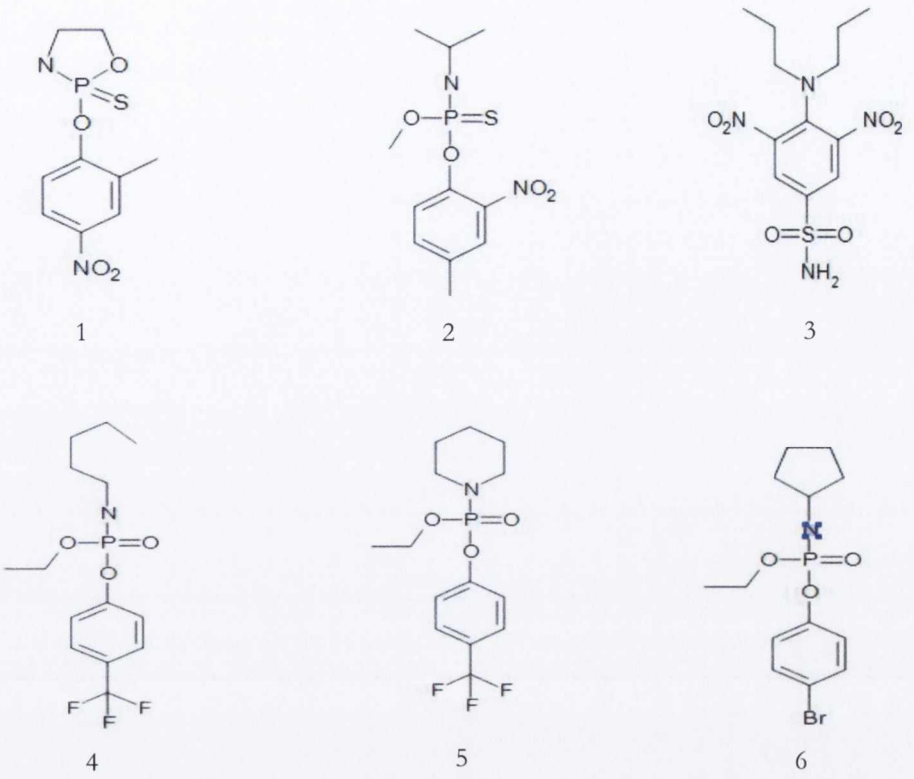


FIGURE 4.8: Compounds making up the pharmacophore training set. APM is compound 2 and OZN is compound 3.

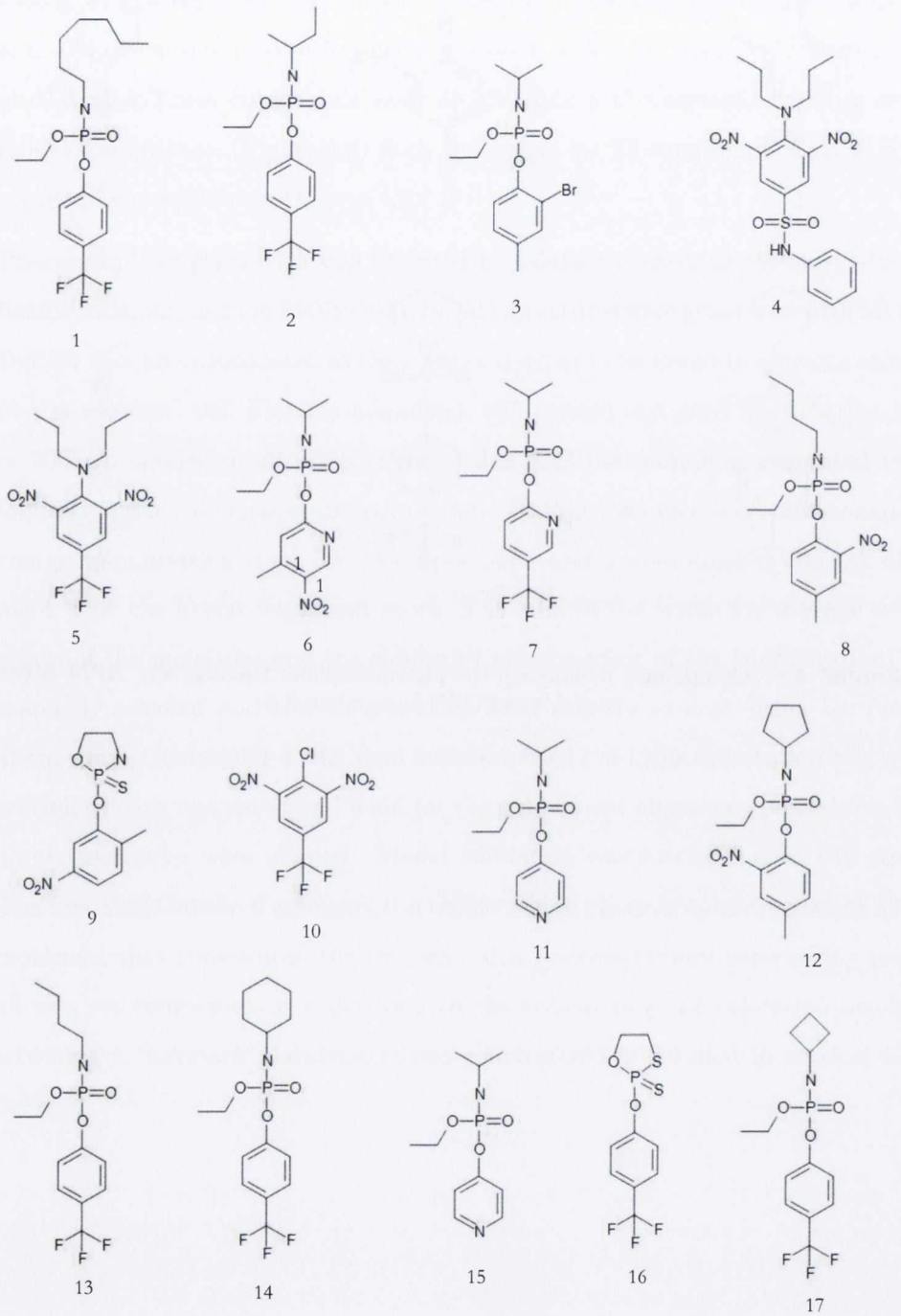


FIGURE 4.9: Compounds making up the pharmacophore test set.

4.10.3 ROCS Query Generation

The energy minimised structure of APM was use as the ROCS query. As only the shape of the molecules were of interest, ShapeTanimoto was used to rank hit molecules.

4.10.4 Preparation of Validation Database

The lack of a set of known inactive molecules to employ as decoys in the validation database necessitated the use of a set of unknown decoys. This validation set was prepared so as to possess similar physicochemical descriptors as the known active compounds.

Lipinski descriptors were calculated for APM analogues in MOE. Pipeline Pilot was employed to select 15% of the Maybridge vendor database as a starting point for decoy database preparation. Using Pipeline Pilot again, the minimum and maximum Lipinski values of the APM analogues were used to filter the 15% Maybridge vendor database to obtain a chemically similar subset (Figure 4.10). Compounds within this subset that were above a Tanimoto similarity of 0.65 to APM were removed from the set. Conformers were generated using OMEGA with a maximum conformer generation value of 90 with the rest of the values at default settings. The final validation database contained 25 active compounds and 780 decoy compounds.

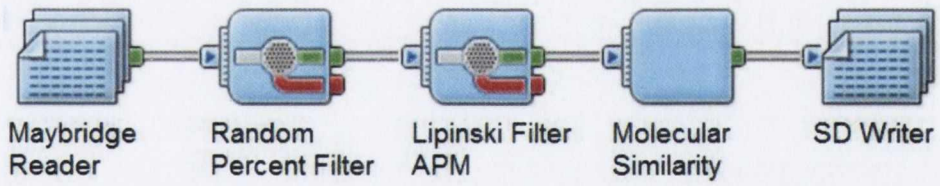


FIGURE 4.10: Pipeline Pilot Protocol for Validation Database Preparation The protocol reads in the Maybridge chemical database and then applies a random 15% filter. A second filter was then added using the minimum and maximum Lipinski values of the APM analogues and used to filter the 15% Maybridge vendor database to obtain a chemically similar subset. Compounds within this subset that were above a Tanimoto similarity of 0.65 to APM were removed from the set .

4.10.5 Validation of Ligand-Based Approaches

Computational approaches were validated using a ROC curve protocol built in Pipeline Pilot. Briefly, this protocol reads in a database of unique, ranked molecules

from a virtual screen, plots a ROC curve based on this data, and then calculates the area under the curve.

4.11 Results and Discussion

4.12 Navigating Parasitic Chemotherapy Space

Figure 4.11 displays the relative positions of anti-trypanosomal and anti-malarial medicinal chemistry space. Both sets occupy very close areas of chemistry space with a tight core and a number of entities lying outside this zone. The compounds lying in the extremities of the graph have high molecular weights and atom counts in comparison to the rest of the dataset.

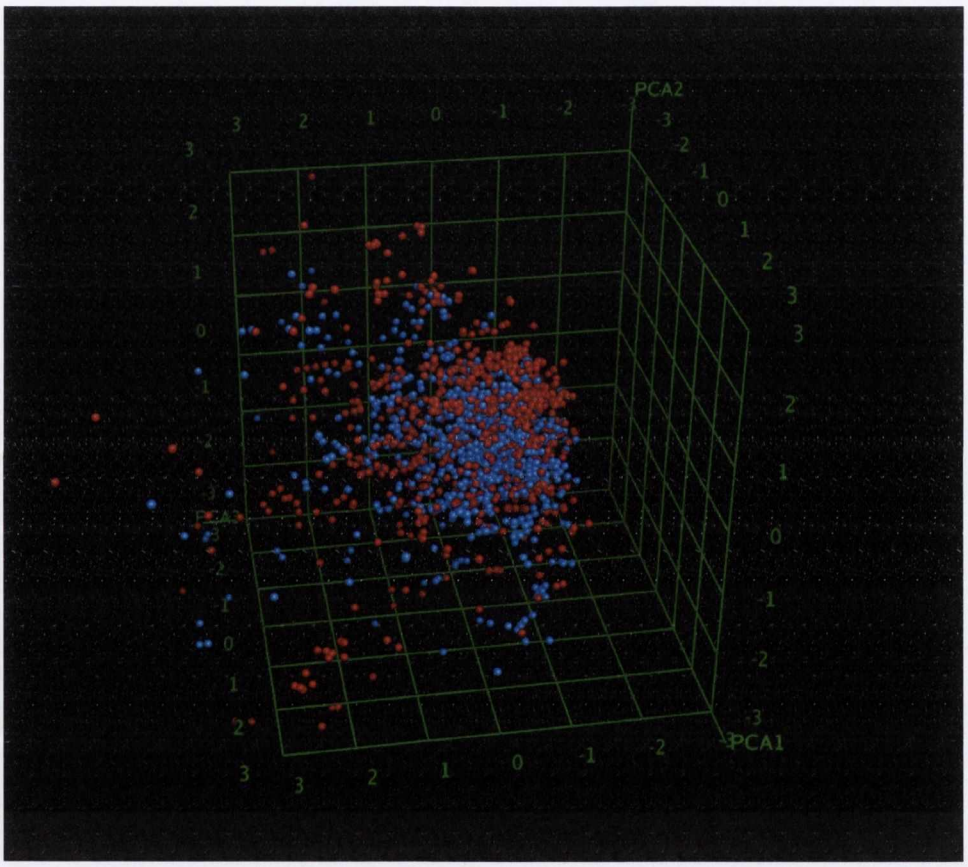


FIGURE 4.11: PCA of anti-trypanosomal (bue spheres) and anti-malarial (red spheres) medicinal chemistry space.

When a database of drug-like compounds was added to the PCA analysis, Figure 4.12, it was clear that many compounds that have passed different phases of clinical trials are often not ones that are considered to be traditional drug-like molecules. What is illustrated is the relatively compact nature of RO5 drug-like space when compared with wider medicinal chemistry space, especially considering that over 20% of the antiparasitic compounds included span a considerable breadth of space. Such a general spatial distribution of acitve drugs outside the boundaries of the traditional drug-like space has implications for the creation of an all-encompassing malaria or trypanosome-generic filtering rules for database pre-processing, and it would perhaps be advantageous to adopt a target-by-target focus that could be used to craft tailored chemiinformatic filters biased to specific drug targets within the parasites, generating a target-specific medicinal chemistry space. For this type of study to carry meaningful results would require obtaining a sufficiently large database of molecules with known molecular targets, a non-trivial assignment.

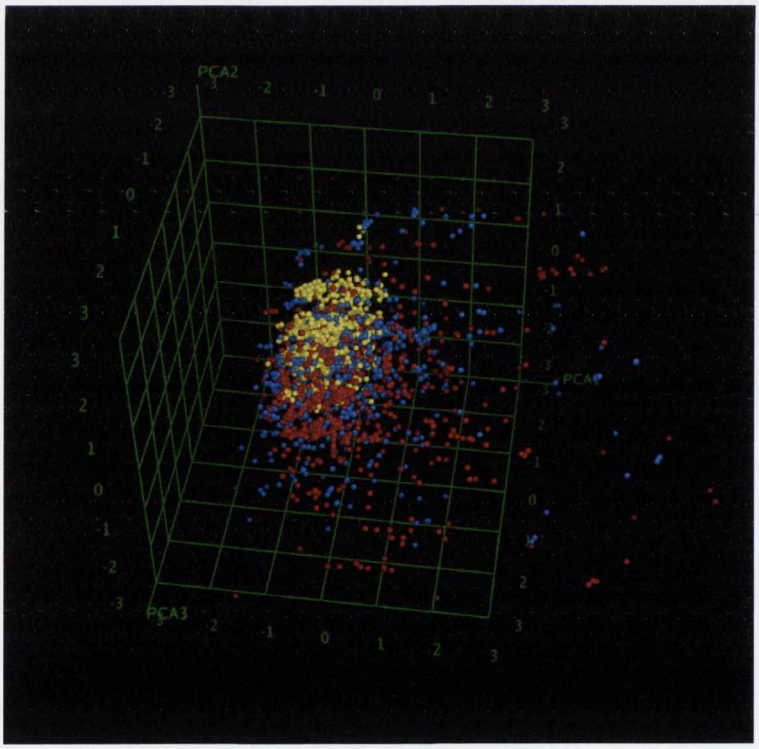


FIGURE 4.12: Anti-trypanosomal (bue spheres), anti-malarial (red spheres) and "drug-like" medicinal chemistry space (yellow spheres).

4.13 Pharmacophore Design and Validation

Using MOE pharmacophore elucidator a 3 feature pharmacophore was built, consisting of a hydrophobic feature on the ring nitro group, a hydrophobic/aromatic feature on the aromatic ring and a h-bond acceptor feature on the number 2 oxygen atom. When tested against the validation set this pharmacophore returned 780 decoys and the internal ranking of actives was poor, as shown with an AUC of 0.344. A second hydrophobic feature was added at the hydrophobic head region of APM. Screening the validation database with this second pharmacophore returned 25 actives and 765 decoys; the internal active ranking remained poor with an AUC of 0.410. PH4-3 was created by removing the hydrophobic feature from pharmacophore 2 and then adding a H-bond donor/acceptor feature around the amine nitrogen at the APM head region (Figure 4.13) An AUC of 0.829 was achieved after employing this new pharmacophore in a screen of the validation database, returning 25 actives and 157 decoy compounds (Figure 4.14). While this was a large improvement upon the previous pharmacophores, the internal ranking of actives remained unacceptable. The addition of a sixth feature, a third hydrophobic feature at oxygen 2 carbon, offered no improvement on PH4-3. The addition of volume constraints around PH4-3 proved time consuming and inefficient in improving AUCs and it was decided that ROCS would be employed to post-process the pharmacophore output in an attempt to improve the ranking of actives.

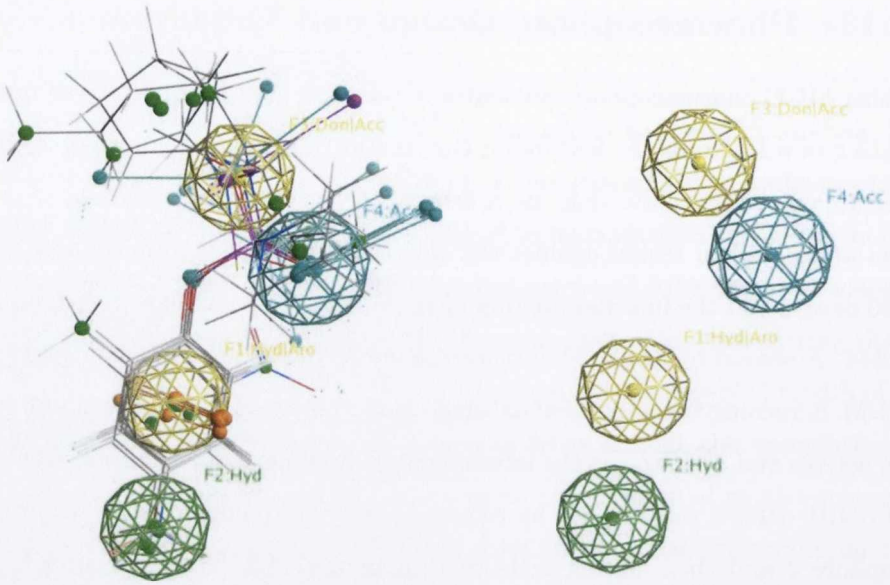


FIGURE 4.13: Final pharmacophore, showing the alignment of 5 APM analogues and the pharmacophore itself.

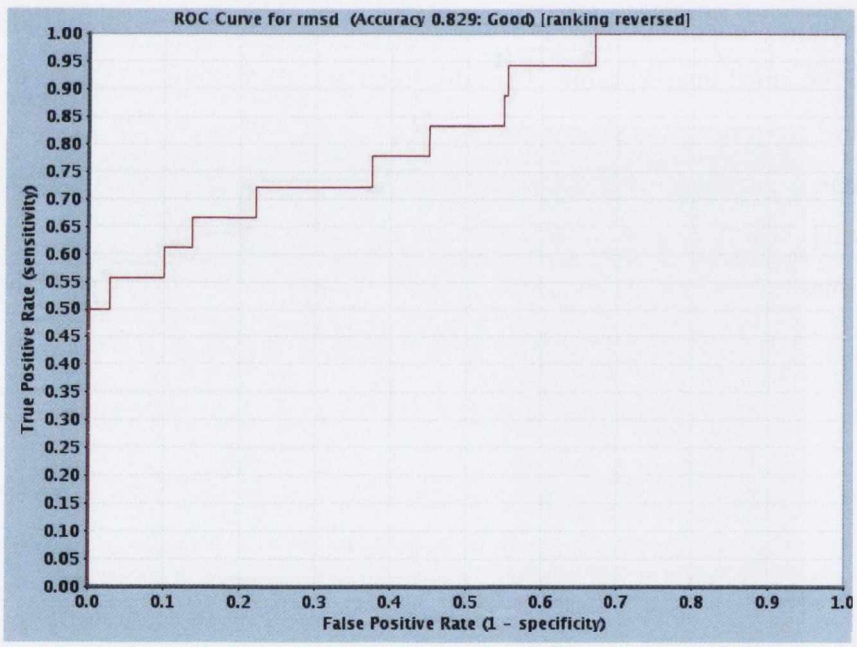


FIGURE 4.14: ROC curve calculated after searching the decoy database with pharmacophore 3.

4.14 ROCS Query Generation and Validation

An energy minimised structure of APM was used as the query for ROCS screening of the PH4-3 output from the validation database screen. Using ShapeTanimoto to rank the output, ROCS was able to generate an AUC of 0.935 after re-ranking PH4-3's output database, an increase of more than 10% (Figure 4.15). This figure was superior to both pharmacophore and ROCS screening on their own, illustrating the benefit of adopting a tiered approach to virtual screening.

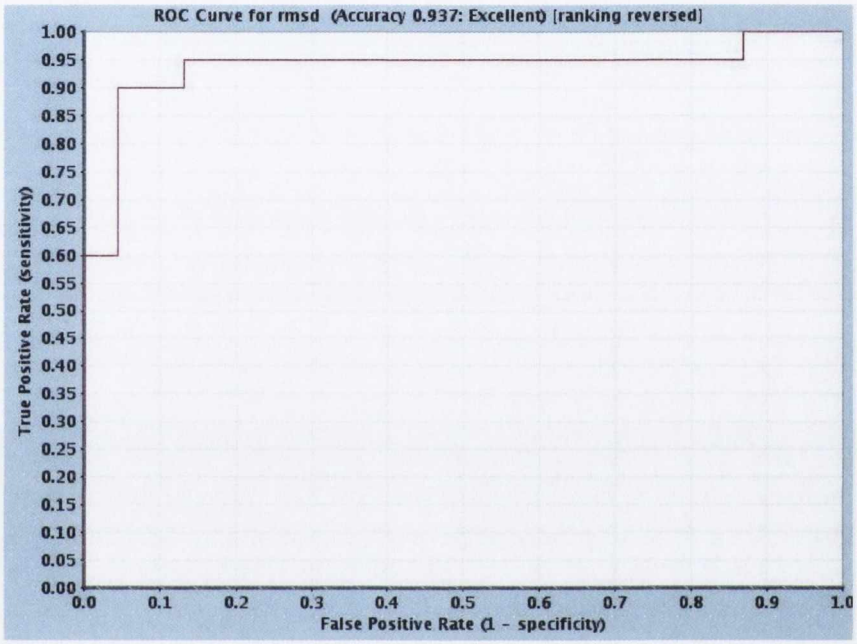


FIGURE 4.15: ROC curve calculated ater searching the decoy database with pharmacophore 3 and then re-ranking the pharmacophore output with ROCS searching.

4.15 Vendor Database Screening

The SPECS vendor database (169,239 compounds) was selected to screen for compounds using a tiered screening process involving an initial PH4 filter followed by a post-processing step with ROCS. Prior to screening the SPECS database was reduced in size using a Pipeline Pilot protocol. Briefly, this protocol filtered the incoming SPECS entries according to a set of descriptor values detailed in Table 4.3 and then removed compounds that had a Tanimoto similarity above 80% to a set of known promiscuous inhibitors (Figure 4.16). Each compound passing this

filter was submitted for conformer generation with Omega with a maximum of 50 conformers generated per molecule.

TABLE 4.3: Descriptors used to filter the SPECS database prior to virtual screening

Descriptor	Value
Molecular Weight	300 - 500 Da
Octanol/water partition coefficent	0.9 - 5
Number of H-bond donors	0-3
Number of H-bond acceptors	1-6

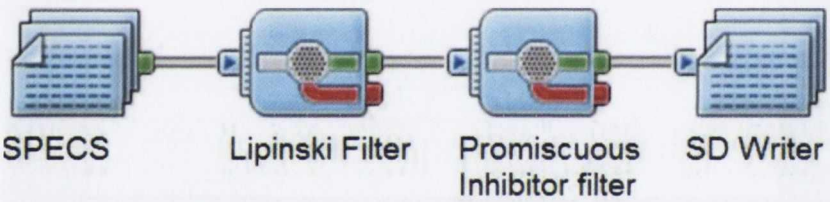


FIGURE 4.16: Pipeline Pilot Protocol for Vendor Database Preparation. This protocol filtered the incoming SPECS entries according to a set of descriptor values detailed in table 4.3 and then removed compounds that had a Tanimoto similarity above 80% to a set of known promiscuous inhibitors.

Searching the resulting database of enumerated SPECS compounds returned 8777 individual hits. Each of these were again submitted to conformer generation as in the previous step and re-ranked using ROCS, returning 200 scored hits. The 200 hits were clustered in Pipeline Pilot using a maximum dissimilarity method to select cluster centers and 15 compounds from the clusters were chosen for experimental testing (Figure 4.17). Molecules were chosen based on them having no previous experimental history as anti-parasitic compounds as well as their availability for purchase.

Figure 4.18 and 4.19 below illustrate the ordered compounds and their alignment

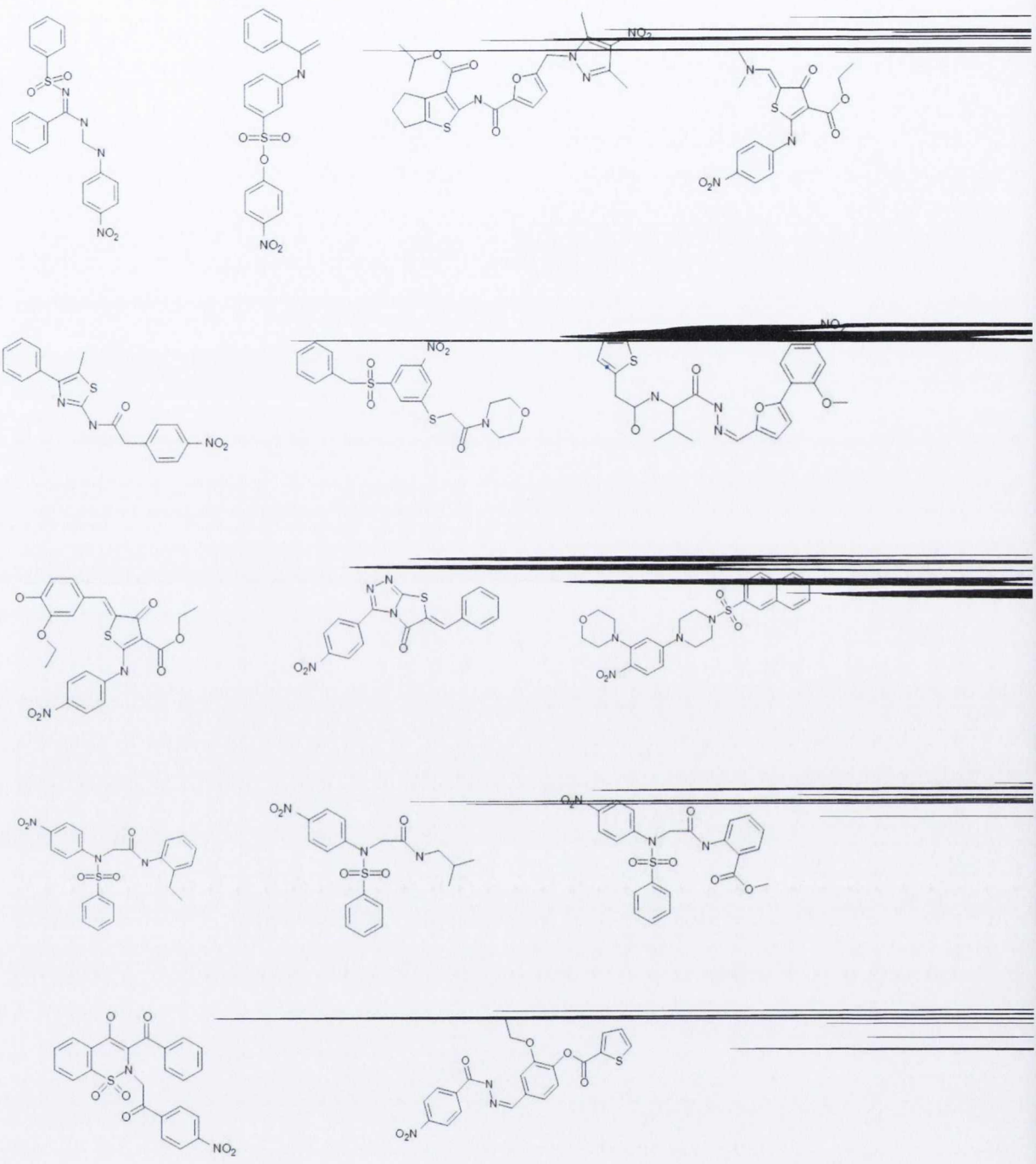


FIGURE 4.17: Compounds ordered from SPECS vendor database.

with the herbicide pharmacophore while figure 2.13 shows a selection of the ordered compounds and their overlap with the herbicide ROCS query structure Figure 2.16 shows the 2D structures of these compounds. The nitrophenyl moiety of compound one matched to the aromatic and hydrophobic pharmacophore features (nitro group = hydrophobic), while the acceptor and acceptor/donor features

mapped to the oxygen atoms of the sulfonamide group. In compound 2 the same feature mapping arrangement as in compound 1 was selected. With compound 3, the hydrophobic feature mapped to a methyl group on the pyrazole ring and the aromatic group aligned with the furan ring. The acceptor features map to the oxygen atoms of the peptide carbonyl oxygen and the carbonyl oxygen in the ester. Although compounds 4 and 8, which have identical orientations within the pharmacophore, have the same nitrophenyl group as compounds 1 and 2, these groups don't match the pharmacophore alignment of 1 and 2. Instead, the hydrophobic feature maps to the tertiary amines and the aromatic feature maps to the thiophene rings. The acceptor features map to the carbonyl oxygens on the thiophene ring and the ester group. Compound 5 repeats the arrangement of the nitrophenyl features as in compounds 1 and 2, while acceptor and acceptor/donor features map to the peptide bond's carbonyl oxygen and amine group respectively. Compound 6's morpholine group matched the hydrophobic and hydrophobic/aromatic features, while the sulfonamide oxygen atoms mapped to the acceptor and donor/acceptor features. The hydrophobic feature mapped to the thiazole ring in compound 7, while the acceptor and donor/acceptor feature mapped to a carbonyl oxygen and secondary amine in the linker region between the thiazole and furan rings. Compound 9's nitrophenyl group follows the mapping as in previous examples and the acceptor and donor/acceptor features mapped to the nitrogen atoms of the imidazole ring. Compound 10's napthalene group mapped to the hydrophobic and hydrophobic/aromatic features while the oxygen atoms of the sulfonamide group mapped to the acceptor and donor/acceptor features, identical to compounds 1 and 2, an arrangement that is repeated in compounds 11, 12 and 13. The hydrophobic and hydrophobic/aromatic features map to the nitrophenyl group in compound 14, while the acceptor and donor/acceptor features map to a carbonyl oxygen and a sulfonamide oxygen.Compound 15 followed the nitrophenyl feature mapping pattern and the acceptor and donor/acceptor features mapped to the secondary amine and tertiary amine respectively.

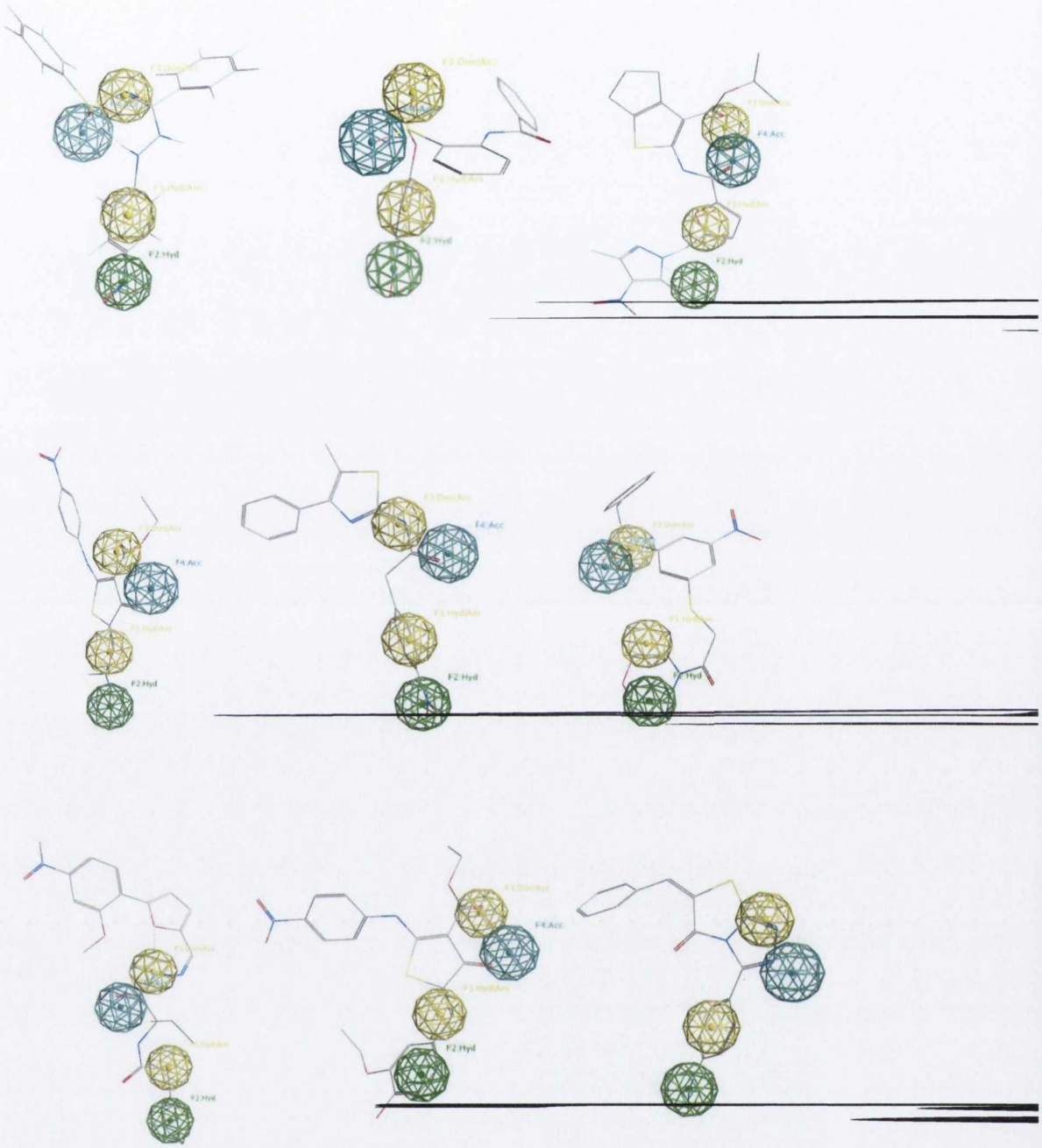


FIGURE 4.18: Compounds ordered from SPECS vendor database aligned with pharmacophore 3.

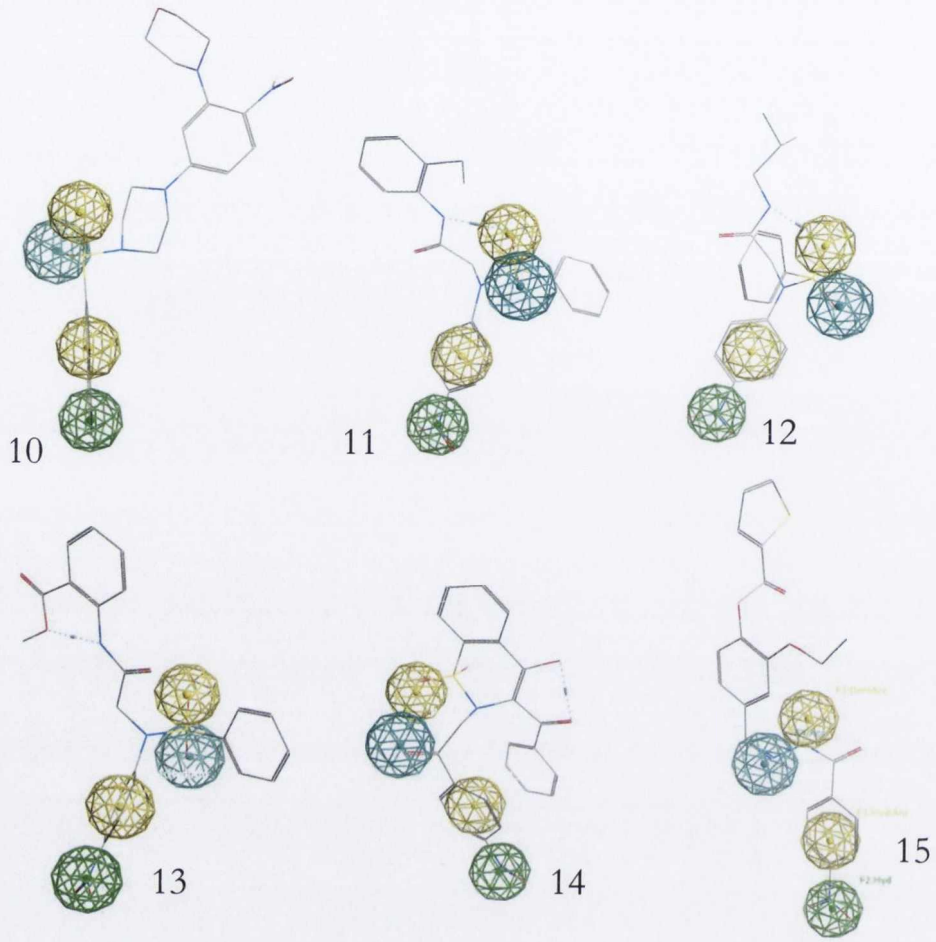


FIGURE 4.19: Compounds ordered from SPECS vendor database aligned with pharmacophore 3.

4.16 PCA Analysis of Compounds Identified Through VS

The 15 compounds identified through ligand-based virtual screening were analysed by PCA to compare their position in chemistry space relative to the APM herbicide analogues and wider 'drug-like' space. Figure 4.20 illustrates the distinct regions of chemical space occupied by the APM series in blue and the 15 vs compounds in blue. The 15 compounds occupy more or less a similar region of space, representing a move form the herbicide chemical scaffold to a new chemical scaffold. These compounds are situated in the middle of 'drug-like' space as represented by a susbset of drug-like compounds from the Maybridge vendor database (Figure 4.21).

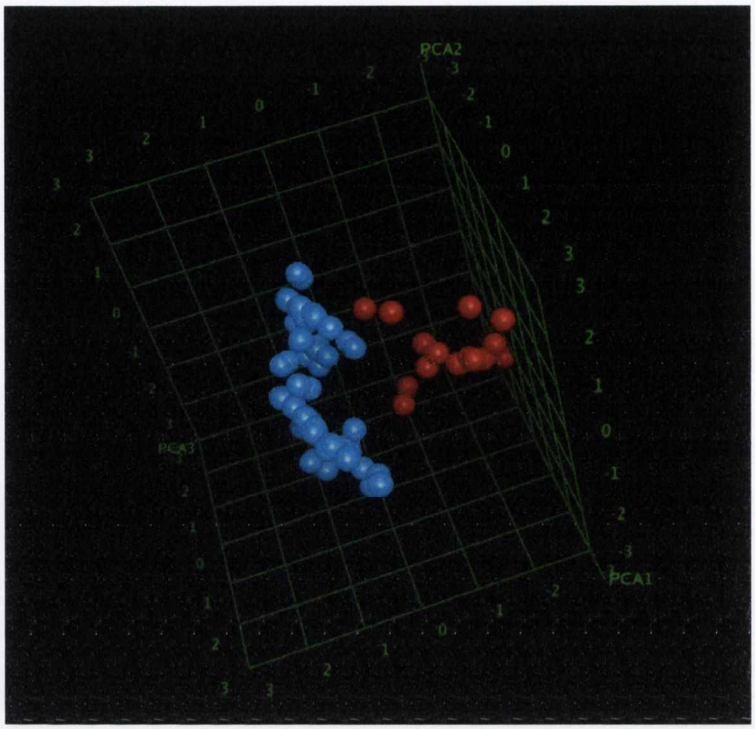


FIGURE 4.20: Chemistry Space of the APM herbicides (blue) versus the 15 compounds ordered from SPECS (red).

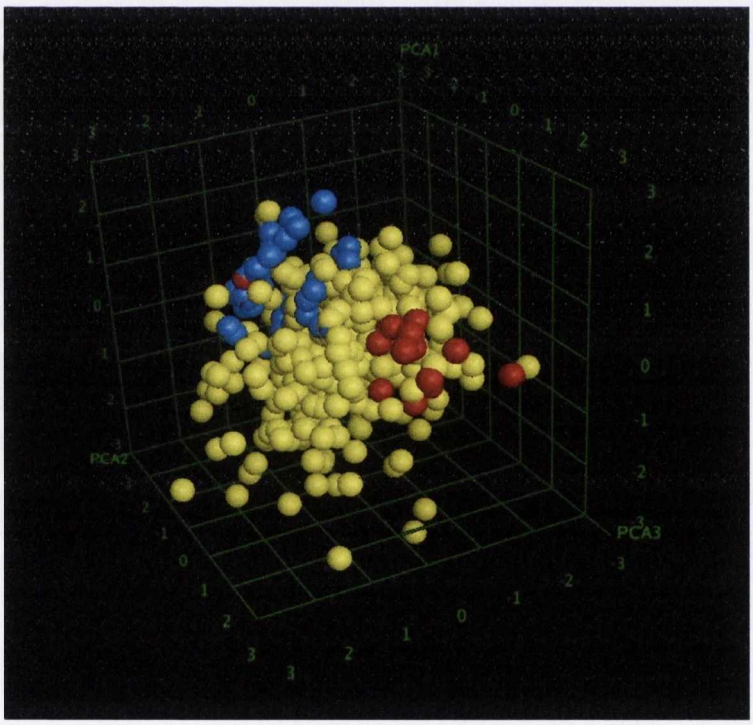


FIGURE 4.21: Chemistry Space of the APM herbicides (blue) versus the 15 compounds ordered from SPECS (red) and drug-like compounds in yellow.

4.17 Conclusion

As outlined in section 1.6 of Chapter 1 two related families of organophosphorous herbicides have been shown to cytotoxic to several protozoal parasites, including *P. falciparum* and *T. brucei*, but are associated with unwanted properties including low solubility and their potential to act as inhibitors of autonomic transmission. These facts considered a ligand-based drug design approach was chosen to aid in the identification of novel compounds based on the herbicide scaffolds. The ultimate aim of this was to move from the herbicide chemical class, and away from any potential anti-cholinesterase effects, while retaining the parasite-selective cytotoxic nature of the herbicides.

In this chapter the implementation of a tiered virtual screening approach, utilising a three-dimensional pharmacophore filter followed by a shape-based re-ranking with ROCS, has been used to identify 15 novel compounds from the SPECS viendor database. Pharmacophore modeling approaches such as the one detailed here are the primary recourse for projects for which no direct experimental structural data is available. There are many examples in the literature detailing pharmacophore design and inhibitor discovery based solely on the energy minimised structures of active compounds when protein-ligand binding poses are unknown[200].

A set of three-dimensional pharmacophore models were generated using MOE pharmacophore editor in order to probe the common features of a set of phosphorothioamidate compounds that have been shown to kill a number of parasites in what is thought to be a tubulin-targeted manner. Due to the lack of any structural information regarding the compound binding site, a purely ligand-based approach was undertaken. Among the pharmacophore hypotheses generated, one (ph4-3) displayed a fairly good ROC score when screened against a decoy set, however not good enough to be used as the final method in a virtual screen. The addition of a ROCS shape-based re-ranking step improved the ROC score by more than 10%, proving the validity of the approach. After using this protocol to screen the SPECS vendor database, 15 compounds were identified to be brought forward for biological validation. These compounds represent a move away from the herbicide scaffold to a new chemistry space that is more drug-like and potentially less dangerous that the phosphorothioamidate herbicide structure.

The main area of debate in this chapter, and indeed in all pharmacophore screening approaches, is the concept of model validation. At this step we are asking ourselves 'do we have a model fit for screening', and this question relies on which definition of the term 'pharmacophore' one is using. In this specific case, we term pharmacophore as meaning a collection of features that are important for a given biological activity, but not necessarily vital for supramolecular interactions with a target protein. This terminology was necessary as no structural information was available and although a series of APM analogues was available for the study, their biological data was restricted to whole cell IC_{50} assays and not on target information, meaning that even though the pharmacophores generated were intended to identify new tubulin inhibitors the lack of on target affinity constants made this approach problematic. Of course this works both ways; on-target biochemical assays give very uncomplicated information (with the exception of agonist/antagonist data, allosteric modulation) regarding ligand-protein binding affinity, but what is lost here is the cellular context of the drug action, the ultimate functionality, and for this reason it is important to merge the two assay types for optimum drug discovery output. With the pharmacophore terminology defined, a theoretical concept in truth, the limitation becomes material in the quality of the available data sets, specifically the knowledge of active/inactive data. In this case the availability of known inactive compounds was extremely limited, making validation pharmacophore hypotheses more difficult. As such it was necessary to create a database of decoy molecules for the validation procedure, a process with the potential for introducing unavoidable bias in the study. The problem in these situations is how one defines 'decoy' in relation the known actives; making the decoys too dissimilar to the actives results in a validation process that is unchallenging, making the decoys too similar increases the chances of removing key information from the pharmacophore hypotheses. Decoy selection in the absence of known inactives is a serious area of debate in pharmacophore design; the prevailing approach in the past has been to seed actives in a database of randomly selected compounds (assumed inactive) with the model validated by its ability to select actives over decoys. The problems inherent with this approach are obvious, mainly that randomly selected decoys will often plainly be inactive due to reasons that do not require a revolutionary pharmacophore model to decipher, including having molecular weights far larger than the known actives. Recent studies have

shown that decoy data set composition has a massive effect on validation methods, and differences in simple descriptors such as molecular weight between decoys and actives can result in significant artificial enrichment. It is therefore more logical to select decoy compounds that share some identity with the active compounds. However, following this method increases the chances of selecting potentially active compounds as decoys, a problem that can only be avoided when one knows for certain the inactivity of the decoys. In this work, the ultimate validation of the protocols will be the biological characterisations of the compounds identified.

Chapter 5

Biological Characterisation of Computationally Designed Compounds

5.1 Introduction

Any virtual screening method, no matter how statistically robust or reliably validated *in silico*, has to be assessed through *in vitro* assays for final validation. This study focuses on the aplication of *in vitro* experimental methods toward the identification of a novel scaffold in an attempt to identify compound distinct from the herbicide classes of parasitic microtubule disrupting compounds described in Chapter 1. These compounds are desirable in their ability to affect the viability of parasitic species such as *Plasmodia* and *Trypanosoma* while displaying negligible effects against mammalian cell lines. The herbicides have also been previously shown to disrupt parasitic tubulin specifically, a characteristic they do not exhibit against mammalian tubulin. However, as discussed in Chapter 1, these compounds possess traits that make them unsuitable for use in humans, including poor solubility in the case of the dinitroanilines and potential anti-cholinesterase activities of the phosphorothioamidates. In an attempt to further characterise the leads identified through ligand-based virtual screening as described in Chapter 4, these compounds were assayed for their ability to exert negative effects on parasitic cultures *in vitro* though cell viability assays. These types of assay are used to

determine the ability of cells to continue to grow or recover in the presence or absence of certain stimuli, in this case the presence of putative inhibitor compounds. In this study three different methods of cell viability were employed: (1) a parasite lactate dehydrogenase (pLDH) assay to assess *P. falciparum* viability, (2) an ATP-based measurement of cell viability for *T. brucei*, and (3) an Alamar blue cell viability assay to measure mammalian cell viability.

5.2 Methods

5.2.1 Chemicals & Reagents

All the chemicals and reagents were from Sigma-Aldrich, Dublin, Ireland unless otherwise stated. Trifluralin, oryzalin, amiprophos-methyl (Fluka Chemie AG, Switerland), novel APM derivatives (Specs, The Netherlands), novel trifluralin derivatives (a gift from Epichem, Murdoch, Australia), taxol, vinblastine and colchicine were all dissolved in dimethylsulphoxide (DMSO) to a stock concentration of 20 mM. Aliquots of these samples were made and stored at -70 oC. Bovine tubulin (α-β) was obtained from Cytoskeleton (Denver, CO, USA). SYPRO® orange dye, Albumax II and RPMI 1640 medium were from Invitrogen (Carlsbad, CA, USA). CellTiter-Glo was purchased from Promega (Southampton, UK).

5.2.2 Culture of *Plasmodium falciparum*

P. falciparum 3D7 strain was maintained in continuous culture in human erythrocytes (whole blood obtained from the Irish National Blood Centre, St. James's Hospital, Dublin and erythrocytes extracted as described by Read and Hyde[201]) according to the method of Trager and Jensen[202]. Parasites were cultured in washed human erythrocytes at 2.5% haematocrit in RPMI 1640 medium supplemented with 0.5g (w/v) Albumax II, 25 mM HEPES, 25 μg/ml gentamicin, 0.18g (w/v) sodium bicarbonate and 50μg/ml hypoxanthine. Parasites were maintained in Petri dishes, in a candle jar with reduced O_2 tension at 37°C. Culture medium was replaced depending on the parasitaemia. Parasitaemia was monitored by microscopic examination of Giemsa-stained smears.

5.2.3 *Plasmodium falciparum* viability assay

To assess the effects of drug compounds on asynchronous populations of P. falciparum, asynchronous cultures at 2% haematocrit and 0.8% parasitemia were grown in 96-well flat-bottom microtitre plates in culture medium supplemented with the appropriate concentration of the inhibitor. Inhibitors were diluted from the stock (20 mM) solutions into culture medium, and subsequently serially diluted two-fold in wells of the microtitre plates from an initial high concentration

to sub-inhibitory concentrations. Following 48h and 72h incubations at 37°C the effects of the inhibitors on growth of P. falciparum was determined using the parasite lactate dehydrogenase (pLDH)- based assay described by Makler *et al.*[203]. Absorbance readings were obtained and a dose-response curve constructed for each inhibitor and results are expressed as IC_{50} values. The half maximal inhibitory concentration (IC_{50}) is a measure of the effectiveness of a compound at inhibiting the parasite growth based on the pLDH activity.

5.2.4 Culture of *Trypanosoma brucei*

The bloodstream forms of *Trypanosoma brucei* were acquired from the cultures of Dr H.P. Voorhies and Dr Andrew Knox of Trinity College, Dublin 2.

Bloodstream forms of *T. brucei* were maintained in sterile HMI-9 media supplemented with inactivated fetal calf serum (10%), guanosine (0.02%), adenosine (0.025%), sodium bicarbonate (0.2%), 2- beta-mercaptoethanol (7ul/500mls) and pH adjusted to 7.4. Bloodstream form *T. brucei* cutures were incubated at 37 °C in a 5% CO_2, 95% O_2 incubator and harvested at a cell densitiy between $1.5x10^5$ and $2x10^6$ cells per ml. Cells were maintained in T25 and T75 flasks and counted daily to ensure optimal cell concentrations for appropriate experimentation.

5.2.5 *Trypansoma brucei* Cell Viability Assay

The assessment of compound-mediated killing of *T. brucei* was conducted using the CellTiter Glo kit. *T. brucei* parasites were cultured to a concentration of $2x10^5$ cells per ml in T75 flasks and 100 ul of this was transfered to 96 well plates, giving a final concentration of $2x10^4$ cells per ml. Compounds were made up to 6mM and 0.6mM stock solutions in 100% DMSO and diluted 1 -in-20 in HMI-9 media. 5μl of diluted compound was added to each well, giving a final DMSO concentration of 0.16%, a concentration used for the vehicle control cells. Untreated wells were used as a negative control and 5ul of 10% Triton-X-100 was used as a positive control. Plates were incubated for 24h before the addition of 50ul of CellTiter-Glo reagent to each well. This reagent allows the measurement of the number of viable cells based on the presence of ATP in metabolically active cells. Problems with this assay have been noted in the past, specifically in relation to some drugs lowering

ATP levels but not killing cells, and this potential problem was addressed by performing a cell count on drug-treated parasites to confirm ATP measurements. Plates were read on a LUMIstar Galaxy multiplate luminometer.

5.2.6 HeLa cell culture

The HeLa cell line is a Human Papilloma Virus (HPV) positive cervical adenocarcinoma cell model. This cell line was derived from a pre-treatment biopsy of Henrietta Lacks, a 31 year old African American cervical cancer patient who eventually died, in 1951. These cells represent the first immortalised *in vitro* human cell line and have since been used extensively in cancer cell biology research. HeLa cells were cultured in Dulbeccos Modified Eagles Medium (DMEM) plus Glutamax supplemented with 10% (v/v) Fetal Bovine Serum (FBS) and 100μg/ml PenStrep. Cells were maintained in 75cm^2 tissue culture flasks, containing 18ml supplemented media, at a temperature of 37°C in a 5% CO_2 / 95% O_2 humidified atmosphere. Cells were split at a ratio of 1 in 6, three times per week and used in experimentation until a passage number of no greater than 35 passages.

5.2.7 HEK293 cell culture

Human embryonic kidney cells (HEK) are a cell line derived from human embryonic kidney cells. These cells were originally generated in the 1970s through transformation of normal human embyonic kidney cell cultures with sheared adenovirus 5 DNA. Cells were cultured in DMEM plus Glutamax supplemented with 10% (v/v) FBS and 100ug/ml PenStrep. Cells were maintained in 75cm^2 tissue culture flasks, containing 18ml supplemented media, at a temperature of 37°C in a 5% CO_2 / 95% O_2 humidified atmosphere. Cells were split at a ratio of 1 in 6, three times per week and used in experimentation until a passage number of no greater than 35 passages.

5.2.8 Cell Cryo-storage

Low passage number stocks of each cell line were stored in liquid nitrogen to ensure a constant supply of cell lines for frequent experimentation. To generate cryo-stocks of each cell type, 5x10^6 cells of a low passage number were taken and

resuspended in 1.5ml of freezing media composed of 60% tissue culture media, 30% (v/v) FBS and 10% (v/v) DMSO. Suspended cells were placed in cryovials and incubated in a polycarbonate container containing iso-propanol and left overnight in a -80°C freezer; this freezing method ensures that the cells cool at a rate of -1°C per minute, protecting cellular integrity. Cells were then transferred to liquid nitrogen and stored at -180°C.

5.2.9 Alamar Blue cell viability assay

Alamar Blue^{TM} is a non-toxic, water soluble dye that can be employed to calculate the viability of cultured cells in a single-step, fluorescence-based assay. The assay is based on the ability of the oxidised form of Alamar Blue to enter the cytosol where it is reduced in viable cells with active mitochondria by cytochromes and oxidative phosphorylation substrates FMNH2, FADH2, NADH, and NADPH. The reduced form of Alamar Blue is fluorescent, with excitation at 544nm and emission at 590nm, and the redox event produces an associated colorimetric change in media from the dark oxidised form to a pink reduced form.

Depending on cell type, 0.5x10^4 - 1.5x10^5 cells per well were seeded in 96 well plates and left for 18h to adhere in the case of adherent cells. After 18h for mammalian cells, cells were treated with compounds and controls in triplicate and approximately 6h prior to endpoint 20μl of Alamar Blue was added to each well and left in the dark at 37°C in a 5% CO_2 / 95% O_2 humidified atmosphere. Fluorescence intensity (FI) was read on a Spectramax Gemini fluorometric plate reader using a SOFTmax Pro version 4.0 (Molecular Devices) software package. For data analysis, raw FI values of samples (Sample FI) were blanked by subtracting the FI of oxidised Alamar Blue of wells containing only Alamar Blue and media (Blank FI). Alamar Blue data was normalised by using the FI values of untreated control as representative of 100% viability (Untreated FI). Percentage viability of treated samples was generated using the following equation:

$$[(SampleFI)-(BlankFI)]/[(UntreatedFI)-(BlankFI)]x100/1 = \%CellViability \quad (5.1)$$

5.2.10 Thermal melt/shift assay

Samples were prepared in one of 4 buffers to determine the optimal buffer which yielded the best melting curve for the proteins in question. These buffers were: buffer 1 (100 mM HEPES, 150 mM NaCl, pH 7.5), buffer 2 (100 mM potassium phosphate pH 7.0), buffer 3 (100 mM sodium phosphate pH 7.5) or buffer 4 (100 mM sodium citrate pH 5.5). Samples were pipetted into 0.2 ml thin-walled PCR tubes (final volume: 50μl per tube). The final concentration of protein was 1 μM. The final concentration of SYPRO Orange was 3.2X from a stock solution of 5000X. Initial screening of compound-protein interactions was done at compound concentrations of 100μM. Experiments were performed in a Rotor Gene-3000 thermal cycler (Corbett Research, Sydney, Australia). Samples were heated from 30°C to 95°C at a rate of 2° per minute. Fluorescence readings were taken for each sample at each 0.2° increment (excitation 470 nm; emission 585 nm). A melting temperature (T_m) was determined by obtaining the first derivative to the curve and taking the curves maximal point. Any compounds that were shown to raise the tubulin T_m by § 2°C were then tested for a dose-dependent stabilisation of the protein at concentrations of 100, 50, 10, 1 and 0.5μM. Compounds causing a dose-dependent increase were considered as tubulin binders.

5.3 Results

5.4 Parasitic Viability Assays

5.4.1 *Plasmodium falciparum* Cell Viability Assay

All P. falciparum compound testing was performed by Mr Finian Doyle in the School of Microbiology, Trinity College Dublin.

The 15 compounds identified from ligand-based virtual screening were submitted for a cell viability assay on *P. falciparum* parasites. The results from this assay are presented in Table 5.1. Although the majority of the compounds did not show any improvement upon the activity of APM, compound 3 showed a roughly 2 fold improvement in $IC_{50} value, 1.58\mu$M, while compound 11 (5.45μM) also displayed an acceptable IC_{50}.

TABLE 5.1: Mean 50% inhibitory concentration data for the novel scaffolds identified through ligand-based virtual screening. The assay measures the lactate dehydrogenase activity of asynchronous *P. falciparum* 3D7 strain parasites as a measure for parasite growth and survival. The data presented is for 72-h cultures.

Compound	Geometric Mean IC50 μM
APM	3.24
1	-
2	18.2
3	1.58
4	-
5	37.0
6	50.7
7	-
8	51.3
9	24.5
10	-
11	5.45
12	59.4
13	58.9
14	40.7
15	62.5

5.4.2 *Trypanosoma brucei Cell Viability Assay*

Compounds 1-15 were tested for their ability to effect cell survival of bloodstream form *Trypanosoma brucei* parasites after 24h incubation at a concentration of 10 micromolar using the Cell Titre Glo cell survival assay method. Cells were also incubated with vehicle control and triton-x-100 positive controls.

From the initial results obtained, a number of compounds appeared to be exerting an effect on the survival of bloodstream form *T brucei*. At the 10μM concentration cutoff point, compounds 1, 3, 5, 7, 8, 10, 11, 13, 14, and 15 had a statistically significant ($P<0.001$) effect on T brucei survival as compared to control untreated (Figure 5.1). Of these, the results of compounds 5, 8, 10, 11, and 14 warranted further investigation through a dose-response analysis using the same assay. The results are presented in Figure 5.2. Compounds 5, 8, 10, 11 and 14 exhibited 24h IC_{50}s of 1.9μM, 1.4μM, 31.8nM, 4.6μM, and 0.9μM respectively. Compound 11 thus exhibited IC_{50}s of 4.6μM and 5.45 μM against *T. brucei* and *P. falciparum* respectively.

The 15 compounds were also assayed against HeLa and HEK293 cell lines to ascertain their effects, if any, on the proliferation of mammalian cells. As figures 5.3 and 5.4 show, none of the compounds tested displayed any significant killing effects against the cell lines tested. In contrast, etoposide (2μM) significantly reduced cellular viability to 24% and 12% in HEK and HeLa cells respectively. Table 5.2 summarises the results for the most interesting compounds after initial biological characterisation.

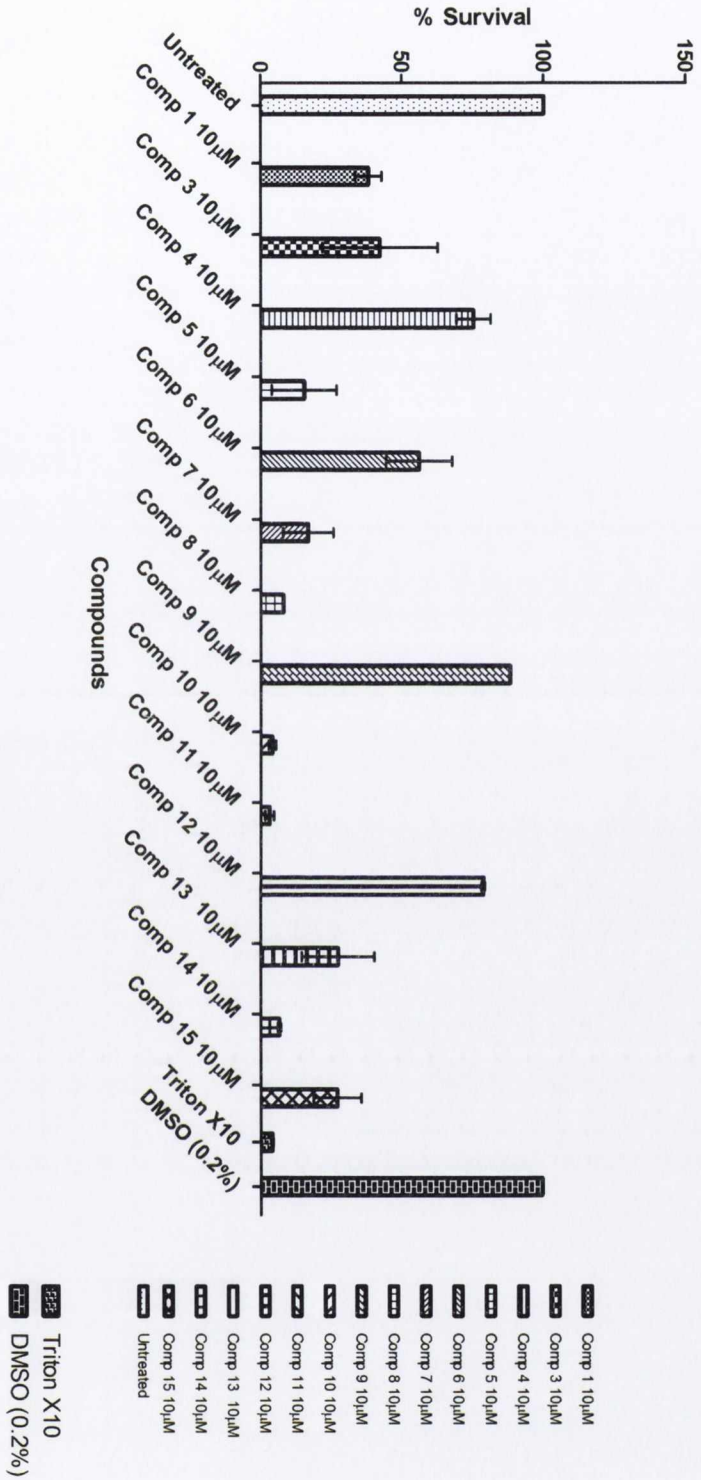


FIGURE 5.1: 24 h mean % survival data for the 15 compounds identified through virtual screening tested at 10μM versus bloodstream form of *T brucei*.

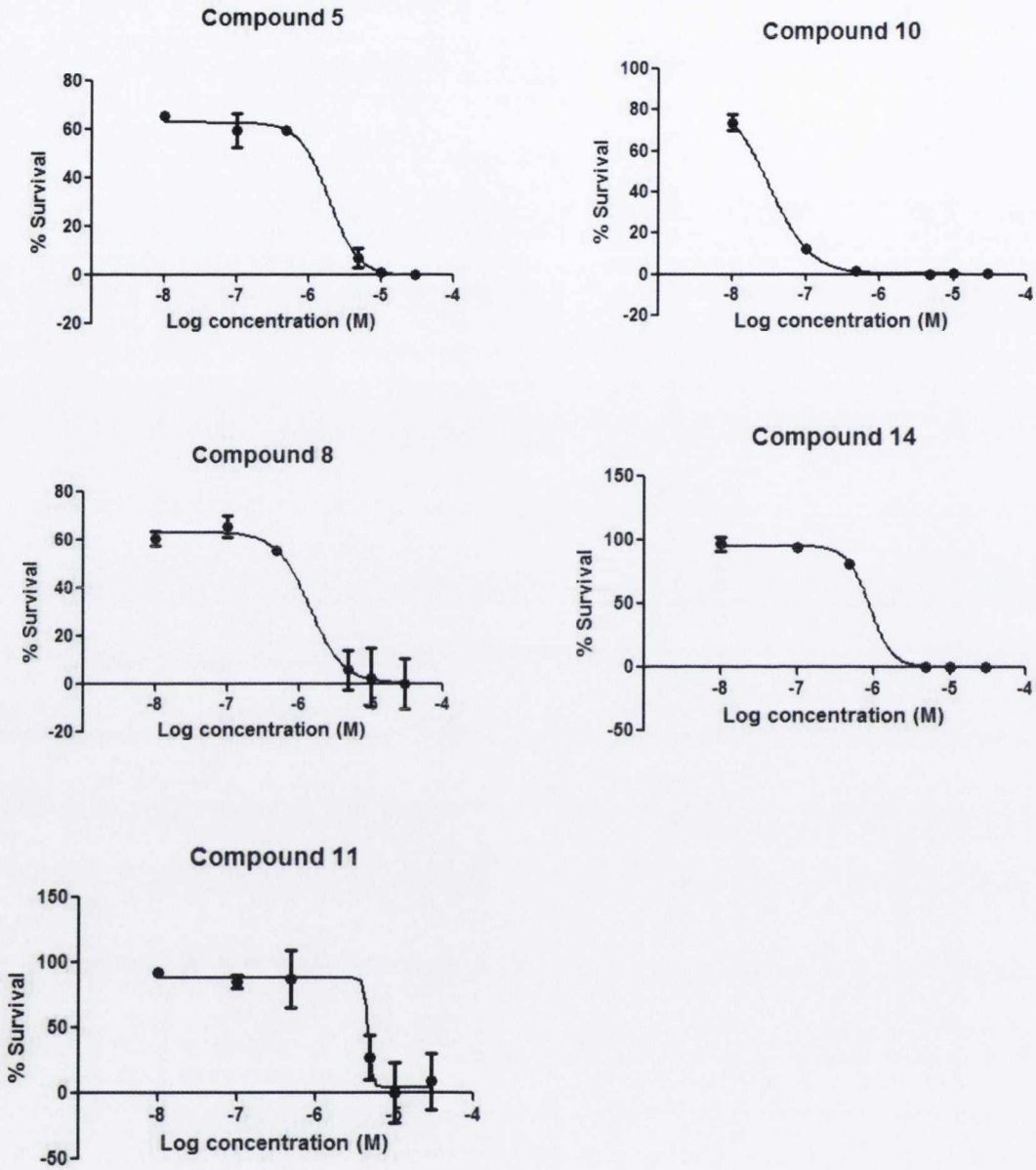
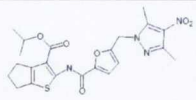
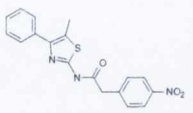
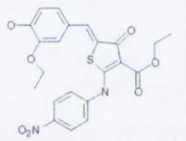
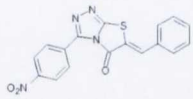
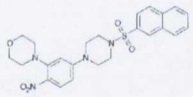
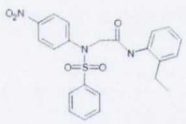
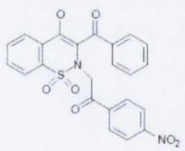


FIGURE 5.2: 24h dose-response curves of compounds 5, 8, 10, 11, and 14 against bloodstream form of *T brucei*.

TABLE 5.2: The biological effects of the most interesting compounds identified through ligand-based virtual screening. Included are the 72h effects of compounds on *P. falciparum*, the 24h effects on *T. brucei*, as well as 24h effects of compounds on HeLa and HEK293 mammalian cell lines (- no effect)

Compound Number	Structure	*P. falciparum* 72h IC_{50} (μM)	*T. brucei* 24h IC_{50} (μM)	HeLa 24h (10μM)	HEK293 24h (10μM)
3		1.58	NA	-	-
5		37	1.9	-	-
8		51	1.4	-	-
9		24	-	-	-
10		NA	0.031	-	-
11		5.45	4.6	-	-
14		40	0.9	-	-

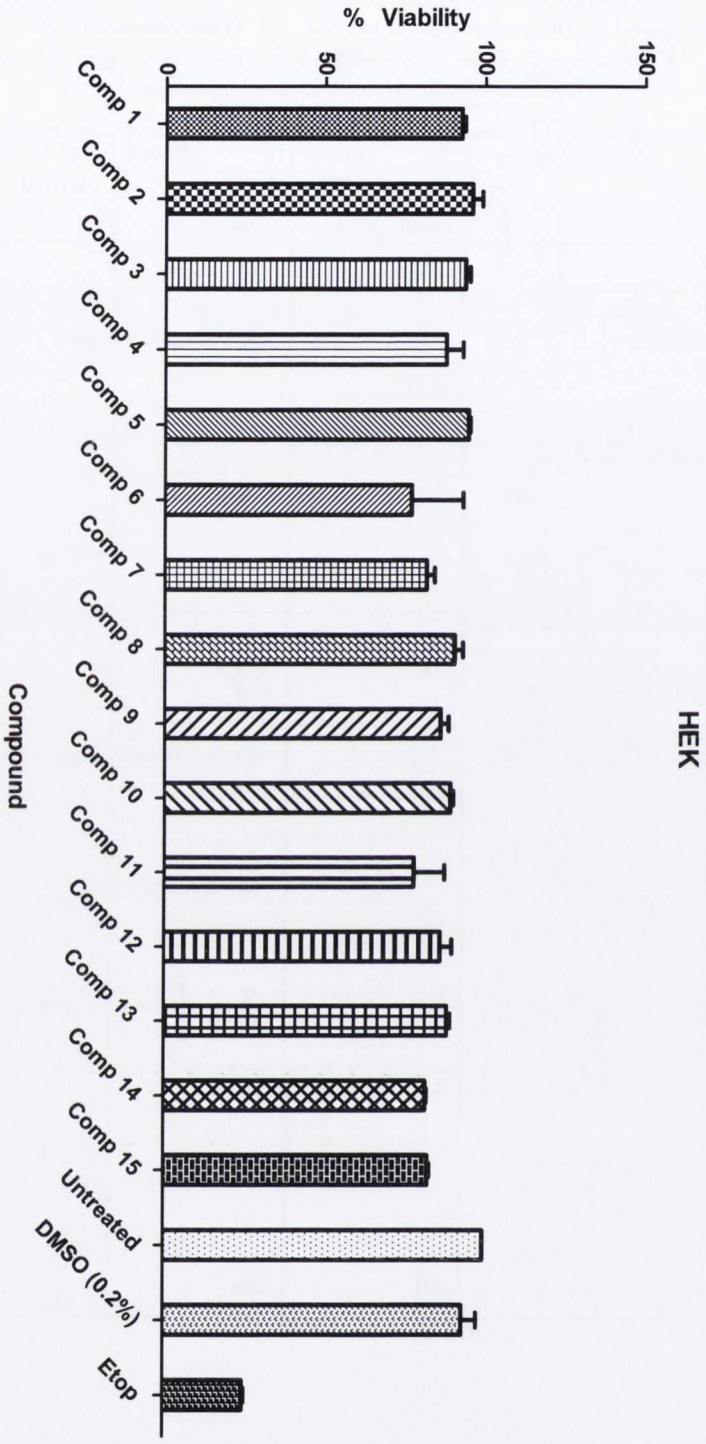


FIGURE 5.3: Graph of the effects of compounds 1-15 on HEK293 cell line as tested with the Alamar blue cell viability assay.

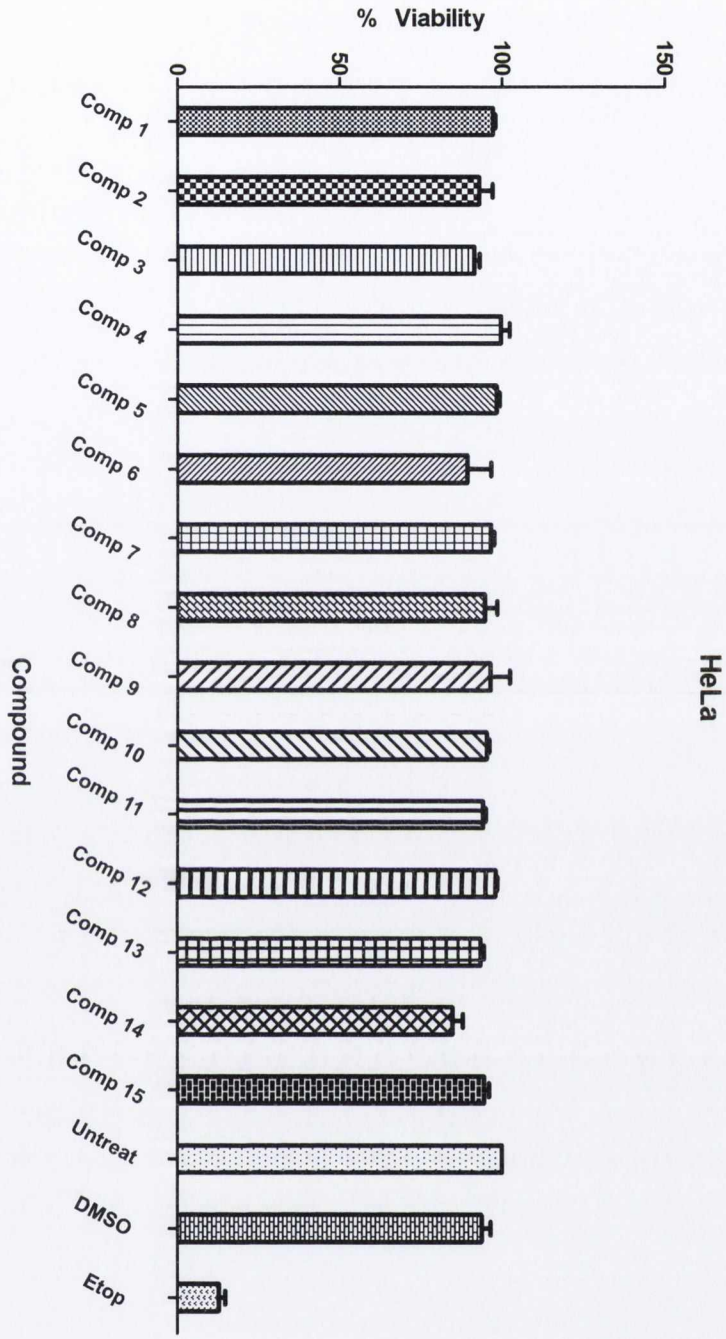


FIGURE 5.4: Graph of the effects of compounds 1-15 on HeLa cell line as tested with the Alamar blue cell viability assay.

5.4.3 Thermal Melt Assay

Thermal melt assays were performed by Mr Finian Doyle in the School of Microbiology, Trinity College Dublin.

All compounds were tested in a thermal shift assay to determine whether any interaction with bovine brain tubulin could be detected. By measuring the T_m of a protein in the presence and absence of a ligand, this assay can determine whether a ligand can bind to the protein. When coupled with binding data for the compounds with recombinant *P. falciparum* tubulin, these results will provide an indication as to whether the compounds show any specificity for parasite tubulin over that of the host.

Thermal melt assays were performed for bovine brain tubulin in the presence of the tubulin ligands vinblastine or colchicine alone and together. These assays were used to determine whether the assay was applicable for use with tubulin. Each ligand has a separate binding site on tubulin so it was expected that either ligand alone would bind to the bovine brain tubulin, thereby stabilising the protein and increasing the T_m. Both ligands together should further stabilise the protein resulting in a larger increase in the Tm (Fig.5.5, E and F). The data in Fig. 5.5 shows that the assays performed as expected and could be of use for examining possible *P. falciparum* tubulin interactions.

To determine the statistical significance of this data a T test was performed. When analysed using Graphpad Prism software it was shown that the binding of vinblastine and colchicine to bovine brain tubulin was statistically significant (P value < 0.0001) while no significant difference was seen with bovine in the absence or presence of APM (P value > 0.9). This confirms that the assay is reliable and robust.

Panel B of Fig. 5.5 shows a dose-dependent increase in binding of vinblastine to bovine brain tubulin. The T_m values were: 55.9 °C, 55.9 °C, 57.5 °C, 59.1 °C and 59.5 °C for compound concentrations of 0.5 μM, 1 μM, 10 μM, 50 μM and 100 μM respectively. The T_m value for bovine brain tubulin in the absence of ligand was 54.7 °C. Panel D shows a dose-dependent increase in binding of colchicine to bovine brain tubulin. The T_m values were: 56.0 °C, 56.5 °C, 57.3 °C, 57.4 °C and 57.5 °C for compound concentrations of 0.5 μM, 1 μM, 10 μM, 50 μM and 100 μM

respectively. The T_m value for bovine brain tubulin in the absence of ligand was 54.9 °C. Panel F shows a dose-dependent increase in binding of both vinblastine and colchicine to bovine brain tubulin. The T_m values were: 56.3 °C, 57.0 °C, 60.2 °C, 61.4 °C and 61.6 °C for compound concentrations of 0.5 μM, 1 ţM, 10 μM, 50 μM and 100 μM respectively. The T_m value in the absence of ligand was 54.7 °C. Panel H shows the T_m curves for the addition of APM to bovine brain tubulin. The T_m values were: 54.8 °C, 54.7 °C and 54.5 °C for compound concentrations of 0.5 μM, 10 μM and 100 μM respectively. The T_m value in the absence of ligand was 54.8 °C.

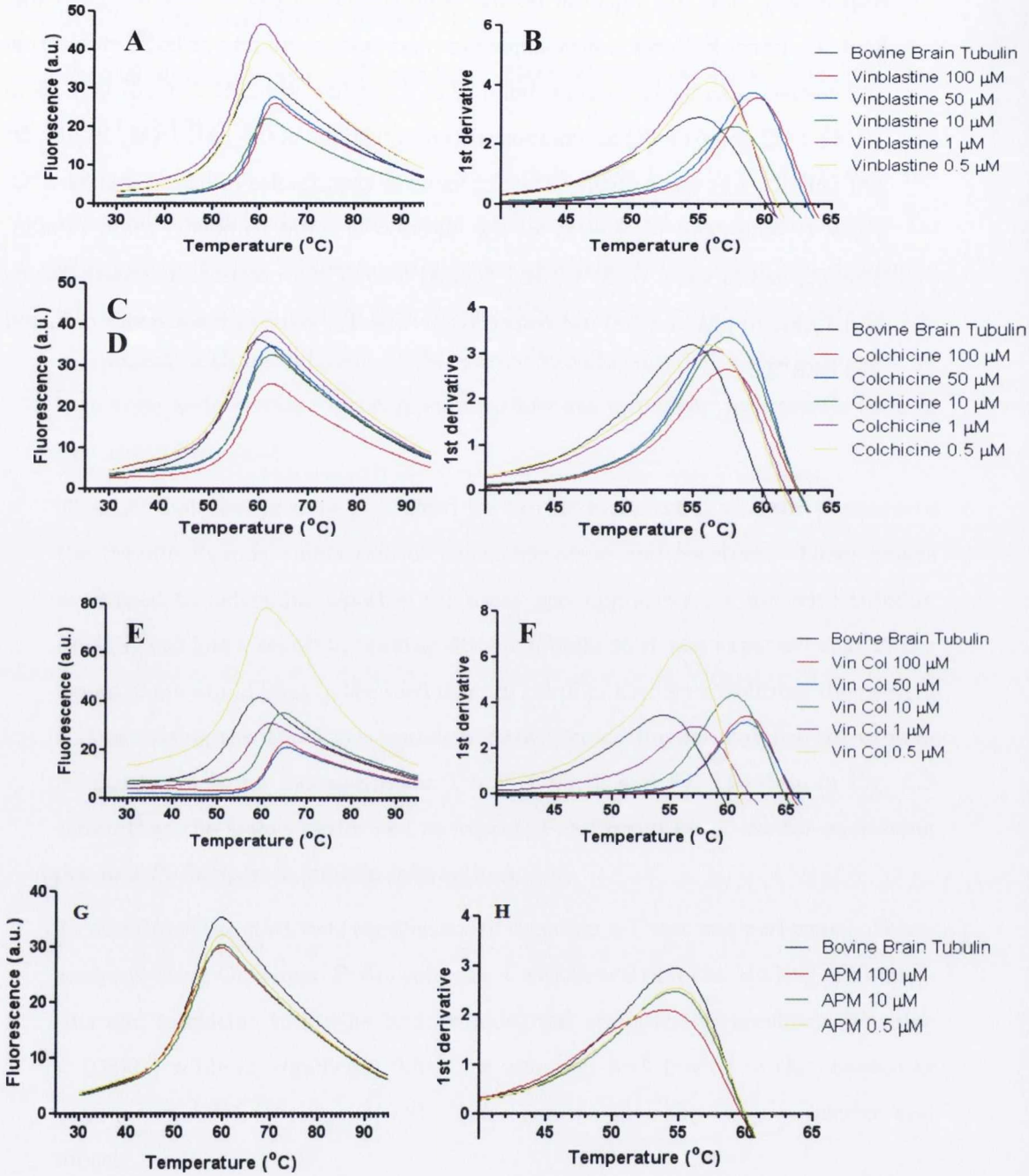


FIGURE 5.5: Thermal melt profiles of bovine brain tubulin (1μM) in the presence of a range of concentrations of vinblastine (**A** and **B**), colchicine (**C** and **D**), vinblastine plus colchicine (**E** and **F**) and APM (**G** and **H**). Bovine brain tubulin in the absence of any compound is indicated in the black curves. First derivative plots of the raw data curves are on the right

An initial determination of the Kd for the interaction between vinblastine and bovine brain tubulin (1μM) at 37 °C was made using the multiparameter fit method[204,205]. The value was determined as 0.1 μM. This value corresponds well with values reported in the literature (0.09-2 μM)[206 108] for this interaction. The data in Figures 5.6 and 5.7 (below) show that the majority of the compounds do not bind to bovine brain tubulin, as indicated by the lack of shift in T_m. The two exceptions were compound 6, which shows no detectable melting curve at 100 μM and compound 2, which shows a shift of 6.9 °C. Upon further investigation, using varying concentrations of both compounds on bovine brain tubulin, it was found that for compound 6 at lower concentrations (< 50 μM), normal melt curves were produced whereas there was a gradual reduction in fluorescence detected at higher concentration up to 100 μM, where almost no signal was detected. This could be due to having such an excess in ligand that it is no longer soluble. For compound 2, no interaction was detected at concentrations of 10 μM or less. From this data compound 11 and 3 show no detectable binding to tubulin which coupled with their anti-malarial activity, shows that they could be of promise. These compounds will need to be tested for binding to *P. falciparum* and *T. brucei* tubulin in future work.

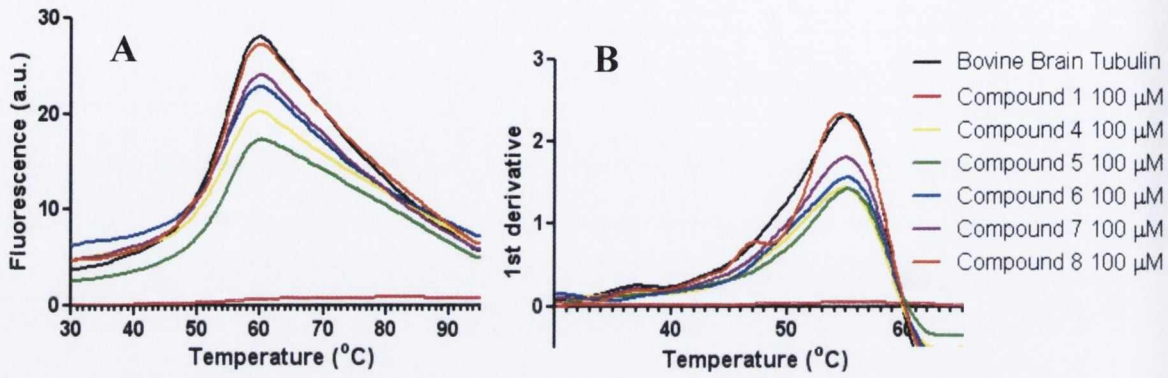


FIGURE 5.6: (A) Thermal melt profiles of bovine brain tubulin (1 μM) in the presence of a range of 100 μM of compounds 1,4,5,6,7, and 8 or in the absence of any compound (black line). (B) First derivative plots of the raw data curves in (A). The Tm is defined as the temperature where the first derivative is maximal. The Tm values were: C1 not defined, C4 54.6 oC, C5 55.1 oC, C6 54.9 oC, C7 54.9 oC, C8 54.6 oC and bovine brain tubulin 54.9 oC

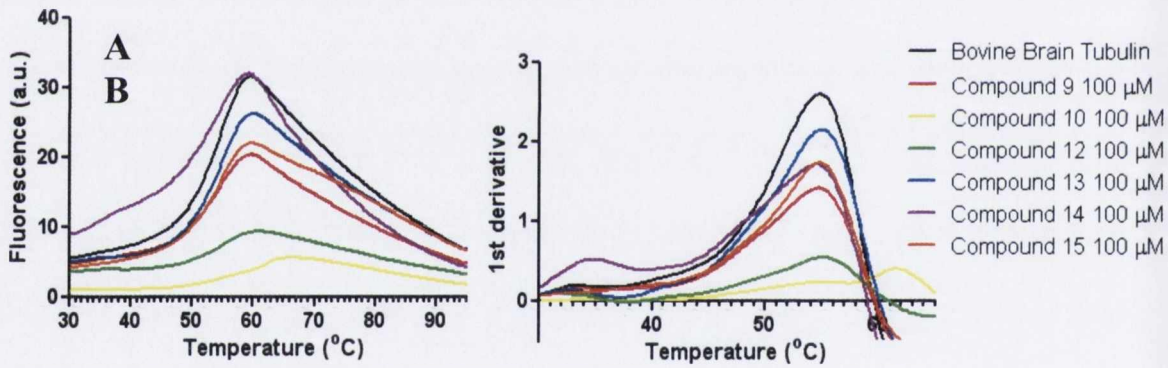


FIGURE 5.7: (A) Thermal melt profiles of bovine brain tubulin (1 μM) in the presence of 100 μM of compounds (C) 9, 10, 12, 13, 14 and 15 or in the absence of any compound (black line). (B) First derivative plots of the raw data curves in (A). The Tm values were: C9 54.7 oC, C10 61.7 oC, C12 55.0 oC, C13 54.9 oC, C14 54.8 oC, C15 54.8 oC and bovine brain tubulin 54.8 oC.

5.5 Discussion and Conclusion

This chapter detailed the biological validation of 15 compounds identified through ligand-based virtual screening in Chapter 4. Of the 15 compounds tested, one compound, compound 11, displayed the ability to inhibit the proliferation of *P. falciparum* and *T. brucei* parasites *in vitro* with IC_{50}s of 5.45 μM and 4.6 μM respectively. The result of the *T. brucei* assay is particularly interesting as subsequent work done with the same bloodstream form parasite strain revealed them to be resistant to treatment with eflornithine. Compounds 5, 8, 10, 11, and 14 also displayed significant ability to inhibit the survival of *T. brucei* parasites *in vitro*. Compound 10 was by far the most active compound identified with an IC_{50} of 31 nM versus the *T. brucei* parasites *in vitro*, more potent than nifurtimox (2.9 μM) and eflornithine (22 μM). Compound 3 was the most active of the compounds tested against *P. falciparum* with a calculated IC_{50} of 1.58 μM, although this compound had no effect on *T. brucei* parasites up to 10 μM. Cell viability assays carried out on HEK293 and HeLa cell lines showed that none of the 15 compounds displayed toxicity towards mammalian cells. Future work will involve carrying out dose-response analysis of the active compounds against mammalian cell lines at higher concentrations to establish IC_{50}s. In turn, this will allow us to generate selectivity profiles for the active compounds. Thermal melt assays were employed in order to ascertain whether any of the compounds interacted with purified bovine tubulin. Results showed that only compounds 2 and 6 interact with bovine brain tubulin, although a shift was only seen at high concentrations. Taken altogether, these results point towards compound 11 as a potential parasite-selective inhibitor. It will be necessary to undertake further work to ascertain whether or not the parasite-killing compounds are exerting their effect through disrupting the tubulin protein dynamics. However, as previously discussed, efforts to purify a working *P. falciparum* tubulin protein proved unsuccessful and purification of *T. brucei* tubulin was beyond the remit of this project. An alternative method to deduce whether or not these compounds are eliciting their effects through disrupting tubulin polymerisation dynamics would be to use indirect immunofluorescence staining of chemically fixed, compound-treated cells with tubulin antibody, allowing the examination of the distribution of microtubules in the cytoplasm at the moment of fixation.

All of the compounds that displayed antiparasitic activity are entirely novel chemical entities that have no previous descriptions as antiparasitic drugs in the literature and, as such, these results validate the choice of virtual screening campaign undertaken in Chapter 4. Analogues of each of these compounds have been identified through similarity searches and future work will involve generating SAR data around these new compounds. Taken together, this chapter validates the computational drug design approach made in chapter 4. The classic method of identifying new potential drugs involves conducting millions of tests on small molecules in an automated process. However, this method has been shown to have several drawbacks, including high cost, inefficiency, and high false positive numbers. Recent work has divulged that, in some cases, hit rates from High-Throughput Screening are as low as 0.021%[207]. In this work, we have achieved an anti-malarial hit rate of 13.3% and an anti-trypanosomal hit rate of 0f 33%, far higher than would be expected from random compound screening.

The biological results showed no pattern with respect to compound properties such as molecular weight or logP. However, the most promising compound, compound 11, possessed a very similar molecular surface to APM (Figure 5.8). Although two anlalogues of compound 11 were tested, compounds 12 and 13, it is difficult to obtain meaningful SAR from such limited data (Figure 5.9). However, from the limited data obtained, it can be deduced that an aromatic ring (2-ethylphenyl) substituent (compound 11) results in an IC_{50} an order of magnitude better than an isobutyl substituent (compound 12) or a benzoate substituent (compound 13) in the same position. These results, although limited, might indicate that a hydrophobic aromatic ring in this position confers better activity compared to a small hydrophobic substituent or a polar aromatic group as in compounds 12 and 13 respectively. It will be necessary to test further analogues of compound 11 to confirm or rebuke this observation.

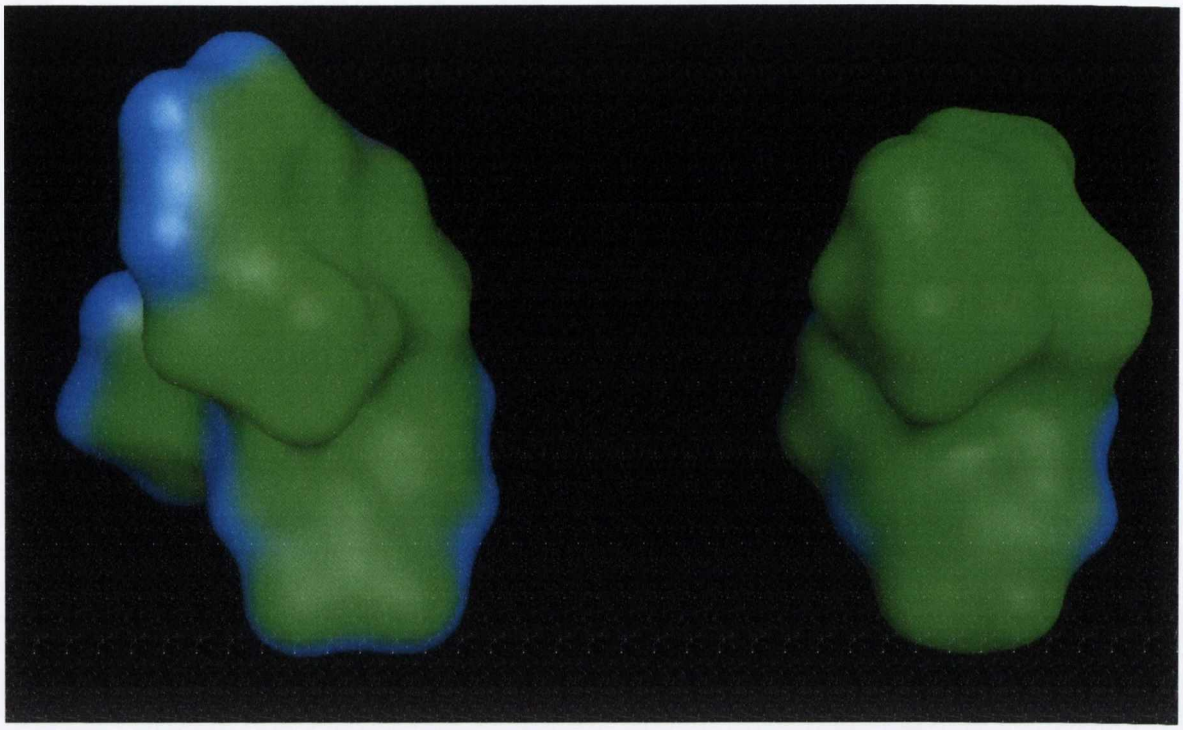
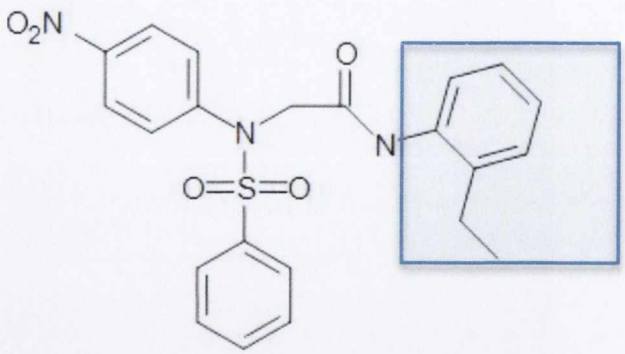


FIGURE 5.8: Molecular Surfaces of APM (right) and Compound 11. The green colour represents areas of high hydrophobicity whilst blue areas denote mild polar regions.

IC_{50}s (μM)

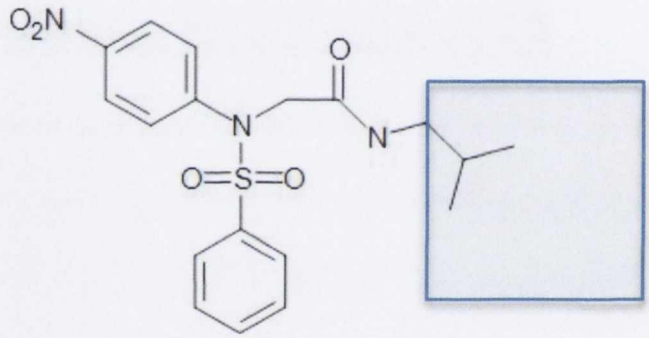
T *brucei* = 4.6

P *falciparum* = 5.45



T *brucei* = -

P *falciparum* = 59.4



T *brucei* = -

P *falciparum* = 58.9

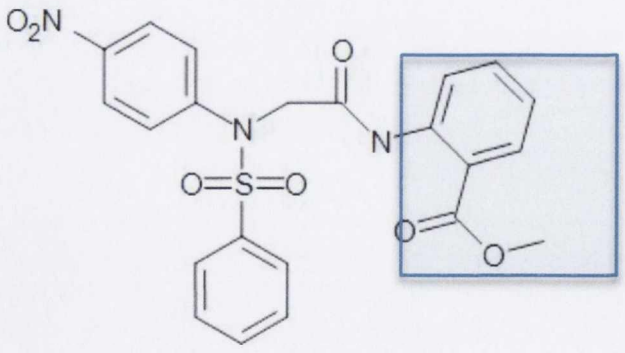


FIGURE 5.9: Biological Activities and SAR of Compounds 11 (top), 12 (middle) and 13 (bottom). The boxed regions identify the the substitutions between the three compounds.

Chapter 6

General Discussion

Approximately one billion people suffer from neglected diseases worldwide, roughly one-sixth of the entire population of the world. These diseases flourish in the poorest regions of the world and are often endemic in remote areas where vectors can thrive and infections go untreated due to a lack of hospitals, medical equipment, and the availability of cheap drugs. Current chemotherapy for both malaria and trypanosomiasis have been repeatedly shown to be unsafe, too expensive, and worst of all, increasingly marred through mounting levels of drug-resistance amongst parasites. Between the years 1975 to 1999, roughly one in 100 drugs that reached the market was for the treatment of a neglected disease, a situation thought to have been brought about by the lack of commercial incentives to develop new drugs for use in the developing world.

Of the drugs available to treat *P. faciparum*, nearly all have unwanted side effects. Mefloquine has been associated with causing anxiety disorders, psychosis, toxic encephalopathy and cardiovascular problems. The drug primaquine has been recorded as causing anorexia, vomiting, and anaemia while artemisinin has been shown to cause abdominal bleeding and vomiting. Halofantrine has been associated with causing cardiac arrhythmias. With respect to chemotherapy for Trypanosomiasis, drugs such as suramin, pentamidine, melarsoprol and eflornithine have long been linked with debilitating side effects. Suramin has been shown to cause nausea, vomiting and loss of consciousness, while pentamidine is associated with increased chest pain, hypotension and nausea. Melarsoprol is a highly toxic drug that causes convulsions, fever, and myocardial damage. The drug eflornithine has been linked

with suppression of bone marrow and anaemia. Along with increasing instances of drug resistance for both these diseases, it is clear that the identification of new, potent, low-cost drugs is of immediate importance.

In the work presented several computational approaches were undertaken aiming to identify the position of a parasite-selective, herbicide binding site on the tubulin protein, followed by the ligand-based design of novel antipararasitic compounds based on these herbicide chemical scaffolds. In Chapter 2 we developed homology models of the malarial and trypanosomal tubulin heterodimers based on the 3.70Å X-ray structure of porcine tubulin, 1TUB. In the context of our starting objectives, we have achieved what we set out to do to the extent that we performed a thorough computational analysis of the potential herbicide-binding sites on parasitic tubulin in Chapter 3, and in the process have identified a novel, previously unidentified putative binding site. As previously discussed, the validity of this binding site can ultimately only be established through further experimental work. The results also cast light on previous publications in the field. The current accepted binding site for the herbicide compounds is known as the 'Morrissette' site and incorrectly labelled as the N-loop, was shown to be one of the poorer pockets in our parasite tubulin models according to binding free energy calculations, with the ligands eventually migrating out of the pocket. This is unsurprising when one considers that when Morrissette *et al.* generated their *T. gondii* homology model they modeled a missing loop region in α-tubulin on the corresponding region in β-tubulin. However, analysis of these regions in several recently published tubulin X-ray structures revealed that β-tubulin possesses a nine residue α-helix whereas -tubulin contains a much shorter three residue helix. The addition of this larger helix in the α-subunit of the *T. gondii* homology model may have served to introduce a larger binding site in this region, directly influencing the results of the docking analysis carried out by Morrissette *et al.*

A further aim of the project was to employ ligand-based virtual screening for the identification of novel antiparasitic compounds while moving away from the herbicide chemical scaffolds. In Chapter 4 a tiered virtual screening approach utilising a pharmacophore filter followed by a ROCS shape-based re-ranking was developed for virtual screening of virtual libraries. Pharmacophore searching on its own proved unable to discriminate active compounds from decoy compounds to a high

enough degree during the pharmacophore validation and so an additional ligand-based screening tool was used to re-screen the pharmacophore hits. This second step, a ROCS shape-based re-ranking, improved the ranking of active compounds by over 10% during the validation step, a large increase on pharmacophore screening alone. This tiered-screening approach was applied to the Maybridge vendor database and 15 compounds were identified for biological characterisation in parasitic and mammalian cell models in Chapter 5. Of the 15 compounds tested, one compound, compound 11, inhibited the proliferation of *P. falciparum* and *T. brucei* parasites *in vitro* with IC_{50}s of 5.45μM and 4.6μM respectively. Compounds 5, 8, 10, 11, and 14 also displayed significant ability to inhibit the survival of *T brucei* parasites *in vitro*. Compound 3 was the most active of the compounds tested against *P. falciparum* with a calculated IC_{50} of 1.58μM. Cell proliferation assays using HEK293 and HeLa cell lines showed that none of the 15 compounds were toxic to mammalian cells and thermal melt experiments on purified bovine tubulin gave no evidence of compounds 3, 5, 8, 10, 11, or 14 interacting with the protein. None of these compounds have been previously described in the literature and so represent the opening of 6 new avenues of anti-malarial small molecule drug discovery.

The originality in this study is represented by the first molecular dynamics approach of more that 2.5ns towards finding a putative binding site on P. falciparum and *T. brucei* tubulin for the dinitroaniline and phosphorothioamidate compounds. While simulations up to 2ns are generally valid, in our own approaches we have observed that equilibration times often take longer than 3-5ns and the MM-PBSA-derived binding free energies generally stabilise after 4ns also. Failure to take into account a sufficiently long simulation time may result in misleading figures. The second point of originality in this study is the first instance of computational ligand-based drug design based on the phosphorothioamidate chemical class. No previous studies have attempted to move away from the herbicide scaffold to a more 'drug-like' chemical class to date, an approach taken in this work with successful results.

Future work related to this current study will follow the same pattern, namely structure-based and ligand-based routes. With respect to the structure-based side

it will be beneficial to re-make homology models for the parasitic tubulin heterodimers using the recently published tubulin crystal structures 4F6R and 4F61. These structures are completely resolved and contain the loop region that had previously been untenable to homology modeling due to its absence in experimentally resolved structures. These new models will remove the bias present in all previous attempts at modeling parasitic tubulin, including the bias present in the current study. Further to this it would be interesting to perform unguided molecular dynamics docking simulations to shed additional light on the putative ligand-binding process. In these simulations the protein is placed in a water box and a ligand, or ligands, of interest is placed is placed in a random location in the water and multiple simulations are then run over a sufficiently long timescales. This technique assumes no prior knowledge of the ligand-binding site and instead lets the ligand move around the protein until it finds a suitable pocket. This work is true blind docking, allowing the ligand to migrate to its preferred binding pocket. Multiple simulations, perhaps in the 100s and with timescales above 50ns, would be amenable to statistical and cluster analysis and allow us to say with a greater degree of certainty than the current study where the most energetically favourable putative ligand-binding pocket resides on the protein.

The data produced from the ligand-based biological validation will allow further ligand-based design to be executed. Similarity searches and partial ligand substructure searches using the most active compounds identified in **Chapter 5** have identified a number of compounds present in commercial databases whose biological characterisation will allow us to expand our knowledge of why the compounds exhibit antiparasitic activity. A detailed SAR investigation will allow us take this biological data and use it to design more potent inhibitors. Coupled to the ligand-based design approaches it is important to determine the mechanism of action of the newly identified compounds. It will be necessary to establish whether these compounds are causing cell death through interfering with microtubule polymerisation or through another mechanism completely.

Bibliography

[1] P. Bhargava and R. Singh, "Developments in diagnosis and antileishmanial drugs," *Interdiscip Perspect Infect Dis*, vol. 2012, p. 626838, 2012.

[2] R. Brun, R. Don, R. T. Jacobs, M. Z. Wang, and M. P. Barrett, "Development of novel drugs for human african trypanosomiasis," *Future Microbiol*, vol. 6, pp. 677–91, Jun 2011.

[3] A. A. Siddiqui, B. A. Siddiqui, and L. Ganley-Leal, "Schistosomiasis vaccines," *Hum Vaccin*, vol. 7, pp. 1192–7, Nov 2011.

[4] D. S. Nsagha, E. A. Bamgboye, J. C. N. Assob, A. L. Njunda, H. L. F. Kamga, A.-C. Zoung-Kanyi Bissek, E. N. Tabah, A. B. O. O. Oyediran, and A. K. Njamnshi, "Elimination of leprosy as a public health problem by 2000 ad: an epidemiological perspective," *Pan Afr Med J*, vol. 9, p. 4, 2011.

[5] N. R. de Silva, S. Brooker, P. J. Hotez, A. Montresor, D. Engels, and L. Savioli, "Soil-transmitted helminth infections: updating the global picture," *Trends Parasitol*, vol. 19, pp. 547–51, Dec 2003.

[6] K. Schautteet, E. De Clercq, and D. Vanrompay, "Chlamydia trachomatis vaccine research through the years," *Infect Dis Obstet Gynecol*, vol. 2011, p. 963513, 2011.

[7] P. Trouiller, P. Olliaro, E. Torreele, J. Orbinski, R. Laing, and N. Ford, "Drug development for neglected diseases: a deficient market and a public-health policy failure," *Lancet*, vol. 359, pp. 2188–94, Jun 2002.

[8] B. Stirner, "Stimulating research and development of pharmaceutical products for neglected diseases," *Eur J Health Law*, vol. 15, pp. 391–409, Dec 2008.

[9] R. I. Chima, C. A. Goodman, and A. Mills, "The economic impact of malaria in africa: a critical review of the evidence," *Health Policy*, vol. 63, pp. 17–36, Jan 2003.

[10] J. Sachs and P. Malaney, "The economic and social burden of malaria," *Nature*, vol. 415, pp. 680–5, Feb 2002.

[11] R. Aguilar, Y. Dong, E. Warr, and G. Dimopoulos, "Anopheles infection responses; laboratory models versus field malaria transmission systems," *Acta Trop*, vol. 95, pp. 285–91, Sep 2005.

[12] W. H. Organization, *World Malaria Report 2008*, 2008.

[13] J. C. Beier, "Malaria parasite development in mosquitoes," *Annu Rev Entomol*, vol. 43, pp. 519–43, 1998.

[14] P. Baldacci and R. Ménard, "The elusive malaria sporozoite in the mammalian host," *Mol Microbiol*, vol. 54, pp. 298–306, Oct 2004.

[15] J. W. Barnwell, "Vesicle-mediated transport of membrane and proteins in malaria-infected erythrocytes," *Blood Cells*, vol. 16, no. 2-3, pp. 379–95, 1990.

[16] I. A. Clark and W. B. Cowden, "The pathophysiology of falciparum malaria," *Pharmacol Ther*, vol. 99, pp. 221–60, Aug 2003.

[17] A. Sturm, S. Graewe, B. Franke-Fayard, S. Retzlaff, S. Bolte, B. Roppenser, M. Aepfelbacher, C. Janse, and V. Heussler, "Alteration of the parasite plasma membrane and the parasitophorous vacuole membrane during exo-erythrocytic development of malaria parasites," *Protist*, vol. 160, pp. 51–63, Feb 2009.

[18] R. R. Stanway, T. Witt, B. Zobiak, M. Aepfelbacher, and V. T. Heussler, "Gfp-targeting allows visualization of the apicoplast throughout the life cycle of live malaria parasites," *Biol Cell*, vol. 101, pp. 415–30, 5 p following 430, Jul 2009.

[19] L. L. Hsiao, R. J. Howard, M. Aikawa, and T. F. Taraschi, "Modification of host cell membrane lipid composition by the intra-erythrocytic human

malaria parasite plasmodium falciparum," *Biochem J*, vol. 274 (Pt 1), pp. 121–32, Feb 1991.

[20] S. D. Srinivas and S. K. Puri, "Time course of in vitro maturation of intra-erythrocytic malaria parasite: a comparison between plasmodium falciparum and plasmodium knowlesi," *Mem Inst Oswaldo Cruz*, vol. 97, pp. 901–3, Sep 2002.

[21] D. E. Arnot, E. Ronander, and D. C. Bengtsson, "The progression of the intra-erythrocytic cell cycle of plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division during schizogony," *Int J Parasitol*, vol. 41, pp. 71–80, Jan 2011.

[22] E. S. Bergmann-Leitner, E. H. Duncan, and E. Angov, "Msp-1p42-specific antibodies affect growth and development of intra-erythrocytic parasites of plasmodium falciparum," *Malar J*, vol. 8, p. 183, 2009.

[23] D. E. Arnot and K. Gull, "The plasmodium cell-cycle: facts and questions," *Ann Trop Med Parasitol*, vol. 92, pp. 361–5, Jun 1998.

[24] D. A. Baker, "Malaria gametocytogenesis," *Mol Biochem Parasitol*, vol. 172, pp. 57–65, Aug 2010.

[25] D. F. Wirth, "The parasite genome: Biological revelations," *Nature*, vol. 419, pp. 495–496, October 2002.

[26] M. J. Mackinnon and K. Marsh, "The selection landscape of malaria parasites," *Science*, vol. 328, pp. 866–71, May 2010.

[27] M. Foley and L. Tilley, "Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents," *Pharmacol Ther*, vol. 79, pp. 55–87, Jul 1998.

[28] V. R. Solomon and H. Lee, "Chloroquine and its analogs: a new promise of an old drug for effective and safe cancer therapies," *Eur J Pharmacol*, vol. 625, pp. 220–33, Dec 2009.

[29] A. F. Slater, "Chloroquine: mechanism of drug action and resistance in plasmodium falciparum," *Pharmacol Ther*, vol. 57, no. 2-3, pp. 203–35, 1993.

[30] A. F. Slater and A. Cerami, "Inhibition by chloroquine of a novel haem polymerase enzyme activity in malaria trophozoites," *Nature*, vol. 355, pp. 167–9, Jan 1992.

[31] H. Ginsburg, O. Famin, J. Zhang, and M. Krugliak, "Inhibition of glutathione-dependent degradation of heme by chloroquine and amodiaquine as a possible basis for their antimalarial mode of action," *Biochem Pharmacol*, vol. 56, pp. 1305–13, Nov 1998.

[32] S. Bhattacharyya, K. Solakyildirim, Z. Zhang, R. J. Linhardt, and J. K. Tobacman, "Chloroquine reduces arylsulphatase b activity and increases chondroitin-4-sulphate: implications for mechanisms of action and resistance," *Malar J*, vol. 8, p. 303, 2009.

[33] P. Preechapornkul, M. Imwong, K. Chotivanich, W. Pongtavornpinyo, A. M. Dondorp, N. P. J. Day, N. J. White, and S. Pukrittayakamee, "Plasmodium falciparum pfmdr1 amplification, mefloquine resistance, and parasite fitness," *Antimicrob Agents Chemother*, vol. 53, pp. 1509–15, Apr 2009.

[34] P. Schlagenhauf, M. Adamcova, L. Regep, M. T. Schaerer, and H.-G. Rhein, "The position of mefloquine as a 21st century malaria chemoprophylaxis," *Malar J*, vol. 9, p. 357, 2010.

[35] T. Weinke, M. Trautmann, T. Held, G. Weber, D. Eichenlaub, K. Fleischer, W. Kern, and H. D. Pohle, "Neuropsychiatric side effects after the use of mefloquine," *Am J Trop Med Hyg*, vol. 45, pp. 86–91, Jul 1991.

[36] D. Stürchler, J. Handschin, D. Kaiser, L. Kerr, M. L. Mittelholzer, R. Reber, and M. Fernex, "Neuropsychiatric side effects of mefloquine," *N Engl J Med*, vol. 322, pp. 1752–3, Jun 1990.

[37] A. Sowunmi, L. A. Salako, A. M. Oduola, O. Walker, J. A. Akindele, and O. A. Ogundahunsi, "Neuropsychiatric side effects of mefloquine in africans," *Trans R Soc Trop Med Hyg*, vol. 87, no. 4, pp. 462–3, 1993.

[38] J. Karbwang and K. Na Bangchang, "Clinical pharmacokinetics of halofantrine," *Clin Pharmacokinet*, vol. 27, pp. 104–19, Aug 1994.

[39] J. E. Touze, L. Fourcade, F. Peyron, P. Heno, and J. C. Deharo, "Is halofantrine still advisable in malaria attacks?," *Ann Trop Med Parasitol*, vol. 91, pp. 867–73, Oct 1997.

[40] J. E. Touze, P. Heno, L. Fourcade, J. C. Deharo, G. Thomas, S. Bohan, P. Paule, P. Riviere, E. Kouassi, and A. Buguet, "The effects of antimalarial drugs on ventricular repolarization," *Am J Trop Med Hyg*, vol. 67, pp. 54–60, Jul 2002.

[41] F. O. ter Kuile, G. Dolan, F. Nosten, M. D. Edstein, C. Luxemburger, L. Phaipun, T. Chongsuphajaisiddhi, H. K. Webster, and N. J. White, "Halofantrine versus mefloquine in treatment of multidrug-resistant falciparum malaria," *Lancet*, vol. 341, pp. 1044–9, Apr 1993.

[42] O. Bouchaud, P. Imbert, J. E. Touze, A. N. O. Dodoo, M. Danis, and F. Legros, "Fatal cardiotoxicity related to halofantrine: a review based on a worldwide safety data base," *Malar J*, vol. 8, p. 289, 2009.

[43] G. O. Adjei, B. Q. Goka, F. Binka, and J. A. Kurtzhals, "Artemether-lumefantrine: an oral antimalarial for uncomplicated malaria in children," *Expert Rev Anti Infect Ther*, vol. 7, pp. 669–81, Aug 2009.

[44] G. Kokwaro, L. Mwai, and A. Nzila, "Artemether/lumefantrine in the treatment of uncomplicated falciparum malaria," *Expert Opin Pharmacother*, vol. 8, pp. 75–94, Jan 2007.

[45] M. Makanga and S. Krudsood, "The clinical efficacy of artemether/lumefantrine (coartem)," *Malar J*, vol. 8 Suppl 1, p. S5, 2009.

[46] A. Djimdé and G. Lefèvre, "Understanding the pharmacokinetics of coartem," *Malar J*, vol. 8 Suppl 1, p. S4, 2009.

[47] N. Vale, R. Moreira, and P. Gomes, "Primaquine revisited six decades after its discovery," *Eur J Med Chem*, vol. 44, pp. 937–53, Mar 2009.

[48] K. Kaur, M. Jain, R. P. Reddy, and R. Jain, "Quinolines and structurally related heterocycles as antimalarials," *Eur J Med Chem*, vol. 45, pp. 3245–64, Aug 2010.

[49] L. Carvalho, J. R. Luque-Ortega, J. I. Manzano, S. Castanys, L. Rivas, and F. Gamarro, "Tafenoquine, an antiplasmodial 8-aminoquinoline, targets leishmania respiratory complex iii and induces apoptosis," *Antimicrob Agents Chemother*, vol. 54, pp. 5344–51, Dec 2010.

[50] M. Crockett and K. C. Kain, "Tafenoquine: a promising new antimalarial agent," *Expert Opin Investig Drugs*, vol. 16, pp. 705–15, May 2007.

[51] D. S. Walsh, C. Eamsila, T. Sasiprapha, S. Sangkharomya, P. Khaewsathien, P. Supakalin, D. B. Tang, P. Jarasrumgsichol, C. Cherdchu, M. D. Edstein, K. H. Rieckmann, and T. G. Brewer, "Efficacy of monthly tafenoquine for prophylaxis of plasmodium vivax and multidrug-resistant p. falciparum malaria," *J Infect Dis*, vol. 190, pp. 1456–63, Oct 2004.

[52] M. Schlitzer, "Malaria chemotherapeutics part i: History of antimalarial drug development, currently used therapeutics, and drugs in clinical development," *ChemMedChem*, vol. 2, pp. 944–86, Jul 2007.

[53] D. Rathore, T. F. McCutchan, M. Sullivan, and S. Kumar, "Antimalarial drugs: current status and new developments," *Expert Opin Investig Drugs*, vol. 14, pp. 871–83, Jul 2005.

[54] U. Eckstein-Ludwig, R. J. Webb, I. D. A. Van Goethem, J. M. East, A. G. Lee, M. Kimura, P. M. O'Neill, P. G. Bray, S. A. Ward, and S. Krishna, "Artemisinins target the serca of plasmodium falciparum," *Nature*, vol. 424, pp. 957–61, Aug 2003.

[55] M. A. Travassos and M. K. Laufer, "Resistance to antimalarial drugs: molecular, pharmacologic, and clinical considerations," *Pediatr Res*, vol. 65, pp. 64R–70R, May 2009.

[56] I. T. Jovel, R. E. Mejía, E. Banegas, R. Piedade, J. Alger, G. Fontecha, P. E. Ferreira, M. I. Veiga, I. G. Enamorado, A. Bjorkman, and J. Ursing, "Drug resistance associated genetic polymorphisms in plasmodium falciparum and plasmodium vivax collected in honduras, central america," *Malar J*, vol. 10, p. 376, 2011.

[57] D. J. Johnson, D. A. Fidock, M. Mungthin, V. Lakshmanan, A. B. S. Sidhu, P. G. Bray, and S. A. Ward, "Evidence for a central role for pfcrt in conferring

plasmodium falciparum resistance to diverse antimalarial agents," *Mol Cell*, vol. 15, pp. 867–77, Sep 2004.

[58] D. J. Johnson, A. Owen, N. Plant, P. G. Bray, and S. A. Ward, "Drug-regulated expression of plasmodium falciparum p-glycoprotein homologue 1: a putative role for nuclear receptors," *Antimicrob Agents Chemother*, vol. 52, pp. 1438–45, Apr 2008.

[59] M. L. Ibrahim, N. Khim, H. H. Adam, F. Ariey, and J.-B. Duchemin, "Polymorphism of pfatpase in niger: detection of three new point mutations," *Malar J*, vol. 8, p. 28, 2009.

[60] R. Jambou, E. Legrand, M. Niang, N. Khim, P. Lim, B. Volney, M. T. Ekala, C. Bouchier, P. Esterre, T. Fandeur, and O. Mercereau-Puijalon, "Resistance of plasmodium falciparum field isolates to in-vitro artemether and point mutations of the serca-type pfatpase6," *Lancet*, vol. 366, pp. 1960–3, Dec 2005.

[61] K. Stuart, R. Brun, S. Croft, A. Fairlamb, R. E. Gürtler, J. McKerrow, S. Reed, and R. Tarleton, "Kinetoplastids: related protozoan pathogens, different diseases," *J Clin Invest*, vol. 118, pp. 1301–10, Apr 2008.

[62] V. Jamonneau, H. Ilboudo, J. Kaboré, D. Kaba, M. Koffi, P. Solano, A. Garcia, D. Courtin, C. Laveissière, K. Lingue, P. Büscher, and B. Bucheton, "Untreated human infections by trypanosoma brucei gambiense are not 100% fatal," *PLoS Negl Trop Dis*, vol. 6, no. 6, p. e1691, 2012.

[63] W. H. Organization, "Trypanosomiasis, human african (sleeping sickness)."

[64] WHO, "First who report on neglected tropical diseases," 2010.

[65] K. R. Matthews, "The developmental cell biology of trypanosoma brucei," *J Cell Sci*, vol. 118, pp. 283–90, Jan 2005.

[66] K. R. Matthews, J. R. Ellis, and A. Paterou, "Molecular regulation of the life cycle of african trypanosomes," *Trends Parasitol*, vol. 20, pp. 40–7, Jan 2004.

[67] K. Vickerman, "Developmental cycles and biology of pathogenic trypanosomes," *Br Med Bull*, vol. 41, pp. 105–14, Apr 1985.

[68] T. E. Voogd, E. L. Vansterkenburg, J. Wilting, and L. H. Janssen, "Recent research on the biological activity of suramin," *Pharmacol Rev*, vol. 45, pp. 177–203, Jun 1993.

[69] M. P. Barrett, D. W. Boykin, R. Brun, and R. R. Tidwell, "Human african trypanosomiasis: pharmacological re-engagement with a neglected disease," *Br J Pharmacol*, vol. 152, pp. 1155–71, Dec 2007.

[70] F. Hawking, "Concentration of bayer 205 (germanin) in human blood and cerebrospinal fluid after treatment," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 34, pp. 37–52, June 1940.

[71] J. M. Collins, R. W. Klecker, Jr, R. Yarchoan, H. C. Lane, A. S. Fauci, R. R. Redfield, S. Broder, and C. E. Myers, "Clinical pharmacokinetics of suramin in patients with htlv-iii/lav infection," *J Clin Pharmacol*, vol. 26, pp. 22–6, Jan 1986.

[72] C. C. Wang, "Molecular mechanisms and therapeutic approaches to the treatment of african trypanosomiasis," *Annu Rev Pharmacol Toxicol*, vol. 35, pp. 93–127, 1995.

[73] M. Sands, M. A. Kron, and R. B. Brown, "Pentamidine: a review," *Rev Infect Dis*, vol. 7, no. 5, pp. 625–34, 1985.

[74] T. A. Shapiro, "Inhibition of topoisomerases in african trypanosomes," *Acta Trop*, vol. 54, pp. 251–60, Sep 1993.

[75] C. A. Lanteri, R. R. Tidwell, and S. R. Meshnick, "The mitochondrion is a site of trypanocidal action of the aromatic diamidine db75 in bloodstream forms of trypanosoma brucei," *Antimicrob Agents Chemother*, vol. 52, pp. 875–82, Mar 2008.

[76] A. J. Nok, "Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human african trypanosomiasis," *Parasitol Res*, vol. 90, pp. 71–9, May 2003.

[77] A. H. Fairlamb, G. B. Henderson, and A. Cerami, "Trypanothione is the primary target for arsenical drugs against african trypanosomes," *Proc Natl Acad Sci U S A*, vol. 86, pp. 2607–11, Apr 1989.

[78] J. Pépin and F. Milord, "The treatment of human african trypanosomiasis," *Adv Parasitol*, vol. 33, pp. 1–47, 1994.

[79] A. Moore and M. Richer, "Re-emergence of epidemic sleeping sickness in southern sudan," *Trop Med Int Health*, vol. 6, pp. 342–7, May 2001.

[80] D. Legros, C. Fournier, M. Gastellu Etchegorry, F. Maiso, and E. Szumilin, "[therapeutic failure of melarsoprol among patients treated for late stage t.b. gambiense human african trypanosomiasis in uganda]," *Bull Soc Pathol Exot*, vol. 92, pp. 171–2, Jul 1999.

[81] S. M. Landfear, B. Ullman, N. S. Carter, and M. A. Sanchez, "Nucleoside and nucleobase transporters in parasitic protozoa," *Eukaryot Cell*, vol. 3, pp. 245–54, Apr 2004.

[82] E. Matovu, F. Geiser, V. Schneider, P. Mäser, J. C. Enyaru, R. Kaminsky, S. Gallati, and T. Seebeck, "Genetic variants of the tbat1 adenosine transporter from african trypanosomes in relapse infections following melarsoprol therapy," *Mol Biochem Parasitol*, vol. 117, pp. 73–81, Sep 2001.

[83] N. Baker, L. Glover, J. C. Munday, D. Aguinaga Andrés, M. P. Barrett, H. P. de Koning, and D. Horn, "Aquaglyceroporin 2 controls susceptibility to melarsoprol and pentamidine in african trypanosomes," *Proc Natl Acad Sci U S A*, vol. 109, pp. 10996–1001, Jul 2012.

[84] B. S. Hall, C. Bot, and S. R. Wilkinson, "Nifurtimox activation by trypanosomal type i nitroreductases generates cytotoxic nitrile metabolites," *J Biol Chem*, vol. 286, pp. 13088–95, Apr 2011.

[85] S. R. Wilkinson, M. C. Taylor, D. Horn, J. M. Kelly, and I. Cheeseman, "A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes," *Proc Natl Acad Sci U S A*, vol. 105, pp. 5022–7, Apr 2008.

[86] R. Docampo and A. O. Stoppani, "Generation of superoxide anion and hydrogen peroxide induced by nifurtimox in trypanosoma cruzi," *Arch Biochem Biophys*, vol. 197, pp. 317–21, Oct 1979.

[87] R. Docampo, S. N. Moreno, A. O. Stoppani, W. Leon, F. S. Cruz, F. Villalta, and R. F. Muniz, "Mechanism of nifurtimox toxicity in different forms of trypanosoma cruzi," *Biochem Pharmacol*, vol. 30, pp. 1947–51, Jul 1981.

[88] I. M. Vincent, D. J. Creek, K. Burgess, D. J. Woods, R. J. S. Burchmore, and M. P. Barrett, "Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against trypanosoma brucei," *PLoS Negl Trop Dis*, vol. 6, no. 5, p. e1618, 2012.

[89] G. Priotto, L. Pinoges, I. B. Fursa, B. Burke, N. Nicolay, G. Grillet, C. Hewison, and M. Balasegaram, "Safety and effectiveness of first line eflornithine for trypanosoma brucei gambiense sleeping sickness in sudan: cohort study," *BMJ*, vol. 336, pp. 705–8, Mar 2008.

[90] N. V. Grishin, A. L. Osterman, H. B. Brooks, M. A. Phillips, and E. J. Goldsmith, "X-ray structure of ornithine decarboxylase from trypanosoma brucei: the native structure and the structure in complex with alpha-difluoromethylornithine," *Biochemistry*, vol. 38, pp. 15174–84, Nov 1999.

[91] M. Iten, H. Mett, A. Evans, J. C. Enyaru, R. Brun, and R. Kaminsky, "Alterations in ornithine decarboxylase characteristics account for tolerance of trypanosoma brucei rhodesiense to d,l-alpha-difluoromethylornithine," *Antimicrob Agents Chemother*, vol. 41, pp. 1922–5, Sep 1997.

[92] B. J. Berger, N. S. Carter, and A. H. Fairlamb, "Characterisation of pentamidine-resistant trypanosoma brucei brucei," *Mol Biochem Parasitol*, vol. 69, pp. 289–98, Feb 1995.

[93] D. J. Bridges, M. K. Gould, B. Nerima, P. Mäser, R. J. S. Burchmore, and H. P. de Koning, "Loss of the high-affinity pentamidine transporter is responsible for high levels of cross-resistance between arsenical and diamidine drugs in african trypanosomes," *Mol Pharmacol*, vol. 71, pp. 1098–108, Apr 2007.

[94] K. A. Werbovetz, "Tubulin as an antiprotozoal drug target," *Mini Rev Med Chem*, vol. 2, pp. 519–29, Dec 2002.

[95] E. Nogales, "Structural insights into microtubule function," *Annu Rev Biochem*, vol. 69, pp. 277–302, 2000.

[96] C. M. Waterman-Storer and E. D. Salmon, "Microtubule dynamics: treadmilling comes around again," *Curr Biol*, vol. 7, pp. R369–72, Jun 1997.

[97] A. L. Risinger, F. J. Giles, and S. L. Mooberry, "Microtubule dynamics as a target in oncology," *Cancer Treat Rev*, vol. 35, pp. 255–61, May 2009.

[98] O. Valiron, N. Caudron, and D. Job, "Microtubule dynamics," *Cell Mol Life Sci*, vol. 58, pp. 2069–84, Dec 2001.

[99] R. H. Wade, "On and around microtubules: an overview," *Mol Biotechnol*, vol. 43, pp. 177–91, Oct 2009.

[100] H. Y. Kueh and T. J. Mitchison, "Structural plasticity in actin and tubulin polymer dynamics," *Science*, vol. 325, pp. 960–3, Aug 2009.

[101] C. Conde and A. Cáceres, "Microtubule assembly, organization and dynamics in axons and dendrites," *Nat Rev Neurosci*, vol. 10, pp. 319–32, May 2009.

[102] R. B. G. Ravelli, B. Gigant, P. A. Curmi, I. Jourdain, S. Lachkar, A. Sobel, and M. Knossow, "Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain," *Nature*, vol. 428, pp. 198–202, Mar 2004.

[103] P. A. Friedman and E. G. Platzer, "Interaction of anthelmintic benzimidazoles and benzimidazole derivatives with bovine brain tubulin," *Biochim Biophys Acta*, vol. 544, pp. 605–14, Dec 1978.

[104] C. Rappl, P. Barbier, V. Bourgarel-Rey, C. Grégoire, R. Gilli, M. Carre, S. Combes, J.-P. Finet, and V. Peyrot, "Interaction of 4-arylcoumarin analogues of combretastatins with microtubule network of hbl100 cells and binding to tubulin," *Biochemistry*, vol. 45, pp. 9210–8, Aug 2006.

[105] F. Bellina, S. Cauteruccio, S. Monti, and R. Rossi, "Novel imidazole-based combretastatin a-4 analogues: evaluation of their in vitro antitumor activity and molecular modeling study of their binding to the colchicine site of tubulin," *Bioorg Med Chem Lett*, vol. 16, pp. 5757–62, Nov 2006.

[106] J. Griggs, J. C. Metcalfe, and R. Hesketh, "Targeting tumour vasculature: the development of combretastatin a4," *Lancet Oncol*, vol. 2, pp. 82–7, Feb 2001.

[107] A. Cirla and J. Mann, "Combretastatins: from natural products to drug discovery," *Nat Prod Rep*, vol. 20, pp. 558–64, Dec 2003.

[108] B. Gigant, C. Wang, R. B. G. Ravelli, F. Roussi, M. O. Steinmetz, P. A. Curmi, A. Sobel, and M. Knossow, "Structural basis for the regulation of tubulin by vinblastine," *Nature*, vol. 435, pp. 519–22, May 2005.

[109] A. Cormier, M. Marchand, R. B. G. Ravelli, M. Knossow, and B. Gigant, "Structural insight into the inhibition of tubulin by vinca domain peptide ligands," *EMBO Rep*, vol. 9, pp. 1101–6, Nov 2008.

[110] C. Dumontet and M. A. Jordan, "Microtubule-binding agents: a dynamic field of cancer therapeutics," *Nat Rev Drug Discov*, vol. 9, pp. 790–803, Oct 2010.

[111] J. F. Díaz, I. Barasoain, and J. M. Andreu, "Fast kinetics of taxol binding to microtubules. effects of solution variables and microtubule-associated proteins," *J Biol Chem*, vol. 278, pp. 8407–19, Mar 2003.

[112] J. L. Ross and D. K. Fygenson, "Mobility of taxol in microtubule bundles," *Biophys J*, vol. 84, pp. 3959–67, Jun 2003.

[113] M. Kikkawa and N. Hirokawa, "High-resolution cryo-em maps show the nucleotide binding pocket of kif1a in open and closed conformations," *EMBO J*, vol. 25, pp. 4187–94, Sep 2006.

[114] G. Bhattacharya, M. M. Salem, and K. A. Werbovetz, "Antileishmanial dinitroaniline sulfonamides with activity against parasite tubulin," *Bioorg Med Chem Lett*, vol. 12, pp. 2395–8, Sep 2002.

[115] Y. M. Traub-Cseko, J. M. Ramalho-Ortigão, A. P. Dantas, S. L. de Castro, H. S. Barbosa, and K. H. Downing, "Dinitroaniline herbicides against protozoan parasites: the case of trypanosoma cruzi," *Trends Parasitol*, vol. 17, pp. 136–41, Mar 2001.

[116] T. G. George, M. M. Endeshaw, R. E. Morgan, K. V. Mahasenan, D. A. Delfín, M. S. Mukherjee, A. J. Yakovich, J. Fotie, C. Li, and K. A. Werbovetz, "Synthesis, biological evaluation, and molecular modeling of

3,5-substituted-n1-phenyl-n4,n4-di-n-butylsulfanilamides as antikinetoplastid antimicrotubule agents," *Bioorg Med Chem*, vol. 15, pp. 6071–9, Sep 2007.

[117] B. J. Fennell, J. A. Naughton, E. Dempsey, and A. Bell, "Cellular and molecular actions of dinitroaniline and phosphorothioamidate herbicides on plasmodium falciparum: tubulin as a specific antimalarial target," *Mol Biochem Parasitol*, vol. 145, pp. 226–38, Feb 2006.

[118] G. S. Dow, A. Armson, M. R. Boddy, T. Itenge, D. McCarthy, J. E. Parkin, R. C. A. Thompson, and J. A. Reynoldson, "Plasmodium: assessment of the antimalarial potential of trifluralin and related compounds using a rat model of malaria, rattus norvegicus," *Exp Parasitol*, vol. 100, pp. 155–60, Mar 2002.

[119] J. A. Naughton, R. Hughes, P. Bray, and A. Bell, "Accumulation of the antimalarial microtubule inhibitors trifluralin and vinblastine by plasmodium falciparum," *Biochem Pharmacol*, vol. 75, pp. 1580–7, Apr 2008.

[120] C. Délye, Y. Menchari, S. Michel, and H. Darmency, "Molecular bases for sensitivity to tubulin-binding herbicides in green foxtail," *Plant Physiol*, vol. 136, pp. 3920–32, Dec 2004.

[121] Y. B. Blume, A. Y. Nyporko, A. I. Yemets, and W. V. Baird, "Structural modeling of the interaction of plant alpha-tubulin with dinitroaniline and phosphoroamidate herbicides," *Cell Biol Int*, vol. 27, no. 3, pp. 171–4, 2003.

[122] N. S. Morrissette, A. Mitra, D. Sept, and L. D. Sibley, "Dinitroanilines bind alpha-tubulin to disrupt microtubules," *Mol Biol Cell*, vol. 15, pp. 1960–8, Apr 2004.

[123] S. Lyons-Abbott, D. L. Sackett, D. Wloga, J. Gaertig, R. E. Morgan, K. A. Werbovetz, and N. S. Morrissette, "Îś-tubulin mutations alter oryzalin affinity and microtubule assembly properties to confer dinitroaniline resistance," *Eukaryot Cell*, vol. 9, pp. 1825–34, Dec 2010.

[124] J. V. Murthy, H. H. Kim, V. R. Hanesworth, J. D. Hugdahl, and L. C. Morejohn, "Competitive inhibition of high-affinity oryzalin binding to plant

tubulin by the phosphoric amide herbicide amiprophos-methyl," *Plant Physiol*, vol. 105, pp. 309–320, May 1994.

[125] B. J. Bogitsh, O. L. Middleton, and R. Ribeiro-Rodrigues, "Effects of the antitubulin drug trifluralin on the proliferation and metacyclogenesis of trypanosoma cruzi epimastigotes," *Parasitol Res*, vol. 85, pp. 475–80, Jun 1999.

[126] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, and C. Zardecki, "The protein data bank," *Acta Crystallogr D Biol Crystallogr*, vol. 58, pp. 899–907, Jun 2002.

[127] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using modeller," *Curr Protoc Protein Sci*, vol. Chapter 2, p. Unit 2.9, Nov 2007.

[128] W. R. Pearson and D. J. Lipman, "Improved tools for biological sequence comparison," *Proc Natl Acad Sci U S A*, vol. 85, pp. 2444–8, Apr 1988.

[129] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J Mol Biol*, vol. 215, pp. 403–10, Oct 1990.

[130] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J Mol Biol*, vol. 48, pp. 443–53, Mar 1970.

[131] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *J Mol Biol*, vol. 147, pp. 195–7, Mar 1981.

[132] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, and A. Sali, "Protein structure modeling with modeller," *Methods Mol Biol*, vol. 426, pp. 145–59, 2008.

[133] A. Sali and J. P. Overington, "Derivation of rules for comparative protein modeling from a database of protein structure alignments," *Protein Sci*, vol. 3, pp. 1582–96, Sep 1994.

[134] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187–217, 1983.

[135] August 2001.

[136] A. Fiser, R. K. Do, and A. Sali, "Modeling of loops in protein structures," *Protein Sci*, vol. 9, pp. 1753–73, Sep 2000.

[137] A. Fiser and A. Sali, "Modeller: generation and refinement of homology-based protein structure models," *Methods Enzymol*, vol. 374, pp. 461–91, 2003.

[138] D. S. M. R. A. Laskowski, M. W. MacArthur and J. M. Thornton, "Procheck: a program to check the stereochemical quality of protein structures," *Journal of Applied Crystalography*, vol. 26, pp. 283–291, 1993.

[139] G. N. RAMACHANDRAN, C. RAMAKRISHNAN, and V. SASISEKHARAN, "Stereochemistry of polypeptide chain configurations," *J Mol Biol*, vol. 7, pp. 95–9, Jul 1963.

[140] A. L. Morris, M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton, "Stereochemical quality of protein structure coordinates," *Proteins*, vol. 12, pp. 345–64, Apr 1992.

[141] R. A. E. R. Huber, "Accurate bond and angle parameters for x-ray protein structure refinement," *Acta Crystallographica*, vol. 47, pp. 392–400, 1991.

[142] N. Sreerama and R. W. Woody, "A self-consistent method for the analysis of protein secondary structure from circular dichroism," *Anal Biochem*, vol. 209, pp. 32–44, Feb 1993.

[143] September 1995.

[144] J. Cooper, "Side chain conformations," 1995.

[145] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, "Swiss-model: An automated protein homology-modeling server," *Nucleic Acids Res*, vol. 31, pp. 3381–5, Jul 2003.

[146] J. Löwe, H. Li, K. H. Downing, and E. Nogales, "Refined structure of alpha beta-tubulin at 3.5 a resolution," *J Mol Biol*, vol. 313, pp. 1045–57, Nov 2001.

[147] G. Vriend, "What if: a molecular modeling and drug design program," *Journal of Molecular Graphics*, vol. 8, pp. 52–56, March 1990.

[148] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The swiss-model workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, pp. 195–201, Jan 2006.

[149] "National center for biotechnology information."

[150] S. P. Holloway, P. F. Sims, C. J. Delves, J. G. Scaife, and J. E. Hyde, "Isolation of alpha-tubulin genes from the human malaria parasite, plasmodium falciparum: sequence analysis of alpha-tubulin," *Mol Microbiol*, vol. 3, pp. 1501–10, Nov 1989.

[151] K. Sen and G. N. Godson, "Isolation of alpha- and beta-tubulin genes of plasmodium falciparum using a single oligonucleotide probe," *Mol Biochem Parasitol*, vol. 39, pp. 173–82, Mar 1990.

[152] B. E. Kimmel, S. Samson, J. Wu, R. Hirschberg, and L. R. Yarbrough, "Tubulin genes of the african trypanosome trypanosoma brucei rhodesiense:nucleotide sequence of a 3.7-kb fragment containing genes for alpha and beta tubulins," *Gene*, vol. 35, no. 3, pp. 237–48, 1985.

[153] C. C. Group, *MOE v2011.10.* www.chemcomp.com, Montreal, 2011.

[154] B. Gigant, P. A. Curmi, C. Martin-Barbey, E. Charbaut, S. Lachkar, L. Lebeau, S. Siavoshian, A. Sobel, and M. Knossow, "The 4 a x-ray structure of a tubulin:stathmin-like domain complex," *Cell*, vol. 102, pp. 809–16, Sep 2000.

[155] J. H. Nettles, H. Li, B. Cornett, J. M. Krahn, J. P. Snyder, and K. H. Downing, "The binding mode of epothilone a on alpha,beta-tubulin by electron crystallography," *Science*, vol. 305, pp. 866–9, Aug 2004.

[156] E. Nogales, S. G. Wolf, and K. H. Downing, "Structure of the alpha beta tubulin dimer by electron crystallography," *Nature*, vol. 391, pp. 199–203, Jan 1998.

[157] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega," *Mol Syst Biol*, vol. 7, p. 539, 2011.

[158] R. Datta, I. Das, B. Sen, A. Chakraborty, S. Adak, C. Mandal, and A. K. Datta, "Homology-model-guided site-specific mutagenesis reveals the mechanisms of substrate binding and product-regulation of adenosine kinase from leishmania donovani," *Biochem J*, vol. 394, pp. 35–42, Feb 2006.

[159] C. de Graaf, C. Oostenbrink, P. H. J. Keizers, B. M. A. van Vugt-Lussenburg, R. A. B. van Waterschoot, R. A. Tschirret-Guth, J. N. M. Commandeur, and N. P. E. Vermeulen, "Molecular modeling-guided site-directed mutagenesis of cytochrome p450 2d6," *Curr Drug Metab*, vol. 8, pp. 59–77, Jan 2007.

[160] P. Traxler, P. Furet, H. Mett, E. Buchdunger, T. Meyer, and N. Lydon, "Design and synthesis of novel tyrosine kinase inhibitors using a pharmacophore model of the atp-binding site of the egf-r," *J Pharm Belg*, vol. 52, no. 2, pp. 88–96, 1997.

[161] S. Ghosh, X. P. Liu, Y. Zheng, and F. M. Uckun, "Rational design of potent and selective egfr tyrosine kinase inhibitors as anticancer agents," *Curr Cancer Drug Targets*, vol. 1, pp. 129–40, Aug 2001.

[162] J. A. M. Jacob D Durrant, "Molecular dynamics simulations and drug discovery," *BMC Biology*, vol. 9, no. 71, 2011.

[163] D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The amber biomolecular simulation programs," *J Comput Chem*, vol. 26, pp. 1668–88, Dec 2005.

[164] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J Comput Chem*, vol. 25, pp. 1157–74, Jul 2004.

[165] W. C. P. A. K. Christopher I. Bayly, Piotr Cieplak, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model," *Journal of Physical Chemistry*, vol. 97, pp. 10269–10280, October 1993.

[166] A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation," *J Comput Chem*, vol. 23, pp. 1623–41, Dec 2002.

[167] D. G. Levitt and L. J. Banaszak, "Pocket: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids," *J Mol Graph*, vol. 10, pp. 229–34, Dec 1992.

[168] M. Hendlich, F. Rippmann, and G. Barnickel, "Ligsite: automatic and efficient detection of potential small molecule-binding sites in proteins," *J Mol Graph Model*, vol. 15, pp. 359–63, 389, Dec 1997.

[169] G. P. Brady, Jr and P. F. Stouten, "Fast prediction and visualization of protein binding pockets with pass," *J Comput Aided Mol Des*, vol. 14, pp. 383–401, May 2000.

[170] P. J. Goodford, "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules," *J Med Chem*, vol. 28, pp. 849–57, Jul 1985.

[171] T. Halgren, "New method for fast and accurate binding-site identification and analysis," *Chem Biol Drug Des*, vol. 69, pp. 146–8, Feb 2007.

[172] P. Schmidtke, A. Bidon-Chanal, F. J. Luque, and X. Barril, "Mdpocket: open-source cavity detection and characterization on molecular dynamics trajectories," *Bioinformatics*, vol. 27, pp. 3276–85, Dec 2011.

[173] V. Le Guilloux, P. Schmidtke, and P. Tuffery, "Fpocket: an open source platform for ligand pocket detection," *BMC Bioinformatics*, vol. 10, p. 168, 2009.

[174] A. J. S. Knox, M. J. Meegan, G. Carta, and D. G. Lloyd, "Considerations in compound database preparation–"hidden" impact on virtual screening results," *J Chem Inf Model*, vol. 45, no. 6, pp. 1908–19, 2005.

[175] D. T. Moustakas, P. T. Lang, S. Pegg, E. Pettersen, I. D. Kuntz, N. Brooijmans, and R. C. Rizzo, "Development and validation of a modular, extensible docking program: Dock 5," *J Comput Aided Mol Des*, vol. 20, no. 10-11, pp. 601–19, 2006.

[176] G. Jones, P. Willett, and R. C. Glen, "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation," *J Mol Biol*, vol. 245, pp. 43–53, Jan 1995.

[177] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "Autodock4 and autodocktools4: Automated docking with selective receptor flexibility," *J Comput Chem*, vol. 30, pp. 2785–91, Dec 2009.

[178] I. Muegge and Y. C. Martin, "A general and fast scoring function for protein-ligand interactions: a simplified potential approach," *J Med Chem*, vol. 42, pp. 791–804, Mar 1999.

[179] H. Gohlke, M. Hendlich, and G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions," *J Mol Biol*, vol. 295, pp. 337–56, Jan 2000.

[180] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, "An all atom force field for simulations of proteins and nucleic acids," *Journal of Computational Chemistry*, vol. 7, no. 2, pp. 230–252, 1986.

[181] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, "A semiempirical free energy force field with charge-based desolvation," *J Comput Chem*, vol. 28, pp. 1145–52, Apr 2007.

[182] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," *J Comput Aided Mol Des*, vol. 11, pp. 425–45, Sep 1997.

[183] G. L. Warren, C. W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head, "A critical assessment

of docking programs and scoring functions," *J Med Chem*, vol. 49, pp. 5912–31, Oct 2006.

[184] Y. Pan, N. Huang, S. Cho, and A. D. MacKerell, Jr, "Consideration of molecular weight during compound selection in virtual target-based database screening," *J Chem Inf Comput Sci*, vol. 43, no. 1, pp. 267–72, 2003.

[185] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: Equilibria, structure, spectra, and dynamics," *Chem Rev*, vol. 99, pp. 2161–2200, Aug 1999.

[186] R. Walker, "http://ambermd.org/tutorials/advanced/tutorial3/."

[187] B. Kuhn, P. Gerber, T. Schulz-Gasch, and M. Stahl, "Validation and use of the mm-pbsa approach for drug discovery," *J Med Chem*, vol. 48, pp. 4040–8, Jun 2005.

[188] J. A. GRANT, M. A. GALLARDO, and B. T. PICKUP, "A fast method of molecular shape comparison: A simple application of a gaussian description of molecular shape," *Journal of Computational Chemistry*, vol. 17, no. 14, pp. 1653–1666, 1996.

[189] G. S. Paul Hawkins, "Ligand-based design workflow openeye scientific software."

[190] P. Labute, "Lowmodemd–implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops," *J Chem Inf Model*, vol. 50, pp. 792–800, May 2010.

[191] Schneider, Neidhart, Giller, and Schmid, ""scaffold-hopping" by topological pharmacophore search: A contribution to virtual screening," *Angew Chem Int Ed Engl*, vol. 38, pp. 2894–2896, Oct 1999.

[192] D. K. A. Huafeng Xu, "Retrospect and prospect of virtual screening in drug discovery," *Current Topics in Medicinal Chemistry*, vol. 2, pp. 1305–1320, 2002.

[193] M. A. M. W. Patrick Walters, Matthew T. Stahl, "Virtual screening – an overview," *Drug Discovery Today*, vol. 3, no. 4, pp. 160–178, 1998.

[194] Y. C. M. Robert D. Brown, "Use of structureâĹŠactivity data to compare structure-based clustering methods and descriptors for use in compound selection," *Journal of Chemical Information and Modeling*, vol. 36, no. 3, pp. 572–584, 1996.

[195] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, and T. Langer, "Evaluation of the performance of 3d virtual screening protocols: Rmsd comparisons, enrichment assessments, and decoy selection–what can we learn from earlier mistakes?," *J Comput Aided Mol Des*, vol. 22, no. 3-4, pp. 213–28, 2008.

[196] A. S. A. N. Paul Hawkins, G Warren, "How to do an evaluation: pitfalls and traps," *Journal of Computer Aided Molecular Design*, vol. 22, no. 3-4, pp. 179–190, 2008.

[197] J. O. Ebalunode, Z. Ouyang, J. Liang, and W. Zheng, "Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques," *J Chem Inf Model*, vol. 48, pp. 889–901, Apr 2008.

[198] T. Rodrigues, R. Moreira, J. Gut, P. J. Rosenthal, P. M. O Neill, G. A. Biagini, F. Lopes, D. J. V. A. dos Santos, and R. C. Guedes, "Identification of new antimalarial leads by use of virtual screening against cytochrome bcâĆĄ," *Bioorg Med Chem*, vol. 19, pp. 6302–8, Nov 2011.

[199] K. A. Werbovetz, D. L. Sackett, D. Delfín, G. Bhattacharya, M. Salem, T. Obrzut, D. Rattendi, and C. Bacchi, "Selective antimicrotubule activity of n1-phenyl-3,5-dinitro-n4,n4-di-n-propylsulfanilamide (gb-ii-5) against kinetoplastid parasites," *Mol Pharmacol*, vol. 64, pp. 1325–33, Dec 2003.

[200] N. Neamati, H. Hong, S. Sunder, G. W. Milne, and Y. Pommier, "Potent inhibitors of human immunodeficiency virus type 1 integrase: identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors," *Mol Pharmacol*, vol. 52, pp. 1041–55, Dec 1997.

[201] M. Read and J. E. Hyde, "Simple in vitro cultivation of the malaria parasite plasmodium falciparum (erythrocytic stages) suitable for large-scale preparations," *Methods Mol Biol*, vol. 21, pp. 43–55, 1993.

[202] W. Trager and J. B. Jensen, "Human malaria parasites in continuous culture. 1976," *J Parasitol*, vol. 91, pp. 484–6, Jun 2005.

[203] M. T. Makler, J. M. Ries, J. A. Williams, J. E. Bancroft, R. C. Piper, B. L. Gibbins, and D. J. Hinrichs, "Parasite lactate dehydrogenase as an assay for plasmodium falciparum drug sensitivity," *Am J Trop Med Hyg*, vol. 48, pp. 739–41, Jun 1993.

[204] M.-C. Lo, A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas, and G. Ellestad, "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery," *Anal Biochem*, vol. 332, pp. 153–9, Sep 2004.

[205] D. Matulis, J. K. Kranz, F. R. Salemme, and M. J. Todd, "Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using thermofluor," *Biochemistry*, vol. 44, pp. 5258–66, Apr 2005.

[206] D. Panda, M. A. Jordan, K. C. Chu, and L. Wilson, "Differential effects of vinblastine on polymerization and dynamics at opposite microtubule ends," *J Biol Chem*, vol. 271, pp. 29807–12, Nov 1996.

[207] T. N. Doman, S. L. McGovern, B. J. Witherbee, T. P. Kasten, R. Kurumbail, W. C. Stallings, D. T. Connolly, and B. K. Shoichet, "Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1b," *J Med Chem*, vol. 45, pp. 2213–21, May 2002.