UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

How Point Mutations Confer Resistance in BCR-Abl:

A Computational Study

by

Genevieve Sian Clapton

Thesis for the degree of Doctor of Philosophy

February 2015

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Chemistry

Thesis for the degree of Doctor of Philosophy

HOW POINT MUTATIONS CONFER RESISTANCE IN BCR-ABL:

A COMPUTATIONAL STUDY

Genevieve Sian Clapton

Imatinib is a first generation BCR-Abl tyrosine kinase inhibitor widely used in the treatment of CML. A significant cause of relapse in patients receiving imatinib therapy can be attributed to the development of point mutations in the BCR-Abl tyrosine kinase and resulting imatinib-resistance. Nilotinib is one of several second generation BCR-Abl tyrosine kinase inhibitors developed to treat imatinib-resistant CML. Nilotinib inhibits the majority of imatinib-resistant mutations, however mutations that are resistant to both imatinib and nilotinib persist in patients. These highly resistant mutations include compound mutants, where multiple mutations occur in the same BCR-Abl molecule. Further understanding regarding how mutations confer resistance and the ability to predict the resistance profile of a given compound mutation would enable drug design initiatives, both in CML and other cancers, to be improved in the future.

In this thesis three major computational approaches; molecular dynamics, ligand docking and MM-GBSA have been applied to investigate the effects of single and compound point mutations in the BCR-Abl kinase domain. Wild-type structures of BCR-Abl were studied initially in order to validate methods and act as a benchmark for comparison against the mutant structures. A total of ten mutant structures of inactive BCR-Abl were studied; five in complex with imatinib and five with nilotinib. Using this approach we were able to rank the single mutant structures according to experimental data and test the protocol on two previously unstudied compound mutations.

Table of Contents

TABLE OF CONTENTS ..... **I**

LIST OF TABLES ..... **VII**

LIST OF FIGURES ..... **XI**

ACKNOWLEDGEMENTS ..... **XXI**

1 INTRODUCTION **1**

2 STUDYING POINT MUTATIONS IN KINASES **3**

2.1 INTRODUCTION 3

2.1.1 Kinases 3

2.1.2 Point mutations 3

2.2 EXPERIMENTAL TECHNIQUES 4

2.2.1 X-ray crystallography 4

2.2.2 Biomolecular NMR 5

2.2.3 Hydrogen exchange mass spectrometry 5

2.2.4 Assays 5

2.3 COMPUTATIONAL METHODS 6

2.3.1 Methods that investigate kinase structure and dynamics 6

2.3.2 Methods that investigate protein-ligand binding 7

2.4 CHAPTER SUMMARY 9

3 COMPUTATIONAL METHODS **11**

3.1 INTRODUCTION 11

3.2 FORCE FIELDS 11

3.3 MONTE CARLO SIMULATIONS 12

3.4 MOLECULAR DYNAMICS 13

3.5 THERMODYNAMIC CONDITIONS 16

3.6 PERIODIC BOUNDARY CONDITIONS 18

3.7 SOLVENT MODELS 18

3.7.1 Explicit water models 18

3.7.2 Implicit water models 19

3.8 SIMULATION ANALYSIS TECHNIQUES 20

3.8.1 Root-mean-squared-deviation (RMSD) 20
3.8.2 Root-mean-squared-fluctuations (RMSF) 21
3.8.3 Distance analysis 21
3.8.4 Principal component analysis (PCA) 21
3.8.5 Binding site analysis 23
3.9 FREE ENERGY CALCULATIONS 24
3.9.1 Free energy perturbation 24
3.9.2 Thermodynamic integration 25
3.9.3 MM-PB/GBSA 26
3.9.4 Free energy calculations using pymdgbsa.py 27
3.10 PROTEIN-LIGAND DOCKING 27
3.10.1 Search algorithms 27
3.10.2 Scoring functions 30
3.10.3 Modelling protein flexibility 31
3.10.4 Ensemble docking using Glide 32
3.11 WATER MOLECULE PREDICTION IN PROTEIN BINDING SITES 35
3.11.1 Ptraj 36
3.11.2 JAWS 36
3.12 CHAPTER SUMMARY 37
4 THE BCR-ABL KINASE & CHRONIC MYELOID LEUKAEMIA 39
4.1 INTRODUCTION 39
4.2 THE C-ABL KINASE 40
4.2.1 The different domains of c-Abl 41
4.2.2 The inactive (autoinhibited) c-Abl structure 44
4.2.3 The active c-Abl structure 48
4.3 THE BCR-ABL FUSION KINASE 54
4.4 CHRONIC MYELOID LEUKAEMIA (CML) 55
4.4.1 What is CML? 55
4.4.2 The three phases of CML 56
4.5 CURRENT SMALL MOLECULE INHIBITORS OF BCR-ABL USED IN THE TREATMENT OF CML 57
4.5.1 Imatinib 58
4.5.2 Nilotinib 59
4.5.3 Dasatinib 60
4.5.4 Bosutinib 60

4.5.5 Ponatinib 61
4.6 Resistance to tyrosine kinase inhibitors 62
4.6.1 Possible causes of drug resistance 62
4.6.2 Point mutations 62
4.6.3 Treating drug resistance 65
4.6.4 Compound mutations 66
4.7 Chapter summary 66
5 ACTIVE BCR-ABL 67
5.1 Aim 67
5.2 Molecular dynamics study of active Abl 68
5.2.1 Description of protein structure 69
5.2.2 Description of ligand binding 69
5.2.3 Review of expected characteristics 70
5.2.4 Protein structure preparation and MD parameters 73
5.3 Results: Active Abl kinase domain in complex with VX-680 (2F4J.pdb) 74
5.3.1 Structural stability and flexibility 74
5.3.2 Protein-ligand binding 80
5.4 Results: Active Abl kinase domain pseudo apo structure (2F4J$_{P\text{-}APO}$) 85
5.4.1 Structural stability and flexibility 85
5.5 Chapter summary 91
6 INACTIVE BCR-ABL 93
6.1 Aim 93
6.2 Molecular dynamics study of inactive WT Abl in complex with imatinib and nilotinib 94
6.2.1 Description of protein structures 94
6.2.2 Description of ligand binding 95
6.2.3 Review of expected characteristics 96
6.2.4 Protein structure preparation and MD parameters 99
6.3 Results: WT Abl kinase domain in complex with imatinib (1IEP.pdb) 99
6.3.1 Structural stability and flexibility 99
6.3.2 Protein-ligand binding 106
6.4 Results: WT Abl kinase domain pseudo apo structure (1IEP$_{P\text{-}APO}$) 110
6.4.1 Structural stability and flexibility 110
6.5 Results: WT Abl kinase domain in complex with nilotinib (3CS9.pdb) 118

6.5.1 Structural stability and flexibility 118
6.5.2 Protein-ligand binding 124
6.6 SUMMARY OF MOLECULAR DYNAMICS STUDY 127
6.7 ENSEMBLE DOCKING 130
6.7.1 Methodology 130
6.7.2 Results 131
6.8 MM-GBSA CALCULATIONS 131
6.8.1 Methodology 131
6.8.2 Results 132
6.9 CHAPTER SUMMARY 132
7 SINGLE MUTATION STUDY 136
7.1 AIM 136
7.2 MOLECULAR DYNAMICS STUDY OF MUTATIONS F317L, Y253F AND T315I 137
7.2.1 Review of the mutations 137
7.2.2 Protein structure preparation and MD parameters 139
7.3 RESULTS: F317L 139
7.3.1 Imatinib 140
7.3.2 Nilotinib 146
7.4 RESULTS: Y253F 152
7.4.1 Imatinib 153
7.4.2 Nilotinib 160
7.5 RESULTS: T315I 168
7.5.1 Imatinib 168
7.5.2 Nilotinib 174
7.6 SUMMARY OF MOLECULAR DYNAMICS STUDY 179
7.7 ENSEMBLE DOCKING 180
7.7.1 Methodology 180
7.7.2 Results 181
7.7.3 Discussion 183
7.8 MDGBSA CALCULATIONS 183
7.8.1 Methodology 183
7.8.2 Results 184
7.8.3 Discussion 185
7.9 WATER MOLECULE PREDICTION 186

7.9.1 Observational analysis 186
7.9.2 Water molecule prediction using Amber (PTRAJ) 188
7.9.3 Free energy calculations to determine the stability of W282 188
7.10 CHAPTER SUMMARY 190
8 COMPOUND MUTATION STUDY 194
8.1 AIM 194
8.2 MOLECULAR DYNAMICS STUDY OF COMPOUND MUTATIONS 196
8.2.1 Review of the mutations 196
8.2.2 Protein structure preparation and MD parameters 198
8.3 RESULTS: DOUBLE COMPOUND MUTATION 198
8.3.1 Imatinib 198
8.3.2 Nilotinib 205
8.4 RESULTS: TRIPLE COMPOUND MUTATION 211
8.4.1 Imatinib 211
8.4.2 Nilotinib 215
8.5 SUMMARY OF MOLECULAR DYNAMICS STUDY 222
8.6 ENSEMBLE DOCKING 223
8.6.1 Methodology 223
8.6.2 Results 223
8.6.3 Discussion 228
8.7 MMGBSA CALCULATIONS 229
8.7.1 Methodology 229
8.7.2 Results 230
8.7.3 Discussion 234
8.8 CHAPTER SUMMARY 234
9 CONCLUSIONS 238
APPENDIX A: THE SEQUENCE AND SECONDARY STRUCTURE OF 2F4J.PDB 242
APPENDIX B: THE SEQUENCE AND SECONDARY STRUCTURE OF 1IEP.PDB 243
APPENDIX C: THE SEQUENCE AND SECONDARY STRUCTURE OF 3CS9.PDB 244
APPENDIX D: 1ATP VS 2F4J (NEEDLEMAN-WUNSCH SEQUENCE ALIGNMENT) 245
APPENDIX E: 2F4J VS 1IEP (NEEDLEMAN-WUNSCH SEQUENCE ALIGNMENT) 246

APPENDIX F: 1IEP VS 3CS9 (NEEDLEMAN-WUNSCH SEQUENCE ALIGNMENT) 247

APPENDIX G: SUMMARY OF DOCKING STUDIES (BEST SCORE ACHIEVED)..... 248

BIBLIOGRAPHY..... 250

List of Tables

Table 4.5.1: Approved indications for TKI therapy in CML [298, 302-305]..... 57
Table 4.6.1 Treatment options based on BCR-ABL kinase domain mutation status. Adapted from references [293, 327] 65
Table 5.2.1: Distance between R-spine residues (2F4J.pdb) 71
Table 5.2.2: C-spine residues for PKA (1ATP.pdb) and Abl (2F4J.pdb) 71
Table 5.2.3: Distance between C-spine residues (2F4J.pdb) 72
Table 5.3.1: Mean distance between R-spine residues 78
Table 5.3.2: Mean distance between C-spine residues 79
Table 5.3.3: Hydrogen bond analysis: 2F4J (4 Å, 30°); S = sidechain, M = main chain. 79
Table 5.4.1: Mean distance between R-spine residues 88
Table 5.4.2: Mean distance between C-spine residues 88
Table 5.4.3: Hydrogen bond analysis: $2F4J_{p\text{-}apo}$ vs. 2F4J (4 Å, 30°); S = sidechain, M = main chain. 89
Table 6.2.1: Distance between R-spine residues (1IEP.pdb and 3CS9.pdb) 97
Table 6.2.2: Distance between C-spine residues (1IEP.pdb and 3CS9.pdb) 98
Table 6.3.1: Mean distance between R-spine residues 103
Table 6.3.2: Mean distance between C-spine residues 104
Table 6.3.3: Hydrogen bond analysis: 1IEP (4 Å, 30°); S = sidechain, M = main chain..... 105
Table 6.4.1: Mean distance between R-spine residues 114
Table 6.4.2: Mean distance between C-spine residues 114
Table 6.4.3: Hydrogen bond analysis: $1IEP_{p\text{-}apo}$ vs. 1IEP (4 Å, 30°); S = sidechain, M = main chain. 115
Table 6.5.1: Mean distance between R-spine residues 122
Table 6.5.2: Mean distance between C-spine residues 122
Table 6.5.3: Hydrogen bond analysis: 3CS9 vs 1IEP (4 Å, 30°); S = sidechain, M = main chain. 123
Table 6.7.1: GlideScore for imatinib and nilotinib docked to $1IEP_{WT}$ and $3CS9_{WT}$ (best achieved) vs IC_{50} values. IC_{50} values from Zabriskie et al [331]. 131
Table 6.8.1: MM-GBSA results for imatinib and nilotinib ($1IEP_{WT}$ and $3CS9_{WT}$)..... 132
Table 7.1.1. Sensitivity of Bcr-Abl kinase domain mutants to imatinib and nilotinib (based on assays using Ba/F3 cells. The IC_{50} value is the concentration of inhibitor resulting in a 50% reduction in cell viability; NR, not recorded. Adapted from references: Azam et al, 2003 [44], O'Hare et al, 2009 [333] Weisberg et al, 2005 [32] and Zabriskie et al, 2014 [331]) 136
Table 7.1.2. Sensitivity scale for imatinib and nilotinib. Adapted from reference [331]..... 136
Table 7.3.1: Mean distance between R-spine residues 141
Table 7.3.2: Hydrogen bond analysis: $1IEP_{F317L}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 142
Table 7.3.3: Mean distance between R-spine residues 147
Table 7.3.4: Hydrogen bond analysis: $3CS9_{F317L}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 148
Table 7.4.1: Mean distance between R-spine residues 155

Table 7.4.2: Hydrogen bond analysis: $1IEP_{Y253F}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 156
Table 7.4.3: Mean distance between R-spine residues 161
Table 7.4.4: Hydrogen bond analysis: $3CS9_{Y253F}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 162
Table 7.5.1: Mean distance between R-spine residues 170
Table 7.5.2: Hydrogen bond analysis: $1IEP_{T315I}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 170
Table 7.5.3: Mean distance between R-spine residues 176
Table 7.5.4: Hydrogen bond analysis: $3CS9_{T315I}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 176
Table 7.7.1: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 181
Table 7.7.2: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 181
Table 7.7.3: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 182
Table 7.7.4: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 182
Table 7.7.5: IC_{50} values of Abl TKIs vs. molecular weight. IC_{50} values from Zabriskie et al [331]. 183
Table 7.8.1: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I) with imatinib 184
Table 7.8.2: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 184
Table 7.8.3: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I) with nilotinib 185
Table 7.8.4: MM-GBSA results for 3CS9 with nilotinib (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 185
Table 7.9.1: Stability of hydrogen bonding network with W_1 and W_2. 1IEP simulation 187
Table 7.9.2: Ptraj demonstrates how waters cluster around imatinib in the 1IEP simulation . 188
Table 7.10.1: Results summary – Imatinib, single mutations study. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 191

Table 7.10.2: Results summary – Nilotinib, single mutations study. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 191
Table 8.1.1. Sensitivity of Bcr-Abl kinase domain mutants to imatinib and nilotinib (based on assays using Ba/F3 cells. The IC_{50} value is the concentration of inhibitor resulting in a 50% reduction in cell viability; NR, not recorded. Adapted from references: Zabriskie et al, 2014 [331]) 195
Table 8.1.2. Sensitivity scale for imatinib, nilotinib and ponatinib. Adapted from reference [331] 195
Table 8.3.1: Mean distance between R-spine residues 200
Table 8.3.2: Hydrogen bond analysis: $1IEP_{YF}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 201
Table 8.3.3: Mean distance between R-spine residues 206
Table 8.3.4: Hydrogen bond analysis: $3CS9_{YF}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 206
Table 8.4.1: Mean distance between R-spine residues 212
Table 8.4.2: Hydrogen bond analysis: $1IEP_{TYF}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 212
Table 8.4.3: Mean distance between R-spine residues 217
Table 8.4.4: Hydrogen bond analysis: $3CS9_{TYF}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain. 218
Table 8.6.1: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 224
Table 8.6.2: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 224
Table 8.6.3: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 225
Table 8.6.4: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 225
Table 8.6.5: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 226

Table 8.6.6: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM[331]). IC_{50} values from O'Hare et al, 2009 [333]. 227
Table 8.6.7: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 227
Table 8.6.8: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 228
Table 8.7.1: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF) with imatinib 230
Table 8.7.2: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 230
Table 8.7.3: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF) with nilotinib 231
Table 8.7.4: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF) with nilotinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 231
Table 8.7.5: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF, TYF) with imatinib..... 232
Table 8.7.6: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF, TYF) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 232
Table 8.7.7: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) with nilotinib..... 233
Table 8.7.8: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) with nilotinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 233
Table 8.8.1: Results summary – Imatinib, single and compound mutation studies. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 236
Table 8.8.2: Results summary – Nilotinib, single and compound mutation studies. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333]. 236

List of Figures

Figure 2.3.1: Simulation-based mutational study process 9
Figure 3.10.1: Glide docking "funnel", showing the Glide docking hierarchy. Adapted from reference [97]. 33
Figure 4.1.1: Translocation of the Abl gene and the BCR gene 39
Figure 4.2.1: Structure of adenosine triphosphate (ATP) 40
Figure 4.2.2: Kinase phosphorylation - ATP binds to the kinase pocket, the kinase can then phosphorylate a substrate that can interact with the downstream effector molecules. Adapted from reference [242]..... 40
Figure 4.2.3: The structural regions of c-Abl kinase, adapted from references [249, 250] 41
Figure 4.2.4: Schematic representation of inactive c-Abl (1b) kinase. Adapted from reference [251, 252] 42
Figure 4.2.5: Secondary structure of Abl kinase domain with labelled helices and beta sheets (inactive structure, 1IEP.pdb) 43
Figure 4.2.6: Schematic representation of inactive c-Abl (1b) kinase. Adapted from reference [251, 252] 45
Figure 4.2.7: Key characteristics of the inactive Abl conformation [1IEP.pdb]. 46
Figure 4.2.8: The c-Abl A-loop and DFG-loop in the inactive “Asp-out” conformation [1IEP.pdb]. 47
Figure 4.2.9: Isolated view of the DFG-loop a) Inactive “Asp-out” conformation [1IEP.pdb]. b) Active “Asp-in” conformation [2F4J.pdb]. 47
Figure 4.2.10: Key characteristics of the active Abl conformation [2F4J.pdb]. 49
Figure 4.2.11: Schematic representation of active c-Abl (1b) kinase. Adapted from reference [251, 252] 51
Figure 4.2.12: The c-Abl DFG-loop in the active “Asp-in” conformation [2GQG.pdb]..... 51
Figure 4.2.13: The residues of the R-spine [2F4J.pdb]..... 53
Figure 4.2.14: The residues of the C-spine [1ATP.pdb]..... 53
Figure 4.3.1: The structural domain differences between c-Abl kinase and BCR-Abl kinase. Adapted from references [249, 250] 54
Figure 4.3.2: Schematic representation of the BCR-Abl kinase. Adapted from reference [251, 252] 55
Figure 4.5.1: (a) Imatinib binds to the ATP pocket of BCR-Abl (competing with ATP), therefore the substrate cannot be phosphorylated. Adapted from reference [242]. (b) The structure of imatinib 58
Figure 4.5.2: The structure of nilotinib. 59
Figure 4.5.2 :The structure of dasatinib..... 60
Figure 4.5.4: The structure of bosutinib 61
Figure 4.5.5: The structure of ponatinib 61
Figure 4.6.1: Kinase domain mutations: IC_{50} as measured in transformed BaF_3 cells. ND, not determined. IC_{50} for native BCR-Abl is 0.6 µM (imatinib). Mutations in bold have been reported in patients [44]. 63
Figure 5.2.1 Position of the H396P mutation in 2F4J.pdb 69

Figure 5.2.2: The VX-680 binding site in 2F4J.pdb. Hydrogen bond distances (Å) labelled. Adapted from reference [31]. 70
Figure 5.2.3: a) Extended conformation of the A-loop in the 2F4J.pdb crystal structure; b) Alternative Asp-in conformation of the DFG-loop in the 2F4J.pdb crystal structure 70
Figure 5.2.4: Residues in the R-spine (2F4J.pdb) 71
Figure 5.2.5: Residues in the C-spine (2F4J.pdb) 72
Figure 5.2.6: Conformation of the P-loop in the 2F4J crystal structure (ligand not shown) 72
Figure 5.3.1: The key motifs of the active structure (2F4J.pdb) 74
Figure 5.3.2: RSMD results for 2F4J.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green 75
Figure 5.3.3: RMSF results for 2F4J.pdb. 75
Figure 5.3.4: RMSF (black) vs B-factors (red) for 2F4J.pdb 76
Figure 5.3.5: PCA projections for the first three principal components for active BCR-Abl (2F4J). 76
Figure 5.3.6: The change in P-loop conformation (2F4J.pdb): a) Initial structure; b) Final structure post MD simulation 77
Figure 5.3.7: The regulatory and catalytic spines in the active conformation (2F4J.pdb) 78
Figure 5.3.8: Binding site analysis, average pocket overlaid on the starting structure (2F4J.pdb). Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown). 80
Figure 5.3.9: RMSD of VX-680 (2F4J.pdb) 81
Figure 5.3.10: Index of atoms in VX-680 81
Figure 5.3.11: RMSF of ligand VX-680 (2F4J.pdb) 82
Figure 5.3.12: The VX-680 binding site in 2F4J.pdb. Hydrogen bond distances (Å) labelled. Adapted from reference [31]. 82
Figure 5.3.13: Protein-ligand interactions by residue (2F4J.pdb) 83
Figure 5.3.14: Protein-ligand interaction diagram for VX-680 (2F4J.pdb) 84
Figure 5.4.1: RSMD results for $2F4J_{p\text{-}apo}$; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green 85
Figure 5.4.2: RMSF results for active BCR-Abl with VX-680 (2F4J, black) and active BCR-Abl pseudo -apo structure (2F4Jp-apo, grey) 86
Figure 5.4.3: PCA projections for the first three principal components for the active BCR-Abl apo structure (2F4Jp-apo). 87
5.4.4: The change in P-loop conformation: a) Initial structure ($2F4J_{p\text{-}apo}$); b) Final structure post MD simulation ($2F4J_{p\text{-}apo}$); c) Final structure post MD simulation (2F4J) 87
Figure 5.4.5: Distance between Glu286 and Lys271 90
Figure 5.4.6: Binding site analysis, average pocket overlaid on the starting structure ($2F4J_{p\text{-}apo}$) a) Most conserved region of pocket (pocket present 60-80% of simulation). b) Less conserved regions of pocket (pocket present 20-80% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown). 90
Figure 6.2.1: The imatinib binding site in 1IEP.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [30] 95
Figure 6.2.2: The nilotinib binding site in 3CS9.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [262] 96
Figure 6.2.3: Conformation of the A-loop, DFG-loop and P-loop in the 1IEP crystal structure. . 96

Figure 6.2.4: Conformation of the A-loop, DFG-loop and P-loop in the 3CS9 crystal structure. 97
Figure 6.2.5: Residues in the R-spine (1IEP.pdb) 97
Figure 6.2.6: Residues in the C-spine (1IEP.pdb) 98
Figure 6.3.1: The key motifs of the inactive structure (1IEP.pdb) 99
Figure 6.3.2: RSMD results for 1IEP.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green. 100
Figure 6.3.3: Conformation of the F and H-helices; (a) Initial structure; (b) Final after MD production..... 100
Figure 6.3.4: RMSF results for 1IEP (black) compared to 2F4J (red). 101
Figure 6.3.5: RMSF (black) vs. B-factors (red) for 1IEP.pdb 101
Figure 6.3.6: PCA projections for the first three principal components for BCR-Abl apo structure ($1IEP_{apo}$). 102
Figure 6.3.7: Fluctuations of residues preceding the αC-helix. (a) Initial structure; (b) Final structure post MD simulation (1IEP.pdb) 102
Figure 6.3.8: Catalytic (a) and regulatory (b) spines in the inactive conformation (1IEP.pdb). 103
Figure 6.3.9: Binding site analysis, average pocket overlaid on the starting structure (1IEP.pdb) (a) Most conserved region of pocket (pocket present 40-100% of simulation). (b) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown). 105
Figure 6.3.10: RMSD of imatinib (1IEP.pdb) 106
Figure 6.3.11: Index of atoms in imatinib (1IEP.pdb) 106
Figure 6.3.12: RMSF of ligand imatinib (1IEP.pdb) 107
Figure 6.3.13: The imatinib binding site in 1IEP.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [30] 108
Figure 6.3.14: Protein-ligand interactions by residue (1IEP.pdb) 109
Figure 6.3.15: Protein-ligand interaction diagram for imatinib (1IEP.pdb) 109
Figure 6.4.1: RSMD results for $1IEP_{p\text{-}apo}$; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green. 110
Figure 6.4.2: The change in P-loop conformation: a) Initial structure ($1IEP_{p\text{-}apo}$); b) Final structure post MD simulation ($1IEP_{p\text{-}apo}$)..... 110
Figure 6.4.3: RMSD results for inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl apo structure ($1IEP_{p\text{-}apo}$, purple)..... 111
Figure 6.4.4: Comparison of RMSF plots of inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl apo structure ($1IEP_{p\text{-}apo}$, purple). 111
Figure 6.4.5: Figure 6.28: Conformation of F-helix; a) Initial structure (1IEP_p-apo); b) Final structure post MD simulation (1IEP_p-apo); c) Final structure post MD simulation (1IEP)..... 112
Figure 6.4.6: Conformation of H-helix; a) Initial structure (1IEP_p-apo); b) Final structure post MD simulation (1IEP_p-apo); c) Final structure post MD simulation (1IEP)..... 112
Figure 6.4.7: RMSF (black) vs B-factors (red) for 3CS9.pdb 112
Figure 6.4.8: PCA projections for the first three principal components for inactive BCR-Abl apo structure ($1IEP_{p\text{-}apo}$). 113
Figure 6.4.9: Conformation of loop section preceding C-helix (1IEPp-apo); a) Initial structure b) Final structure post MD simulation 113

Figure 6.4.10: The change in C-helix conformation: a) Initial structure (IEPp-apo); b) Final structure post MD simulation (1IEPp-apo); c) Final structure post MD simulation (1IEP) 114
Figure 6.4.11: Loss of interaction between Lys271 and Glu286 in the $1IEP_{p\text{-}apo}$ simulation; a) initial position b) rotated position 116
Figure 6.4.12: Distance between Lys271 and Glu286 over the 20 ns simulation for $1IEP_{p\text{-}apo}$ .. 116
Figure 6.4.13: Binding site analysis, average pocket overlaid on the starting structure ($1IEP_{p\text{-}apo}$) a) Most conserved region of pocket (pocket present 40-100% of simulation). B) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown). 117
Figure 6.5.1: The key motifs of the inactive structure (3CS9.pdb) 118
Figure 6.5.2: RSMD results for 3CS9.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green 119
Figure 6.5.3: RMSD results for inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl with nilotinib (3CS9, blue) 119
Figure 6.5.4: Conformation of the C-helix (3CS9); a) Initial structure b) Final structure post MD simulation 119
Figure 6.5.5: Comparison of RMSF plots of inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl with nilotinib (3CS9, blue). 120
Figure 6.5.6: Additional regions of the kinase with increased RMSF compared to 1IEP.pdb (3CS9.pdb): β5-sheet, residues 80-85, magenta; αF-helix, residues 180-190, orange; αH-helix, residues 220-230, cyan. 120
Figure 6.5.7: PCA projections for the first three principal components for inactive BCR-Abl with nilotinib (3CS9). 121
Figure 6.5.8: Catalytic (a) and regulatory (b) spines in the inactive conformation (3CS9.pdb) 122
Figure 6.5.9: Binding site analysis, average pocket overlaid on the starting structure (3CS9.pdb) a) Most conserved region of pocket (pocket present 40-100% of simulation). B) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown). 124
Figure 6.5.10: RMSD of nilotinib (3CS9.pdb) 124
Figure 6.5.11: Index of atoms in nilotinib (3CS9.pdb) 125
Figure 6.5.12: RMSF of ligand nilotinib (3CS9.pdb) 125
Figure 6.5.13: The nilotinib binding site in 3CS9.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [262] 126
Figure 6.5.14: Protein-ligand interactions by residue (3CS9.pdb) 126
Figure 6.5.15: Protein-ligand interaction diagram for nilotinib (3CS9.pdb) 127
Figure 6.7.1: Docking study methodology 130
Figure 7.2.1: (a) The position of Phe317 in WT Abl (1IEP.pdb); (b) The position of Leu317 in mutant Abl ($1IEP_{F317L}$) 138
Figure 7.2.2: (a) The position of Tyr 253 in WT Abl (1IEP.pdb); (b) The position of Phe 253 in mutant Abl ($1IEP_{Y253F}$) 138
Figure 7.2.3: (a) The position of Thr315 in WT Abl (1IEP.pdb); (b) The position of Ile315 in mutant Abl ($1IEP_{T315I}$) 139
Figure 7.3.1: RMSF results for $1IEP_{F317L}$ 140
Figure 7.3.2: P-loop conformation: initial (a) and during simulation (b). WT shown in grey, F317L mutated structure in green 141

Figure 7.3.3: Principal component analysis for 1IEPF317L (PC1, projected on structure) 141
Figure 7.3.4: RMSD of imatinib ($1IEP_{F317L}$)..... 143
Figure 7.3.5: Index of atoms in imatinib ($1IEP_{F317L}$)..... 143
Figure 7.3.6: RMSF of imatinib ($1IEP_{F317L}$) 144
Figure 7.3.7: Comparison of imatinib binding site in the WT and F317L mutant structures; grey/silver, WT; purple/cyan, F317L (5 ns) 144
Figure 7.3.8: Protein-ligand interactions by residue ($1IEP_{F317L}$) 145
Figure 7.3.9: RMSF results for $3CS9_{WT}$ and $3CS9_{F317L}$ 146
Figure 7.3.10: Principal component analysis for $3CS9_{F317L}$ (PC1, projected on structure)..... 147
Figure 7.3.11: Glu286-Arg386 bond formation ($3CS9_{F317L}$) 149
Figure 7.3.12: RMSD of nilotinib ($3CS9_{F317L}$)..... 150
Figure 7.3.13: Index of atoms in nilotinib ($3CS9_{F317L}$)..... 150
Figure 7.3.14: RMSF of nilotinib ($3CS9_{F317L}$) 151
Figure 7.3.15: Comparison of nilotinib binding site in the WT and F317L mutant structures; grey/silver, WT; purple/cyan, F317L (5 ns) 151
Figure 7.3.16: Protein-ligand interactions by residue ($3CS9_{F317L}$) 152
Figure 7.4.1: RMSF results for $1IEP_{WT}$ and $1IEP_{Y253F}$ 153
Figure 7.4.2: RMSD of the P-loop ($1IEP_{WT}$ and $1IEP_{Y253F}$)..... 154
Figure 7.4.3: Fluctuations of the P-loop: initial (a) and during simulation (b) ($1IEP_{Y253F}$) 154
Figure 7.4.4: RMSD of the DFG-loop ($1IEP_{WT}$ and $1IEP_{Y253F}$) 155
Figure 7.4.5: (a) Initial conformation of DFG loop (b) Conformation shift of DFG loop during simulation (between ~4 ns and 11 ns)..... 155
Figure 7.4.6: RMSD of imatinib ($1IEP_{Y253F}$) 157
Figure 7.4.7: Index of atoms in imatinib ($1IEP_{Y253F}$)..... 157
Figure 7.4.8: RMSF of imatinib ($1IEP_{Y253F}$) 158
Figure 7.4.9: Comparison of imatinib binding site in the WT and Y253F mutant structures; grey/silver, WT; purple/cyan, Y253F (10 ns)..... 158
Figure 7.4.10: Protein -ligand interactions by residue ($1IEP_{Y253F}$) 159
Figure 7.4.11: RMSF results for $3CS9_{WT}$ and $3CS9_{Y253F}$ 160
Figure 7.4.11: Principal component analysis for $3CS9_{Y253F}$ (PC1, projected on structure)..... 160
Figure 7.4.13: RMSD of the P-loop ($3CS9_{WT}$ and $3CS9_{Y253F}$)..... 161
Figure 7.4.14: Partial rotation of the C-helix in $3CS9_{Y253F}$ 163
Figure 7.4.15: Full rotation of the C-helix in $3CS9_{Y253F}$ 163
Figure 7.4.16: RMSD for nilotinib ($3CS9_{Y253F}$) 164
Figure 7.4.17: Index of atoms for nilotinib ($3CS9_{Y253F}$)..... 165
Figure 7.4.18: RMSF of nilotinib ($3CS9_{Y253F}$) 165
Figure 7.4.19: Comparison of nilotinib binding site in the WT and Y253F mutant structures; grey/silver, WT; purple/cyan, Y253F (15 ns)..... 165
Figure 7.4.20: Protein-ligand interactions by residue for nilotinib ($3CS9_{Y253F}$) 166
Figure 7.5.1: RMSF results for $1IEP_{WT}$ and $1IEP_{T315I}$..... 168
Figure 7.5.2: RMSD of the P-loop ($1IEP_{WT}$ and $1IEP_{T315I}$) 169
Figure 7.5.3: P-loop conformation: initial (a) and during simulation (b). ($1IEP_{T315I}$)..... 169
Figure 7.5.4: RMSD of imatinib ($1IEP_{T315I}$)..... 171
Figure 7.5.5: Index of atoms in imatinib ($1IEP_{T315I}$) 171

Figure 7.5.6: RMSF of imatinib ($1IEP_{T315I}$) 172
Figure 7.5.7: Comparison of imatinib binding site in the WT and T315I mutant structures; grey/silver, WT; purple/cyan, T315I (10 ns) 172
Figure 7.5.8: Protein-ligand interactions by residue ($1IEP_{T315I}$) 173
Figure 7.5.9: RMSF results for $3CS9_{WT}$ and $3CS9_{T315I}$ 174
Figure 7.5.10: RMSD of the DFG-loop ($3CS9_{WT}$ and $3CS9_{T315I}$) 175
Figure 7.5.11: Principal component analysis for $3CS9_{Y253F}$ (PC1, projected on structure) 175
Figure 7.5.12: (a) interactions between Lys271, Glu286 and Arg386; (b) Glu286-Arg386 bond formation ($3CS9_{T315I}$) 177
Figure 7.5.13: RMSD of nilotinib ($3CS9_{T315I}$) 177
Figure 7.5.14: Index of atoms for nilotinib ($3CS9_{T315I}$) 178
Figure 7.5.15: RMSF of nilotinib ($3CS9_{T315I}$) 178
Figure 7.5.16: Comparison of nilotinib binding site in the WT and T315I mutant structures; grey/silver, WT; purple/cyan, T315I (17 ns) 178
Figure 7.5.17: Protein-ligand interactions by residue for nilotinib ($3CS9_{T315I}$) 179
Figure 7.9.1: Hydrogen bonding network with W_1 and W_2 in the 1IEP crystal structure; a) underside of binding site b) parallel to binding site 186
Figure 7.9.2: Hydrogen bonding network with W_1 in the 1IEP WT simulation; underside of binding site 187
Figure 7.9.3: Hydrogen bonding network with W1 and W2 in the 1IEP WT simulation; a) underside of binding site b) parallel to binding site 187
Figure 7.9.4: Binding free energy of W_2 determined by FDTI (pink) and JAWS (blue). 189
Figure 8.2.1: Location of the residues in the double mutant structure (a) The position of Tyr253 and Phe317 in WT Abl (1IEP.pdb); (b) The position of Phe253 and Leu317 in mutant Abl ($1IEP_{YF}$) 196
Figure 8.2.2: (a) The position of Tyr253, Phe317 and Thr315 in WT Abl (1IEP.pdb); (b) The position of Phe253, Leu317 and Ile 315 in mutant Abl ($1IEP_{TYF}$) 197
Figure 8.3.1: RMSF results for $1IEP_{WT}$ and $1IEP_{YF}$ 198
Figure 8.3.2: P-loop conformation: (a) initial and (b) during simulation 199
Figure 8.3.3: RMSD of the DFG-loop ($1IEP_{YF}$) 199
Figure 8.3.4: Different conformations of the DFG loop in $1IEP_{YF}$; Double mutant in cyan, WT in grey 200
Figure 8.3.5: The positions of Glu286, Arg386 and Lys271; (a) facing the αC-Helix (b) rotated 90° 201
Figure 8.3.6: RMSD of imatinib ($1IEP_{YF}$) 202
Figure 8.3.7: Index of atoms in imatinib ($1IEP_{YF}$) 202
Figure 8.3.8: RMSF of imatinib ($1IEP_{YF}$) 203
Figure 8.3.9: Comparison of imatinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (15 ns) 203
Figure 8.3.10: Protein-ligand interactions by residue ($1IEP_{YF}$) 204
Figure 8.3.11: RMSF results for $3CS9_{WT}$ and $3CS9_{YF}$ 205
Figure 8.3.12: Conformation of the P-loop (a) initial (b) final 205
Figure 8.3.13: Conformation of Glu286 (a) initial (b) fully rotated 207
Figure 8.3.14: RMSD for nilotinib ($3CS9_{YF}$) 208
Figure 8.3.15: Index of atoms for nilotinib ($3CS9_{YF}$) 208

Figure 8.3.16: RMSF for nilotinib ($3CS9_{YF}$)..... 209
Figure 8.3.17: Comparison of nilotinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (15 ns)..... 209
Figure 8.3.18: Protein-ligand interactions by residue (nilotinib, $3CS9_{YF}$)..... 210
Figure 8.4.1: RMSF results for $1IEP_{WT}$ and $1IEP_{TYF}$..... 211
Figure 8.4.2: Different conformations of the P-loop in $1IEP_{TYF}$ (a) initial (b) final 211
Figure 8.4.3: RMSD of imatinib ($1IEP_{TYF}$)..... 213
Figure 8.4.4: Index of atoms for imatinib 213
Figure 8.4.5: RMSF of imatinib ($1IEP_{TYF}$)..... 214
Figure 8.4.6: Comparison of imatinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (6 ns)..... 214
Figure 8.4.7: Protein-ligand interactions by residue ($1IEP_{TYF}$)..... 215
Figure 8.4.8: RMSF results for $3CS9_{WT}$ and $3CS9_{TYF}$..... 216
Figure 8.4.9: Different conformations of the P-loop in $1IEP_{TYF}$ (a) initial (b) final 216
Figure 8.4.10: Different conformations of the DFG-loop ($1IEP_{TYF}$)..... 217
Figure 8.4.11: Denaturation of the αC-helix (a); Position of Glu286 forming a hydrogen bond with Arg386..... 218
Figure 8.4.12: RMSD of nilotinib ($3CS9_{TYF}$)..... 219
Figure 8.4.13: Index of atoms for nilotinib ($3CS9_{TYF}$) 220
Figure 8.4.14: RMSF of nilotinib ($3CS9_{TYF}$)..... 220
Figure 8.4.15: Comparison of nilotinib binding site in the WT and TYF mutant structures; grey/silver, WT; purple/cyan, TYF (15 ns)..... 221
Figure 8.4.16: Protein-ligand interactions by residue (nilotinib, $3CS9_{TYF}$) 221
Figure 8.6.1: Summary of docking study using 3CS9 as the receptor..... 229

DECLARATION OF AUTHORSHIP

I, GENEVIEVE SIAN CLAPTON, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

HOW POINT MUTATIONS CONFER RESISTANCE IN BCR-ABL: A COMPUTATIONAL STUDY

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:

Date:

Acknowledgements

I would like to take this opportunity to say a big thank you to Professor Jonathan Essex for all his support and guidance throughout this research.

Thank you to my industrial supervisor Dr Katrin Spiegel for her encouragement and suggestions and to Dr Ian Craig and the rest of the CADD team at Novartis for sharing their office with me while I completed the docking and MM-GBSA studies on site at Horsham.

Thank you to the Graduate School staff and entire Essex research group for making my time at Southampton so enjoyable; to past member Dr Sarah Williams for always taking the time to answer my questions in the first year of my PhD and to Dr Michael Bodnarchuk for running the JAWs calculations discussed in Chapter 6.

Many thanks also to Ivan Walton and David Baker for their support regarding the computer systems and supercomputer Iridis on which I ran the molecular dynamics simulations.

Thank you to Dr Dmitry Lupyan at Schrödinger for allowing me access to the Simulations Interactions Diagram (SID) program which provided some valuable insights into the systems studied.

A special thank you to my amazing fiancé Graeme, and hugely supportive friends and family for all their love and encouragement over the course of my PhD.

Finally, thank you to BBSRC and Novartis for funding this project.

Chapter 1

1 Introduction

Drug resistance remains a major challenge for targeted cancer therapies such as tyrosine kinase inhibitors (TKIs). TKI resistance can be caused by a range of factors, one of the most significant being the development of point mutations in the kinase receptor. Further insight into how point mutations confer resistance is required to support the development of new TKIs that are effective against a range of mutated receptors [1-5].

The effects of point mutations can be assessed in a number of different ways. Experimentally this usually involves *in vitro* and *in vivo* techniques as well as x-ray crystallography and biomolecular NMR. *In vitro* trials typically involve a study where cells containing the mutant kinase are subjected to different inhibitors and the level of kinase activity or cell growth/survival is recorded. Cells with a good survival or growth rate, and kinases demonstrating a high level of activity indicate that the particular mutation present would confer drug resistance to the inhibitor used in the study [6]. *In vivo* trials evaluate how animal and patient survival is effected by the presence of mutations and response to different drug treatments [7, 8].

Computational techniques can be used to both evaluate what is known experimentally and predict likely outcomes in clinical practice [9-11]. The dynamics of proteins, [12-14] the shape and location of binding pockets [15-17] and the expected interactions between the protein receptor and ligand [18-20] can be investigated. When studying drug resistance, mutated protein structures can be compared with that of the wild-type to identify notable differences [21, 22]; this could include the loss of an interaction or a change in conformation either of which could have consequences on ligand binding [4, 9].

In this thesis three computational approaches; molecular dynamics, [23, 24] ligand docking [25, 26] and MD-GBSA [27, 28] have been applied to investigate the effects of point mutations in the BCR-Abl kinase domain and how they confer resistance.

A total of fifteen structures have been studied, adapted from three crystal structures. Two of these crystal structures were acquired from the Protein Data Bank [29]; the inactive structure of BCR-Abl in complex with imatinib (PDB ID: 1IEP [30]) and the active structure of BCR-Abl in complex with the aurora kinase inhibitor VX-680 (PDB ID: 2F4J [31]). The third crystal structure, inactive BCR-Abl in complex with nilotinib, was originally provided by Novartis but has since been made publicly available (PDB ID: 3CS9 [32]). As no apo structures of the inactive or active BCR-

Abl kinase were available, the ligand was removed from 1IEP and 2F4J to create the two pseudo-apo structures $1IEP_{p\text{-}apo}$ and $2F4J_{p\text{-}apo}$. Mutants were created from the two inactive structures 1IEP and 3CS9 by mutating the following residues in each structure: Threonine 315 (T315I); Tyrosine 253 (Y253F); Phenylalanine 317 (F317L); Tyrosine 253 and Phenylalanine 317 (Y253F_F317L, compound mutant 1); Threonine 315, Tyrosine 253 and Phenylalanine 317 (T315I_Y253F_F317L, compound mutant 2).

The final fifteen structures are listed below:

Crystal Structure	*Pseudo-apo*	*Mutant*
2F4J	$2F4J_{p\text{-}apo}$	none
$1IEP_{WT}$	$1IEP_{p\text{-}apo}$	$1IEP_{T315I}$, $1IEP_{Y253F}$, $1IEP_{F317L}$, $1IEP_{Y253F\_F317L}$, $1IEP_{T315I\_Y253F\_F317L}$
$3CS9_{WT}$	none	$3CS9_{T315I}$, $3CS9_{Y253F}$, $3CS9_{F317L}$, $3CS9_{Y253F\_F317L}$, $3CS9_{T315I\_Y253F\_F317L}$

The following chapters detail the current research methods used to study kinase point mutations (Chapter 2), the computational methods used in this research (Chapter 3) and detailed background information on the BCR-Abl tyrosine kinase and CML (Chapter 4). Chapter 5 is concerned with the structure and dynamics of active BCR-Abl and is used as a benchmark to compare to the WT inactive structure. In Chapter 6, molecular dynamics is used to study the inactive WT structure and compare the binding affinity of inhibitors imatinib and nilotinib via protein-ligand docking and free energy calculations. Chapter 7 models the drug resistant effects of the three single mutations (T315I, Y253F and F317L) on imatinib and nilotinib using molecular dynamics, protein-ligand docking and free energy calculations. Following on from this, Chapter 8 investigates how resistance is conferred when multiple mutations are present, with a study of two compound mutations, Y253F/F317L and T315I/Y253F/F317L.

Chapter 2

2 Studying point mutations in kinases

2.1 Introduction

2.1.1 Kinases

Proteins play a major role in various biological processes, including cell proliferation and survival. These processes are controlled via signal transduction pathways where one protein is activated by another protein, normally via phosphorylation; proteins that phosphorylate other proteins, via the addition of a phosphate group, are called kinases. Deregulated kinase activity is a significant cause of disease [33, 34] and in recent years kinases have played a critical role as targets for cancer therapy [35-38].

2.1.2 Point mutations

Point mutations cause a change in the structure of kinases by randomly replacing one of the single base nucleotides in DNA with another. For example, the code for a lysine residue at DNA level is TTC. If one of the T nucleotides is replaced with a G and the code becomes TGC, the resulting residue is a threonine [39]. Such spontaneous changes in the sequence of an oncogenic (deregulated) kinase have highlighted two major issues for researchers in drug development:

- If a specific mutation is highly prevalent in a particular oncogenic kinase (compared to the cellular kinase), this mutated variant could potentially be a target for inhibition. For example, around 50% of melanoma patients harbour a mutation of the BRAF kinase at position V600; the discovery of which has prompted the development of several inhibitors of the BRAF mutant kinase alongside companion diagnostic tests to identify patients with this disease variant [40-42].
- A mutation can cause a conformational change in the binding site, resulting in the diminished efficacy of inhibitors designed to bind to the original “wild type” (or WT) structure of the kinase. Therefore drug designers need to ensure that their inhibitors are effective against a number of mutations to minimise the occurrence of drug resistance [1, 4, 5]. Two examples of inhibitors susceptible to mutation conferred resistance are imatinib, which inhibits kinases Abl and c-KIT (for the treatment of chronic

- myeloid leukaemia [43, 44] and gastrointestinal stromal tumours [45, 46] respectively), and the EGFR kinase inhibitor gefitinib, which is used in the treatment of non-small cell lung cancer [3, 47, 48].

This thesis is concerned with the latter issue – where the presence of mutations results in drug resistance. Mutations in kinases such as Abl and EGFR typically cluster around four specific regions: the phosphate-binding or glycine rich loop (P-loop), the inhibitor binding site, the catalytic domain, and the activation loop (A-loop) [21, 43, 44, 49]. Mutations can affect inhibition both directly, through steric clashes and the elimination of critical residue contacts, and indirectly, by shifting in the thermodynamic equilibrium towards the active conformation [44]. Furthermore mutations can confer resistance through increasing ATP affinity, as in the case of mutation T790M in EGFR [50].

2.2 Experimental techniques

A range of experimental techniques are employed by researchers to investigate the effect of point mutations on the activity and structure of kinases and their ability to bind to inhibitors. A detailed discussion of all the experimental methods used in this field is outside the scope of this research; however a number of experimental techniques that are frequently used alongside computational methods are broadly discussed below.

2.2.1 X-ray crystallography

X-ray crystallography is used extensively in research to determine kinase structure and provide additional information on mechanisms of action or the specificity of protein-ligand interactions [51, 52]. In x-ray crystallography an x-radiation beam is passed through a singular, high quality protein crystal to create a diffraction pattern. The position and intensity of each spot of the diffraction pattern, together with the phase of the waves which formed each spot, must be determined in order to produce an electron density map, the basis for the 3D model. The known amino acid sequence of the protein is then fitted to the electron density map and subject to a number of refinements, such as the assignment of B-factors; this produces a set of X, Y, Z Cartesian coordinates for all non-hydrogen atoms. The resolution of a model is important and is related to the certainty of the position of atoms in the model; the greater level of certainty, the smaller the resolution value (a structure with a resolution of around 2Å is considered good) and the higher the quality of the model [53, 54].

2.2.2 Biomolecular NMR

NMR methods complement x-ray crystallography and are used by researchers to determine the structure and dynamics of proteins and protein-ligand complexes [55]. Although using solution NMR to determine the 3D structure of large protein-ligand complexes takes time, partial structural information can be quickly determined and useful dynamics data can be derived from the frequencies of internal motions of backbone or side-chain atoms; perturbations in chemical shift cluster around the regions of conformational change enabling the researcher to see the dynamic effects of a mutation or deduce the mechanism of action of a ligand [56, 57]. A similar pattern of perturbation in chemical shift between two structures or complexes indicates dynamic similarities. In addition to structural information, potential targets can be evaluated through NMR assay development and small scale NMR screening [58].

2.2.3 Hydrogen exchange mass spectrometry

Hydrogen exchange (HDX) mass spectrometry is widely used for studying the structure and dynamics of proteins [59, 60]. In HDX mass spectrometry the protein is exposed to D_2O, which causes rapid amide H → D exchange in regions of disorder (i.e. where the secondary structure is not tightly folded) and slow isotope exchange in the more ordered areas which are less exposed. Exchange events are modulated by the conformational fluctuations of the protein as regions become more or less exposed to D_2O depending on their conformation [61].

2.2.4 Assays

Researchers frequently utilise assays to measure the activity of a mutated kinase in the presence of an inhibitor and compare it to that of the wild-type [3, 37, 44, 62-65]. Assays can measure activity on a kinase level, monitoring the rate of phosphorylation and autophosphorylation, or cellular level, measuring cell growth or proliferation. Inhibition is generally measured as an IC_{50} value, the half maximal inhibitory concentration, which represents the concentration of drug required for 50% inhibition (of phosphorylation/cell growth etc.) *in vitro.*

Mutations within the kinase can be generated using a number of approaches; culturing cells in increasing concentrations of an inhibitor in order for drug resistant mutations to evolve; implementing random mutations in the initial kinase and expressing the mutant clones in cells grown in the presence of the inhibitor; or introducing individual mutations using site-directed mutagenesis. Due to their growth properties Ba/F3 cells (a murine interleukin-3 dependent pro-B cell line) are frequently used as a model system in mutagenesis studies and assays [66].

Additional approaches for determining binding affinity are traditional affinity chromatography [67, 68] or measuring the ability of an inhibitor to alter the melting temperature of a kinase [69].

2.3 Computational methods

In addition to the experimental techniques discussed in section 2.2, a number of computational methods are regularly used to obtain data on kinase structure, dynamics and potential as a drug target, "druggability".

2.3.1 Methods that investigate kinase structure and dynamics

As discussed in section 2.2.1 x-ray crystallography is regularly used by researchers to obtain information on protein structure. Unfortunately, the crystal structure only provides information on one specific conformation of the protein; it does not reveal data on other likely conformations and therefore the dynamic effects of mutations. Computational simulations, particularly molecular dynamics (MD) (see section 3.4) and Monte Carlo (MC) (see section 3.3) simulations are regularly used to provide insight into the dynamics of protein kinases.

2.3.1.1 Biomolecular simulation

The starting point for simulations is normally a crystal structure PDB file [29], which provides the initial structural information required by the MD/MC program. Several programs are available to mutate residues in the initial crystal structure to produce mutant versions [70, 71]. It is useful to keep the original unmutated crystal structure (known as the wild-type or WT) as a control. Each structure must then go through a series of preparatory processes (e.g. the addition of hydrogen and solvent atoms). Individual simulations are completed for each of the structures, providing a trajectory describing the motion of each system. These trajectories can then be analysed to determine the dynamic differences between the mutant structures and the WT structure [9, 10, 21, 22].

Monte Carlo methods use random numbers and statistical probability to model outcomes involving significant uncertainty. MC allows random moves from one configuration to another, where the outcome of each change depends on the preceding state. Although not time-dependent like MD simulations, MC simulations create similar trajectories that sample the flexibility and dynamics of the protein [24].

Molecular dynamics (MD) simulations calculate the time-dependent behaviour of a molecular system, based on classical molecular mechanics [24]. Although MD samples the general motion and flexibility of a protein system, it is not always possible to capture the large movements associated with conformational change. These events typically occur on a ms timescale (whereas most standard MD simulations are less than 50 ns in length) so would require extensive simulation time using standard MD. Targeted MD simulation methods, in which an external force is applied to a specific section of the protein structure, can be used to drive a desired conformational transition. In this method, the selected protein section is steered towards a position a specific root mean square (RMS) distance from the target structure. Both the receptor and the solvent are unrestrained and move in response to the steering force. The RMS distance is gradually decreased over the course of the simulation, resulting in a simulated structure close to the target structure and a trajectory capturing the transition [14, 72].

2.3.1.2 Binding site analysis

The analysis of protein binding sites has various applications in the field of structural based ligand-design, including the estimation of druggability (see section 3.8.5). Trajectories from MD or MC simulations are normally used since they sample a greater number of binding site conformations compared to a static structure. The majority of pocket analysis methods are either geometric or energy-based. Geometric methods use spherical probes or grid points to assess the surface of a protein and find any cavities, whereas energy-based methods search the protein for favourable interaction energies via the calculation of van der Waals interaction energies. Pockets are then ranked according to volume or total interaction energy [73-75].

2.3.2 Methods that investigate protein-ligand binding

Since mutations in a receptor can disrupt ligand binding, modelling the effect of specific mutations on binding affinity is of great value to researchers, enabling them to learn more about why certain mutations confer resistance and to predict the likely resistance profile for inhibitors currently in development. Free energy calculations (see section 3.9) and protein-ligand docking (see section 3.10) and are two popular techniques used to estimate binding affinities.

2.3.2.1 Free energy calculations

A number of approaches have been developed to calculate the relative free energies between two systems. One of the most widely used approximate methods is the Molecular Mechanics/Generalised Born (Poisson Boltzmann) Solvation Area (MM/GB(PB)SA) method.

The MM-GBSA and MM-PBSA methods estimate binding free energy by calculating the average free energy difference between the bound state (complex) and the unbound state of a solvated ligand and receptor [76-78].

A trajectory for the complex is determined through MD or MC simulations; for each snapshot of the trajectory the free energy is calculated for the receptor, ligand and complex and the binding free energy in found by subtracting the free energies of the receptor and ligand from the free energy of the complex:

$$\Delta G_{bind} = G_{com} - G_{rec} - G_{lig}$$

Where com, rec and lig stand for complex, receptor and ligand, respectively.

Mutations can be implemented during the preparation stage of the MD simulations [70, 71], thus creating a trajectory for the mutant receptor that can then be used to calculate its binding free energy with any number of ligands. Trajectory snapshots can also be used as a receptor ensemble for docking studies, with the benefit that using such an ensemble of conformations simulates protein flexibility [79, 80].

2.3.2.2 Protein-ligand docking studies

Traditionally docking studies involve multiple ligands being docked to multiple receptors in order to find a suitable complex; in mutational studies the method is generally targeted towards a specific receptor. As with the free energy calculations, mutated versions of the receptor can be created (containing single or multiple mutations per structure) [70, 71] and are docked to one or more inhibitors, so that data can be collected on how the efficacy of a drug might change depending on the location of a mutation [81-83]. Although a single average structure of the mutated receptor can be used it is more useful to submit an ensemble of conformations derived from an MD/MC simulation. Docking protocols are generally made up of a search algorithm (that predicts how the ligand fits to the protein by searching through the different possible orientations and conformations) and a scoring function (that evaluates and ranks the predicted conformations by estimating the binding affinity of each complex) [26, 84].

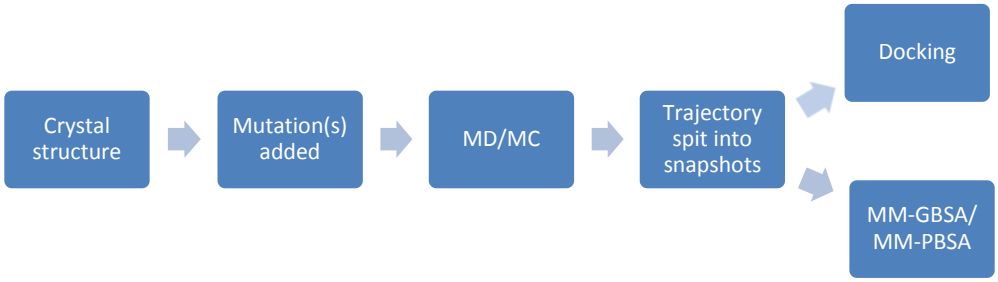


Figure 2.3.1: Simulation-based mutational study process

2.4 Chapter summary

A range of experimental and computational methods can be used to study the effects of point mutations. Traditionally x-ray crystallography, NMR spectroscopy, hydrogen exchange spectroscopy, and assays have been used by researchers to learn more about the structure, dynamics and druggability of a mutant protein kinase. Computational methods have been developed alongside these techniques to offer further insight and clarification on kinase behaviour and ligand-binding. The use of protein-ligand docking, free-energy calculations and biomolecular simulations are now an established part of the drug design process, capable of replicating experimental data. In the next chapter, these computational methods will be discussed in greater detail.

Chapter 3

3 Computational methods

3.1 Introduction

In the previous chapter computational methods used in the study of kinase mutations have been discussed. In this chapter Monte Carlo (MC), molecular dynamics (MD), protein-ligand docking, free binding energy calculations and the analysis methods used in this research are discussed in further detail.

3.2 Force fields

In molecular mechanics, force fields refer to the functional form and parameter sets that describe a system's potential energy. There are a number of different force fields available (examples include AMBER [27, 85, 86] and CHARMM [87-89]), however most include terms for the intra- and intermolecular forces, as shown in (3.1 [24, 90, 91]

$$U(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \quad \textbf{(3.1)}$$

In (3.1, $U(r^N)$ is the potential energy as a function of positions r of N particles; the bonds (stretching motions), angles (bending motions) and torsions (change in energy due to bond rotations) terms describe the intramolecular interactions, and the last term describes the intermolecular interactions. The intermolecular interactions term is described using a Lennard-Jones potential for van der Waals interactions (first part) and a Coulomb potential for the electrostatic contribution (second part) [24].

Both the bonds and angles terms are harmonic potentials based on Hooke's Law, where k_i is the stretching constant (i.e. the stiffness of the bond or angle). The bonds term gives the contribution to the energy when the bond length l_i deviates from the equilibrium value $l_{i,0}$. Similarly, the angles term gives the contribution to the energy when the angle θ_i deviates from

the equilibrium value $\theta_{i,0}$. For the torsions term, V_n is the constant reference energy, also referred to as the rotational barrier height, n is the multiplicity term (this denotes the number of minima in the function as the bond rotates through 360°), ω is the torsion angle, and γ is the phase factor (i.e. the phase shift along the rotational axis) [24, 90, 92].

Van der Waals interactions are described in (3.1 by the Lennard-Jones potential; ε_{ij} is the well-depth (i.e. the strength of attraction between the two particles), σ_{ij} is the van der Waals radius, (i.e. the distance where the intermolecular potential between the two particles is zero) and r_{ij} is the inter-atomic distance. Van der Waals interactions are short-range and quickly weaken as atoms move apart; repulsion occurs when the distance between atoms (i and j) becomes slightly less than the sum of their contact radii (σ) [24, 90, 93].

The electrostatic interactions term is described by a Coulomb potential, an effective pair potential that describes the electrostatic interaction between non-bonded pairs of atoms. In (3.1, q is the partial atomic charge, ε_0 is the electronic permittivity of free space and r is the distance between charges [24, 90, 93]. If information on long-range electrostatics are required, switching or Particle Mesh Ewald (PME) methods can be applied to calculations [24, 90, 94].

Force field parameters are frequently checked for accuracy against experimental results and quantum mechanics calculations for smaller molecules [24].

In this study, the all-atom force fields AMBER99 [85, 95] and AMBER03 [96] have been used.

3.3 Monte Carlo simulations

Monte Carlo (MC) methods use random numbers and statistical probability to model outcomes involving substantial uncertainty. Atoms or molecules are moved at random to create new configurations where each new configuration is dependent on the one that came before it. If the new configuration is energetically more favourable than the prior configuration, it is accepted. If the new configuration is energetically less favourable than the prior configuration, the Boltzmann factor of the difference in potential energy between the two configurations is calculated. The Boltzmann factor is then compared to a random number (between 0 and 1); if the Boltzmann factor is higher than the random number the new configuration is accepted, however if it is lower than the random number, then the new configuration is rejected [24].

In this research MC methods are used by the protein-ligand docking program Glide to assess ligand pose conformation (see section 3.10.4) [97-99].

MC applications are widely used by researchers to model biomolecular systems. Since only the energy of the current and suggested configurations decide acceptance, MC simulations are able to access substantial phase space compared to other techniques such as molecular dynamics (MD). This avoids problems where simulations are trapped in local energy minima [24]. Unfortunately time-dependent properties cannot be calculated using MC methods. If it is necessary to investigate a systems dynamics over time, other methods such as MD should be utilised [24, 90].

3.4 Molecular dynamics

Molecular dynamics (MD) is a method of simulating the behaviour of a system over time, based on classical molecular mechanics and Newton's second law ((3.2); where F_i is the force on atom i , m_i is the mass of atom i, and a_i is the acceleration of atom I [24, 100].

$$F_i = m_i a_i \tag{3.2}$$

The force F_i can also be described as the negative of the gradient of the potential energy, V ((3.3) [24, 100].

$$F_i = -\Delta V \tag{3.3}$$

Combining these equations and expressing F_i and a_i as derivatives, relates the derivative of the potential energy V to the position of atom r_i at a given time point ((3.4) [24, 100].

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{3.4}$$

Since a biomolecular system is made up of a large number of atoms, a finite difference method is used to solve the equations of motions simultaneously via Taylor series expansions. This method splits the integration into small steps of fixed time intervals (δt) and the force on each atom at time t is determined as the vector sum of the atom's interactions with other atoms in the system; where r is the position of the atom, v is the velocity, a the acceleration and b is the

first derivative of acceleration with respect to time. Within a given step the force is assumed to be constant. ((3.5, (3.6 and (3.7) [24]:

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \frac{1}{24}\delta t^4 c(t) + \cdots$$

(3.5)

$$v(t+\delta t) = v(t) + \delta t a(t) + \tfrac{1}{2}\delta t^2 b(t) + \tfrac{1}{6}\delta t^3 c(t) + \cdots$$

(3.6)

$$\boldsymbol{a(t+\delta t) = a(t) + \delta t b(t) + \frac{1}{2}\delta t^2 c(t) + \cdots}$$

(3.7)

Having determined the force acting on each atom, the acceleration can be calculated; the acceleration values, together with the positions and velocities of the atoms at time t, then allow the positions and velocities at time t + δt to be determined. The atoms will now be in a new position and the forces are re-calculated, resulting in new positions and velocities at time t + 2δt etc. [24].

One of the simplest and most widely used finite difference methods is the Verlet algorithm ((3.8) [101]. The Verlet algorithm calculates the positions of atoms at t + δt, r(t + δt) ((3.8) from the positions and accelerations at time t, and r(t - δt), i.e. the previous time step ((3.9) [24].

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \cdots$$

(3.8)

$$r(t-\delta) = r(t) - \delta t v(t) + \frac{1}{2}\delta t^2 a(t) - \cdots$$

(3.9)

Adding the two equations gives:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t)$$

(3.10)

A drawback of this algorithm is that since the velocities do not appear explicitly in the equation they must be calculated separately after the positions have been determined at the end of the step. In addition, the Verlet algorithm is not "self-starting" since it requires the positions of particles in the previous and current time steps, to calculate the next; therefore when t = 0, the positions for r(t - δt) must be determined through another method such as the Taylor series [24].

Various alternatives to the Verlet algorithm [101] have been developed, including the velocity Verlet [102] Beeman [103] and leap-frog [104] algorithms [24].

The velocity Verlet [102] algorithm calculates the atom positions, velocities and accelerations seemingly simultaneously and is the integrator used in the MD simulations presented in this thesis. First the positions of the atoms at t + δt are calculated using the velocities and accelerations at time t ((3.11). This enables the calculation of the new velocities using the acceleration at time t and t + δt ((3.12) [24].

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t)$$

(3.11)

$$v(t+\delta t) = v(t) + \frac{1}{2}\delta t[a(t) + a(t+\delta t)]$$

(3.12)

The velocities at t + ½ δt are then calculated using (3.13:

$$v(t+\delta t) = v(t) + \frac{1}{2}\delta t a(t)$$

(3.13)

The new forces can then be determined from the current positions, which gives a(t + δt), and finally the velocities at time t + δt are calculated using (3.14.

$$v(t + \delta t) = v\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t a(t + \delta t) \quad \textit{(3.14)}$$

Speed and computational efficiency, in addition to stability and accuracy, are important factors to consider when selecting an integration algorithm [24]. Relatedly, the timestep used in the simulation should not be too big as this results in the forces changing too suddenly and causes instability within the algorithm; small timesteps (e.g. 2 fs) are therefore typically used, but this then means that lengthy simulations are required to sample enough phase space [24, 90, 92].

One way of balancing stability and efficiency is to use the SHAKE [105] and RATTLE [106] algorithms dampen the high-frequency vibrations by applying constraints to the bonds, fixing them to their equilibrium values but enabling the other degrees of freedom to vary under the intramolecular and intermolecular forces present [24, 90, 92].

In this thesis NAMD [107] was used to complete the MD simulations.

In order for simulations to be comparable with experimental conditions, simulation parameters (e.g. temperature and pressure) must be appropriate [24, 92].

3.5 Thermodynamic conditions

The accuracy of a simulation can be checked by comparing the thermodynamic properties of a simulation, such as temperature, volume, pressure and energy, with available experimental data. These properties are used to define a particular thermodynamic state, which is referred to as an ensemble. A number of different ensembles can be used, notably the microcanonical ensemble, the canonical ensemble and the isothermal-isobaric ensemble [108].

The microcanonical (NVE) ensemble describes a system where the number of atoms (N), the volume (V) and the total energy (E) are fixed. This ensemble simulates a completely isolated system without heat exchange. Since there is no change in the total energy of the system the generated trajectories are considered as exchanges in kinetic and potential energy. In a canonical (NVT) ensemble, the number of atoms (N), the volume (V) and the temperature (T) are kept constant but not the energy. This simulates a system in thermal equilibrium with an external heat bath or thermostat (at temperature T) via the exchange of heat energy. The heat

capacity of the thermostat is theorised to be large enough to maintain a fixed temperature for the coupled system. The temperature is also fixed in the isobaric-isothermal (NPT) ensemble along with the number of atoms (N) and the pressure (P) [108].

Since, in reality, keeping the temperature constant which would effectively mean fixing the total kinetic energy, thermostat methods actually work by controlling the average temperature of the system, thereby approximating the canonical ensemble. Popular methods include the Berendsen thermostat [109], Nosé-Hoover thermostat [110], Andersen thermostat [111] and Langevin dynamics [112]. In this research Langevin dynamics was used to control the temperature of the MD simulations.

The Langevin thermostat [112, 113] simulates a system where the real particles are surrounded by many smaller virtual ones. These smaller particles influence the real system through random collisions and the application of a frictional drag force that decreases the velocities of the real system particles. In order to calculate this, the Langevin thermostat relies the Langevin equations of motion rather than Newton's equations of motion. At each time step all particles receive a Gaussian distributed random force and have their velocities decreased using a constant frictional force:

$$\vec{F}_i - \gamma_i v_i + \vec{R}_i(t) = m_i \vec{a}_i$$

(3.15)

Where $\overrightarrow{F_i}$ is the sum of all forces exerted on atom i, γ is the frictional force acting on the velocity of atom i and $\vec{R}_i$ is the random force [114].

This thermostat essentially reproduces the dynamics of a solute system surrounded by solvent if the friction coefficients are representative of the solvent viscosity. In explicit-solvent simulations, too small a value for the frictional coefficient will result in loose coupling and ineffective temperature control, i.e. if $\gamma \rightarrow 0$ the simulation will sample a microcanonical ensemble (NVE). If the coefficient is too high, resulting in tight coupling, the dynamics of the system are likely to be unrealistically altered by the large stochastic and frictional forces [115].

Similar to the temperature control in molecular dynamics simulations, it may be desirable to maintain the system at constant pressure. The isobaric-isothermal (NPT) ensemble is an adaption of the canonical ensemble and simulates a system in contact with a thermostat (at temperature T) and a barostat (at pressure P). As with the NVT ensemble the system exchanges heat with the thermostat but it also exchanges volume and work with the barostat [108, 115]. Simulations that utilise pressure control require the use of periodic boundary conditions.

3.6 Periodic boundary conditions

Calculating the forces for a large number of solvent molecules is computationally expensive; periodic boundary conditions (PBC) create the illusion of a bulk environment by replicating the cubic box that contains the protein and solvent in all directions to give a periodic array. When a particle leaves one box, an identical particle enters from the other side, therefore there are no surfaces and no surface effects. Given that all the boxes are identical, only the coordinates of the central box are recorded [24, 90, 92].

3.7 Solvent models

Solvent models allow researchers interested in simulating solvated systems to estimate free energy interactions between the solvent and solute. There are two main types of solvent model, explicit and implicit. In explicit models individual water molecules are positioned around the solute molecule and although no additional terms are added to the force field, new parameters are defined for the water molecules. In implicit models, also known as continuum models, the water molecules are represented as a continuous medium and the solvation effects are calculated by extra terms in the force field calculations [116, 117].

3.7.1 Explicit water models

There are a number of different approaches to modelling explicit water models [117]; the more simplistic water models comprise of rigid water molecules with defined interaction points. The pairwise electrostatic and van der Waals interactions in these models are modelled by Coulomb and Lennard Jones potentials.

The exact partial charge distribution, geometry and Lennard-Jones parameters depend on the model type and the number of interaction points. In models with three interaction points such as Simple Point Charge (SPC [118]) or Transferable Interaction Potential (TIP3P [119]), the interaction between two water molecules is calculated via the Lennard Jones function using one interaction point centred on the oxygen atom; van der Waals interactions involving the hydrogen atoms are not included in the calculation [24]. More advanced rigid models have 4 or 5 interaction points, utilising dummy atoms to improve the electrostatic distribution around the water molecule e.g. TIP4P [119], TIP5P [120, 121] and ST2 [122].

The MD simulations within this study have used the TIP3P explicit solvent model for water [119].

Since there is considerable computational expense associated with simulating explicit water molecules, an alternative option is to use implicit solvent models.

3.7.2 Implicit water models

Implicit models characterise the solvent as a continuous medium (with the average properties of the actual solvent) which surrounds the solute molecule from the point of the van der Waals surface [123]. In these models the energy component for the solvent is known as the free energy of solvation, ΔG_{solv} and consists of three distinct contributions: a free-energy term for the creation of the cavity in the solvent that the solute will occupy, ΔG_{cav}; a free-energy term for the electrostatic interactions between the solute and the solvent, ΔG_{elec}; and a free-energy term for the short-range van der Waals like interactions between the solute and the solvent, ΔG_{vdw}.

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{elec} + \Delta G_{vdw} \quad \textbf{(3.16)}$$

Several continuum models have been described including; Solvent Accessible Surface Area (SASA) [124], Poisson Boltzmann Equation (PB/PBE) [125, 126], generalized Born (GB), [127, 128] and generalized Born Surface Area (GBSA) [129].

The simplest implicit solvent model is one where the solvent-induced interaction is estimated using the solvent-accessible surface area (SASA) of the macromolecule and a parameter value that assumes the free energy of interaction between solute and solvent can be considered as a sum of energies of atomic groups [24].

$$\Delta G_{solv} = \sum_i \sigma_i A_i \quad \textbf{(3.17)}$$

where A is the accessible surface area of molecule i and σ is the atomic solvation parameter for atom i based on experimental values [124]. Although successfully used in a number of applications in molecular modeling, the SASA method is not deemed as rigorous for calculating the electrostatic contribution as other methods such as those based on the Poisson Boltzmann equation [125, 126].

The Poisson Boltzmann equation (PBE) is a differential equation that describes the electrostatic interactions between molecules in ionic solutions [24, 125, 126, 130]:

$$\nabla \cdot [\epsilon(r)\nabla\emptyset(r)] - \kappa' sinh[\emptyset(r)] = -4\pi\rho(r)$$

(3.18)

Although theoretically accurate, PBE methods are computationally expensive and so other approximation methods, such as Generalised Born (GB) models, have been developed for the calculation of continuum electrostatics. The GB approximation is based on the PB theory, but models the molecule as a series of spheres where the internal dielectric is different to that of the solvent [76, 129].

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f(r_{ij}, a_{ij})}$$

(3.19)

Where ε is the dielectric constant of the solvent, q_i is the charge on atom i, q_j is the charge on atom j, r_{ij} is the distance between atoms i and j and a_{ij} is the born radii [24].

The generalized Born model is often used in combination with the SASA estimate of the non-polar solvation energy. Although not as rigorous as PBE, this method, known as the GBSA model allows the solvation energy to be quickly calculated and so has become a popular model for MD simulations [28, 131].

In this research the GBSA model has been used to estimate binding affinity values (see section 3.10).

3.8 Simulation analysis techniques

3.8.1 Root-mean-squared-deviation (RMSD)

Root-mean-squared-deviation is a calculation of the average distance between a reference structure and selected atoms, where δ is the distance between N (backbone) atoms [24].

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

(3.20)

As an average is taken over all the atoms, calculating the RMSD for an MD simulation will provide time specific values. This enables researchers to see in the resulting (RMSD vs. time) graph the point in the simulation where RMSD was highest or lowest.

In this thesis RMSD was calculated using the *g_rms* function in GROMACS [132] and is one of the methods used to compare mutant structures of BCR-Abl with the wild-type structure.

3.8.2 Root-mean-squared-fluctuations (RMSF)

Root-mean-squared-fluctuation is a calculation of the deviation of the position of an atom i from a reference structure, such as the average structure from the MD trajectory [133, 134].

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{r}(x_i - \bar{x})^2}$$

(3.21)

Since an average is taken over time, the RMSF provides a value for each atom or residue in a protein structure. This enables researchers to highlight the atoms or residues with the highest (or lowest) mobility over the course of the simulation. RMSF values are frequently compared to B-factor values to check the quality of a simulation. B-factors are a measure of disorder and thermal motion captured at the crystallisation stage that are provided alongside the x, y, z coordinates of a PDB file. Atoms with large b-factor values are typically positioned along the parts of the protein that are most flexible, e.g. loops, and should therefore be reasonably aligned with calculated RMSF values [135, 136]. Since crystal packing effects can cause inaccuracies in the calculation of B-factors, it is useful to review and compare several different crystal structures of the same protein [137].

In this thesis RMSF was calculated using the *g_rmsf* function in GROMACS [132] and is used to emphasise which regions of the kinase structure are most flexible during the MD simulation.

3.8.3 Distance analysis

In order to further analyse the protein structures, it is often useful to measure the distance between two residues (or a protein residue and a ligand) over the course of a simulation.

In this thesis distance analysis measurements were performed in VMD [138] and used to study interactions between individual protein residues (e.g. salt bridges).

3.8.4 Principal component analysis (PCA)

PCA simplifies a dataset by reducing its dimensions to the representative principal components, with the first principal component (eigenvector with the highest associated eigenvalue)

contributing the most to the variance in the dataset. The process can be broken down into several key steps; creating and centring the dataset, calculating the covariance matrix, diagonalization of the matrix (finding the eigenvectors) and finally ordering the eigenvectors by eigenvalue [24, 90].

First the number of dimensions for each atom is calculated (number of dimensions is equal to the number of atoms multiplied by 3) and the average structure is subtracted across each dimension so the mean of the dataset is equal to zero. This removes the centre of mass rotation and translations, which are irrelevant to the internal motions of the system.

A covariance matrix C_{ij} is then created for the positional fluctuations, for coordinates i and j and determines the correlation of the motions:

$$C_{ij} = \langle (x_i - \langle x_i \rangle) \left(x_j - \langle x_j \rangle\right) \rangle$$

(3.22)

where ⟨ ⟩ is the average of the structure over the simulation [139].

When the number of dimensions is greater than two, there is more than one covariance measurement to be calculated. For three dimensions, covariance calculations for (x,y), (x,z) and (y,z) are carried out and the covariance matrix C is a symmetric 3N x 3N matrix, where N refers to the number of atoms.

The covariance matrix is then diagonalized to determine the eigenvalues:

$$A^T C A = \lambda$$

(3.23)

where A is the eigenvectors and λ the eigenvalues [139].

The eigenvectors are then arranged by eigenvalue in descending order. The first principle component of the data set is the eigenvector with the highest eigenvalue.

Since the principal components define a transformation to a new coordinate system the trajectory can be fitted onto the eigenvectors to give the PCs.

$$q(t) = A^T M^{\frac{1}{2}} (x(t) - \langle x \rangle)$$

(3.24)

The trajectory can then be filtered along particular PCs to isolate the motions from specific principal components,. For PC i this is obtained by:

$$x^f(t) = \langle x \rangle + M^{-\frac{1}{2}} A_{*iqi}(t)$$

(3.25)

Where A_{*i} denotes the i^{th} eigenvector of the matrix C_{ij} (the i^{th} column of A) [92].

Generally only the first few of the total number of eigenvectors are considered relevant and useful for simulation analysis. The Kaiser criterion [140] and Cattell Scree test [141] can help determine the appropriate number of eigenvectors to evaluate.

In this thesis, PCA has been performed using the *g_covar* and *g_anaeig* modules of GROMACS 3.2.1 [132].

3.8.5 Binding site analysis

The identification and analysis of protein binding sites has various applications in the field of structural based ligand-design including automated drug target annotation, virtual screening and protein function prediction. The majority of methods available are either geometric or energy-based [142].

Geometric methods such as SURFNET [143], LIGSITE [74, 144], PASS [145], CAST [146, 147] and PocketPicker [148] use spherical probes or grid points to assess the surface of a protein and find any cavities. Although generally speaking these methods rely completely on geometric criteria with no account for the "druggability" of a pocket, this is often enough to be successful, since the majority of binding sites are found in the largest pocket [142].

Energy-based methods, like PocketFinder and Q-SiteFinder [75], search the protein for favourable interaction energies by calculating the van der Waals interaction energy of a methyl probe on a grid. The probes then form clusters throughout the protein; PocketFinder ranks the clusters by volume, while Q-SiteFinder sorts the clusters according to their total interaction energies.

In this thesis PocketAnalyzer [149], an adapted version of the LIGSITE [144] algorithm, was used to compare (rather than predict) differences in the binding sites of the wild type structures.

3.8.5.1 PocketAnalyzer

PocketAnalyzer [149] identifies pockets is similar way to the LIGSITE algorithm [144]; a grid is mapped across the protein and points are designated as part of a pocket if:

- they are not within the van der Waals radius of another protein atom,
- they are sufficiently buried within the protein structure (defined as “degree of buriedness” or *dob*),
- they are neighboured by a certain number of other well-buried atoms (defined as “minimal number of neighbours” or *mnb*),
- they are part of a cluster of points (that meet the above criteria) greater than the “minimal cluster size” or *mcs.*

The default values for the parameters are *dob* = 11, *mnb* = 15, and *mcs* = 50 with a grid spacing of 0.8A, but these parameters can be tailored to suit the protein system under investigation. Once the parameters have been assigned the protein structures must be aligned (in order for the protein to be analysed on the same mapped grid) before the program is run.

3.8.5.2 Simulation Interactions Diagram (SID)

The Simulation Interactions Diagram (SID) is an analysis tool within the Desmond Molecular Dynamics system [150, 151]. The SID calculates information on protein-ligand interactions for a given MD trajectory, including protein and ligand RMSD/RMSF, protein-ligand contacts and ligand torsions. In this thesis the SID has been used to gain further information on how critical protein-ligand interactions are affected by the presence of mutations.

3.9 Free energy calculations

Free energy calculation methods can be broadly put into two categories; rigorous methods and approximate methods. Rigorous methods include free energy perturbation and thermodynamic integration (including finite difference thermodynamic integration and replica exchange thermodynamic integration). One of the most widely used approximate methods is the Molecular Mechanics - Generalised Born/Poisson Boltzmann Solvation Area (MM-GB/PBSA) method.

3.9.1 Free energy perturbation

Free energy perturbation (FEP) calculates free energy via the use of intermediate states between the reference (starting) and the perturbation (final) states [152].

The Zwanzig equation [153] describes that the free energy difference between reference state A and perturbed state B can be expressed as:

$$\Delta G_{A-B} = -k_B T ln \langle exp\left(-\frac{\Delta E}{k_B T}\right)\rangle$$

(3.26)

Where $\langle exp\left(-\frac{\Delta E}{k_B T}\right)\rangle$ represents the ensemble average over system A and ΔE represents the change in energy between states A and B. Since the exponent's contribution to the overall average is dependent on the extent to which the two states' phase space overlaps, states A and B are linked via a coupling parameter, λ, which introduces intermediate states. This works by defining the reference state A as λ=0 and the perturbed state B as λ=1, with many intermediate states in between, each with a discrete λ value from which the free energy of that reference state is calculated.

The energy between each λ intermediate is calculated, and then used in the equation below:

$$\Delta G = \sum_{\lambda}^{1} -k_B T ln \langle -\frac{\Delta E'}{k_B T} \rangle$$

(3.27)

ΔE' is the energy difference between the states λ+Δλ and λ, where Δλ is the interval between two successive λ intermediates [24].

3.9.2 Thermodynamic integration

Thermodynamic integration (TI) calculates the rate of change in free energy with respect to λ across the trajectory and then integrates the gradient to obtain the relative free energy:

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \langle \frac{\delta G}{\delta \lambda} \rangle_{\lambda} \, d\lambda$$

(3.28)

Finite difference thermodynamic integration (FDTI), is an adaption of TI that uses FEP theory (finite difference approximation) to calculate the free energy gradient rather than partial derivatives [154].

Replica exchange thermodynamic integration (RETI) [155, 156] uses a combination of FDTI [154] and Hamiltonian replica exchange [157, 158] to calculate free energy values. The λ coordinate scales the force field terms linearly, resulting in a different Hamiltonian for each λ value and the coordinates between neighbouring λ values are periodically swapped according to the following Metropolis test:

$$rand(0,1) \leq \exp\left[\frac{1}{k_B T}\ (E_B(j) - \ E_B(i) - \ E_A(j) + E_A(i))\right]$$

(3.29)

A and B are the two λ values and i and j are replicas of the system at A and B [155, 156].

3.9.3 MM-PB/GBSA

Approximate methods only take into account the beginning and end of a simulation rather than calculating intermediate values, which makes them much faster to complete. One of the most widely used approximate methods is the Molecular Mechanics Poisson Boltzmann / Generalised Born Solvation Area (MM-PB/GBSA) method [76-78].

In MM-PB/GBSA the endpoint is determined through MD or MC simulations and for each snapshot of the end trajectory the binding free energy is calculated by subtracting the free energies of the receptor and ligand from the free energy of the complex:

$$\Delta G_{bind} = G_{com} - G_{rec} - \ G_{lig}$$

(3.30)

Where com, rec and lig stand for complex, receptor and ligand, respectively. The free energy of each component was calculated using:

$$\Delta G_{bind} = \langle \Delta E_{mm} \rangle + \Delta G_{solv} - T\Delta S$$

(3.31)

Where ΔE_{mm} is the difference in the molecular mechanics energy between the complex and the isolated protein and ligand, and G_{solv} is the difference in the solvation free energy between the complex and the individual components [76].

The solvation free energy is calculated using either a Generalised Born or Poisson Boltzmann solvation model, with a hydrophobic solvent accessible surface area term (see section 3.7.2) [24].

The entropic term ΔS is often ignored when structurally similar ligands are compared, since it is assumed that they will have similar values for ΔS [159].

Although MM-PB/GBSA is not considered to be as accurate as FEP or TI in terms of standalone calculations; it is widely used in studies involving protein-ligand docking, e.g. to rescore docking poses [160]. Although the absolute free energy values calculated may not be in direct agreement

with experimental values, the ranking of a set of ligands should correlate with the ranking based on experimental binding affinities; the more negative the value, the stronger the binding.

In this thesis the free energies were calculated using the python script (pymdgbsa.py) written by Romain Wolf at Novartis [161].

3.9.4 Free energy calculations using pymdgbsa.py

The pymdgbsa.py script uses a MD trajectory to perform an MM-GBSA calculation for receptor-ligand interaction free energies [161]. The script calculates the binding energy between the ligand and the receptor for each frame of the trajectory and then calculates the average. In the most recent version of AmberTools there is a similar MMPBSA.py module that can also be used to calculate free energies [161-163].

3.10 Protein-ligand docking

Protein-ligand docking is a computational technique used to predict the orientation and binding affinity of a ligand in the binding site of a target receptor, and is an important part of the modern drug design process. Generally, docking protocols are made up of two distinct functions; a search algorithm and a scoring function. The search algorithm predicts how the ligand fits to the protein by searching through the different possible orientations and conformations of the protein-ligand pair. The scoring function then evaluates and ranks the conformations or “poses” predicted by the search algorithm by estimating the binding affinity of each complex [26, 84, 164]. Methods have evolved to take into consideration the relative flexibility of the ligand and its receptor, since binding often induces a conformational change in the protein (“induced fit”).

3.10.1 Search algorithms

The individual poses of a ligand are assessed for suitability by calculating the root-mean-square deviation (RMSD) compared to the native pose (if known); the smaller the RMSD the closer the pose is to the native. The accuracy of the search algorithm relies on the simulation being as close to a real system as possible, however this requires an extremely high number of degrees of freedom to be included in the conformational search which severely effects the efficiency of the search algorithm.

The number of degrees of freedom sampled can be reduced by:

- approximating (using an implicit model in the scoring function) or excluding any solvent effects from the docking protocol;

- restricting the flexibility of the ligand and protein to varying extents, e.g. allowing ligand flexibility while keeping the protein rigid, therefore only sampling the conformational space of the ligand.

The main challenge in search algorithm design is limiting these approximations while maintaining efficiency.

Ligand-search docking protocols include systematic docking algorithms, stochastic algorithm and simulation methods.

Systematic search algorithms aim to sample all degrees of freedom, the three main types of systematic search algorithms are; exhaustive conformational search methods, fragmentation methods, and conformational database/ensemble methods [26].

- Exhaustive conformational search methods involve the incremental 360° rotation of all the rotatable bonds in the ligand, until all possible combinations have been considered. Despite the comprehensiveness of sampling associated with this method, the number of combinations increases exponentially according to the number of rotatable bonds (combinatorial explosion), therefore to make the docking process more computationally efficient, geometric constraints are used in the initial screening of poses, and the ligand conformations are then further assessed through a series of refinement processes [26]. This type of hierarchical sampling method is used by Glide [97-99] and FRED [165].

- Fragmentation methods (e.g. LUDI [166], FlexX [167] and DOCK [168]) simulate ligand flexibility by incrementally "growing" the ligands into the active site, using either a "place and join" or "incremental" approach. The "place and join" approach docks several fragments into the active site and then re-links them covalently. The "incremental" approach splits the ligand up into a rigid core-fragment with separate flexible regions, first docking the core and then planting the flexible fragments around it [26].

- Conformational database/ensemble methods (e.g. FLOG [169]) use libraries of 3D coordinates to find suitable ligands for a known receptor. Flexibility is simulated by each

candidate ligand structure having an ensemble of associated conformations already generated within the database [26].

Rather than attempting to systematically sample all possible conformations, stochastic search algorithms, implement a change at random to a single ligand or a population of ligands, and then either accept or reject it based on specific probability criteria. Stochastic search algorithms can be broadly categorised into four main types of method: Monte Carlo methods (MC), genetic or evolutionary algorithm methods (GA/EA), Tabu Search methods and swarm optimisation methods (SO) [26].

- In Monte Carlo methods (e.g. Prodock [170] and ICM [171]), the Boltzmann probability function is used for the acceptance criteria. Since only the energy of the current and suggested states decide acceptance, MC methods are able to access substantial phase space compared to other techniques such as MD; this avoids problems where conformations become trapped in local energy minima [24].

- Genetic or evolutionary algorithms (e.g. GOLD [172, 173] and Autodock 3.0 [174-176]) apply evolution theory to docking. Starting with an initial population of ligand poses, where each conformation is defined by a specific "genotype", (a set of "genes" that describe the ligand's translation, orientation and conformation in relation to the target protein) and "phenotype" (the atomic coordinates). Conformational space is sampled by applying genetic operators, such as mutations, to the population until it satisfies the required fitness criteria [26].

- Tabu search methods (e.g. PRO_LEADS [177, 178]) steer the search to sample a wider area of conformational space by restricting the sampling of previously explored areas. The algorithm records a list of previous pose data and accepts a new conformation depending on the RMSD of the new conformation in relation to all previously recorded conformations [26].

- Swarm optimization (SO) algorithms (e.g. SODOCK [179], Tribe-PSO [180], and PSOAutodock [181]) utilise swarm intelligence to govern the ligands optimal position in conformational space. A number of potential solutions, known as particles, move through space according to a given formulae; each particle is influenced by both its own best known position and the best position found by the other particles, which moves the "swarm" toward the optimal solution.

Simulation methods, notably MD and energy minimization, use Newton's equations of motion to evaluate possible protein-ligand complexes [26].

- MD is a widely used simulation technique despite the previously mentioned problems with overcoming energy barriers and computational expense. Strategies to compensate these limitations, such as the use of periodic high temperature annealing, have been developed [182].

- Energy minimisation methods are normally used in combination with another search algorithm rather than a standalone method (e.g. Prodock, ICM, DOCK 4.0). Techniques include: steepest descend, conjugate-gradient and second-derivative methods (e.g. Newton-Raphson) [24] as well as direct search [183] and least-squares methods [184];.

Flexible ligand search docking methods are relatively successful, particularly in cases involving a predominantly rigid receptor and where there is a strong correlation between the crystallographic structure of the receptor and its actual conformation in the docked complex [185]. Unfortunately, many systems exhibit significant conformational change upon ligand binding, and since such protein flexibility is not modelled in these protocols, the accuracy of the docking results can be severely affected.

3.10.2 Scoring functions

Scoring functions assess and rank the ligand conformations found by the search algorithm, by approximating their relative binding affinity with the target protein. The mathematical methods required to do this are complex and in order to balance the need for accuracy without excessive computational expense, the scoring process must be simplified in some manner. The scoring functions used in protein–ligand docking are made up of three main types: force field-based, empirical, and knowledge-based scoring functions [26].

Force field based scoring functions involve the summation of the internal energies (as described in section 3.2 Force fields) of the ligand as well as the interaction energy between the receptor and the ligand. Difficulties in simulating the solvation and entropic terms, and inaccuracies calculating the long-range effects involved in binding have long been cited as the main limitations with this method, however several scoring functions have been updated in recent years to take these terms into account. Examples of force field scoring functions include GoldScore [186] and AutoDock [174, 175, 182]. In addition, the computationally expensive molecular mechanical force-fields such as AMBER (Assisted Model Building and Energy

Refinement) [27, 96, 187] and OPLS (Optimized Potentials for Liquid Simulations) [188] can also be adapted for scoring [26].

Empirical scoring functions use experimental data to derive coefficients that are then used to predict the free binding energy of a given ligand-receptor complex. Although computationally economical, empirical scoring functions are limited by the original experimental data used to define the coefficient terms [26]. Examples of empirical scoring functions include ChemScore [189], SCORE [190, 191], Fresno [192], and X-SCORE [193].

Knowledge based scoring functions use predefined rule sets to suggest reliable geometries in agreement with experimental data; i.e. poses that most closely resemble experimental structures are ranked the highest, based on the number of certain atom–atom pair contacts and other interactions [26]. This method is limited by what information available from the structural data it uses, structures resolved by X-ray crystallography or NMR will not, for example, provide any data on protonation states. DrugScore [194, 195], Muegge's PMF [196, 197], and SMoG score [198] are examples of knowledge-based scoring functions.

Consensus scoring checks data from various scores, in an effort to avoid the inaccuracies associated with individual scoring functions [26, 199-201]. The way in which the score data is correlated requires careful consideration in order to reduce inaccuracies rather than amplify them; however several studies have shown this approach to be more successful than using the individual scoring functions [202, 203]. X-CSCORE [193], is a consensus scoring method that combines the ChemScore, PMF and FlexX scoring functions with DOCK-like and GOLD-like algorithms.

3.10.3 Modelling protein flexibility

Implementing full protein flexibility remains a challenge, however several docking approaches have been developed that allow for additional flexibility. These include molecular relaxation methods using MD or MC simulations, protein ensemble docking [79, 80], soft docking [204, 205], and side chain flexibility [26, 206].

Applying simulation methods to the protein as well as the ligand greatly increases the conformational space sampled by the search algorithm, however since these types of search algorithm involve calculating a greater number of ligand/protein combinations, they take longer to complete. Most flexible protein docking methods restrict protein flexibility to key regions in order to reduce simulation time [26].

Protein ensemble docking emulates protein flexibility by using a set of protein conformations as the receptor target rather than just a single structure. The conformations can come from various sources including crystal structures, NMR or derived from simulations [80, 207-209].

Soft docking methods model protein flexibility implicitly by softening the interatomic van der Waals interactions in docking calculations, thus allowing a small amount of overlap between the ligand and the protein. Although efficient, a limitation of this approach is that large scale motion can result in the superficial enlargement of the active site by allowing mutually exclusive binding regions to be simultaneously considered [204, 205] .

Side-chain flexibility was initially modelled by representing the receptor conformational space as a set of rotametrics for each side chain based on experimental results [206]. A limitation of this method is that it does not take into account the motion of the protein backbone. In recent years, additional techniques have evolved to take into account continuous or discrete side-chain flexibility [210-212].

3.10.4 Ensemble docking using Glide

In this thesis, protein-ligand docking studies have been completed using the Glide program (Grid-based LIgand Docking with Energetics) [97-99], in Maestro, Schrödinger's commercial molecular modelling package. As part of its default protocol, Glide uses several filters to perform a comprehensive search of available ligand poses (Figure 3.10.1). For every ligand pose assessed, Glide takes into account; the positions and orientation of the ligand in relation to the receptor, the core conformation of the ligand, as well as the conformation of the ligands rotamer groups.

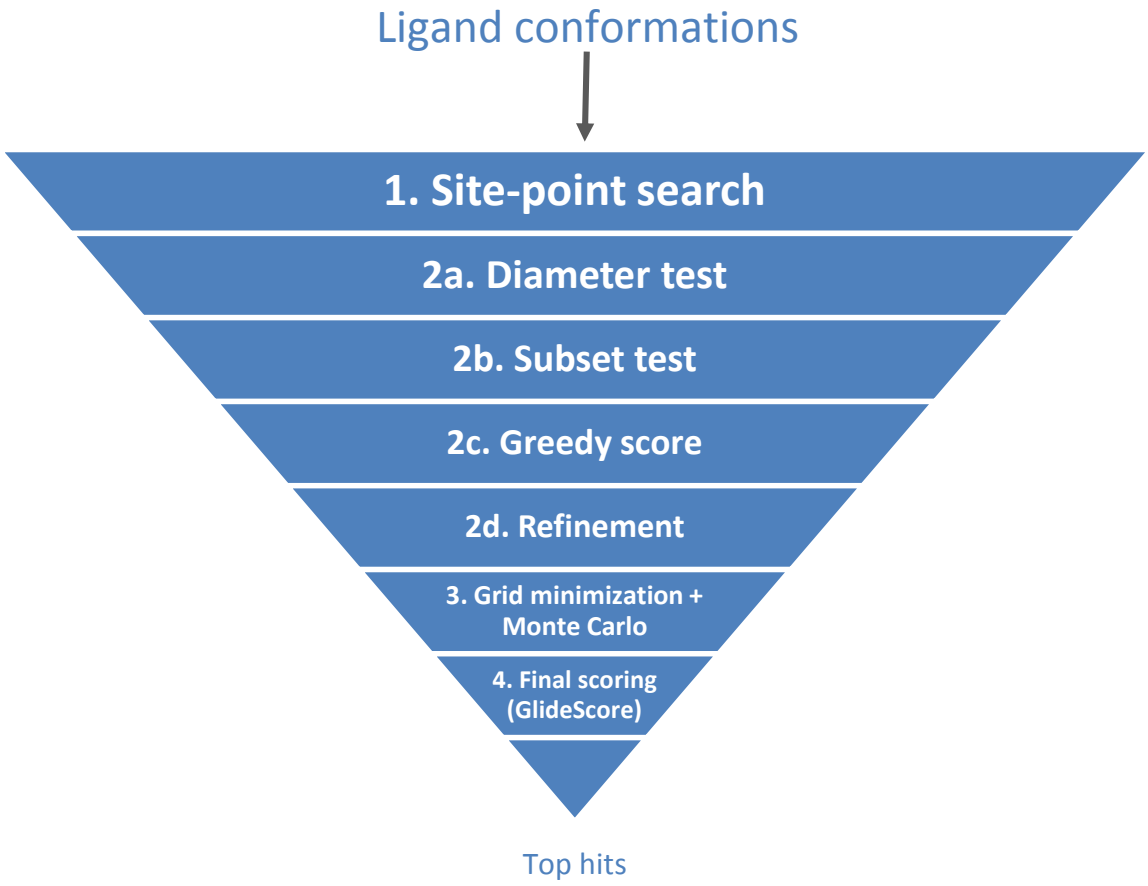


Figure 3.10.1: Glide docking "funnel", showing the Glide docking hierarchy. Adapted from reference [97].

As a pre-processing step the receptor is represented via a number of "site points" on an equally spaced 2 Å grid [97]. The second stage (step 2a – diameter test, Figure 3.10.1) evaluates the atoms within a specified distance of the ligand-diameter axis for a specified number of orientations; an orientation is skipped if it causes too many steric clashes with the receptor. The ligand is then rotated about the ligand diameter and the subset of the atoms that make hydrogen bonds or ligand metal interactions with the receptor are scored (step 2b – subset test, Figure 3.10.1), if this score is acceptable, the remaining ligand-receptor interactions are also scored (step 2c – greedy score, Figure 3.10.1). The scoring for tests 2a – 2c utilises a version of ChemScore [189] where, for each atom type a discrete, predefined score is assigned to 1 $Å^3$ boxes. Like ChemScore, this algorithm recognizes favourable hydrophobic, metal-ligation, and hydrogen-bonding interactions and imposes penalties for steric clashes. The score for each atom is dependent both on its position relative to the receptor and the improved score it could get by moving 1 Å along the X, Y, and/or Z axis. In order to refine the number of poses at this point, the

top greedy-scored poses are rescored as the entire ligand to moves (rigidly) 1 Å along the X and Y axis [97, 98].

Using the poses selected by the screening described in step 2, the ligand is minimized using the OPLS-AA force field [188] together with a distance-dependent dielectric model. Finally, in order to refine the orientation of ligand peripheral groups, a Monte Carlo method is used to examine the torsional minima of the lowest-energy poses selected from the minimisation process [97].

Emodel, (a combination of GlideScore, the ligand-receptor molecular mechanics interaction energy and the ligand strain energy) is utilised to obtain the best ligand pose. Finally, Glide reduces the van der Waals radii of selected ligand atoms so that the ligand fits more easily into the binding pocket. If required, this can be altered via a scaling parameter for increased enrichment [99].

The ChemScore function [189] from which GlideScore is derived can be described as:

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{lr}) + C_{hbond} \sum g(\Delta r)h(\Delta\alpha) + C_{metal} \sum f(r_{lm)} + C_{rotb} H_{rotb}$$

(3.32)

The summation in the second term encompasses all ligand/receptor atom pairs described by ChemScore as lipophilic, and the third term incorporates all ligand-receptor hydrogen-bonding interactions. The functions *f*, *g*, and *h* give a full score of 1.00 for distances or angles that lie within nominal limits and a partial score between 1.00 and 0.00 for distances or angles that fall beyond them (for a given threshold) [97].

GlideScore modifies and extends the ChemScore function as follows:

$$\Delta G_{bind} = C_{lipo-lipo} \sum f(r_{lr}) + C_{hbond-neut-neut} \sum g\,(\Delta r)h(\Delta\alpha) + C_{hbond-neut-charged} \sum g\,(\Delta r)\,h(\Delta\alpha) + C_{hbond-charged-charged} \sum g\,(\Delta r)\,h(\Delta\alpha) + C_{max-metal-ion} \sum f(r_{lm}) + C_{rotb} H_{rotb} + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} + solvation\ terms$$

(3.33)

The lipophilic-lipophilic and hydrogen bonding terms are essentially the same as in ChemScore, but the hydrogen-bonding term is split into different sections depending on whether or not the donor/acceptor is charged.

The metal-ligand interaction term (the fifth term in (3.33) has had several adaptions compared to ChemScore; only considers interactions with anionic acceptor atoms [97, 213]; only counts the single best interaction when two or more metal ligations are found; and assesses the charge on the metal ion in the unligated apo protein. The seventh term uses active site mapping to identify and reward poses where a polar, non-hydrogen-bonding atom lies in a hydrophobic region.

The second major component involves the energetic contributions from the Coulomb and van der Waals interaction energies. The net ionic charge on charged groups such as carboxylates and guanidiniums, as well as the van der Waals interaction energies for the atoms directly involved, is reduced in GlideScore to minimise the disparities in interaction energies sometimes seen with ChemScore [97, 99, 189].

An additional adaption, which has increased the identification of false positives found by the scoring function, is the inclusion of a solvation model. In reality, the charged and polar groups of the ligand and protein would be subject to solvation effects, however many docking protocols either overlook this, or use continuum solvation models to simulate these effects [26, 214]. Unfortunately it is the location of individual water molecules that becomes important in the scenario a ligand bound active site and current continuum solvent models are not able to represent these interactions. Glide docks explicit water molecules into the binding site for each favourable ligand pose and employs empirical scoring terms that measure the exposure of various groups to explicit waters [97-99].

3.11 Water molecule prediction in protein binding sites

Several modelling studies have shown that ligand-bound crystallographic water molecules can make significant contributions to the protein-ligand free energies by establishing a network of hydrogen bonds which stabilize the protein-ligand interaction [214-221]. Unfortunately, data regarding the position of ligand-bound water molecules is often unknown (due to difficulties in resolving water molecules by crystallography) or not taken into account during molecular docking studies [26].

Methods of predicting which water molecules contribute to protein-ligand binding have generally been limited to less rigorous empirical techniques based on interaction energies or by

using time-consuming MC or MD simulations [222-226]. Faster Grand Canonical MC (GCMC) methods have also been described; however the acceptance rate is generally poor and the free energy change in the μVT ensemble is not aligned with the free energy changes under standard NPT conditions [227, 228]. Thermodynamic integration simulations [221, 229-231] have been shown to reliably locate hydration sites, however given the computational expense involved, this technique is not recommended for use on systems with a high number of water molecules. A similar technique, known as JAWS, is a more efficient alternative that features "θ-water" molecules that can appear and disappear on a binding-site grid [226, 232, 233].

In this thesis the *grid* module in *ptraj* (AMBER) [162] and the free-energy based JAWS method [226, 232, 233] have been used to determine the position of water molecules in protein binding sites.

3.11.1 Ptraj

The grid module in ptraj [234] (part of the AMBER toolkit [162]) creates a grid in X-Plor density format with a count of the specified atoms (e.g. water molecules) in each grid cell for a given MD simulation. The "average" output shows the most common position of the water molecules throughout the simulation, which gives an indication of the position of hydration sites within the pocket [234].

3.11.2 JAWS

The JAWS method [226, 232, 233, 235, 236] involves a series of "θ-water" molecules on a grid across the binding-site that can appear and disappear [226]. θ is used as an energy scaling parameter, and this controls the interaction energy $U_i(r)$ between θ-water i and the rest of the system [236].

The initial part of the simulation determines the probability of the molecule being 'water-like' or 'ideal' with a given number (N) of θ water molecules distributed at random on the 3-dimensional grid. Attempted moves in the MC simulation that follows include rigid body translations, rotations and a random variation in θ (in 50% of the moves). After every 100 moves the θ_i values are evaluated and, if greater than 0.95, a counter on the nearest grid point is increased by one. These counters increase as the simulation progresses developing a probability distribution of water occupancies, and resulting in a mapped output of estimated hydration sites. The second part of the simulation checks the predicted hydration sites by calculating the binding free energy for a θ water molecule at each identified site. The θ_i water molecules sample

both high and low values of θ to obtain a reliable free energy estimate. If the binding free energy is negative, a real water molecule is inserted at the hydration site [226].

3.12 Chapter summary

This chapter has outlined the computational methods used in this thesis. Molecular dynamics (MD), using the NAMD software, has been used to simulate the dynamics and flexibility of the wild-type and mutant protein structures. Trajectory (RMSD, RMSF and PCA) and binding site analysis methods (PocketAnalyzer, SID, ptraj and JAWS) have been applied to the subsequent MD trajectories to identify differences between the structures. In order to reproduce what is known experimentally (e.g. poor binding between mutant receptors and the inhibitors compared to the wild-type structure) ensemble protein-ligand docking studies have been completed for the wild-type and mutant structures via the docking program Glide. In order to support the docking results, the binding free energy of the wild-type and mutant structures have been calculated from the MD trajectories using the MM-GBSA method via the perl script mdgbsa.pl.

Chapter 4 describes the biochemical background of the BCR-Abl kinase and chronic myeloid leukaemia.

Chapter 4

4 The BCR-Abl kinase & chronic myeloid leukaemia

4.1 Introduction

Chronic myeloid leukaemia (CML) is caused by an abnormal chromosome called the Philadelphia chromosome, BCR-Abl [237]. The Philadelphia chromosome is formed when two chromosomes (9 and 22) swap places resulting in the Breakpoint Cluster Region gene becoming fused to the Abl gene [238].

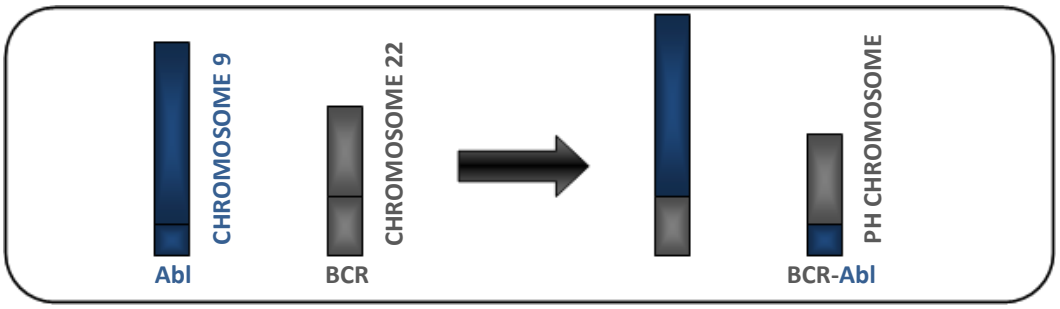


Figure 4.1.1: Translocation of the Abl gene and the BCR gene

The resulting fusion gene creates an oncogenic tyrosine kinase which, unlike its cellular counterpart (c-Abl), is unregulated and interferes with cellular pathways, causing leukemic cells to grow and multiply [238-240].

Several small molecule inhibitors of BCR-Abl are used in the treatment of CML, however point mutations can develop in the kinase resulting in varying levels of drug resistance.

This chapter will discuss the differences between the active and inactive structures of the cellular Abl kinase; the BCR-Abl kinase and how it causes CML; CML treatment through BCR-Abl inhibition and how drug resistance can develop.

4.2 The c-Abl kinase

The c-Abl kinase is a non-receptor protein tyrosine kinase (PTK) belonging to the Src kinase family (SKF). PTKs catalyse the transfer the γ-phosphate group from ATP to a tyrosine residue on a protein substrate (often another tyrosine kinase). This phosphorylation of proteins is an important mechanism in signal transduction for the regulation of cellular activity [241].

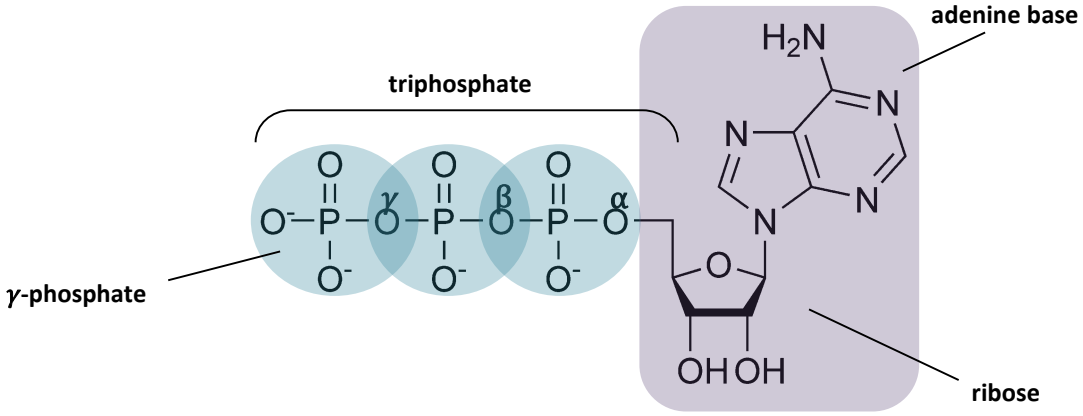


Figure 4.2.1: Structure of adenosine triphosphate (ATP)

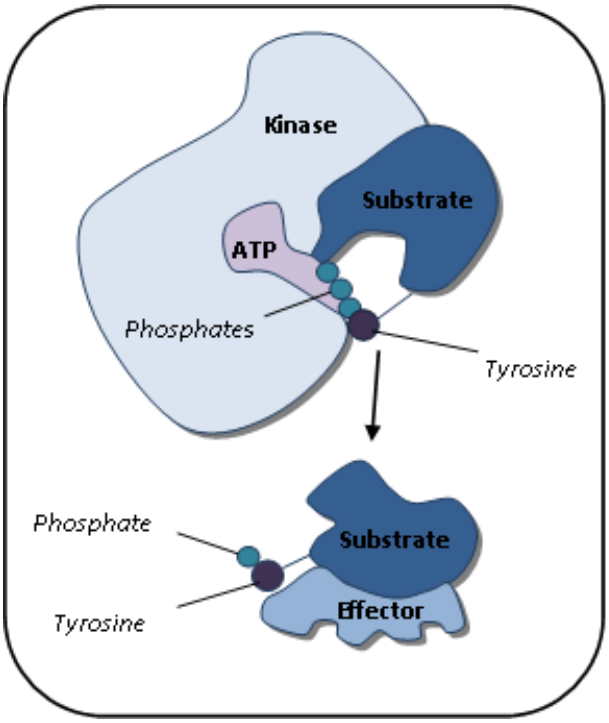


Figure 4.2.2: Kinase phosphorylation - ATP binds to the kinase pocket, the kinase can then phosphorylate a substrate that can interact with the downstream effector molecules. Adapted from reference [242].

C-Abl kinases are found within the cell cytoplasm and nucleus and can shuttle between these two cellular domains. The locality of c-Abl defines its function within the cell. Nuclear c-Abl has been found to play a role in growth inhibition and the apoptosis of cells, and modulates the cellular response induced by DNA damage [243-245]. Cytoplasmic c-Abl is involved in the promotion of DNA synthesis, morphogenesis, mitogenesis, F-actin assembly and receptor trafficking [243, 246-248].

4.2.1 The different domains of c-Abl

The c-Abl kinase consists of three core domains: the kinase domain, the Src Homology 2 (SH2) domain and the Src Homology 3 (SH3) domain. This core is bordered by an autoinhibitory cap at the N-terminus and an exon region at the C-terminus. There are two isoforms of c-Abl in humans; 1a and 1b. Both have the same catalytic core and last exon region, but differ in that 1b has a myristate attached to the autoinhibitory cap [243].

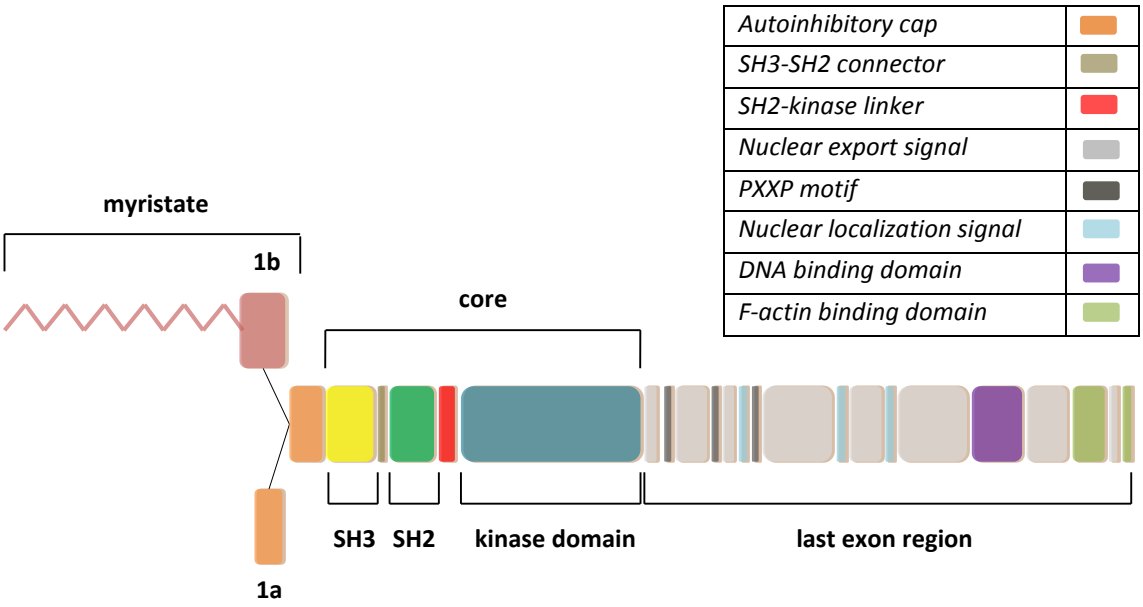


Figure 4.2.3: The structural regions of c-Abl kinase, adapted from references [249, 250]

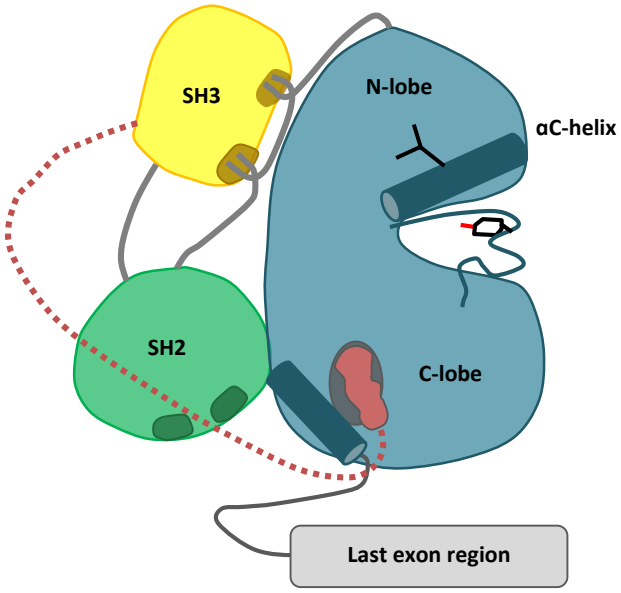


Figure 4.2.4: Schematic representation of inactive c-Abl (1b) kinase. Adapted from reference [251, 252]

4.2.1.1 The autoinhibitory cap

The first 80 residues of c-Abl are known as the N-cap or autoinhibitory cap. The autoinhibitory cap plays a prominent role in the regulation of the kinase via stabilisation of the inactive conformation. In splice 1b this is instigated by a prefixed myristoyl group binding to a hydrophobic pocket in the SH2 domain. In splice 1a, autoinhibition is brought about via cap interaction with the SH3-SH2 connector [243, 249].

4.2.1.2 The SH3 domain

The SH3 domain is a sequence of around 60 residues that are conserved in Src and other non-receptor tyrosine kinases. The SH3 domain helps regulate signal transduction through its ability to interact with the proline-rich sequences of other neighbouring proteins [253, 254].

4.2.1.3 The SH2 domain

The SH2 domain is a sequence of around 100 residues that like the SH3 domain are conserved in several tyrosine kinases. SH2 plays a significant role in signal modulation by binding to phosphorylated tyrosine residues on other kinases [255]. This is made possible by a sequence of

six residues; Phe-Leu-Val-Arg-Glu-Ser, known as the FLVRES sequence, which provides the electrostatic and H-bonding contacts necessary for binding [253, 256].

4.2.1.4 The Kinase Domain

The kinase domain is the central, catalytic part of the kinase that contains the ATP binding site.

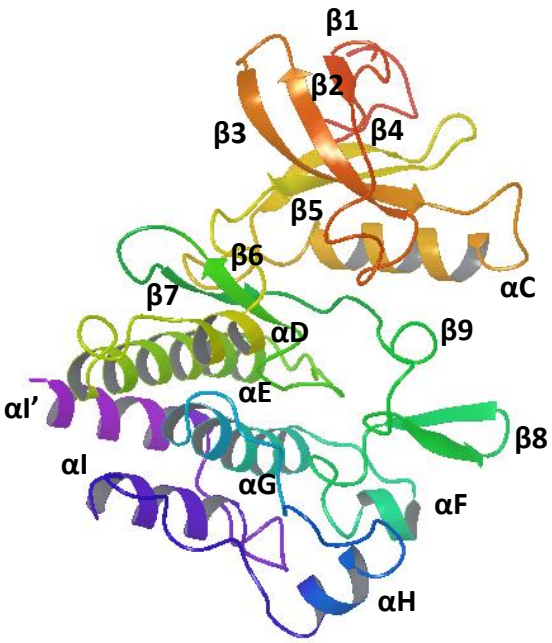


Figure 4.2.5: Secondary structure of Abl kinase domain with labelled helices and beta sheets (inactive structure, 1IEP.pdb)

The kinase domain consists of an N-terminal lobe (containing a 5 stranded β-sheet and an α-helix (αC)), and a larger (predominantly α-helical) C-terminal lobe. ATP binds in the cleft between the lobes, and forms hydrogen bonds with the "hinge region", the portion of the structure that connects the two lobes [35].

4.2.1.5 The last exon region

The last exon region is a long carboxy-terminal extension containing protein-protein interaction sites and binding sites for other kinases such as Crk. It also contains the nuclear export and localisation signals that enable c-Abl to move between the cytoplasm and nucleus. Deletions within the last exon region can sometimes result in increased tyrosine phosphorylation which suggests that the last exon region plays a role in cellular regulation [249, 257].

4.2.2 The inactive (autoinhibited) c-Abl structure

The inactive state of c-Abl is a closed conformation with the A-loop folded into the ATP binding site blocking access to both ATP and substrates. This conformation is facilitated by a regulatory mechanism formed by the neighbouring SH2, SH3 and N-terminal cap domains of Abl kinase. C-Abl uses slightly different mechanisms to autoinhibit itself depending on whether it is splice 1a or 1b [243, 258, 259], but overall relies on six critical actions: the docking of the SH2 domain to the kinase domain; the docking of the SH3 domain to the SH2-kinase linker; the strengthening of the SH3-SH2 clamp by the N-terminal cap; the "DFG flip" to the Asp-out position (kinase domain); the inward folding of the activation loop into the ATP binding site (kinase domain); and the disorder of the regulatory spine or R-spine (a set of residues that form a stack of hydrophobic interactions perpendicular to the ATP binding site, strengthening the open active conformation) [249, 260].

4.2.2.1 Conformational changes outside the kinase domain (SH2, SH3 and the N-terminal cap)

The SH2 domain forms a strong protein–protein interface with the C-lobe of the kinase domain, partially blocking the phospho-tyrosine (pTyr) binding site of the SH2 domain [261]. As well as keeping the conformation rigid, this prevents the SH2 domain from binding to other kinases, which would hinder its interaction with the kinase domain and result in a shift towards the active open conformation [35]. The SH3 domain binds to the linker segment that connects the SH2 and the kinase domain, known as the SH2–kinase linker and the N-lobe of the kinase domain. A small linker, the SH3-SH2 connector, joins the SH3 and SH2 domains and leads to the dynamic coupling of the two domains and the effective formation of a clamp on the kinase domain [249].

In 1a c-Abl the autoinhibitory cap region extends down from the SH3 domain to the SH2 domain forming a "brace", strengthening the clamp function of the SH3-SH2 domains. In 1b c-Abl the same conformation is made possible by the N-terminal myristate binding to a pocket in the C-lobe of the kinase domain. The binding of the myristate (and the extension of the N-cap) causes a conformational change in helix I, which aids the docking of SH2 against the C-terminal lobe of the kinase via hydrogen bonds and a pi-stacking interaction between Tyr158 of the SH2 domain and Tyr 361 on the C-lobe [251, 258].

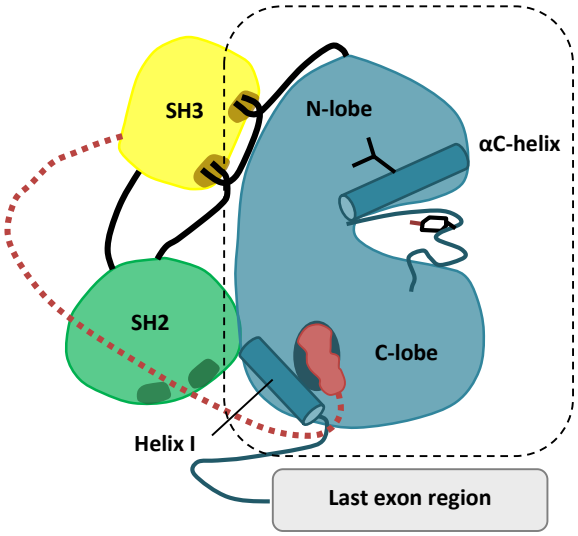


Figure 4.2.6: Schematic representation of inactive c-Abl (1b) kinase. Adapted from reference [251, 252]

4.2.2.2 Conformational changes in the kinase domain

The characteristics of the inactive conformation of the Abl kinase domain include:

- A folded conformation of the A-loop; where the A-loop is folded into the ATP binding site blocking access to both ATP and substrates.
- The DFG-out conformation, where DFG "D"(Asp381) is flipped 180° away from the binding site and the DFG "F" (Phe382) is positioned in the binding site
- Misalignment of the R-spine; a stack of hydrophobic residues that are aligned in active Abl stabilising the active conformation.
- An upturned or kinked conformation of the p-loop (glycine rich loop between β2 and β3) away from the binding site.

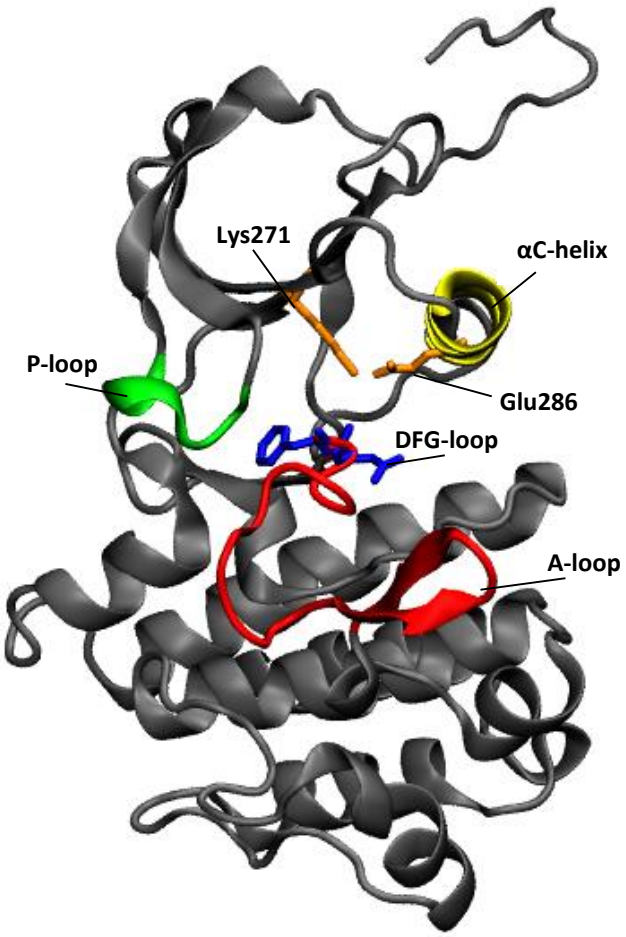


Figure 4.2.7: Key characteristics of the inactive Abl conformation [1IEP.pdb].

On the A-loop, Tyr393 forms a hydrogen bond with Asp363, causing the loop to adopt a folded conformation where it is curled up into the active site, blocking the binding of the substrate and ATP.

The first three residues of the A-loop; Asp-Phe-Gly (DFG) are known as the DFG loop. The positioning of this loop is vital for coordination of ATP, as the Asp381 (D) residue coordinates the position of ATP in the binding site. In the inactive Abl kinase domain, the Asp381 (D) residue flips its position with the Phe382 (F) residue of the DFG motif and is thereby withdrawn from the active site [249] (shown below in Figure 4.16).

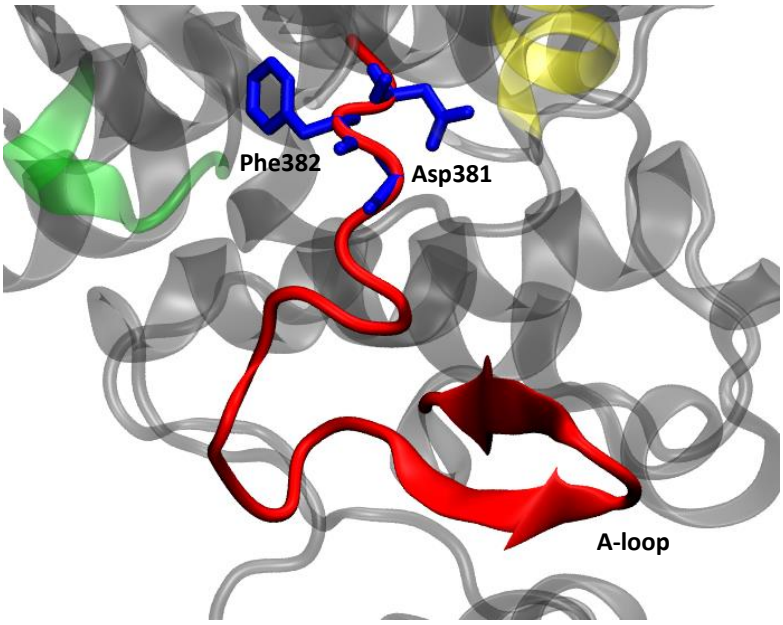


Figure 4.2.8: The c-Abl A-loop and DFG-loop in the inactive "Asp-out" conformation [1IEP.pdb].

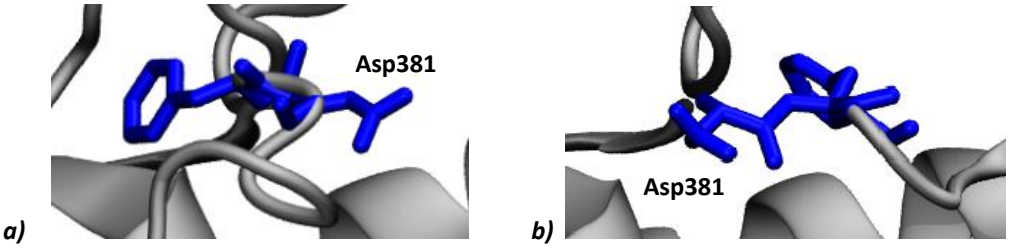


Figure 4.2.9: Isolated view of the DFG-loop a) Inactive "Asp-out" conformation [1IEP.pdb]. b) Active "Asp-in" conformation [2F4J.pdb].

With the A-loop and DFG motif in this configuration the residues that make up the R-spine are displaced and cannot form the hydrophobic stack observed in the active conformation (see section 4.2.3).

The loop between the β1 and β2 sheets contains several glycine residues (in Abl: Gly249-Gly250-Gly251-Gln252-Tyr253-Gly254) and is known as the glycine-rich or P-loop. The P-loop is typically elongated in the active conformation to coordinate the transfer of phosphate ions when ATP binds to Abl. Various kinked conformations of the P-loop have been observed from crystal structures of the inactive state, however since all of these structures are in complex with inhibitors it is unclear whether the kinked conformation is a natural conformation of the inactive form or a result of inhibition [262]. The P-loop is highly flexible and various conformations have been observed in structures that are intermediate between the active and inactive conformation [263].

The inactive conformation of c-Abl is actually closer to that of the open active state than other PTKs, and this helps with the SH2 domain binding to the C-lobe. In most closed confirmations important catalytic residues are removed from the active site via rotation of the αC-helix, preventing ATP and substrate binding [264]. In c-Abl the A-loop is in a slightly more extended conformation than in other inactive kinases and the αC-helix is closer to the normal active position allowing formation of the Lys-Glu salt bridge, which is a conserved motif in active kinases [249].

4.2.3 The active c-Abl structure

Whereas the inactive conformations of different tyrosine kinases are extremely varied, the active conformation is well conserved across the kinase family. The conserved characteristics include:

- An extended A-loop conformation where it uncoils to function as a binding platform for the peptide substrate [265].

- The DFG-in conformation (Asp381-Phe382-Gly383); Asp381 (D) is turned into the ATP binding site where it forms polar contacts (either directly or via magnesium atoms) with the α, β, and γ phosphates of ATP. Asp381 (D) forms a hydrogen bond with glycine residue Gly383 (G), which helps secure the position of Asp381 (D) in the binding site. The DFG Phe382 (F) residue is part of the R-spine and also coordinates the correct positioning of the DFG Asp381 (D) residue by making hydrophobic contacts with the residue Met290 on the C-helix and His361 on the HRD motif (a conserved His-Arg-Asp sequence conserved in most PTKs); Met290 and His361 form part of the R-spine.

- An aligned regulatory spine (R-spine) of four non-consecutive residues, two from the N-lobe and two from the C-lobe that form a stack of hydrophobic interactions thereby strengthening the C-N lobe connection and stabilising the active conformation [260]. In Abl these residues are Leu 301, Met290, Phe382 and His361 [266, 267]. Asp421 on the αF-helix makes a strong hydrogen bond to the main chain of His361, which secures the spine in place.

- A catalytic spine (C-spine) of eight residues that contribute to the positioning of ATP. Details of the residues forming the C-spine are not published for Abl, however in PKA these residues are: Val57, Ala70, Leu173, Leu172, Leu174, Met 128, Leu227 and Met231 [267].

- A salt bridge between a Glu residue on the αC-helix and a Lys on the β3 beta sheet resulting in a tight connection between the C and N lobes (Glu286 and Lys271 in Abl). Although this salt bridge is present in the inactive structure of Abl, in the active conformation Lys271 forms polar contacts with α and β phosphates of ATP, and therefore plays an important role in the correct positioning of ATP in the binding site.

- An extended "tongue-like" conformation of the p-loop (glycine rich loop between β2 and β3) so the tip of the loop is able to coordinate the γ-phosphate for phosphoryl transfer on ATP binding [260]. Various conformations of the p-loop have been observed in structures that are intermediate between the active and inactive conformation [263].

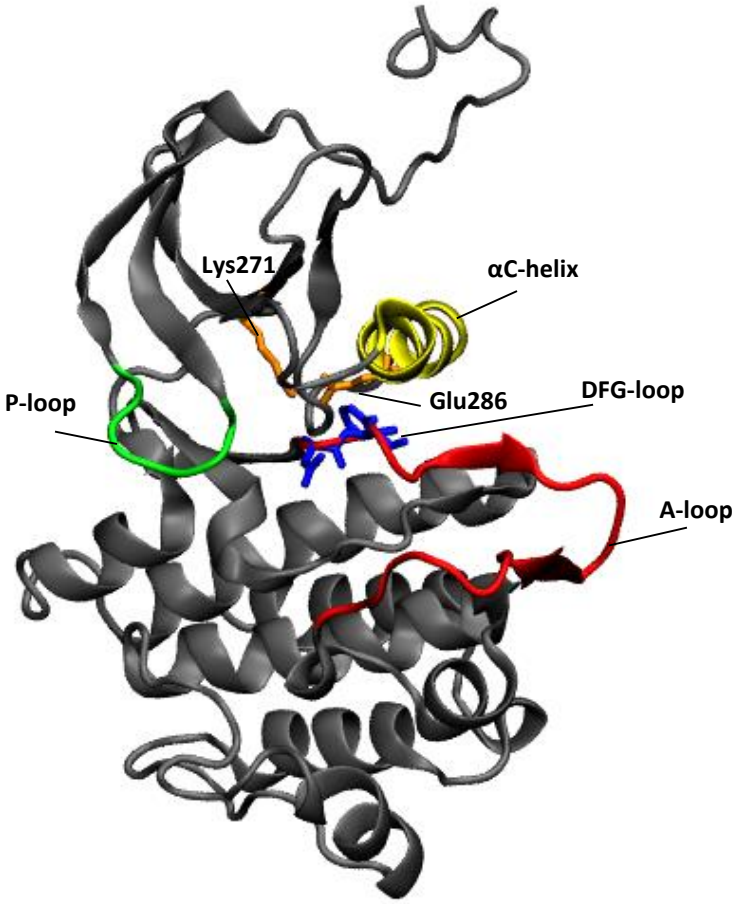


Figure 4.2.10: Key characteristics of the active Abl conformation [2F4J.pdb].

The following sub-sections further discuss how these conformations manifest on activation of Abl.

4.2.3.1 Disruption of the regulatory "clamp" mechanism

Abl activation is thought to be initiated through the removal of the regulatory clamp; once the kinase domain has shifted to the active conformation, phosphorylation stabilises the active structure.

Phosphorylation can occur via a substrate, another PTK such as Src, or c-Abl itself. If phosphorylation occurs via the kinase itself (cis) or another identical c-Abl kinase (trans) the process is known as autophosphorylation. In oncogenic versions of Abl, where the regulatory mechanism is missing, the kinase readily autophosphorylates at Tyr226 and Tyr393 [249, 268] stabilising the active conformation. Mutations in the kinase domain have also been shown to stabilise the active conformation in the absence of the regulatory clamp, demonstrating that the active conformation of Abl is thermodynamically accessible even when the protein is not phosphorylated [31, 264].

Disruption of the regulatory clamp, and subsequent shift towards the active conformation, is thought to be initiated by one or several of the following events: myristoyl displacement, disruption of docking of the SH3 domain to the kinase domain or disruption docking of the SH2 domain and kinase domain. Myristoyl displacement has not been well observed but is thought to be brought about by a fatty acid binding protein which could also affect the docking of the SH2 domain to the kinase domain [251]. Disruption to SH3 docking occurs when a proline-rich ligand is supplied by another protein and is therefore in direct competition with the SH3 linker segment. This ligand would most likely also be phosphorylated by c-Abl and these phosphorylated residues could in turn bind to the SH2 domain. Due to close protein-protein interactions between the SH2 domain and the C-lobe of the kinase domain, ligands are normally unable to access the SH2 phospho-tyrosine binding pocket in inactive c-Abl. If the SH2–kinase domain docking becomes disrupted (due to SH3 ligand binding for example), phosphopeptides are able to bind to the SH2 domain, which further activates c-Abl. Since substrates capable of activating SH2 regularly contain SH3 specific ligands and vice-versa, this suggests that the two mechanisms are very closely tied and that the order in which they occur may often be indistinguishable [249, 251, 264, 265]. These "activating" events are associated with increased catalytic activity in the kinase and result in the effective diminishment, or removal, of the "clamp" mechanism which shifts the kinase towards a more open structure.

4.2.3.2 *Stabilisation of the active conformation via phosphorylation*

As shown in Figure 4.2.11, the active conformation is stabilised by the phosphorylation of one or more tyrosines on the A-loop (in particular Tyr393, defined as Tyr412 in the 1b sequence), the SH2-kinase domain linker section (Tyr226), and also at various other positions on the kinase domain [264, 269].

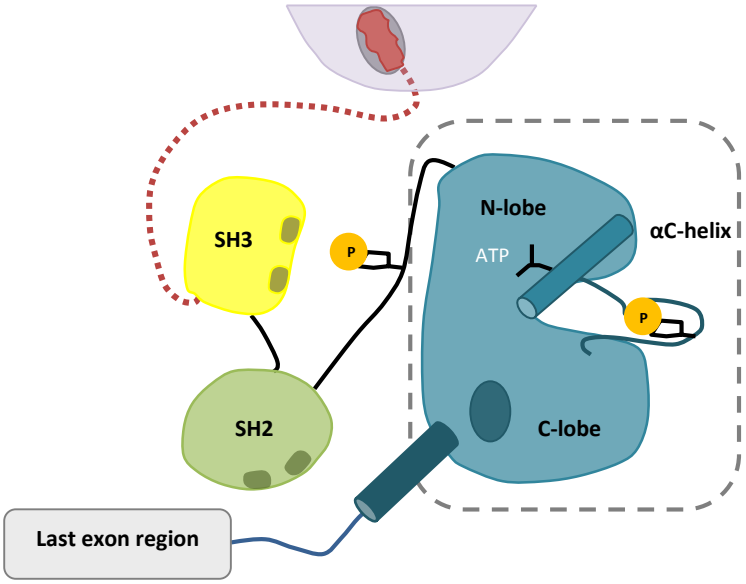


Figure 4.2.11: Schematic representation of active c-Abl (1b) kinase. Adapted from reference [251, 252]

Phosphorylation at Tyr393 appears to support the A-loop conformation and that of the DFG motif by forming a number of hydrogen bonds with the surrounding residues. When Tyr393 is phosphorylated it forms a hydrogen bond with Arg362 (HRD motif) which is held in position on the catalytic loop due to another hydrogen bond with Asp421. The new phospho-Tyr393 also attracts Arg386 which induces the formation of the β6 and β9 sheet. These interactions enable the extended A-loop conformation to be held in place [264, 269].

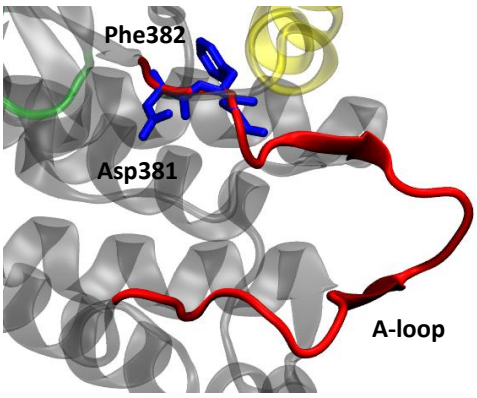


Figure 4.2.12: The c-Abl DFG-loop in the active "Asp-in" conformation [2GQG.pdb].

The carbonyl oxygen of the peptide bond between Leu384 and Ser385 rotates and forms two hydrogen bonds with Arg362. Leu384 and Ser385, together with the DFG loop (i.e. residues 381 – 385) the residues are referred to collectively as the magnesium binding loop due to their role in coordinating magnesium ions in the ATP binding site. The Phe382 and Gly383 peptide bond (of the DFG motif) rotates its carbonyl oxygen, forming a 3-turn, and the amide on Gly383 forms a hydrogen bond with the Asp381 δ-oxygen; this provides the correct Asp-in orientation of Asp381 in the binding site, ready to coordinate the phosphate ions of ATP [35, 249, 264]. Hydrophobic contacts between the DFG Phe382 residue, Met290 on the αC-helix and His361 on the HRD motif (a conserved His-Arg-Asp motif seen in many kinases) further stabilise the position of Asp381. Phe381, Met290 and His361 form part of the regulatory spine which is further discussed below in 4.2.3.3.

4.2.3.3 The alignment of the regulatory and catalytic spines

This active conformation is stabilised by two hydrophobic spines, the regulatory spine or R-spine and the catalytic spine or C-spine. The positioning of each spine is conserved across kinases [267].

- The R-spine is a flexible stack of hydrophobic residues between the DFG loop and phosphorylation sites on the A-loop formed by residues His361, Phe382, Met290 and Leu301 [266, 267]. His361 is part of the rigid C-lobe and its position is coordinated by Asp421 from the αG-helix. As discussed above in 4.2.3.2, the sidechain of His361 forms contacts with Phe382 of the DFG loop; Phe382 has contacts with Met290 (αC-helix) which in turn is packed against Leu301 (β4-sheet, adjacent to αC-helix).

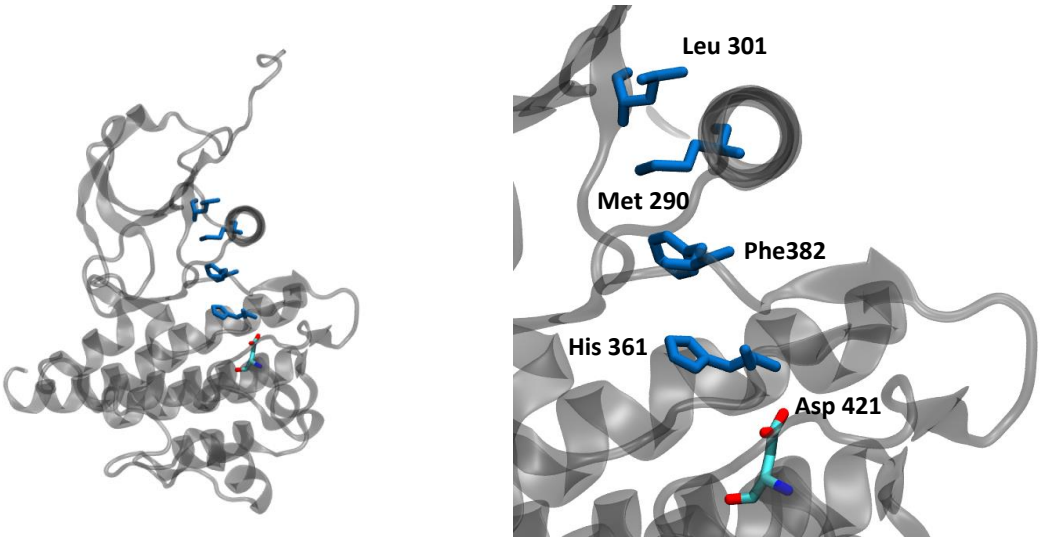


Figure 4.2.13: The residues of the R-spine [2F4J.pdb]

The catalytic spine or C-spine is made up of eight residues that contribute to the positioning of ATP. Residue 1 and 2 are along β3 in the N-lobe and dock with one side of the adenine ring of ATP residue 3 lies in the hinge region on the other side of ATP where it also docks to the adenine ring of ATP. Residue 3 is flanked by residues 4 and 5, which rest on residue 6, a hydrophobic residue along the D-helix. Residues 7 and 8 are positioned on the F-helix anchoring the spine to the base of the C-lobe. Details of the residues forming the C-spine are not published for Abl, however in PKA these residues are: Val57, Ala70, Leu173, Leu172, Leu174, Met 128, Leu227 and Met231 [267].

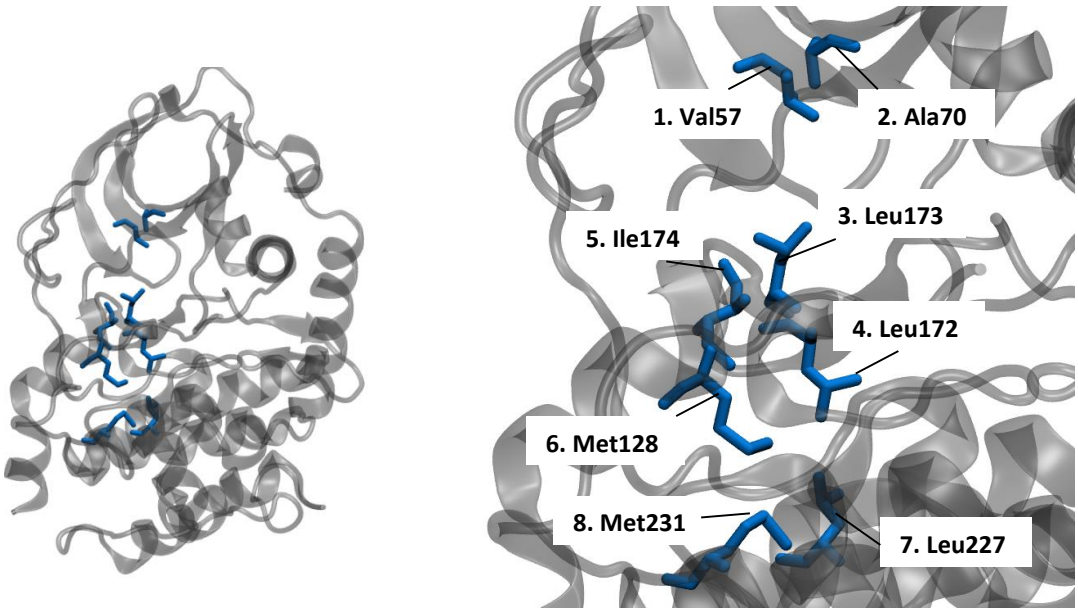


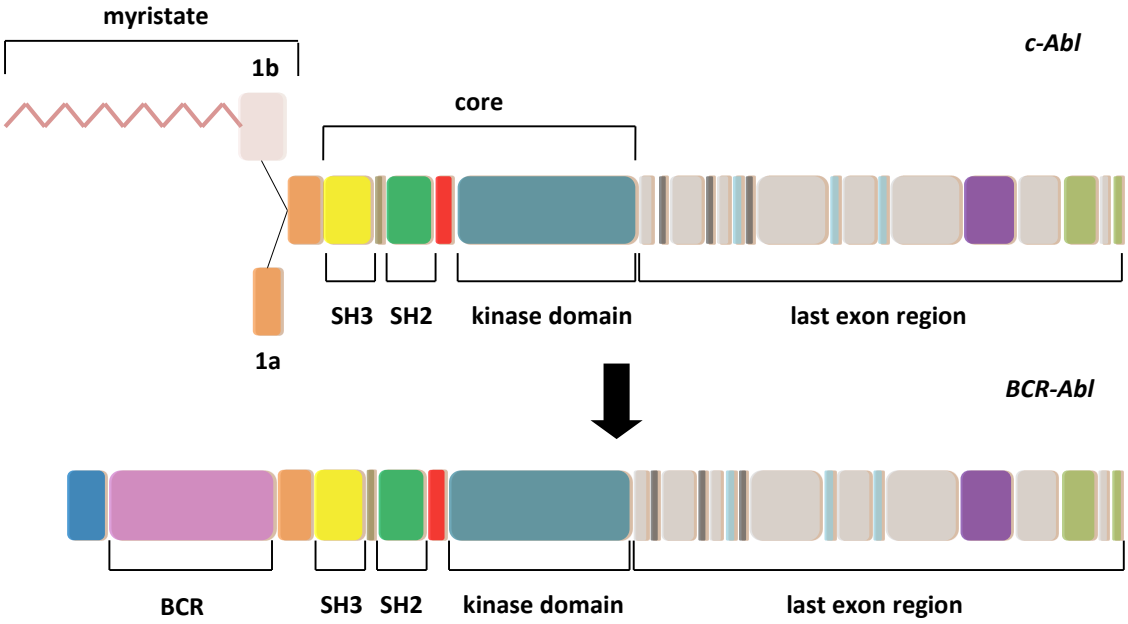
Figure 4.2.14: The residues of the C-spine [1ATP.pdb]

4.2.3.4 *Dephosphorylation effects*

Dephosphorylation on the A-loop leads to the destabilisation of both the A-loop and hydrophobic stack and this relaxation in the connection between the two lobes results in a change in the conformation of the DFG loop. Dephosphorylation and down regulation of activated c-Abl is controlled via PSTPIP1 proteins and ubiquitin dependent proteasomal degradation [264].

4.3 The BCR-Abl fusion kinase

In BCR-Abl the N-terminal regulatory part of Abl is replaced with the N-terminal domains of Bcr, which include an oligomerization domain [270].



Autoinhibitory cap	
SH3-SH2 connector	
SH2-kinase linker	
Nuclear export signal	
PXXP motif	
Nuclear localization signal	
DNA binding domain	
F-actin binding domain	
Oligomerization domain	

Figure 4.3.1: The structural domain differences between c-Abl kinase and BCR-Abl kinase. Adapted from references [249, 250]

With the loss of the myristate (or part of the autoinhibitory cap in the case of isoform 1a) and a large new piece of protein added there instead, the regulatory domains no longer function properly and the BCR-Abl kinase is unable to adopt an autoinhibited conformation. This results in constitutive activation of the kinase domain (where the kinase remains active regardless of metabolic requirements). In addition the oligomerization causes BCR-Abl to form dimers and tetramers, which further increases activation via phosphorylation, as several chains of BCR-Abl are clustered together in close proximity. It has also been found that dimerization alone can be enough to initiate c-Abl activation [271].

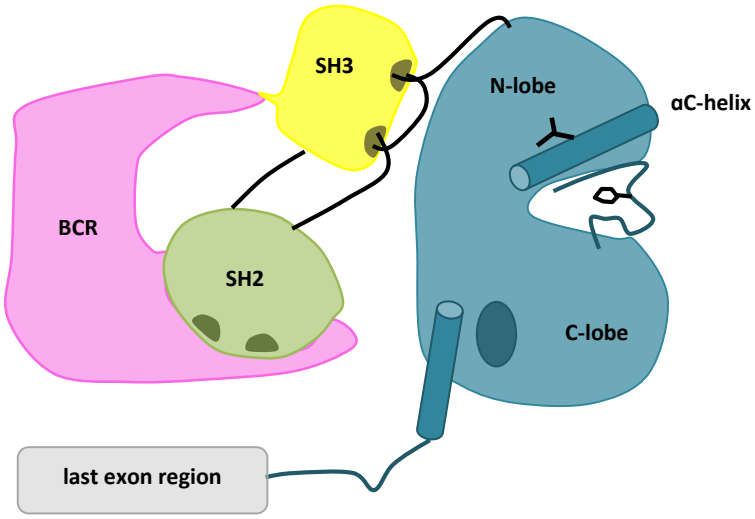


Figure 4.3.2: Schematic representation of the BCR-Abl kinase. Adapted from reference [251, 252]

This results in the uncontrolled proliferation and survival of myeloid cells in the blood, which manifests as CML.

4.4 Chronic myeloid leukaemia (CML)

4.4.1 What is CML?

Chronic myeloid leukaemia (CML) is a rare type of cancer that starts in bone marrow stem cells and moves into the blood. CML rarely occurs in children and of the 700 people diagnosed with CML in the UK each year, most are elderly with a median age of 65 at the time of diagnosis [240, 272].

CML manifests in myeloid stem cells, predominantly young granulocytic cells (white blood cells) but also in cells that would normally develop into red blood cells and platelets. Cellular growth and apoptosis is usually tightly regulated [239] but leukaemic stem cells do not develop correctly and have a longer lifespan than normal cells. This leads to a high percentage of leukaemic cells in the bone marrow and blood which reduces the number of normal cells in circulation and stops the body working as it should [240, 273, 274]. As the name suggests, CML is a chronic disease that progresses over a number of years (unlike acute leukaemia that can rapidly develop in a matter of months [275-277]) due to a "chronic phase" (CP) where most of the leukaemic cells are similar in size to mature healthy cells and their numbers increase at such a slow rate that the patient is often asymptomatic [273]. As CML progresses beyond the chronic phase the leukaemic cells proliferate faster and are less developed than in the chronic phase. If CML is untreated, these immature cells (known as blasts) can rapidly, and fatally, overrun the body [274].

4.4.2 The three phases of CML

The chronic phase is the most stable phase of CML and can last for several years as the leukaemic cells continue to develop at a slow rate. Although 90% of patients are diagnosed during the chronic phase, many of these are by chance as symptoms are mild (e.g. tiredness, slight weight loss) and sometimes none existent. Diagnosis is made on the basis that there are a usually high number of granulocytes in the blood stream, and then confirmed by the identification of the Philadelphia chromosome in the patient's blood and bone marrow cells (Ph+). At this stage fewer than 10% of cells in the bone marrow are immature blast cells. This percentage of blast cells is used to gauge the progress of the leukaemia [240, 278].

In the accelerated phase, symptoms become more apparent (increased tiredness and noticeable weight loss) and some patients may experience pain where their spleen is enlarged. The accelerated phase is much shorter than the chronic phase and typically lasts between 4 to 6 months. In this phase there is an increase in the number of abnormal granulocyte cells and between 10% and 30% of the cells in the bone marrow are blasts [240, 278].

When the number of blast cells in the bone marrow exceeds 30% the leukaemia is said to be in blast crisis. During this phase the leukaemic cells proliferate very rapidly and most of the leukaemic cells produced are immature blasts. The blast crisis phase is the shortest of the phases, typically only lasting a few months [240, 274, 278].

Treatment choices for people with chronic myeloid leukaemia (CML) depends on the phase of the disease, prognostic factors such as their age, and whether or not there is a matched stem cell donor [279]. It is important to accurately identify the phase of the leukaemia when a patient is first diagnosed as this helps physicians to identify the best treatment strategy [280]. CML that has progressed to the accelerated phase or blast crisis is usually harder to treat than in the chronic phase. If the patient is in the chronic phase then further techniques, such as using prognostic scoring [281] and monitoring any cryptogenic or haematological abnormalities in the patient (such as deletions of the derivative chromosome 9), [282-285] can help identify those most at risk of short survival. Modern CML therapy includes small molecule inhibitors, interferon alpha and allogeneic stem-cell transplant.

4.5 Current small molecule inhibitors of BCR-Abl used in the treatment of CML

Since 2001, five small molecule inhibitors (SMIs) have been approved to treat CML. Imatinib, nilotinib and dasatinib are approved for first-line treatment [286-288], and bosutinib and ponatinib are approved for patients with resistance or intolerance to prior therapy [289-293].

Product	FDA approved indication
Imatinib	• Newly diagnosed adult and paediatric patients with Ph+ CML in chronic phase. • Patients with Ph+ CML in blast phase, accelerated phase, or chronic phase after failure of interferon-alpha therapy.
Nilotinib	• Newly diagnosed adult patients with Ph+ CML in chronic phase. • Patients with Ph+ CML in accelerated phase, or chronic phase who are resistant to or intolerant to prior therapy that included imatinib.
Dasatinib	• Newly diagnosed adult patients with Ph+ CML in chronic phase. • Patients with Ph+ CML in myeloid or lymphoid blast phase, accelerated phase, or chronic phase, resistant to or intolerant to prior therapy that included imatinib.
Bosutinib	• Patients with Ph+ CML in chronic, accelerated, or blast phase, resistant or intolerant to prior therapy.
Ponatinib	• Patients with Ph+ CML in chronic, accelerated, or blast phase, resistant or intolerant to prior therapy.

Table 4.5.1: Approved indications for TKI therapy in CML [298, 302-305]

The success of small molecule inhibitors is down to their effectiveness in bringing about complete cytogenetic remission [294, 295] in a high percentage of CML patients in early chronic phase, without resulting in severe adverse events.

These inhibitors will be broadly discussed in the following section, however since the focus of this research is on the first two: imatinib and nilotinib, these will be discussed in greater detail.

4.5.1 Imatinib

The unique conformation the Abl kinase domain's inactive structure made it an excellent target for drug design. Imatinib (known as Gleevec/Glivec in the US and Europe respectively) binds to BCR-Abl (and c-Abl [35, 296]) with an induced fit interaction[1], where the inhibitor binds in the ATP binding site of the kinase and extends beyond this region, through the centre of the molecule into a hydrophobic pocket opened up by displacement of the DFG loop. The aspartate (Asp381) from the DFG loop is flipped out of the position where it would be required to coordinate the magnesium ions in the ATP phosphate binding site.

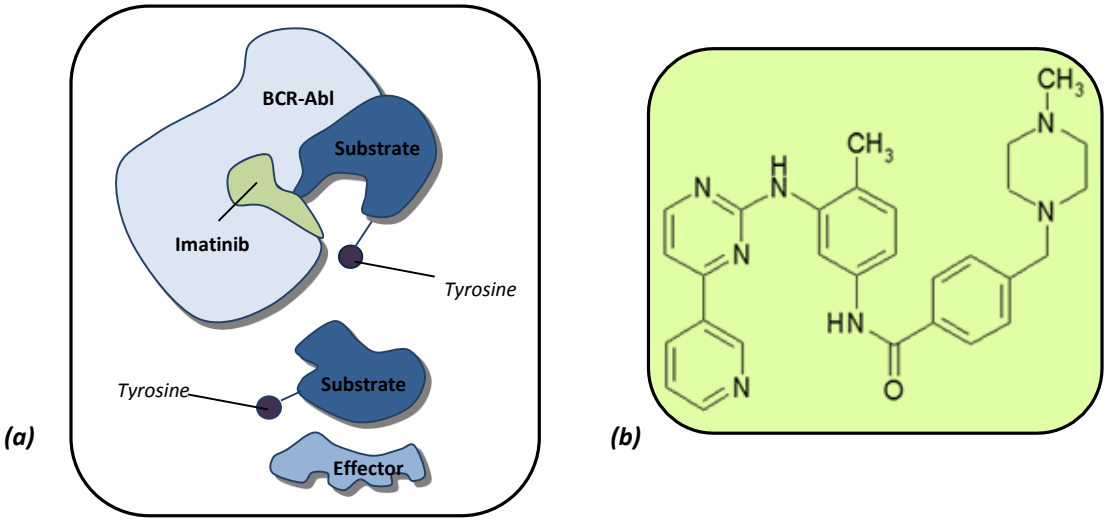


Figure 4.5.1: (a) Imatinib binds to the ATP pocket of BCR-Abl (competing with ATP), therefore the substrate cannot be phosphorylated. Adapted from reference [242]. (b) The structure of imatinib

Imatinib then blocks the path the activated A-loop would normally adopt and the A-loop structure changes to an autoinhibitory conformation in which the unphosphorylated tyrosine binds in and blocks the substrate binding site, as shown in Figure 2.6 [35].

[1] It is important to note that imatinib binds to BCR-Abl (active) and both the active and inactive forms of c-Abl. As the inhibitor binds, it induces the inactive conformation (as described) in active c-Abl/BCR-Abl and stabilises the inactive conformation in inactive c-Abl.

The necessity for Abl to have the ability to adopt this conformation in order to bind to imatinib, together with the requirement that the side chain of the residue in position 315 must be small (such as a threonine) so that the ligand can reach into the hydrophobic pocket near the DFG loop, explains why imatinib has such high selectivity for Abl.

However, the fact that imatinib makes contacts with residues that are not functionally important and that it requires a unique conformation for binding, means that it is susceptible to the development of mutants that are insensitive to treatment with this drug. Further inhibitors of BCR-Abl were later developed; nilotinib is similar in design to imatinib but has increased binding affinity to Abl, and dasatinib which can bind to both the active and inactive forms of Abl. Both inhibitors are efficacious in patients that were previously unresponsive to imatinib treatment, although neither drug can inhibit the T315I gatekeeper mutation [262].

4.5.2 Nilotinib

Based upon the structural data of imatinib, it was thought that a more potent and selective compound could be designed by incorporating alternative binding groups for the N-methylpiperazine group, while retaining an amide substructure to maintain the H-bond interactions to Glu286 and Asp381.

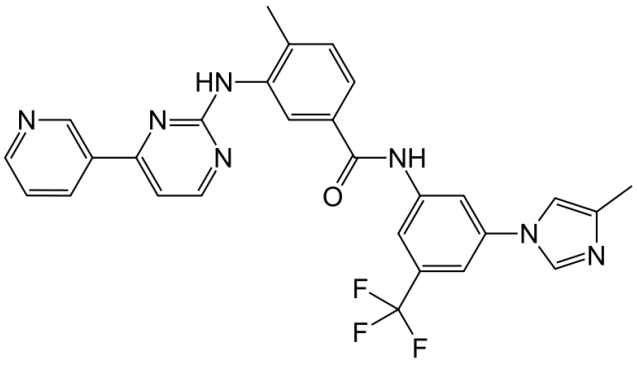


Figure 4.5.2: The structure of nilotinib.

Crystal structure analysis had revealed that much of the binding energy of imatinib was consumed by two energetically unfavourable hydrogen bonds between the *N*-methylpiperazine and Ile360 and His361 residues of Abl. Replacement of the piperazine ring with a trifluorinated imidazole eliminated these two hydrogen bonds. Furthermore, the imidazole ring fits well into a hydrophobic pocket of the kinase. The result was a better topological fit and a molecule that

inhibits ABL with 30-fold increased potency compared with imatinib [297]. More importantly nilotinib was capable of inhibiting most of the imatinib-resistant mutations but not the gatekeeper mutation T315I since the mutation blocks access to the hydrophobic pocket [32].

4.5.3 Dasatinib

Dasatinib (marketed as Sprycel) successfully inhibits a wide range of imatinib-resistant mutations not including the gatekeeper mutation.

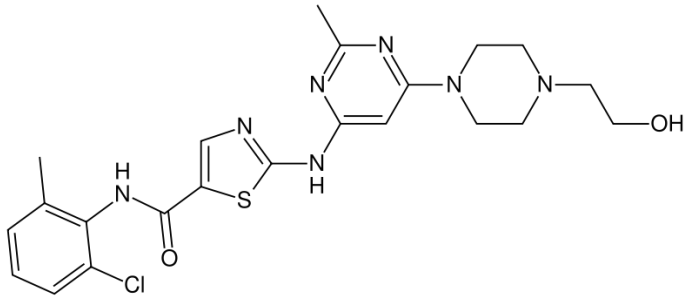


Figure 4.5.3 :The structure of dasatinib.

Dasatinib was initially developed as a Src inhibitor but was found to be a potent inhibitor of multiple kinases including Bcr-Abl. Dasatinib binds to the ATP binding in a very different pose to imatinib and nilotinib, extending in the opposite direction. Dasatinib binds to both the active and inactive conformation and has a greater affinity with Abl compared to imatinib. The activity of dasatinib against mutants of Bcr-Abl is explained by the fact it does not require interaction with some of the residues involved in the mutations [298].

4.5.4 Bosutinib

Bosutinib (marketed as Bosulif) is a Src inhibitor approved for the treatment of CML in adult patients that are either resistant or intolerant to the other therapies available [299, 300].

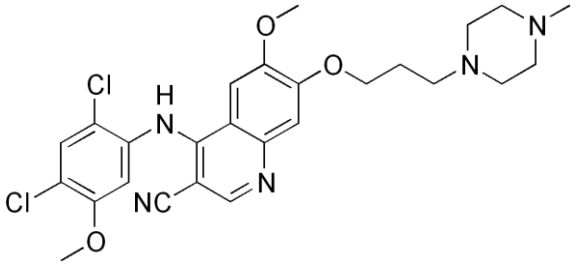


Figure 4.5.4: The structure of bosutinib

Although bosutinib does not make hydrogen bonds with T315, it does make extensive van der Waals contacts, both with T315 and the side chain of V299, explaining why mutations at these residues result in resistance [301].

4.5.5 Ponatinib

In December 2012 the FDA granted accelerated approval to Ponatinib (marketed as Iclusig) for the treatment of adult patients with CML that is resistant or intolerant to prior tyrosine kinase inhibitor therapy.

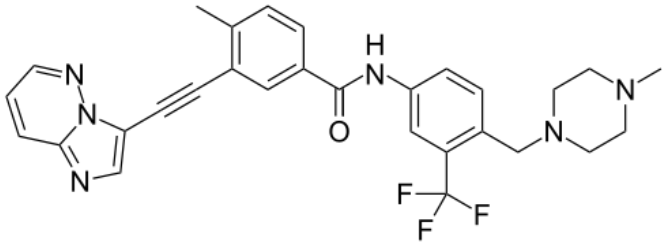


Figure 4.5.5: The structure of ponatinib

Ponatinib was designed to inhibit both wild-type and mutant BCR-Abl and is a mixed kinase inhibitor also capable of inhibiting FGFR, PDGFR, VEGFR. Ponatinib binds to the inactive DFG-out conformation of the kinase domain and has similar contacts to that of imatinib; however superposition of the two ligands shows that while imatinib clashes with the mutated I315 side chain, ponatinib has favourable van der Waals interactions with it [302]. Clinical trials have confirmed that ponatinib is very effective against the gatekeeper mutation T315I, with 72% of patients in the pivotal PACE trial (NCT01207440) achieving a major cytogenetic response [292].

Although broader specificity of TKIs can lead to improved efficacy, the inhibition of several tyrosine kinases often results in a greater number or frequency of adverse events in patients; ponatinib has a black box warning from the FDA on the products label regarding arterial thrombosis and hepatotoxicity [303].

4.6 Resistance to tyrosine kinase inhibitors

4.6.1 Possible causes of drug resistance

Imatinib resistance can be either primary (failure to achieve specific haematological and cytogenetic targets at defined time points) or secondary (the loss of complete cytogenetic response or hematologic response or disease progression) [304, 305]. In around 50% of cases, imatinib resistance is caused by the development of point mutations within the BCR-Abl kinase [306, 307]. Sequencing tests are used to screen patients for mutations following poor results with imatinib treatment [308-310]. Other causes of resistance include; genomic amplification of BCR-Abl [311, 312], clonal changes to the karyotype resulting in an abnormal number of chromosomes [313], or mechanisms altogether independent of the BCR-Abl kinase [314-316].

4.6.2 Point mutations

The mutations that cause drug resistance occur not only around the imatinib binding site but also in other areas of the kinase such as the P-loop, A-loop and SH domains. The most commonly observed clinical mutations are T315I, E255K and M351T, which account for more than 60% of cases [4].

It has long been hypothesised that the location of the mutation plays an important part in how resistance is conferred. Mutations found in the imatinib binding site directly interfere with ligand binding, either by removing a critical interaction with an inhibitor or changing the shape of the pocket in such a way that makes it difficult for the inhibitor to fully access it. Non-binding site mutations appear to distort the closed conformation of the kinase pushing it towards a more open active like structure. As a knock-on effect this distortion often additionally changes the shape of the binding site, further reducing binding affinity.

	Mutations		IC_{50}
	Type 1a	*Type 1b*	*(μM)*
P-loop	**M263V/I**	**M244V/I**	**3.1 V/1.41 I**
	L267R/V	L248R/V	> 20.0 R
	G269E/R	**G250E/R**	**> 20.0 E**
	Q271H	**Q252H**	**2.9**
	Y272H/C/F	**Y253H/C/F**	**17.7 H**
	E274K	**E255K**	**12.1**
	Y276C	Y257C	ND
	E277D	E258D	ND
	S284T/I	S265T/I	ND
	L285M/V	L266M/V	ND
	V287/289A	V268/270A	0.8
	A288V	A269V	1.1
	E294K	E275K	ND
αC-helix	D295V	D276V	1.2
	M297L	M278L	2.2
	E298K	E279K	9.9
	E300K	E281K	2.8
	E301D	E282D	1.3
	F302L	F283L	ND
	L303F	L284F	ND
	V308S	V289S	1.4
	M309L/T	M290L/T	ND
	K310E/R	K291E/R	ND
	E311Q	E292Q	2.5
SH3 contact	K313R	K294R	ND
	Q319H	Q300H	1.6
	L320F	L301F	ND
	F330L/V	**F311L/V**	**1.3 L**
	T334I/S/G	**T315I/S/G**	**>20.0 I/3.8 S**
	E335D	E316D	1.7
	F336L	**F317L**	**2.3**

	Mutations		IC_{50}
	Type 1a	*Type 1b*	*(μM)*
	G340W	G321W	1.5
SH2 contact	N350S	N331S	ND
	V357G	V338G	1.3
	V358A/G	V339A/G	ND
	M362T	**M343T**	**ND**
	A363V	A344V	ND
	Q365H	Q346H	1.39
	M370T/I	**M351T/I**	**4.9 T/1.61 I**
	E371K	E352K	1.3
	F378C	**F359C**	**ND**
	I379F	I360F	ND
	A385D	A366D	0.6
	G391R	G372R	ND
	E392K	E373K	ND
A-loop	**V398E/A**	**V379E/A**	**ND**
	L403M	L384M	2.8
	M407I	M388I	ND
	H415R/P	**H396R/P**	**ND**
	G417R	G398R	1.1
	Y459C	Y440C	ND
	E469K	E450K	2.3
	L470M	L451M	ND
	G482D	G463D	0.6
	M491I	M472I	1.4
	R492L	R473L	ND
	F505S	F486S	9.1
SH2 contact	E513A	E494A	1.2
	E518K/I	499K/I	1.9K
	I521M	I502M	1.2
	E528D	E509D	1.1

Figure 4.6.1: Kinase domain mutations: IC_{50} as measured in transformed BaF_3 cells. ND, not determined. IC_{50} for native BCR-Abl is 0.6 μM (imatinib). Mutations in bold have been reported in patients [44].

4.6.2.1 Binding site mutations

Mutations that directly affect the affinity of imatinib for Abl are T315I/D/N, F317L and G321W. Mutations Y253F and F359C likely have both a direct and indirect effect on imatinib binding to Abl.

The threonine residue at position 315 in Abl is the so-called gatekeeper residue. The size of the residue at this position determines how easily the hydrophobic pocket can be accessed especially in the (DFG-in) active form. In the case of imatinib binding, the hydroxyl group of T315 forms a hydrogen bond to the amine linker between the pyrimidine and phenyl rings of imatinib. Mutation of the threonine therefore results in a loss of this hydrogen bonding interaction, while

the larger sizes of the residues at this position (such as isoleucine in the case of T315I) also result in steric clashes, preventing imatinib from binding to mutant Abl structures [4]. Recent mutagenesis studies have also shown that the gatekeeper mutations in Abl and other kinases (PDGFR, EGFR and Src) have increased activity in cells expressing this mutation. The increased activity is due to the strengthening of the R-spine, which is a conserved feature of active kinases and is disrupted in inactive kinases. The increase in activity can further contribute to drug resistance [260, 317].

If the glycine at position 321 (G321) is mutated to a larger residue such as tryptophan (G321W) this residue would then clash with imatinib, resulting in a loss of binding affinity [262].

The side chain of phenylalanine 317 (F317) forms the ceiling of the ATP binding site near where critical hydrogen bonds are formed, protecting the hinge region from the solvent; mutations at this position can result in a loss of pi stacking interactions between F317 and the pyridine-pyrimidine rings of imatinib [32, 262].

The aromatic ring on the tyrosine at position 253 (Y253) is also thought to form pi stacking interactions with the pyridine-pyrimidine rings of imatinib and mutations to a phenylalanine or a histidine residue at position 253 result in a loss of imatinib binding affinity. Furthermore, Y253 is positioned on the P-loop and changes in the flexibility of this section can have an effect on the kinase dynamics as a whole, including the DFG-out/DFG-in conformation [32, 262].

4.6.2.2 Non-binding site mutations

Mutations situated remotely from the binding site presumably affect the inactive conformational equilibrium of the kinase as well as other binding sites and regulatory domains [2]. In addition to Y253F there have been reports of other P-loop mutations which are likely to greatly affect the dynamics of the glycine-rich loop [2].

Leucine 248 can mutate to both larger (arginine) and smaller residues (valine); where arginine directly clashes with imatinib, valine results in reduced van der Waals interactions, thereby reducing the binding affinity. The E255K/V mutation influences the dynamics of the catalytic domain. Position 255 is located just after the P-loop and with E255 the glutamatic acid side chain is placed within hydrogen bonding distance of a hydroxyl group (from Y257) and a charged nitrogen atom from L247. Mutating this residue to a lysine or valine will significantly disrupt this hydrogen bonding network [4].

In WT Abl the histidine 396 (H396) forms part of the activation loop (A-loop). When imatinib binds to Abl it induces a conformational change where the A-loop takes on a more compact inactive structure. Mutating this residue to a proline would mean that the A-loop would be unable to adopt this conformation, destabilising the inactive conformation [318].

The methionine residue at position 351, is located at the end of the activation loop near where the SH2 domain would bind. Owing to its location it is thought that mutation to a threonine (M351T) has a similar mechanism of resistance to that of H396P, although as this mutation only demonstrates moderate loss of drug sensitivity, it is sensible to presume that destabilisation of the inactive form is not as substantial [319].

4.6.3 Treating drug resistance

There are a number of alternative options available to patients with CML that is resistant to imatinib. Firstly, if the mutations are not too highly resistant, an increased dose of imatinib may be used [320, 321]. If this higher dose is unsuccessful or if the mutations are highly resistant then the use of alternative tyrosine inhibitors can be considered among other treatment options [293]. In 2010 the formerly second line inhibitors nilotinib [322, 323] and dasatinib [324] were approved by the FDA for first line treatment of CML alongside imatinib [287, 288, 325, 326]. Mutational analysis has been shown to influence the choice of inhibitor for CML patients, with NCCN treatment guidelines recommending the choice of one inhibitor over another depending on the presence of some mutations (Table 4.1).

Mutation	Treatment Recommendation
V299L, T315A, F317L/V/I/C	Nilotinib recommended
Y253H, E255K/V, F359V/C/I	Dasatinib recommended
E255K/V, F317L/V/I/C, F359V/C/I, T315A or Y253H	Bosutinib recommended
T315I	Ponatinib recommended

Table 4.6.1 Treatment options based on BCR-ABL kinase domain mutation status. Adapted from references [293, 327]

Although use of these second and third generation tyrosine inhibitors have proved very successful, there is a danger that wide use of these inhibitors could increase the risk of the emergence of the gatekeeper mutation T315I, which is resistant to four out of the five available inhibitors. Similarly, the increased use of ponatinib, which is effective against T315I, may in turn result in the emergence of other highly resistant mutations.

4.6.4 Compound mutations

The sequential use of TKIs has resulted in some patients developing multiple mutations within the same allele; assay data has suggested that a compound mutation can be far more resistant to TKI therapy than each of the constituting mutations, i.e. a TKI may be able inhibit X and Y mutation but not XY [328, 329].

Recent research on ponatinib resistance detailed how several compound mutations involving the gatekeeper mutation conferred resistance to all clinically available TKIs [330, 331].

4.7 Chapter summary

The oncogenic kinase BCR-Abl, is unregulated and interferes with cellular pathways, encouraging leukemic cells to grow and multiply and manifesting as the disease CML.

Five BCR-Abl inhibitors are used in clinical practice to treat CML; imatinib (Gleevec/Glivec), dasatinib (Sprycel), nilotinib (Tasigna), bosutinib (Bosulif), and ponatinib (Iclusig). Unfortunately drug resistance against these inhibitors, particularly imatinib, can occur in a small percentage of CML patients due to the development of point mutations in the BCR-Abl kinase. Focussed research and advances in mutational analysis have shown that some inhibitors are more effective against certain mutations than others. This means that for the majority of mutations, patients can switch to an alternative tyrosine inhibitor and continue with treatment. Unfortunately, the T315I mutation, known as the gatekeeper mutation, is highly resistant to all approved inhibitors except ponatinib, therefore patients with this mutation have very few treatment options compared to overall CML patient population. Furthermore, the sequential use of TKIs has resulted in some patients developing highly resistant compound mutations capable of conferring resistance to all clinically available TKIs.

Since this research was completed PDB structures of Abl in complex with ponatinib have become available (3OXZ.pdb) and been included in computational studies [330-332]. It would have been useful to include a complex with ponatinib in this research given the alternative inhibition profile compared to imatinib and nilotinib, in particular, the ability to bind to the T315I mutant.

This thesis is focused on the effect of five mutations on the Abl kinase domain and how they confer resistance to imatinib and nilotinib; the single mutations T315I, Y253F and F317L, the double mutation Y253F/F317L (YF) and the triple mutation T315I/Y253F/F317L (TYF) have been selected based on the differences in expected mechanisms of resistance and sensitivity to imatinib and nilotinib.

Chapter 5

5 Active BCR-Abl

5.1 Aim

The aim of this chapter is to validate the computational methods used to simulate and analyse the structure and gain insight into the dynamics of the active Abl kinase.

As discussed in section 4.2.3, the conserved characteristics of the active structure include:

- **An extended A-loop conformation** where the A-loop uncoils to function as a binding platform for the peptide substrate [265]; phosphorylation at Tyr393 appears to support the A-loop conformation by forming a number of hydrogen bonds with the surrounding residues. When Tyr393 is phosphorylated it forms a hydrogen bond with Arg362 (HRD motif) which is held in position on the catalytic loop due to another hydrogen bond with Asp421. The new phospho-Tyr393 also attracts Arg386 which induces the formation of the β6 and β9 sheet. These interactions enable the extended A-loop conformation to be held in place.

- **An elongated conformation of the P-loop** (glycine rich loop between β2 and β3) where the tip of the loop is able to coordinate the γ-phosphate for phosphoryl transfer on ATP binding [260]. Various conformations of the P-loop have been observed in structures that are intermediate between the active and inactive conformation [263].

- **An aligned regulatory spine (R-spine)** of four non-consecutive residues, two from the N-lobe and two from the C-lobe that form a stack of hydrophobic interactions thereby strengthening the C-N lobe connection and stabilising the active conformation [260]. In Abl these residues are Leu 301, Met290, Phe382 and His361 [266, 267]. Asp421 on the αF-helix makes a strong hydrogen bond to the main chain of His361, which secures the spine in place.

- **A catalytic spine (C-spine)** of eight residues that contribute to the positioning of ATP. Details of the residues forming the C-spine are not published for Abl, however in PKA these residues are: Val57, Ala70, Leu173, Leu172, Leu174, Met 128, Leu227 and Met231 [267].

- **A salt bridge between a Glu residue on the αC-helix and a Lys on the β3 beta sheet** resulting in a tight connection between the C and N lobes (Glu286 and Lys271 in Abl). Although this

- salt bridge is present in the inactive structure of Abl, in the active conformation Lys271 forms polar contacts with α and β phosphates of ATP, and therefore plays an important role in the correct positioning of ATP in the binding site.

- **The DFG-in conformation** (Asp381-Phe382-Gly383); Asp381 (D) is turned into the ATP binding site where it forms polar contacts (either directly or via magnesium atoms) with the α, β, and γ phosphates of ATP; the peptide bond between Leu384 and Ser385 (the two residues after the DFG loop) rotates and forms two hydrogen bonds with Arg362; this allows Asp381 (D) to form a hydrogen bond with glycine residue Gly383 (G), which helps secure the position of Asp381 (D) in the binding site. The DFG Phe382 (F) residue is part of the R-spine and also coordinates the correct positioning of the DFG Asp381 (D) residue by making hydrophobic contacts with the residue Met290 on the C-helix and His361 on the HRD motif (a conserved His-Arg-Asp sequence conserved in most PTKs); Met290 and His361 form part of the R-spine.

There are several structures of the active Abl kinase domain publicly available in the protein data bank (PDB), all of which are in complex with an inhibitor (e.g. 2QOH, 2GQG, 2F4J). In this research the structure 2F4J.pdb was used to model the active conformation. 2F4J is a 2.0 Å crystal structure of the kinase domain in complex with VX-680, a small molecule inhibitor that favours the binding to the active conformation through hydrogen bonding and steric effects. This structure is not phosphorylated on the activation loop but does contain a mutation, H396P, which destabilises the inactive conformation, facilitating the binding of VX-680 to the active conformation.

Since a crystal structure of the apo active form is not available, a pseudo apo structure was created by removing the ligand VX-680 from the 2F4J crystal structure ($2F4J_{p\text{-}apo}$). It is important to study the apo form since the presence of an inhibitor may restrict some the domain's flexibility.

5.2 Molecular dynamics study of active Abl

A pdb file of the kinase crystal structure was first obtained from the RCSB Protein Data Bank (2F4J.pdb [31]). Since there were no apo structures available, the 2F4J.pdb structure was duplicated and the ligand removed to create a pseudo-apo active structure. This provided two starting crystal structures of the active conformation: 2F4J and $2F4J_{p\text{-}apo.}$

5.2.1 Description of protein structure

In the 2F4J structure H396 has been mutated to a proline (H396P mutation) to destabilise the inactive conformation; this mutation was retained in the simulation. In other active kinases such as Lck, the residue at position 397 has a ϕ value of around -65; e.g. in 3LCK.pdb, Arg397 is at -67.71. In 2F4J, the ϕ value of Pro396 is constrained between -60 to -65 (measured as -58.40 in the crystal structure) [31]. This means that Pro396 is orientated to facilitate the active form rather than the inactive; in the inactive conformation the backbone dihedral angle ϕ of H396 would normally have a value of -174.

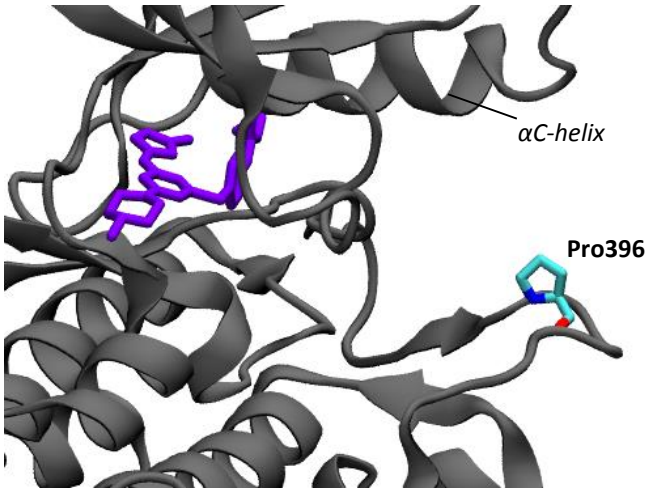


Figure 5.2.1 Position of the H396P mutation in 2F4J.pdb

5.2.2 Description of ligand binding

While the H396P mutation destabilises the inactive conformation, the binding of VX-680 stabilises the active form both sterically and through the formation of the hydrogen bonds with Met318, Glu316 and Asp381 (Figure 5.2.2).

VX-680 is a Y-shaped molecule made up of an N-methyl-piperazine group at the base, a pyrimidine group at the fork, a substituted phenyl group at one arm and a methylpyrazole group at the other arm.

VX-680 is anchored to the kinase domain by four hydrogen bonds. Three of these are formed between two carbonyl groups (Glu316 and Met318) and an amide nitrogen (Met318) in the hinge region on the kinase and three nitrogen atoms, one in the linker between the pyrimidine group and the methylpyrazole group and the other two in the methylpyrazole group. The fourth hydrogen bond is between the nitrogen of the amide group linking the phenyl group to the cyclopropyl substituent and the side chain of the aspartate of the DFG motif (Asp381). This last

bond is very well formed and has a distance of around 3A so likely stabilises the active conformation.

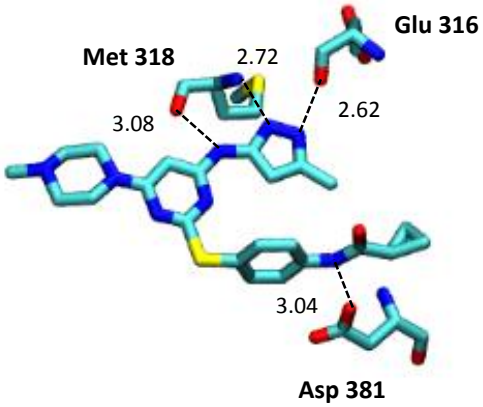


Figure 5.2.2: The VX-680 binding site in 2F4J.pdb. Hydrogen bond distances (Å) labelled. Adapted from reference [31].

5.2.3 Review of expected characteristics

Before the structure was prepared for the molecular dynamics simulation, the crystal structure was reviewed for the expected characteristics of the active conformation, in particular the extended A-loop, DFG-in conformation, regulatory and catalytic spine alignment, and elongated P-loop conformation.

The extended A-loop and DFG-in conformations were present in the 2F4J crystal structure (Figure 5.2.3), however the DFG conformation was not exactly as expected; Since the Asp381 residue forms a hydrogen bond with the ligand, the residue was turned upwards away from Gly383. In this conformation it would be unlikely that Asp381 would form a strong hydrogen bond with Gly383.

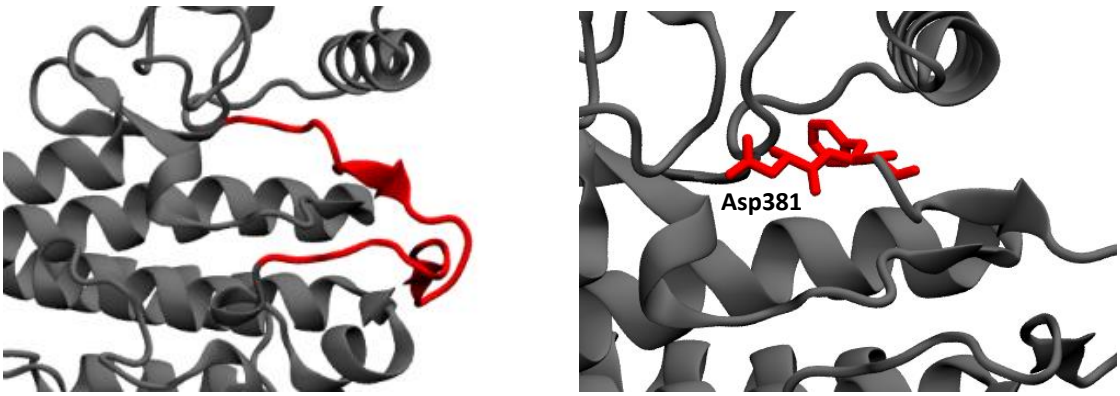


Figure 5.2.3: a) Extended conformation of the A-loop in the 2F4J.pdb crystal structure; b) Alternative Asp-in conformation of the DFG-loop in the 2F4J.pdb crystal structure

The R-spine residues Leu301, Met290, Phe382 and His361 and Asp421 on the αG-helix were found to be aligned in the crystal structure 2F4J.pdb (Table 4.6.1. and Figure 5.2.4).

Residues	Distance (CA-CA, Å)
Leu301 – Met290	6.04
Met290 – Phe382	7.32
Phe382 – His361	6.28
His361 – Asp421	6.73

Table 5.2.1: Distance between R-spine residues (2F4J.pdb)

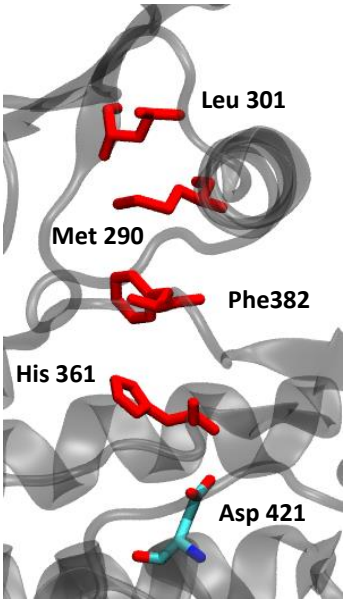


Figure 5.2.4: Residues in the R-spine (2F4J.pdb)

Details of the residues forming the C-spine are not published for Abl, however in PKA these residues are: Val57, Ala70, Leu173, Leu172, Leu174, Met 128, Leu227 and Met231 [267].

Residue	PKA (1ATP.pdb)	Abl (2F4J.pdb)
1	Val57	Val256
2	Ala70	Ala269
3	Leu173	Leu370
4	Leu172	Cys369
5	Ile174	Val371
6	Met128	Leu324
7	Leu227	Leu428
8	Met231	Ile432

Table 5.2.2: C-spine residues for PKA (1ATP.pdb) and Abl (2F4J.pdb)

Using sequence alignment with the PKA structure 1ATP.pdb and Abl structure 2F4J, the corresponding residues for Abl were established as Val256, Ala269, Leu370, Cys369, Val371, Leu324, Leu428 and Ile432 (see section 4.2.3)

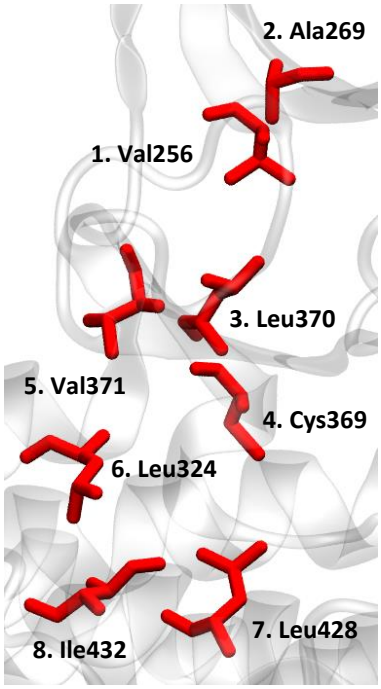


Figure 5.2.5: Residues in the C-spine (2F4J.pdb)

Abl (2F4J.pdb)	Distance (CA-CA, Å)
Ala269 - Val256	6.96
Val256 – Val371	16.39
Val256 – Leu370	13.99
Val371 – Leu370	3.79
Leu370 – Cys369	3.81
Val371 – Leu324	9.59
Leu324 – Cys369	7.80
Leu324 – Ile432	8.05
Cys369 – Leu428	9.79
Ile432 – Leu428	6.31

Table 5.2.3: Distance between C-spine residues (2F4J.pdb)

The residues of the C-spine were similarly aligned in the crystal structure of 2F4J as in the PKA structure 1ATP.pdb (Figure 5.2.5 and Table 5.2.3).

In the crystal structure of 2F4J, the P-loop is in a kinked inactive like conformation and not the elongated active conformation as expected (Figure 5.2.6). This conformation may be a result of ligand binding and given that the P-loop is highly flexible, may change during the course of the simulation.

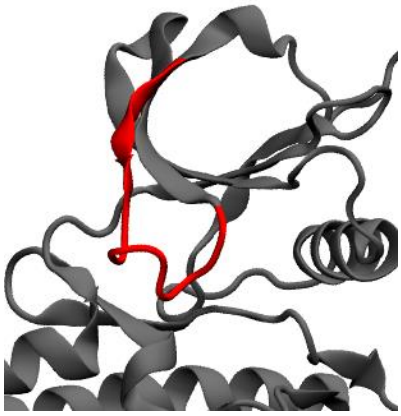


Figure 5.2.6: Conformation of the P-loop in the 2F4J crystal structure (ligand not shown)

5.2.4 Protein structure preparation and MD parameters

Before the molecular dynamics simulation the 2F4J and $2F4J_{p\text{-}apo}$ structures were first checked for additional molecules from the crystallographic process and missing residues. The pH was also noted to evaluate the protonation states (of histidine for example). The ligand in 2F4J (VX-680) was then protonated using the PRODRG website and the protein (combined with the coordinates of the protonated ligand) was protonated in WHATIF. The Amber package program XLEAP was used to solvate and neutralize the structures. Both structures were then solvated in water using the TIP3 model (box size of 10 Å) and neutralised using ions. This provided suitable structures for minimisation in NAMD.

Minimisation was carried out in NAMD using a cut-off of 11 Å with a switching distance of 9 Å. Minimisation is carried out in stages, allowing the system to adjust gradually, preventing any large initial forces which may lead to movement away from the original structure. The solvent and ions were minimized first, with the protein and ligand constrained, for 20000 steps (conjugate gradient), followed by a further 5000 steps (conjugate gradient) on the entire system. The first stage of equilibration was carried out with two runs of NVT ensemble. The maximum temperature for these runs was set at 300 K with a damping parameter of 10 ps^{-1} for the Langevin thermostat. The first run involved running a total of six sequential simulations of 20000 steps each for temperatures 50 K, 100 K, 150 K, 200 K, 250 K and 300 K. The second NVT run had the same maximum temperature of 300 °K but no intervals and ran for a total of 50000 steps. The next stage involved two simulations carried out in the NPT ensemble using the Nose-Hoover Langevin barostat. The first involved a damping parameter of 5 ps^{-1}, a piston period of 200 fs and decay parameter of 100 fs for 50,000 steps. The second, also for 50,000 steps had a damping parameter of $1ps^{-1}$, a piston period of 500 fs and a decay parameter of 300 fs. NPT production was also carried out in NAMD for a period for 20 ns. An Amber99 force field was used with the Langevin temperature set at 300 K and the piston period (the oscillation period) at 500 fs. The simulation was run for 200 blocks of 50,000 steps (each step equal to 2 fs), producing a 20ns trajectory.

5.3 Results: Active Abl kinase domain in complex with VX-680 (2F4J.pdb)

5.3.1 Structural stability and flexibility

RMSD and RMSF of major structural motifs

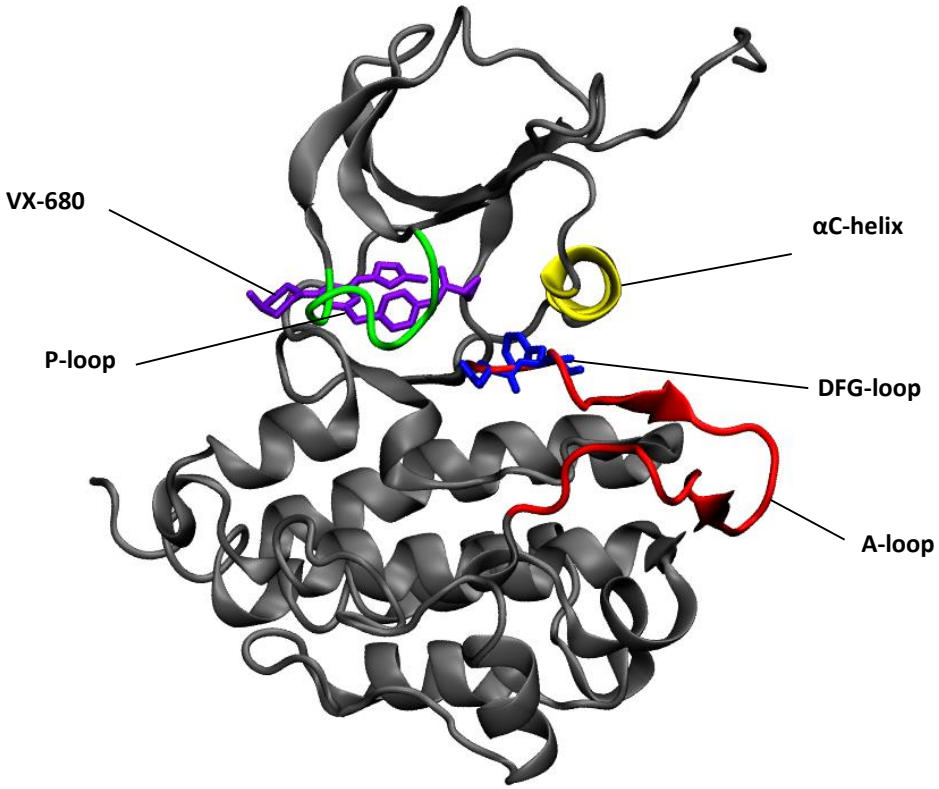


Figure 5.3.1: The key motifs of the active structure (2F4J.pdb)

The RMSD was calculated for the protein, αC-helix (residues 50-68), DFG-loop (residues 155-157), A-loop (residues 155-176) and the residues around the P-loop (residues 18-29) in order to capture any significant movement in these key motifs. After a small initial increase, the RMSD of the protein structure remained between 0.30 nm and 0.4 nm throughout the simulation. As expected the P-loop, A-loop and αC-helix demonstrated increased flexibility compared to the overall protein structure with an overall increase in RMSD and a series of higher fluctuation peaks during the simulation (Figure 5.3.2). The RMSD for the DFG-loop remained consistently low throughout the simulation and was lower than that of overall protein structure, demonstrating the stability of the DFG-loop in this conformation (Figure 5.3.2).

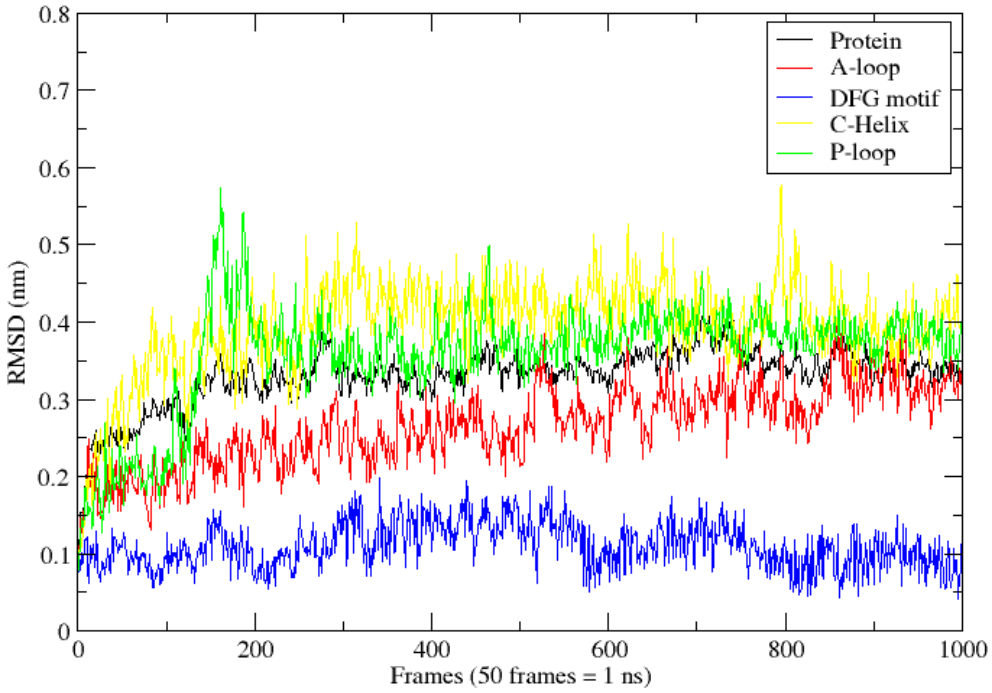


Figure 5.3.2: RSMD results for 2F4J.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green.

The RMSD results were mirrored by the RMSF calculations (Figure 5.3.3), which show large peaks around the P-loop region (residues 23-28) and the A-loop (residues 155-176). The RMSF results also show fluctuation peaks around the αC-helix (residues 54-64), αH-helix (residues 213-233) and αI-helix regions (residues 234-243). Further sequence and secondary structure details for 2F4J.pdb can be found in the appendices of this thesis.

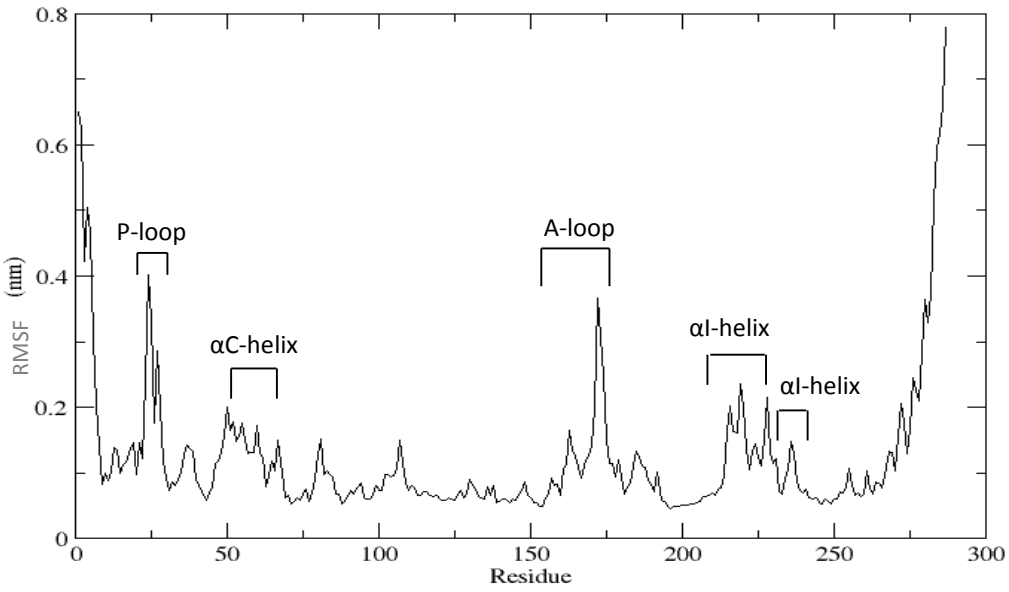


Figure 5.3.3: RMSF results for 2F4J.pdb.

The RMSD and RMSF have highlighted which sections of the active structure are most flexible and are consistent with what we would expect from the B-factor values (Figure 5.3.4).

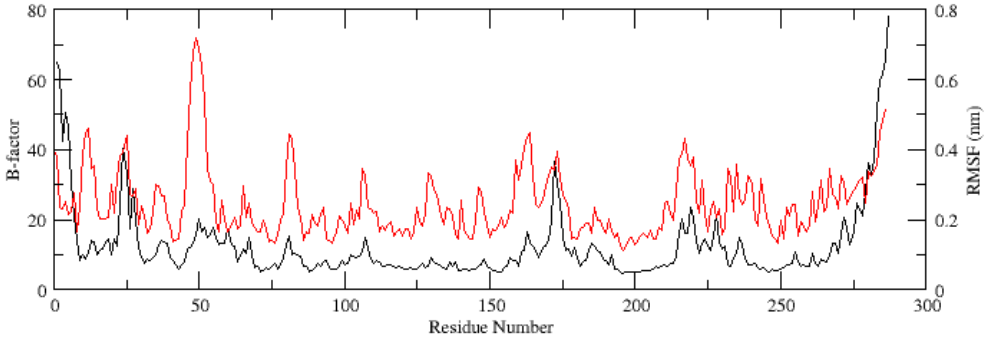


Figure 5.3.4: RMSF (black) vs B-factors (red) for 2F4J.pdb

In order to isolate the nature of motions, PCA was also completed for the simulation.

Principal Component Analysis

The PCA showed that the largest motions (PC1) were found in the P-loop and the second half of the activation loop, which supports the results from the structural analysis (Figure 5.3.5).

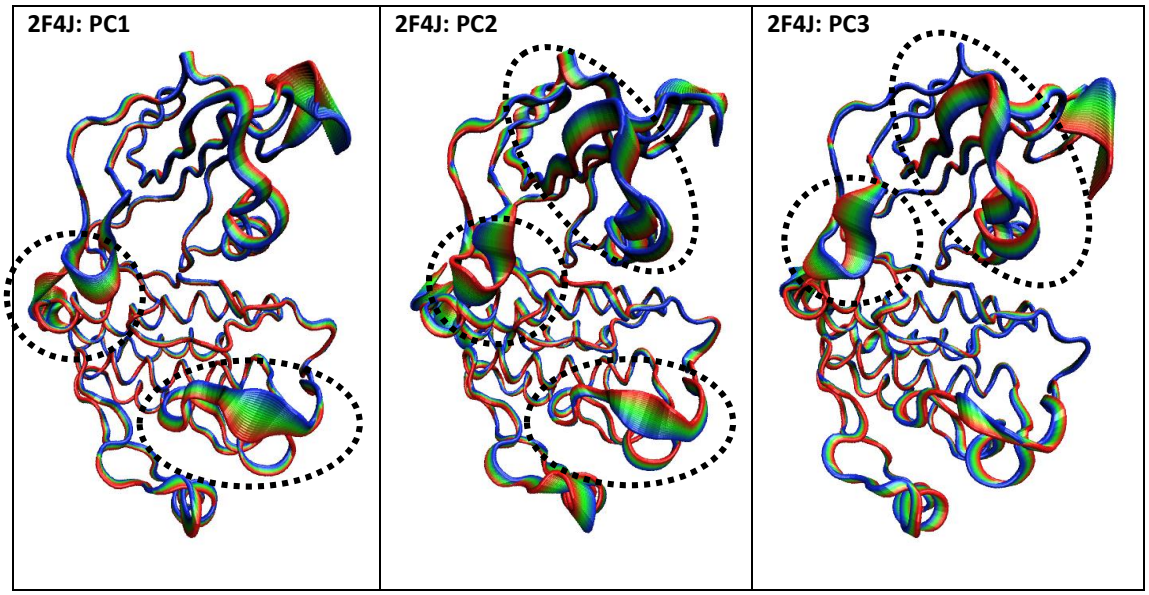


Figure 5.3.5: PCA projections for the first three principal components for active BCR-Abl (2F4J).

The second and third principal components showed motion in the residues surrounding the αC-helix and at the ends of the structure; however the major motions were located around the second half of the A-loop and the base of the P-loop. The P-loop conformation changes significantly during the simulation from the kinked conformation to the elongated conformation (Figure 5.3.6) seen in other active structures of Abl.

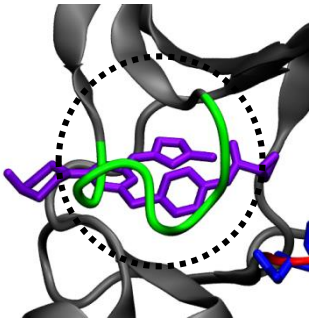
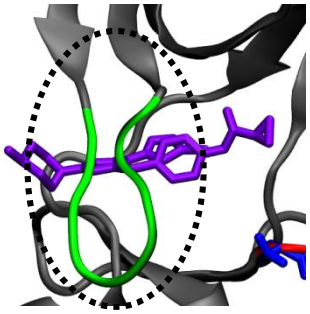
a)  b) 

Figure 5.3.6: The change in P-loop conformation (2F4J.pdb): a) Initial structure; b) Final structure post MD simulation

This conformation of the P-loop is associated with ATP binding, where the tip of the loop is able to coordinate the γ-phosphate for phosphoryl transfer. This conformation is likely stabilised by the p-loop interacting with the ligand.

Analysis of catalytic and regulatory spines

The distances between residues in the two hydrophobic spines, the regulatory spine (R-spine) and the catalytic spine (C-spine) (Figure 5.3.7) were analysed.

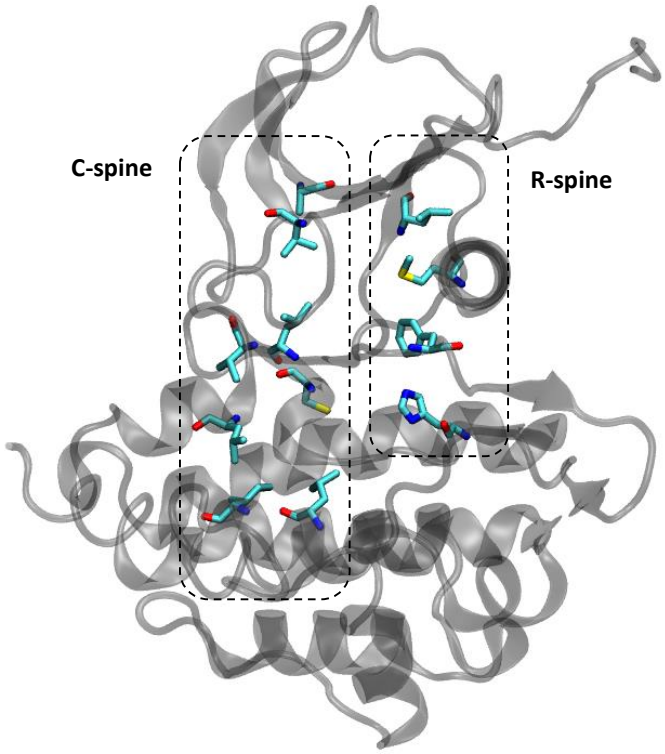


Figure 5.3.7: The regulatory and catalytic spines in the active conformation (2F4J.pdb)

The alignment of both the regulatory (R-) and catalytic (C-) spines was stable throughout the simulation and the average distance between the residues remained similar to those measured in the crystal structure (Table 5.3.1 and Table 5.3.2.).

R-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Leu301 – Met290	6.127	6.04
Met290 – Phe382	7.528	7.32
Phe382 – His361	6.396	6.28

Table 5.3.1: Mean distance between R-spine residues

C-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Ala269 - Val256	7.553	6.96
Val256 – Val371	16.197	16.39
Val256 – Leu370	14.186	13.99
Val371 – Leu370	3.873	3.79
Leu370 – Cys369	3.865	3.81
Val371 – Leu324	9.570	9.59
Leu324 – Cys369	7.803	7.80
Leu324 – Ile432	8.123	8.05
Cys369 – Leu428	9.773	9.79
Ile432 – Leu428	6.236	6.31

Table 5.3.2: Mean distance between C-spine residues

Critical hydrogen-bond analysis

Several hydrogen bonds are associated with the active conformation on Abl; Lys271 and Glu286, Asp381 and Gly383, Asp362 and Asp421, Leu384 and Arg362, and His361 and Asp421. The stability of these hydrogen bonds during the 20 ns simulations was measured in VMD using a cut-off of 4Å and 30°. These interactions are described below in Table 5.3.3.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	29.87
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	5.19
		Gly383(M)	Asp381(M)	9.49
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	96.30
		Arg362(S)	Asp421(S)	0.30
Leu384 – Arg362	Magnesium binding loop forms two hydrogen bonds with R of the HRD motif	Leu384(S)	Arg362(M)	17.58
		Leu384(S)	Arg362(S)	37.06
		Arg362(S)	Leu384(M)	7.19
		Arg362(S)	Leu384(S)	2.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	76.42
		His361(S)	Asp421(S)	4.30

Table 5.3.3: Hydrogen bond analysis: 2F4J (4 Å, 30°); S = sidechain, M = main chain.

The hydrogen bonds between Asp421 and Arg362 and His361 were extremely well formed and present throughout the duration of the simulation. The salt bridge between Lys271 and Glu286 was not as stable as expected, being present for around 30% of the simulation; on further investigation this was due to residues Lys271 and Glu286 both forming interactions with the VX-680 ligand. As predicted the hydrogen bond between Asp381 and Gly383 was not particularly stable, due to the upturned conformation of Asp381, and only present for less than 10% of the simulation.

Binding site analysis

Fluctuations in the size of the binding pocket were measured using PocketAnalyzer. The core pocket was correctly identified beyond the methylpiperazine ring of VX-680, which sticks out of the binding site, behind the P-loop.

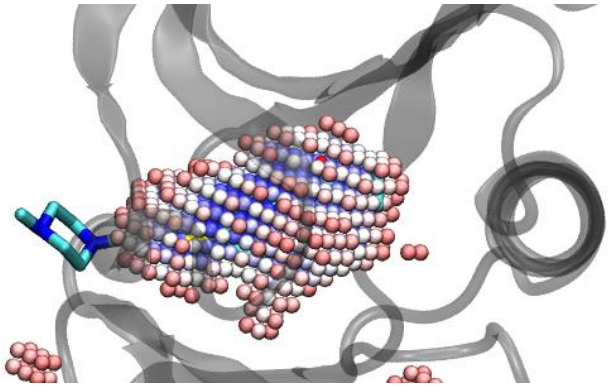


Figure 5.3.8: Binding site analysis, average pocket overlaid on the starting structure (2F4J.pdb). Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown).

The size and shape of the pocket was consistent throughout the simulation. Lower occupancies that account less stable conformations of the pocket (e.g. positions where there is a pocket for less than 40% of the simulation) only showed a small increase in pocket size.

5.3.2 Protein-ligand binding

RMSD and RMSF of VX-680

The binding site analysis indicated that the RMSD of the ligand was relatively low, and this was confirmed by RMSD and RMSF analysis on VX-680.

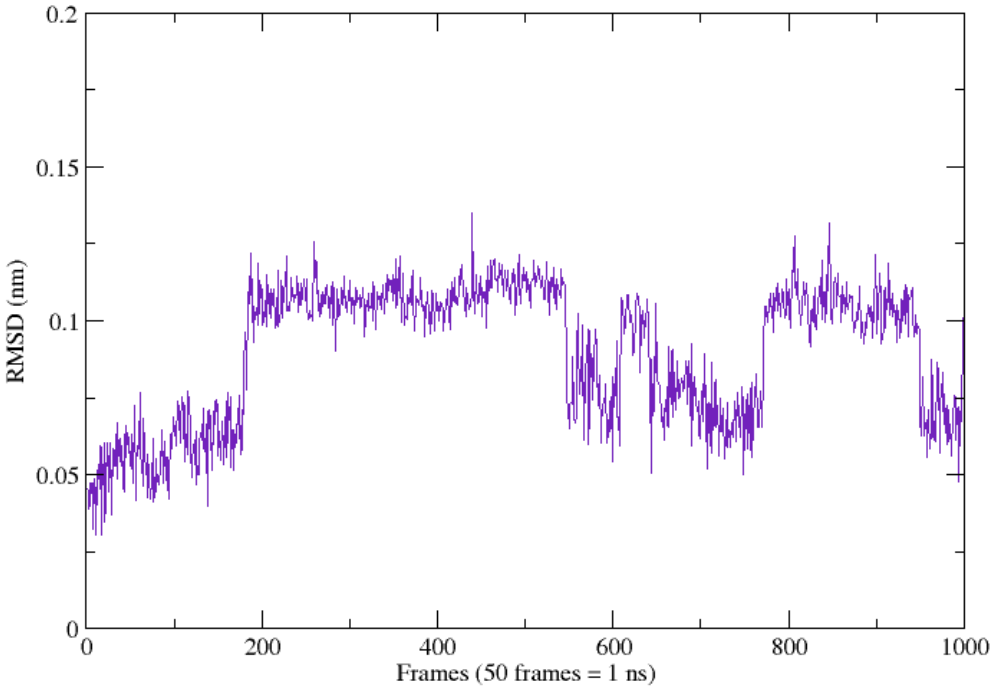


Figure 5.3.9: RMSD of VX-680 (2F4J.pdb)

The ligand fluctuated around 0.5 Å throughout the simulation. These motions were predominantly localized at the two points most exposed to the solvent; the oxygen before the cyclopropyl group (atom 5, Figure 5.3.10) and the methylpiperazine ring (atoms 27-33, Figure 5.3.10) as seen in Figure 5.3.11.

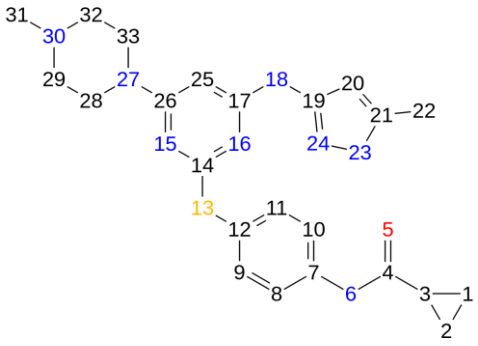


Figure 5.3.10: Index of atoms in VX-680

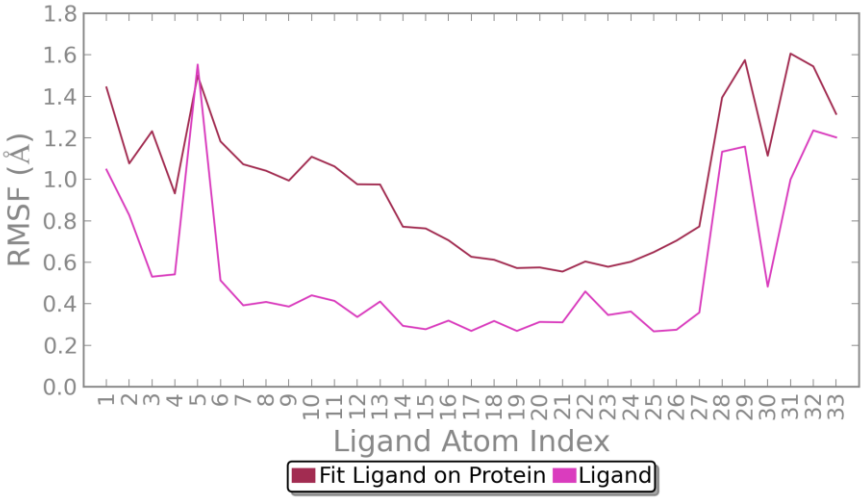


Figure 5.3.11: RMSF of ligand VX-680 (2F4J.pdb)

Protein-ligand interactions

In the crystal structure of 2F4J, VX-680 formed four hydrogen bonds with the protein structure.

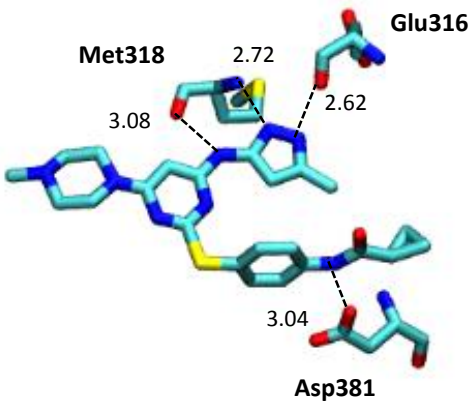


Figure 5.3.12: The VX-680 binding site in 2F4J.pdb. Hydrogen bond distances (Å) labelled. Adapted from reference [31].

Three of the hydrogen bonds were between two carbonyl groups (Glu316 and Met318) and an amide nitrogen (Met318) in the hinge region and three nitrogen atoms, one in the linker between the pyrimidine group and the methylpyrazole group and the other two in the methylpyrazole group. The fourth hydrogen bond is between the nitrogen of the amide group

linking the phenyl group to the cyclopropyl substituent and the side chain of the aspartate of the DFG motif (Asp381).

Protein-ligand interactions during the simulation were analysed using the Simulation Interactions Diagram; the results confirmed that Glu316 and Met318 form strong H-bonds with the ligand (Glu90 and Met 92 in Figure 5.3.13), however the hydrogen bond between Asp381 (Asp155 in Figure 5.3.13) of the DFG loop and the ligand does not appear to be as strong as the crystal structure suggests; interactions are formed via a water bridge, shown in Figure 5.3.14, and the average distance between Asp381 (Asp155 in Figure 5.3.13) and the ligand throughout the simulation was found to be 6.78 Å compared to 3.07 in the crystal structure.

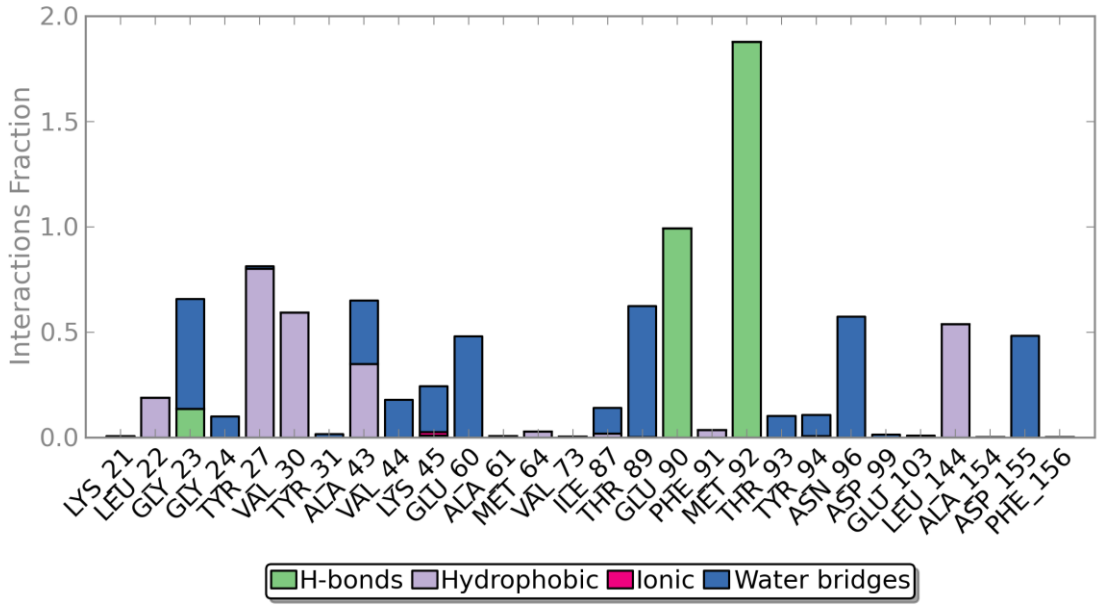


Figure 5.3.13: Protein-ligand interactions by residue (2F4J.pdb)

5: Active BCR-Abl

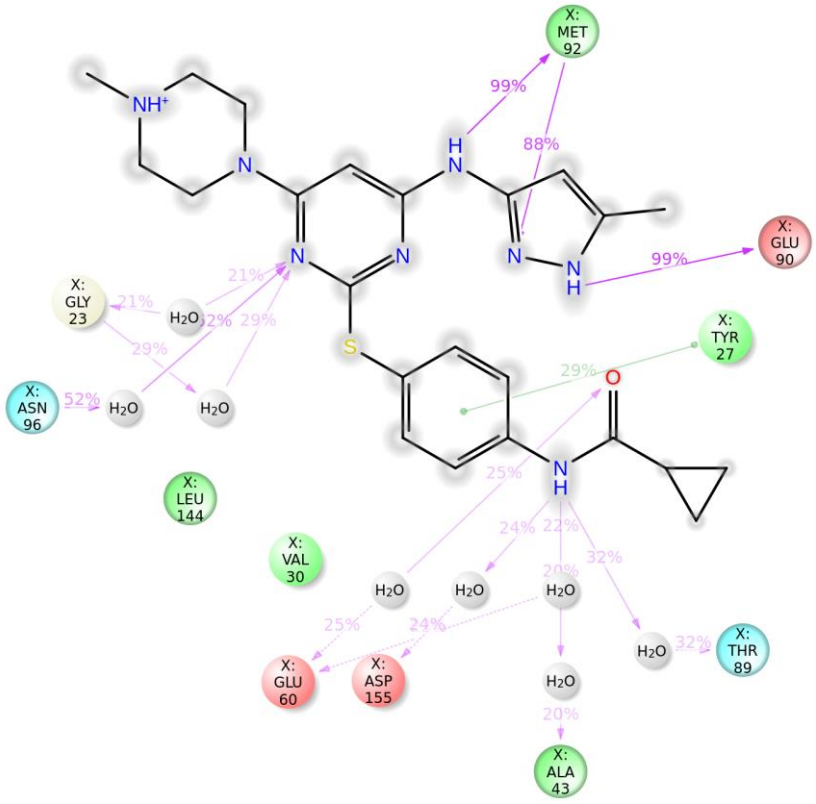


Figure 5.3.14: Protein-ligand interaction diagram for VX-680 (2F4J.pdb)

Additional interactions identified include water bridge interactions with Lys271 and Glu386 (the conserved salt bridge residues) and hydrophobic, hydrogen bond and water bridge interactions with residues around the P-loop (Gly23/249, Tyr27/253 and Val30/256).

5.4 Results: Active Abl kinase domain pseudo apo structure ($2F4J_{p\text{-}apo}$)

5.4.1 Structural stability and flexibility

RMSD and RMSF of major structural motifs

As with the original holo structure, the RMSD was calculated for the protein, αC-helix (residues 50-68), DFG-loop (residues 155-157), A-loop (residues 155-176) and the P-loop (residues 18-29) in order to capture any significant movement in these key motifs (Figure 5.4.1). The results were well aligned with the simulation for 2F4J in complex with VX-680 and with only a slight increase in the RMSD of the overall protein structure and the A-loop compared to the holo structure. The RMSD for the DFG-loop increased slightly during the simulation but was lower than that of overall protein structure.

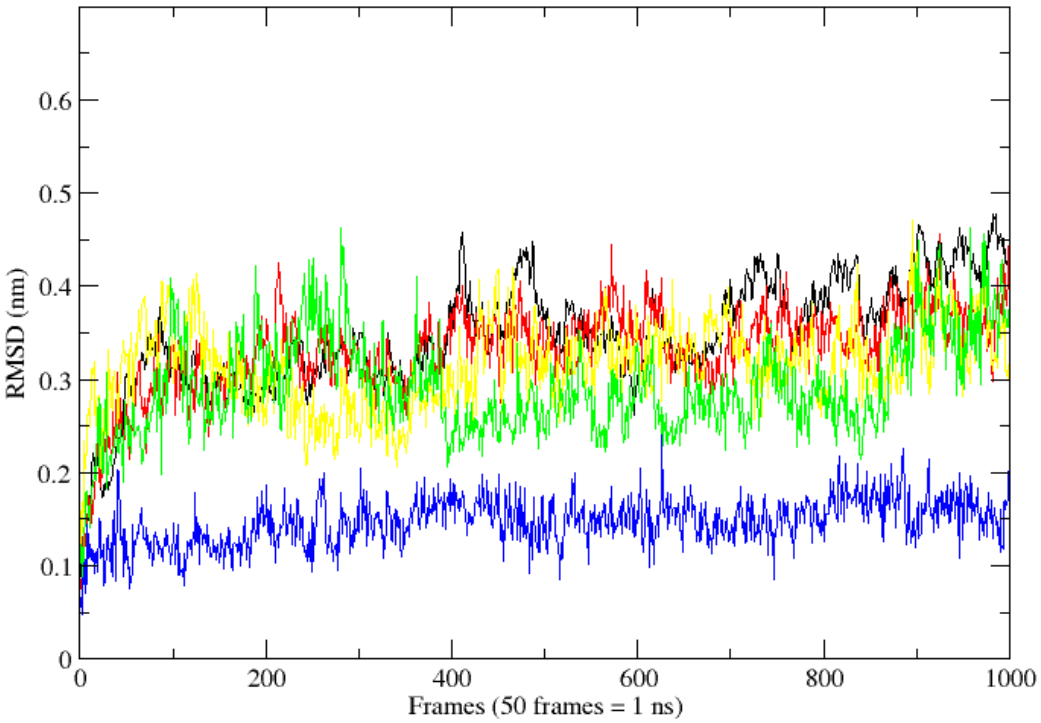


Figure 5.4.1: RSMD results for $2F4J_{p\text{-}apo}$; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green.

The A-loop, P-loop, αC-helix and the loop section at the end of αE-helix (residues 130-135) showed a slight increase in RMSF in the pseudo apo structure compared to the holo structure (Figure 5.4.2). The increase in the RMSD and RMSF in these regions is likely due to the protein structure relaxing into the empty pocket.

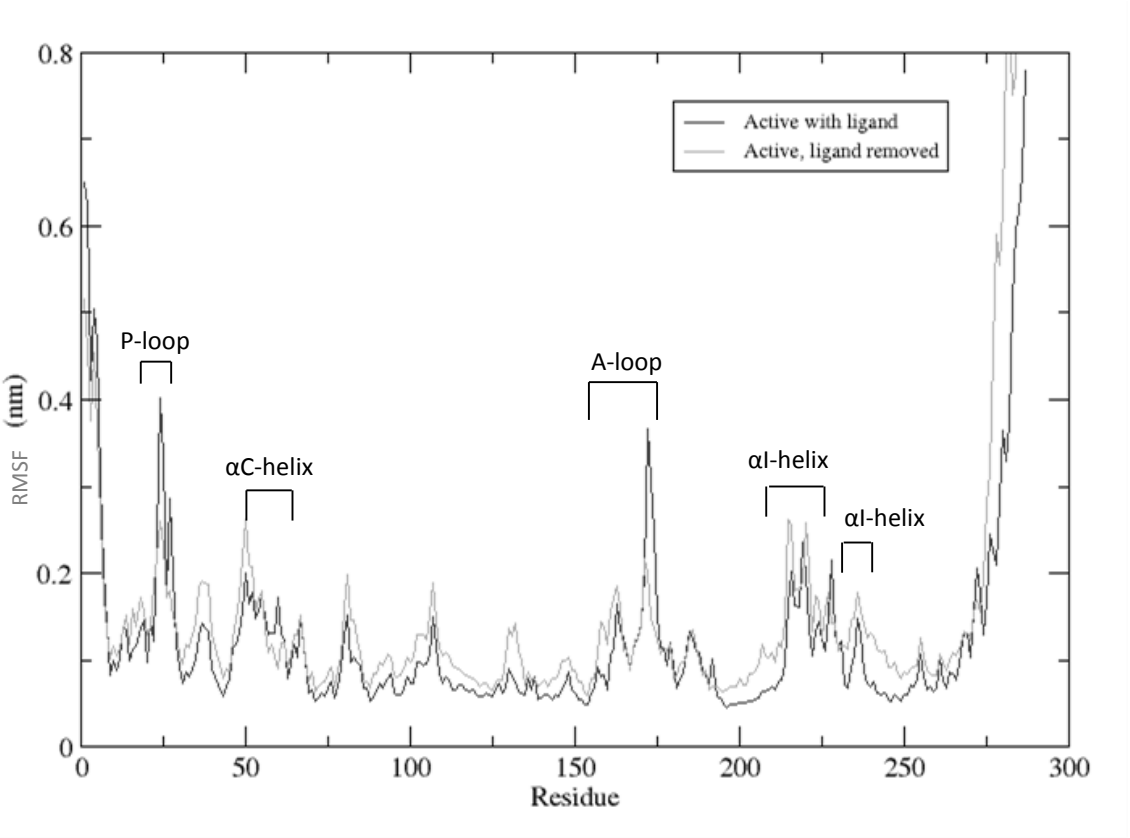


Figure 5.4.2: RMSF results for active BCR-Abl with VX-680 (2F4J, black) and active BCR-Abl pseudo -apo structure (2F4Jp-apo, grey)

Principal component analysis

The PCA analysis (Figure 5.4.3) highlighted increased motion around the αC-helix of the $2F4J_{p\text{-}apo}$ structure, with the loop section just before αC-helix moving downwards into the empty pocket and smaller upward motions in the A-loop. Although there were some motions around the P-loop, these were diminished compared to the holo structure and the loop did not extend into the fully elongated conformation.

In the 2F4J simulation the P-loop forms several interactions with the ligand and DFG loop which causes the initial arrangement into the elongated conformation. The conformation is then stabilised by residues in the P-loop forming hydrogen bonds across the width of the loop.

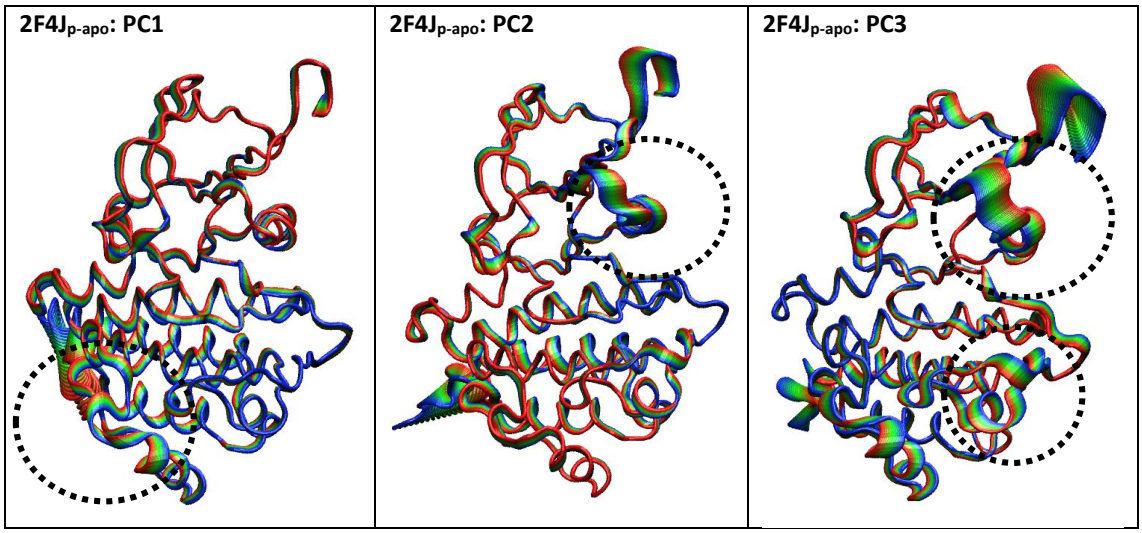
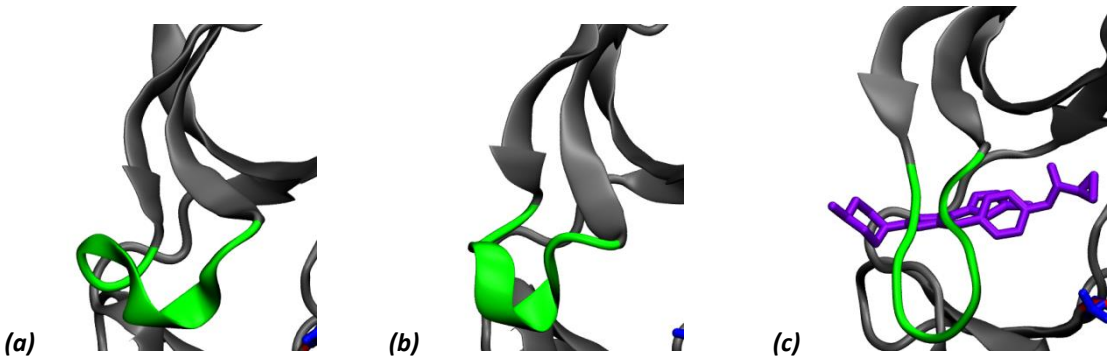


Figure 5.4.3: PCA projections for the first three principal components for the active BCR-Abl apo structure (2F4Jp-apo).

In 2F4J$_{p\text{-}apo}$ the P-loop extends slightly, forming a strong H-bond with Arg367 but maintains a slightly kinked conformation due to residues at the end of the loop forming hydrogen bonds with those higher up.



5.4.4: The change in P-loop conformation: a) Initial structure (2F4J$_{p\text{-}apo}$); b) Final structure post MD simulation (2F4J$_{p\text{-}apo}$); c) Final structure post MD simulation (2F4J)

It is possible that with a longer simulation the P-loop could extend more fully, similar to the conformation achieved in 2F4J.

Analysis of catalytic and regulatory spines

The hydrophobic spines remained well aligned in the pseudo-apo structure and the mean distance between the residues was similar to that of the holo structure. In the C-spine the distance between Val256 and Val371/Leu370 is approximately 2Å larger in the pseudo-apo

structure. This is due to the P-loop conformation not being as extended; therefore Val256, which sits at the end of the P-loop, is positioned slightly further away from residues Val371 and Leu370.

R-spine residues	Mean distance (CA-CA, Å)	
	2F4J_p-apo	2F4J
Leu301 – Met290	6.060	6.127
Met290 – Phe382	7.584	7.528
Phe382 – His361	6.619	6.396

Table 5.4.1: Mean distance between R-spine residues

C-spine residues	Mean distance (CA-CA, Å)	
	2F4J_p-apo	2F4J
Ala269 - Val256	7.387	7.553
Val256 – Val371	14.942	16.197
Val256 – Leu370	12.252	14.186
Val371 – Leu370	3.859	3.873
Leu370 – Cys369	3.863	3.865
Val371 – Leu324	9.627	9.570
Leu324 – Cys369	7.906	7.803
Leu324 – Ile432	8.111	8.123
Cys369 – Leu428	10.259	9.773
Ile432 – Leu428	6.103	6.236

Table 5.4.2: Mean distance between C-spine residues

Critical hydrogen-bond analysis

The critical hydrogen bonds required for the active conformation were conserved in the pseudo apo structure. One of the interactions, the salt bridge between Lys271 and Glu286, appeared to be more stable in this simulation than that for the holo structure; present for 93.71% of the simulation compared to 29.87%.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				2F4J p-apo	2F4J
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	93.71	29.87
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	8.19	5.19
		Gly383(M)	Asp381(M)	4.20	9.49
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	98.90	96.30
		Arg362(S)	Asp421(S)	-	0.30
Leu384 – Arg362	Magnesium binding loop forms two hydrogen bonds with R of the HRD motif	Leu384(S)	Arg362(M)	10.09	17.58
		Leu384(S)	Arg362(S)	38.56	37.06
		Arg362(S)	Leu384(M)	14.39	7.19
		Arg362(S)	Leu384(S)	4.50	2.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	71.33	76.42
		His361(S)	Asp421(S)	8.09	4.30

Table 5.4.3: Hydrogen bond analysis: 2F4J$_{p\text{-}apo}$ vs. 2F4J (4 Å, 30°); S = sidechain, M = main chain.

A review of the interactions between Lys271 and Glu286 during the simulation revealed that the distance between the two residues is consistently short in the 2F4J$_{p\text{-}apo}$ simulation, whereas in 2F4J simulation there are large fluctuations.

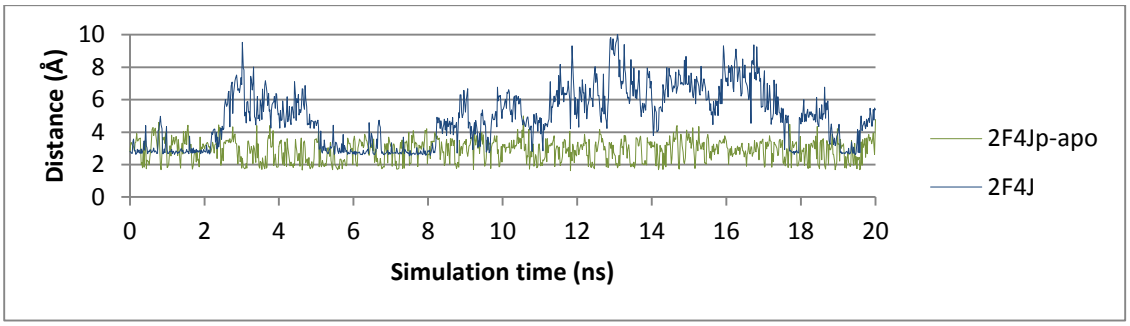


Figure 5.4.5: Distance between Glu286 and Lys271

The Asp381 residue maintained an upturned conformation during the simulation, as in the holo structure, despite the removal of the ligand.

Binding site analysis

Pocket analysis using PocketAnalyzer confirmed that the size of the pocket was reduced in the apo structure; only a fraction of the original pocket was present for over 60% of the simulation, compared to the holo structure.

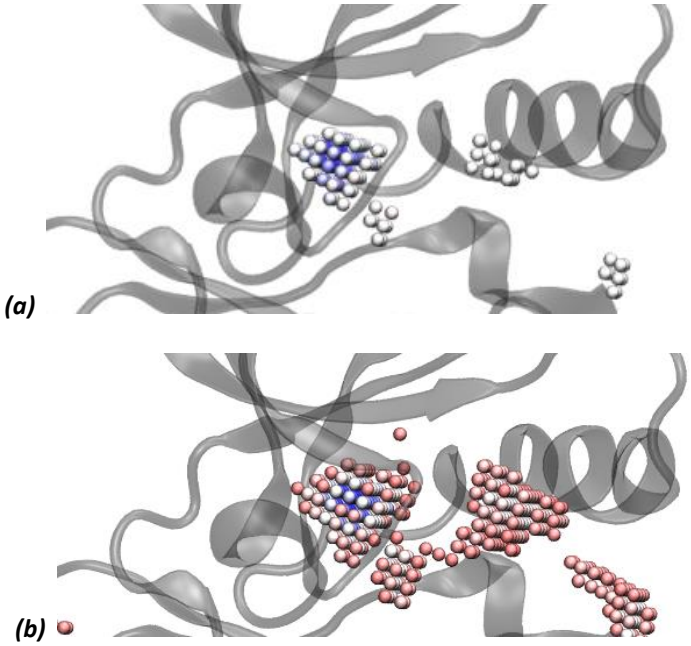


Figure 5.4.6: Binding site analysis, average pocket overlaid on the starting structure ($2F4J_{p\text{-}apo}$) a) Most conserved region of pocket (pocket present 60-80% of simulation). b) Less conserved regions of pocket (pocket present 20-80% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown).

5.5 Chapter summary

In this chapter the structure and dynamics of the active conformation have been investigated using molecular dynamics simulations on two active structures; one in complex with a small inhibitor (2F4J) and one with the inhibitor removed ($2F4J_{p\text{-}apo}$). This work has been completed to confirm the stable characteristics of the active state.

The structural analysis of 2F4J highlighted the flexibility of the P-loop, A-loop and αC-helix regions compared to the overall protein structure. Whereas the A-loop maintained the extended conformation seen in the crystal structure, the P-loop changed from a kinked to elongated conformation during the simulation. This conformation is associated with ATP binding; however it is clear that in the active state the P-loop can move between kinked and elongated conformations.

The R- and C-spines of 2F4J were well conserved throughout the simulation and the average distances between the residues were comparable with the crystal structure.

The critical hydrogen bonds between Asp421 and Arg362 and His361 were extremely stable and present throughout the duration of the simulation, however the salt bridge between Lys271 and Glu286 was destabilised due to these residue forming interactions with the VX-680 ligand. As expected the hydrogen bond between Asp381 and Gly383 was not well formed due to the upturned conformation of Asp381.

Binding site analysis revealed that the size and shape of the pocket was consistent throughout the simulation; this was supported by the RMSD results for the VX-680 ligand which were consistently low during the simulation. Results from the protein-ligand interactions analysis showed that Glu316 and Met318 maintained strong hydrogen bonds with the ligand, however the hydrogen bond with Asp381 did not appear to be as strong as the crystal structure suggested, since the Asp381 moved further away from the ligand during the simulation.

Removing the ligand VX-680 from the 2F4J ($2F4J_{p\text{-}apo}$) resulted in an increase in RMSF and RMSD as the protein relaxed into the empty pocket; PCA analysis highlighted the movement of the loop section just before αC-helix moving downwards into the empty pocket and smaller upward motions in the A-loop. Binding site analysis confirmed that only a fraction of the original pocket was maintained during the simulation for the $2F4J_{p\text{-}apo}$ structure. Structural analysis revealed that the P-loop conformation did not extend fully in the $2F4J_{p\text{-}apo}$ structure and maintained an elongated yet kinked intermediate conformation, supporting the evidence that the P-loop can exhibit various different conformations.

The hydrophobic R- and C-spines remained well aligned in the $2F4J_{p\text{-}apo}$ structure and the mean distance between the residues was similar to that of the holo and crystal structures. The critical hydrogen bonds required for the active conformation were also conserved in the $2F4J_{p\text{-}apo}$ structure. One of the interactions, the salt bridge between Lys271 and Glu286, was actually more stable in this simulation than that of the holo structure, due to the two residues being closer together as the protein relaxed into the empty binding site. Despite the removal of a water bridge interaction between Asp381 and VX-680, the DFG motif maintained an upturned conformation as in the holo structure.

Based on the results of this chapter the following characteristics appear to be stable markers of the active conformation:

- An extended A-loop conformation
- The DFG-in conformation (Asp381-Phe382-Gly383), but not necessarily the hydrogen bond between Asp381 and Gly383.
- An aligned regulatory spine (R-spine, based on the distances between residues)
- An aligned catalytic spine (C-spine, based on the distances between residues)
- An elongated and intermediate conformations of the P-loop

The characteristics of the inactive structure will be reviewed in Chapter 6.

Chapter 6

6 Inactive BCR-Abl

6.1 Aim

The aim of this chapter is to compare the binding modes of imatinib and nilotinib, and gain insight into the dynamics of the inactive Abl kinase structure so that any deviation from this conformation in the mutation studies (Chapters 7 and 8) can be identified.

As discussed in section 4.2.2, the conserved characteristics of the inactive structure include:

- A folded conformation of the A-loop; Tyr393 forms a hydrogen bond with Asp363, causing the loop to adopt a folded conformation where it is curled up into the active site, blocking the binding of the substrate and ATP.

- The DFG Asp-out conformation, where DFG "D"(Asp381) is flipped 180° away from the binding site and the DFG "F" (Phe382) is positioned in the binding site instead [249].

- Misalignment of the R-spine; a stack of hydrophobic residues (Leu 301, Met290, Phe382 and His361) that are aligned in active Abl stabilising the active conformation. In the inactive state, the conformation of the A-loop and DFG loop means that the residues that make up the R-spine are displaced and cannot form the hydrophobic stack observed in the active conformation [249, 260, 266, 267] .

- An upturned or kinked conformation of the P-loop (glycine rich loop between β2 and β3) away from the binding site. The P-loop is typically elongated in the active conformation to coordinate the transfer of phosphate ions when ATP binds to Abl. Various kinked conformations of the P-loop have been observed from crystal structures of the inactive state, however since all of these structures are in complex with inhibitors it is unclear whether the kinked conformation is a natural conformation of the inactive form or a result of inhibition [262]. The P-loop is highly flexible and various conformations have been observed in structures that are intermediate between the active and inactive conformation [263].

Changes in the conformation of the C-spine between the active and inactive states are not well understood. This chapter will identify any changes in the spine's alignment compared to the active conformation. The A-loop in inactive Abl is in a slightly more extended conformation than in other inactive kinases and the αC-helix is closer to the normal active position allowing formation of the Lys-Glu salt bridge between a Glu residue on the αC-helix and a Lys on the β3 beta sheet, which is a conserved motif in active kinases [249]. This chapter will review the interactions between Lys271 and Glu286 to ascertain the differences, if any, in this motif between the active and inactive structures.

There are several structures of the inactive Abl kinase domain publicly available in the protein data bank (PDB), all of which are in complex with an inhibitor. For the inactive Abl kinase domain in complex with imatinib PDB structure 1IEP.pdb was selected. The structure for the inactive Abl kinase domain in complex with nilotinib was originally provided by Novartis but has since been made publicly available (PDB ID: 3CS9.pdb) [32]. Ideally, if time had permitted, multiple simulations using several active and inactive structures would have been completed in order to form a more robust evidence base.

6.2 Molecular dynamics study of inactive WT Abl in complex with imatinib and nilotinib

The PDB files of the kinase crystal structures were obtained from the RCSB Protein Data Bank (1IEP.pdb) and Novartis (3CS9.pdb). The 1IEP structure was duplicated and the ligand was removed from one of the structures to give three starting crystal structures of the inactive conformation: 1IEP, $1IEP_{p\text{-}apo}$ and 3CS9.

6.2.1 Description of protein structures

1IEP is a 2.10 Å crystal structure of the unphosphorylated kinase domain in complex with imatinib (STI-571), a small molecule inhibitor that favours the binding to the inactive conformation of Abl [30]. Since a crystal structure of the apo inactive form was not available, a pseudo apo structure was created by removing imatinib from the 1IEP.pdb crystal structure ($1IEP_{p\text{-}apo}$).

3CS9 is a 2.21 Å crystal structure of the unphosphorylated kinase domain in complex with nilotinib (AMN107), which also inhibits the inactive conformation but is more potent than imatinib [31].

6.2.2 Description of ligand binding

Imatinib

Imatinib (STI-571) binds to the kinase domain with an induced fit interaction; as imatinib binds, Asp381 from the DFG loop is flipped out of the position where it would be required to coordinate the magnesium ions in the ATP phosphate binding site. The inhibitor binds in the ATP binding site and extends through the centre of the kinase into a hydrophobic pocket opened up by displacement of the DFG loop.

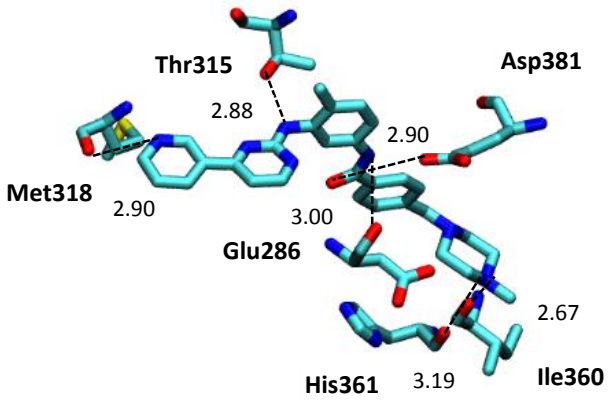


Figure 6.2.1: The imatinib binding site in 1IEP.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [30]

Imatinib forms hydrophobic and van der Waals interactions with the Abl kinase, as well as six hydrogen bonds with residues in the binding site; the amino group of pyrimidine and Thr315, the amide NH and Glu286, the protonated piperazine and Ile360 and His261, the amide CO and Asp381, and the pyridine N and Met318. The P-loop adopts a kinked conformation on ligand binding as it folds over the pyridine and pyrimidine rings of imatinib; this allows residues Tyr253 (P-loop), Phe382 (DFG), Phe317, Leu248, and Leu370 to effectively form a hydrophobic shield around this section of the ligand. The opposite end of the ligand is more exposed to the solvent and forms van der Waals contacts with Ile360, His 361, Asp381 (DFG), Met290 and Val289 [30, 262].

Nilotinib

Nilotinib is based on the imatinib structure, retaining an amide substructure to maintain the H-bond interactions to Glu286 and Asp381, but with alternative binding groups for the N-methylpiperazine group; replacement of the piperazine ring with a trifluorinated imidazole

eliminated two energetically unfavourable hydrogen bonds with Ile360 and His361. The replacement imidazole ring fits well into the hydrophobic pocket resulting in a better topological fit and increased potency compared with imatinib [297].

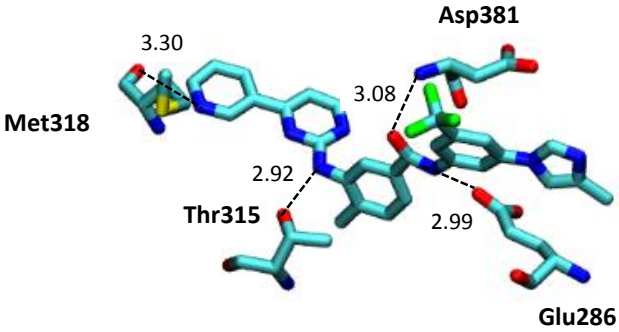


Figure 6.2.2: The nilotinib binding site in 3CS9.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [262]

As with imatinib, nilotinib binds to the inactive conformation of Abl and forms hydrogen bonds with the same residues, save Ile360 and His361: the amino group of pyrimidine and Thr315, the amide NH and Glu286, the amide CO and Asp381 and the pyridine N and Met318.

6.2.3 Review of expected characteristics

Before the structure was prepared for the molecular dynamics simulation, the crystal structure was reviewed for the expected characteristics of the inactive conformation, in particular the folded A-loop, DFG Asp-out conformation, regulatory and catalytic spine alignment, and kinked P-loop conformation.

The conformations of the A-loop, DFG loop and P-loop were as expected in the two inactive structures; the A-loop was folded into the kinase domain, the DFG was in the Asp-out conformation with Ph382 in the binding site, and the P-loop was in a kinked conformation.

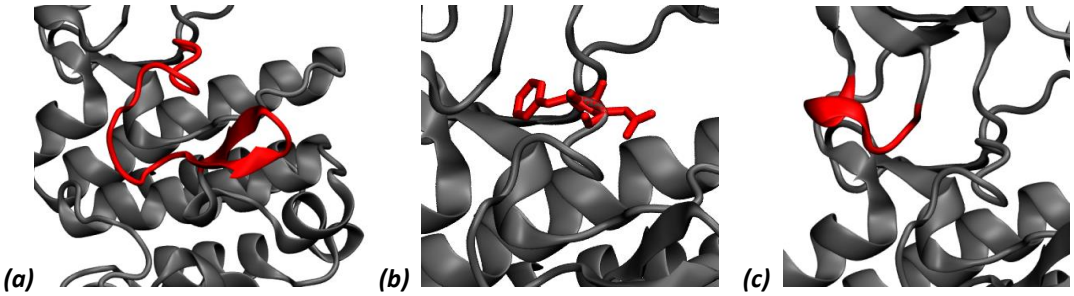


Figure 6.2.3: Conformation of the A-loop, DFG-loop and P-loop in the 1IEP crystal structure.

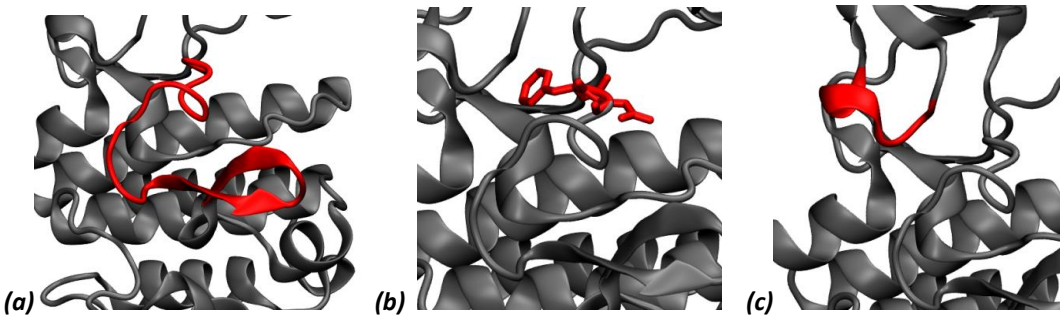


Figure 6.2.4: Conformation of the A-loop, DFG-loop and P-loop in the 3CS9 crystal structure.

The R-spine residues Leu301, Met290, Phe382 and His361 and Asp421 on the αG-helix were found to be misaligned in both of the inactive crystal structures (1IEP.pdb and 3CS9.pdb).

Residues	Distance (CA-CA, Å)		
	1IEP.pdb	3CS9.pdb	2F4J.pdb
Leu301 – Met290	6.418	6.460	6.04
Met290 – Phe382	12.341	12.259	7.32
Phe382 – His361	11.109	11.341	6.28
His361 – Asp421	6.534	6.493	6.73

Table 6.2.1: Distance between R-spine residues (1IEP.pdb and 3CS9.pdb)

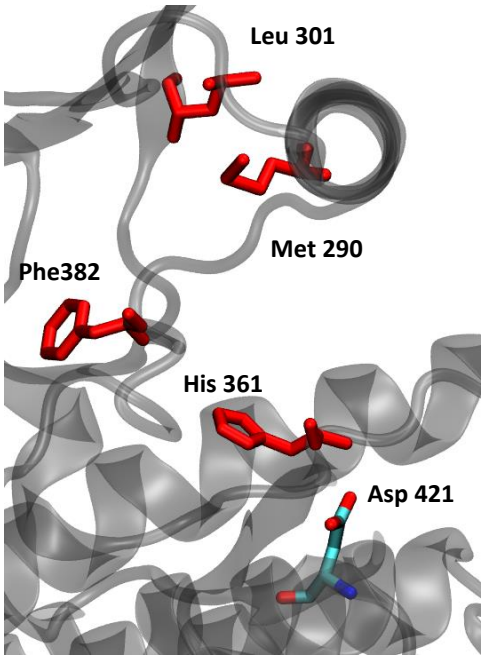


Figure 6.2.5: Residues in the R-spine (1IEP.pdb)

The alignment of the C-spine residues in the inactive crystal structures is comparable to that of the active structure, with only marginal differences in the distances between residues.

Residues	Distance (CA-CA, Å)		
	1IEP.pdb	3CS9.pdb	2F4J.pdb
Ala269 - Val256	6.612	6.590	6.96
Val256 – Val371	17.055	16.864	16.39
Val256 – Leu370	14.403	14.270	13.99
Val371 – Leu370	3.789	3.754	3.79
Leu370 – Cys369	3.805	3.817	3.81
Val371 – Leu324	9.509	9.511	9.59
Leu324 – Cys369	7.605	7.693	7.80
Leu324 – Ile432	8.185	8.151	8.05
Cys369 – Leu428	10.048	9.929	9.79
Ile432 – Leu428	6.201	6.207	6.31

Table 6.2.2: Distance between C-spine residues (1IEP.pdb and 3CS9.pdb)

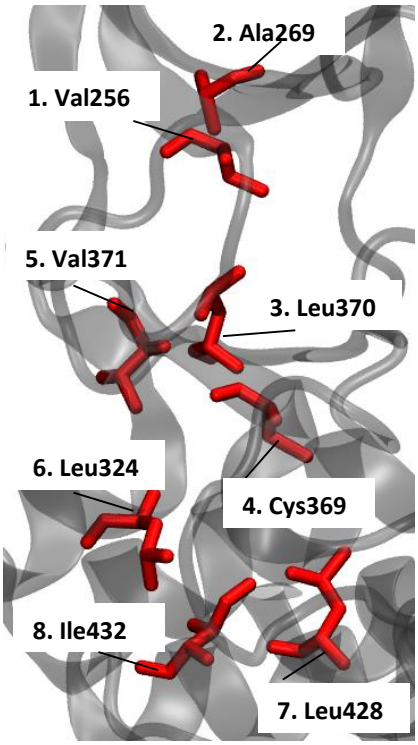


Figure 6.2.6: Residues in the C-spine (1IEP.pdb)

6.2.4 Protein structure preparation and MD parameters

Protein structure preparation of the 1IEP and 3CS9 structures and subsequent molecular dynamics simulations were completed as described in 5.2.4.

6.3 Results: WT Abl kinase domain in complex with imatinib (1IEP.pdb)

6.3.1 Structural stability and flexibility

RMSD and RMSF of major structural motifs

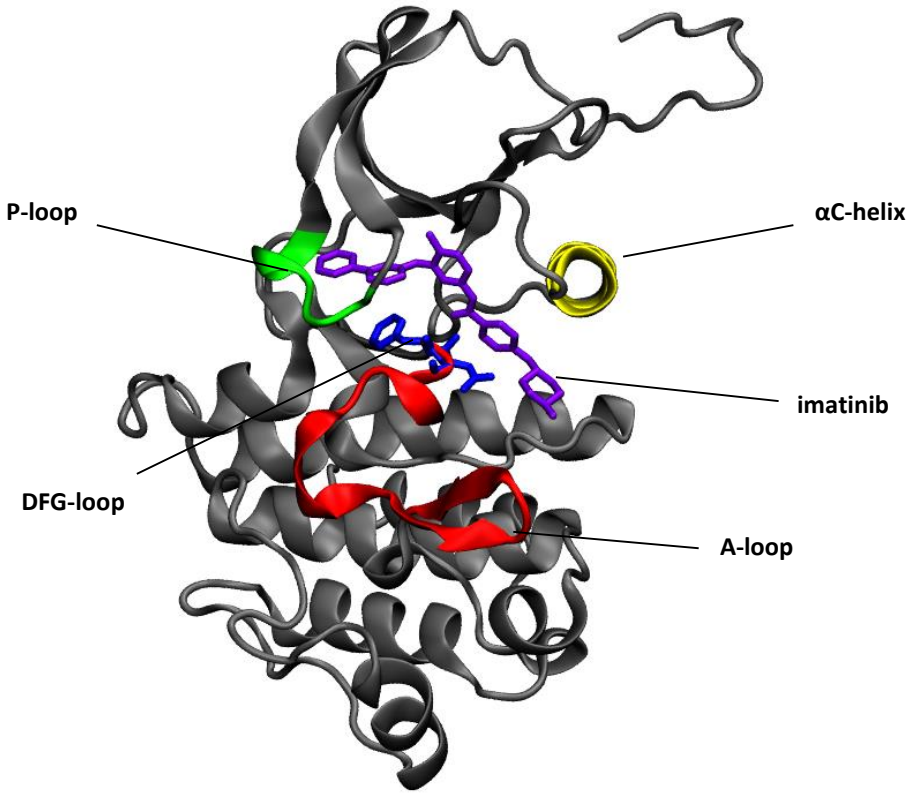


Figure 6.3.1: The key motifs of the inactive structure (1IEP.pdb)

The RMSD was calculated for the protein, αC-helix (residues 52-70), DFG-loop (residues 157-159), A-loop (residues 157-178) and the residues around the P-loop (residues 19-30) in order to capture any significant movement in these key motifs. After a small initial increase, the RMSD of the protein structure remained between 0.20 nm and 0.25 nm throughout the simulation. As with the active structure, (2F4J) the P-loop, A-loop and αC-helix demonstrated increased flexibility compared to the overall protein structure, whereas the RMSD of the DFG-loop

remained consistently low throughout the simulation; demonstrating that the DFG conformation is extremely stable in both the DFG-in (active) and DFG-out (inactive) conformations.

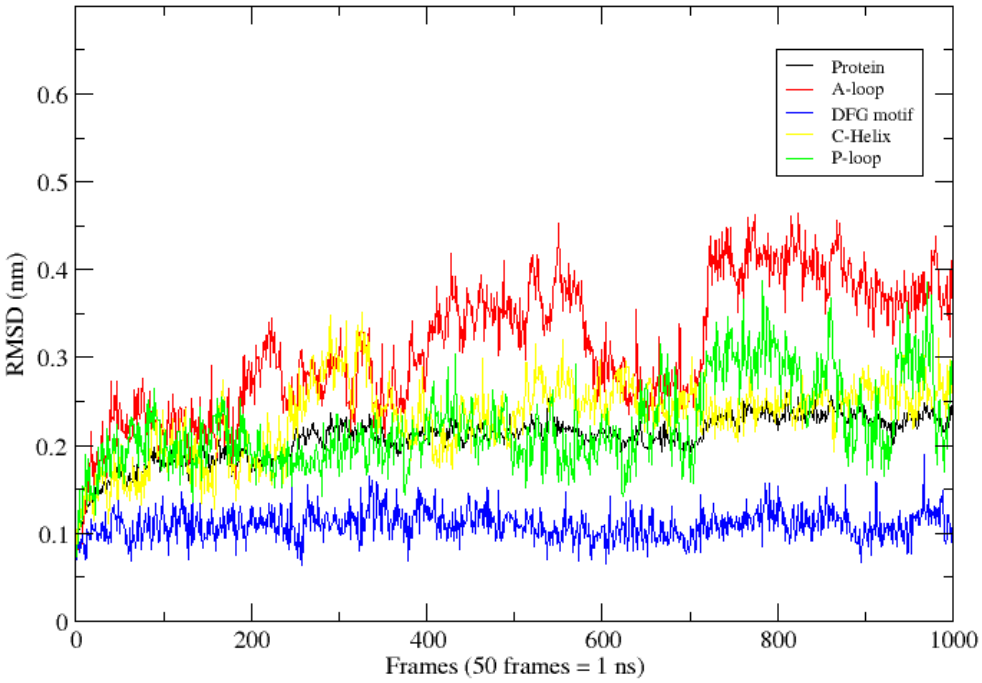


Figure 6.3.2: RSMD results for 1IEP.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green.

The RMSF analysis (Figure 6.3.4) Figure 6.3.4: RMSF results for 1IEP (black) compared to 2F4J (red).was in good agreement with the RMSD results; the highest peaks (0.3 nm) were around residues 157 (A-loop, shaded in red) and 50 (αC-helix, shaded in yellow). The RMSF analysis also highlighted areas of movement in the C-lobe around helices F and H which are connected by flexible loop sections (Figure 6.3.3).

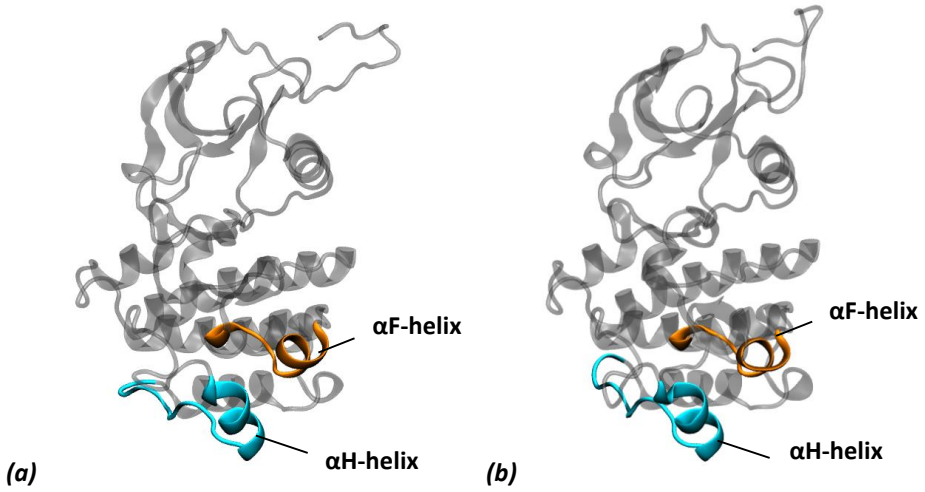


Figure 6.3.3: Conformation of the F and H-helices; (a) Initial structure; (b) Final after MD production

The C-helix region and the first half of the A-loop appear to be more flexible in the inactive structure than the active whereas the P-loop is more rigid (Figure 6.3.4); although there were fluctuations in the P-loop, it remained in a kinked conformation throughout the simulation.

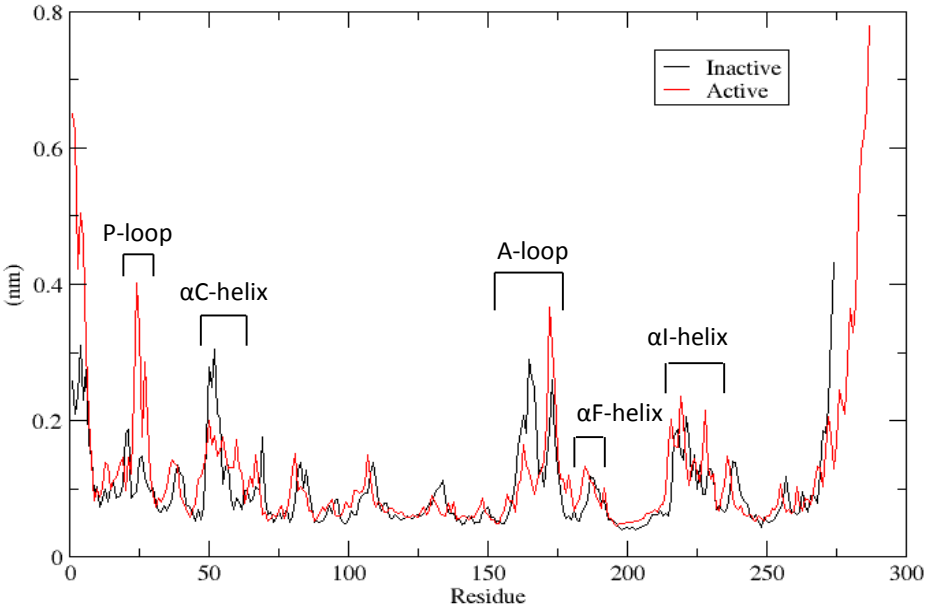


Figure 6.3.4: RMSF results for 1IEP (black) compared to 2F4J (red).

The RMSF values were aligned with the crystallographic B-factor values, i.e. the most flexible regions had high b-factor values.

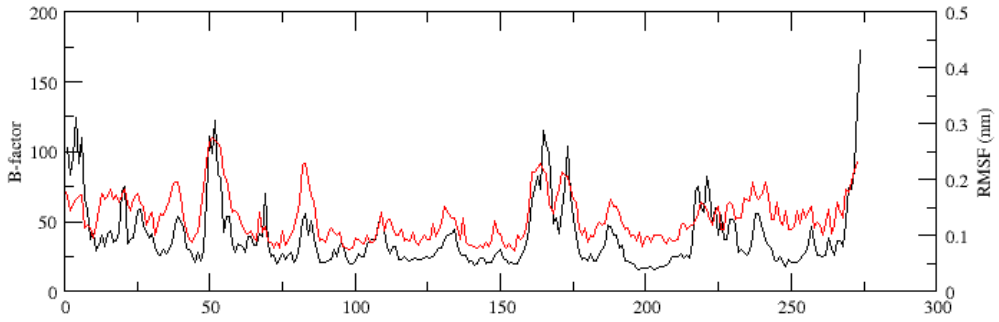


Figure 6.3.5: RMSF (black) vs. B-factors (red) for 1IEP.pdb

Principal Component Analysis

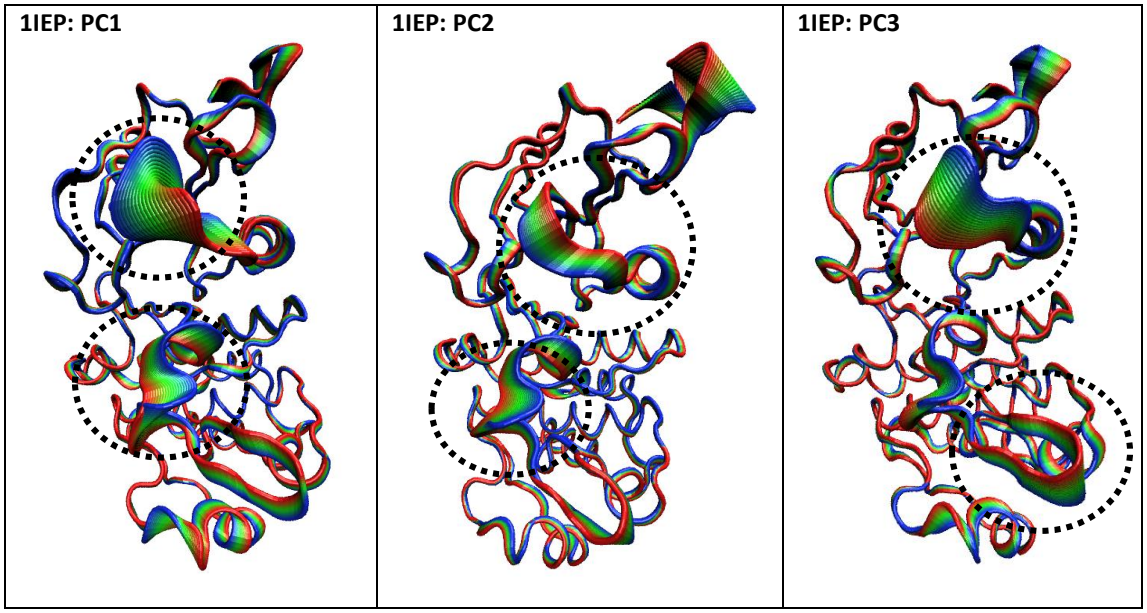


Figure 6.3.6: PCA projections for the first three principal components for BCR-Abl apo structure ($1IEP_{apo}$).

The projections from the PCA show that the largest movements for 1IEP structure (PC1) are in the loop region directly after the C-Helix and the first half of the activation loop. This is likely because the residues close to the C-Helix are in a position where they can fluctuate freely between the top of the binding site and the rest of the N-lobe.

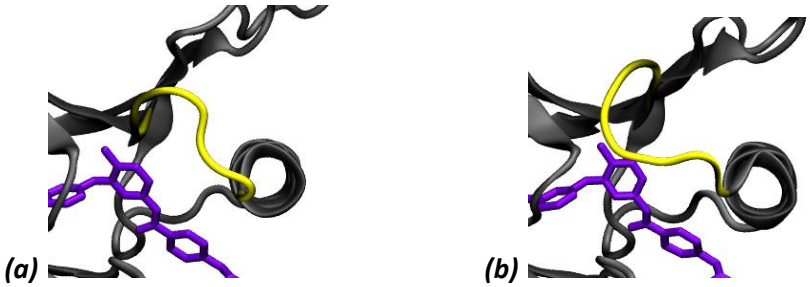


Figure 6.3.7: Fluctuations of residues preceding the αC-helix. (a) Initial structure; (b) Final structure post MD simulation (1IEP.pdb)

The second and third principal components also show motion in the residues surrounding the C-Helix and at the ends of the structures.

Analysis of catalytic and regulatory spines

The distances between residues in the two hydrophobic spines, the regulatory spine (R-spine) and the catalytic spine (C-spine) (Figure 6.3.8) were analysed.

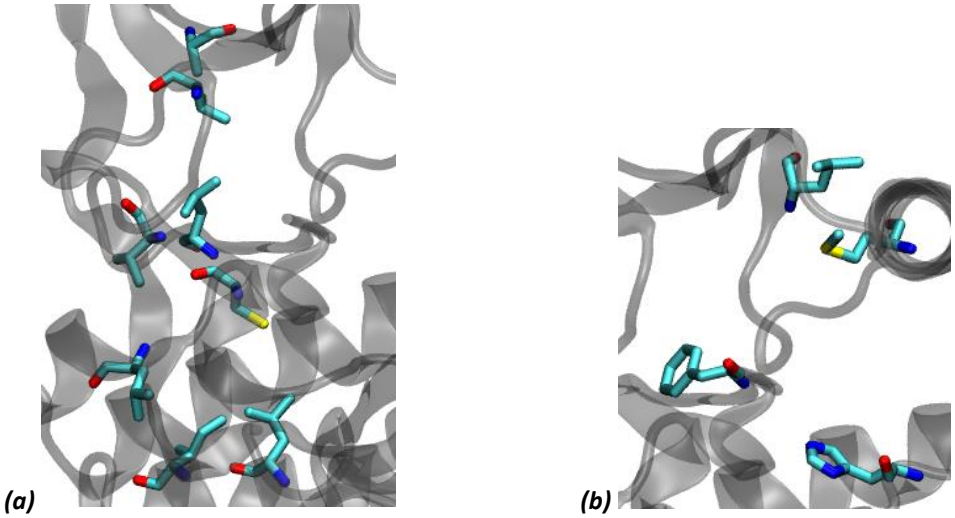


Figure 6.3.8: Catalytic (a) and regulatory (b) spines in the inactive conformation (1IEP.pdb)

The alignment of both the regulatory (R-) and catalytic (C-) spines was stable throughout the simulation and the average distance between the residues remained similar to those of measured for the crystal structure (Table 6.3.1 and Table 6.3.2).

R-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Leu301 – Met290	6.487	6.418
Met290 – Phe382	12.284	12.341
Phe382 – His361	12.302	11.109

Table 6.3.1: Mean distance between R-spine residues

C-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Ala269 - Val256	7.215	6.612
Val256 – Val371	17.273	17.055
Val256 – Leu370	14.756	14.403
Val371 – Leu370	3.879	3.789
Leu370 – Cys369	3.862	3.805
Val371 – Leu324	9.589	9.509
Leu324 – Cys369	7.905	7.605
Leu324 – Ile432	8.270	8.185
Cys369 – Leu428	9.976	10.048
Ile432 – Leu428	6.228	6.201

Table 6.3.2: Mean distance between C-spine residues

The positions of the residues in the R-spine appear to be stable in the 1IEP complex. The misalignment is driven by the positioning of the DFG Phe residue, which was seen to have low RMSD throughout the simulation. The distance between Phe382 and His361 increased slightly during the simulation, likely due to the small fluctuations in the helices of the C-lobe.

The positions of the residues of the C-spine remained close to that of the original crystal structure, suggesting that the C-spine is aligned and stable in both the active and inactive conformations.

Critical hydrogen-bond analysis

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)
Hydrogen bonds associated with the inactive conformation				
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	90.21
Hydrogen bonds associated with the active conformation				
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	45.25
		Gly383(M)	Asp381(M)	0.50
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	80.62
		His361(S)	Asp421(S)	33.47

Table 6.3.3: Hydrogen bond analysis: 1IEP (4 Å, 30°); S = sidechain, M = main chain.

As expected the hydrogen bonds between Tyr393 and Asp363, and Lys271 and Glu286 were extremely stable throughout the simulation (Table 6.3.3).

Binding site analysis

Fluctuations in the size of the binding pocket were measured using PocketAnalyzer.

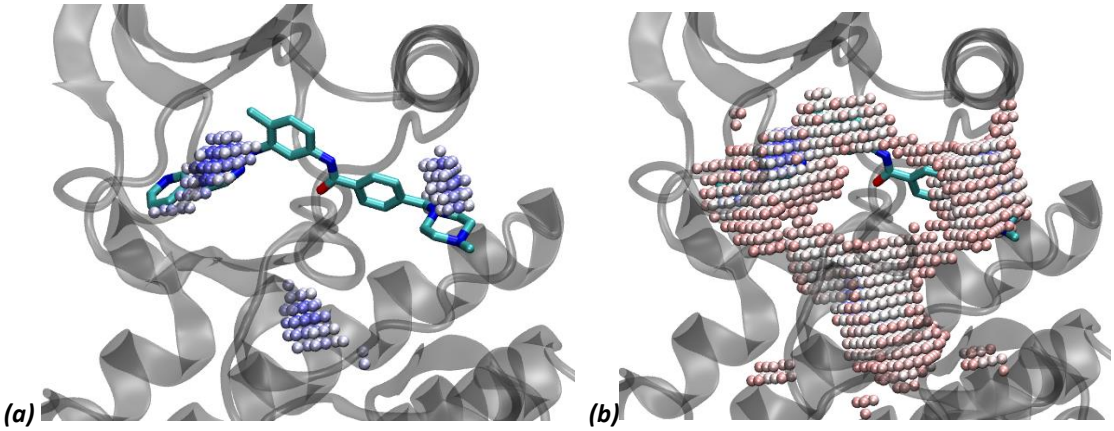


Figure 6.3.9: Binding site analysis, average pocket overlaid on the starting structure (1IEP.pdb) (a) Most conserved region of pocket (pocket present 40-100% of simulation). (b) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown).

The two end portions of the pocket were found to the most stable sections of the pocket throughout the simulation, however the middle section, around the methyl benzene and benzamide rings was less stable and only be identified 20% of the simulation. This is consistent with what is known regarding the central part of the ligand being exposed to the solvent.

6.3.2 Protein-ligand binding

RMSD and RMSF of imatinib

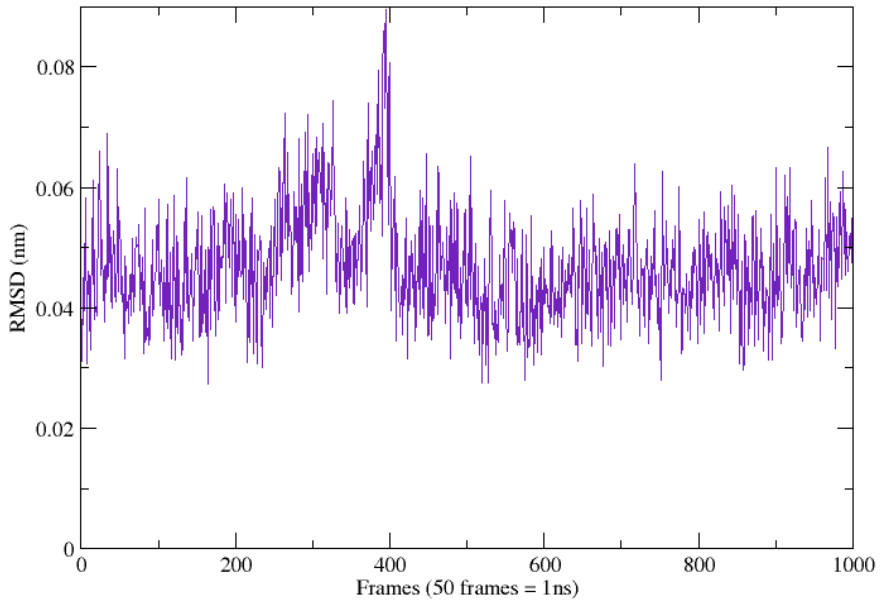


Figure 6.3.10: RMSD of imatinib (1IEP.pdb)

The RMSD and RMSF analysis showed that the ligand was fairly stable in the binding pocket, fluctuating between 0.03 nm and 0.06 nm for the majority of the simulation, with one larger peak after 8 ns (400 frames). The RSMF analysis showed that this was in line with the fluctuations of the overall protein structure (Figure 6.3.11 and Figure 6.3.12)

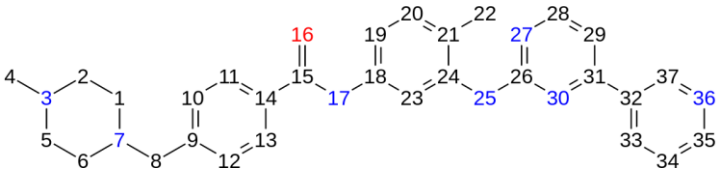


Figure 6.3.11: Index of atoms in imatinib (1IEP.pdb)

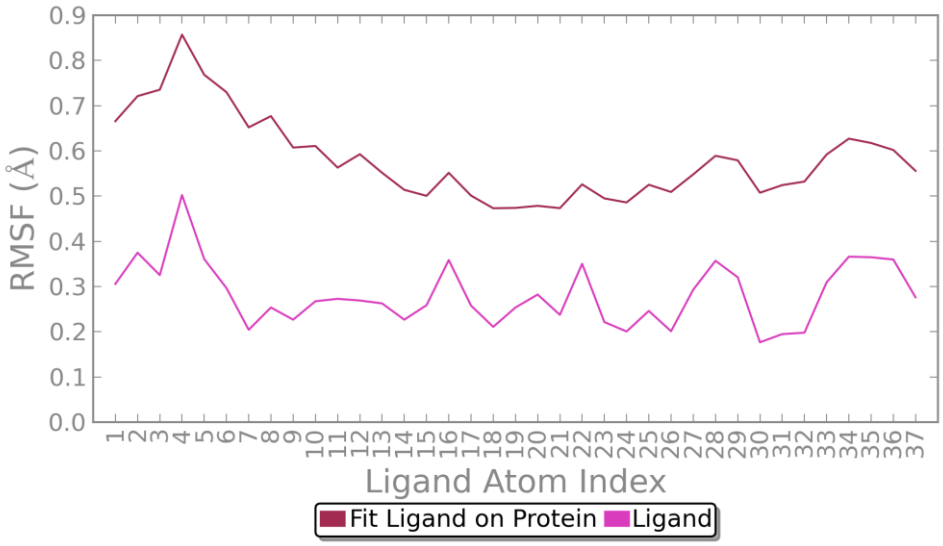


Figure 6.3.12: RMSF of ligand imatinib (1IEP.pdb)

Protein-ligand interactions

In crystal structure 1IEP, imatinib formed six hydrogen bonds with residues in the binding site; the amino group of pyrimidine and Thr315, the amide NH and Glu286, the protonated piperazine and Ile360 and His261, the amide CO and Asp381, and the pyridine N and Met318.

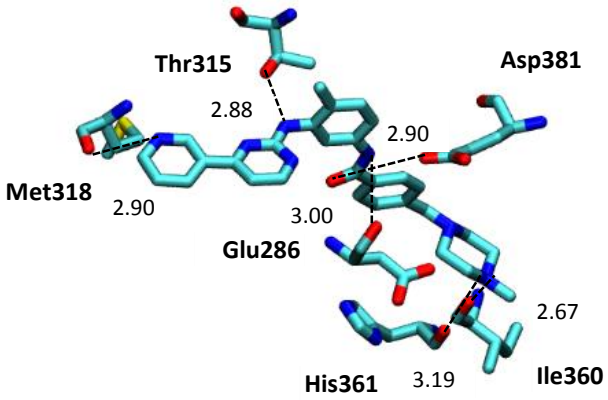


Figure 6.3.13: The imatinib binding site in 1IEP.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [30]

Imatinib is also thought to form van der Waals contacts with Ile360, His 361, Asp381 (DFG), Met290 and Val289.

Protein-ligand interactions during the simulation were analysed using the Simulation Interactions Diagram; the results confirmed that residues Glu286, Thr315 (gatekeeper), Met318, Ile360, Asp381 and His361 maintained bonding with imatinib throughout the simulation (Figure 6.3.15; residues Glu62, Thr91, Met94, Ile136, Asp157 and His 136). The hydrogen bond between Asp381 and imatinib was particularly stable and present for 98% of the simulation. This interaction appears to be stabilised further via interaction with water molecules in the binding site to form water bridges between the residue and the ligand (Figure 6.3.14).

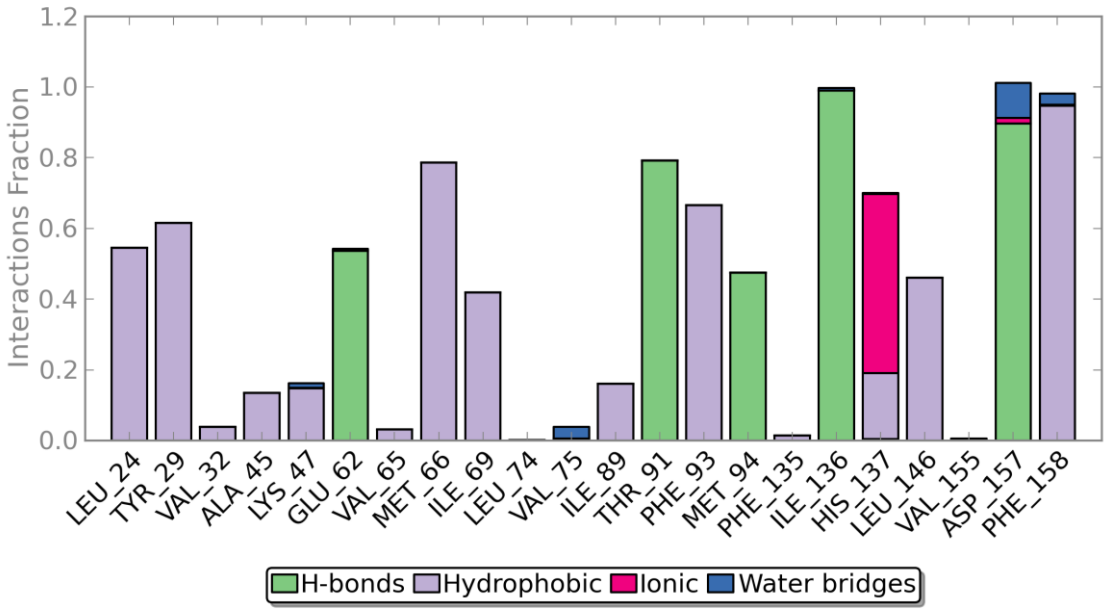


Figure 6.3.14: Protein-ligand interactions by residue (1IEP.pdb)

Strong (>50% if the simulation) hydrophobic interactions with imatinib were also identified for residues Leu248, Tyr253, Met290, Phe317 and Phe382 (Figure 6.3.14 and Figure 6.3.15: residues Leu24, Tyr29, Met66, Phe93 and Phe158).

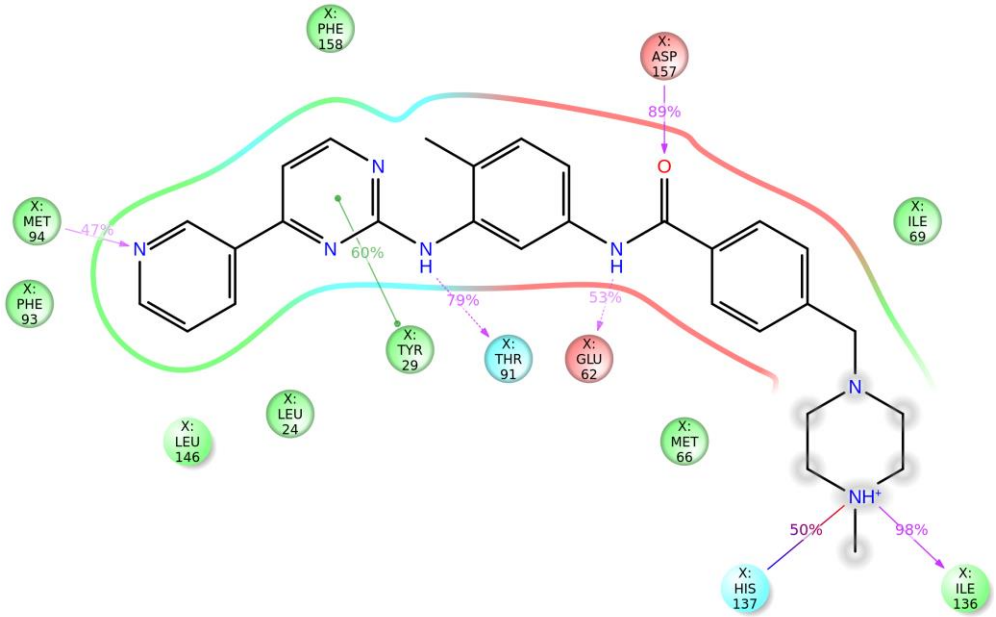


Figure 6.3.15: Protein-ligand interaction diagram for imatinib (1IEP.pdb)

6.4 Results: WT Abl kinase domain pseudo apo structure ($1IEP_{p\text{-}apo}$)

6.4.1 Structural stability and flexibility

RMSD and RMSF of major structural motifs

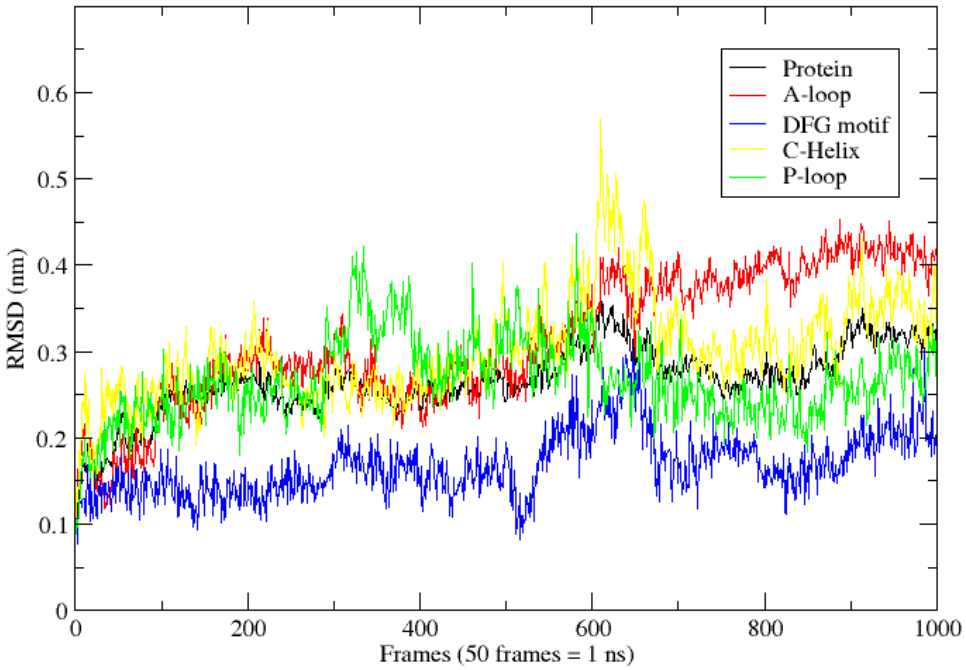


Figure 6.4.1: RSMD results for $1IEP_{p\text{-}apo}$; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green.

As expected removing the ligand increased the overall RMS of the structure and large fluctuations can be seen in the P-loop and C-helix regions (Figure 6.4.1, Figure 6.4.3 and Figure 6.4.4). Despite experiencing an increase in RMSD the P-loop remained in the kinked conformation throughout the simulation.

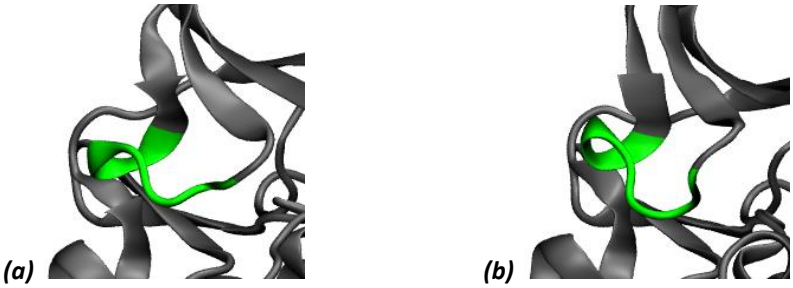


Figure 6.4.2: The change in P-loop conformation: a) Initial structure ($1IEP_{p\text{-}apo}$); b) Final structure post MD simulation ($1IEP_{p\text{-}apo}$)

The RMSD of DFG loop, which was extremely stable in the 1IEP complex structure, was less stable in this simulation and fluctuated between 0.1 nm to 0.3 nm (Figure 6.4.1 and Figure 6.4.3).

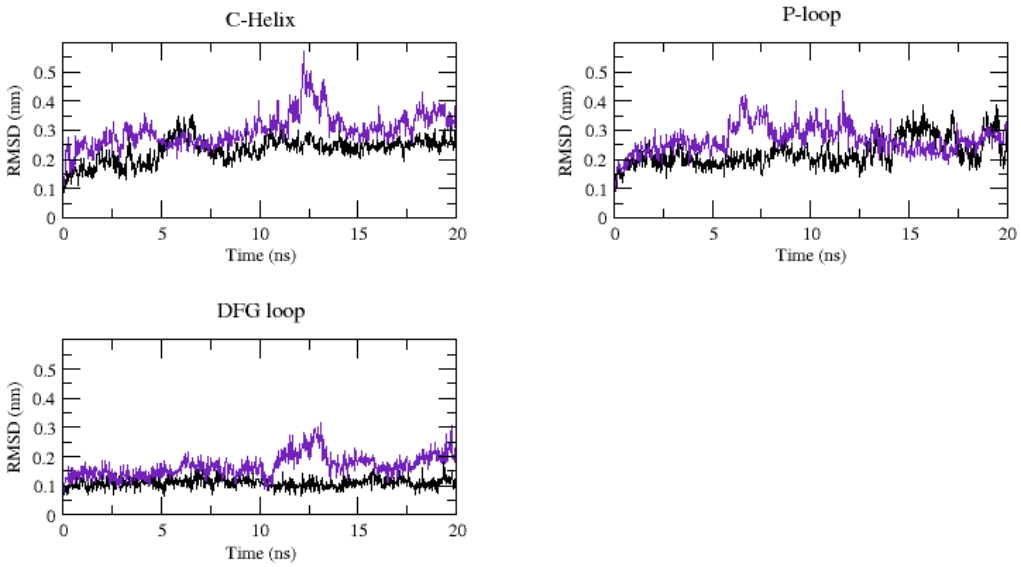


Figure 6.4.3: RMSD results for inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl apo structure (1IEP$_{p\text{-}apo}$, purple)

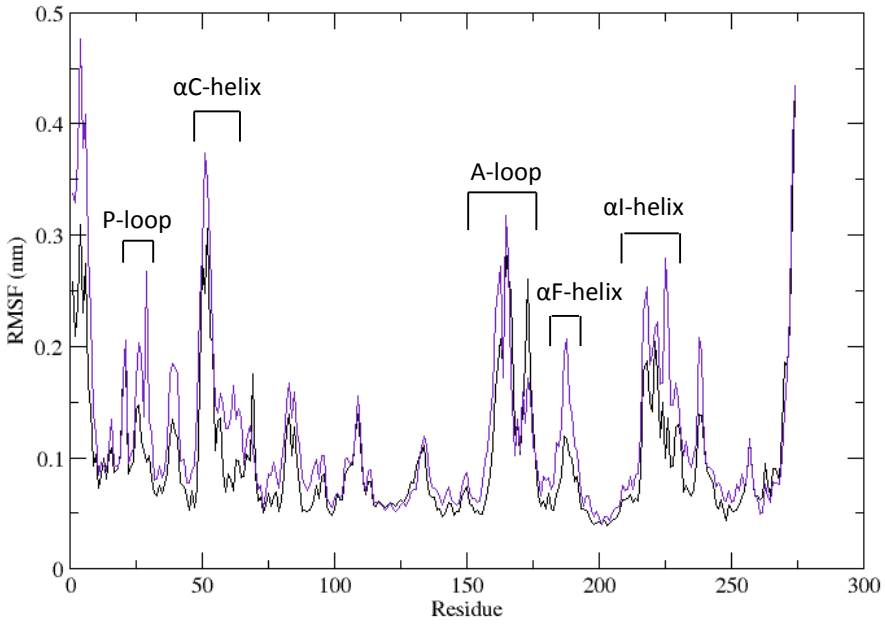


Figure 6.4.4: Comparison of RMSF plots of inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl apo structure (1IEP$_{p\text{-}apo}$, purple).

The RMSF analysis also highlighted regions of increased motion compared to the original 1IEP structure around residues 180 - 190 the F-helix (Figure 6.4.5) and 225-240, the H-helix (Figure 6.4.6).

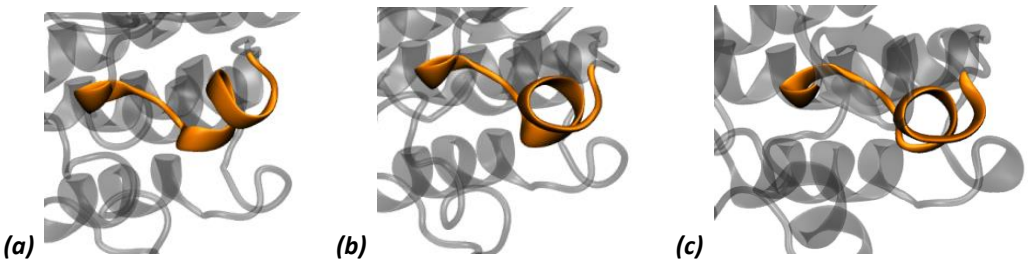


Figure 6.4.5: Figure 6.28: Conformation of F-helix; a) Initial structure (1IEP_p-apo); b) Final structure post MD simulation (1IEP_p-apo); c) Final structure post MD simulation (1IEP)

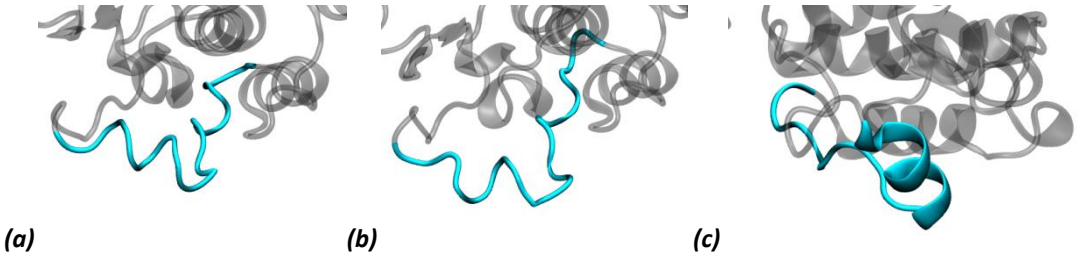


Figure 6.4.6: Conformation of H-helix; a) Initial structure (1IEP_p-apo); b) Final structure post MD simulation (1IEP_p-apo); c) Final structure post MD simulation (1IEP)

The F-helix appears to be exhibiting the same motions as the 1IEP structure, however the H-helix is unfolded in the pseudo-apo structure, which accounts for the higher RMS peak for this region (Figure 6.4.4).

The RMSF values were aligned with the crystallographic B-factor values, i.e. the most flexible regions had high b-factor values.

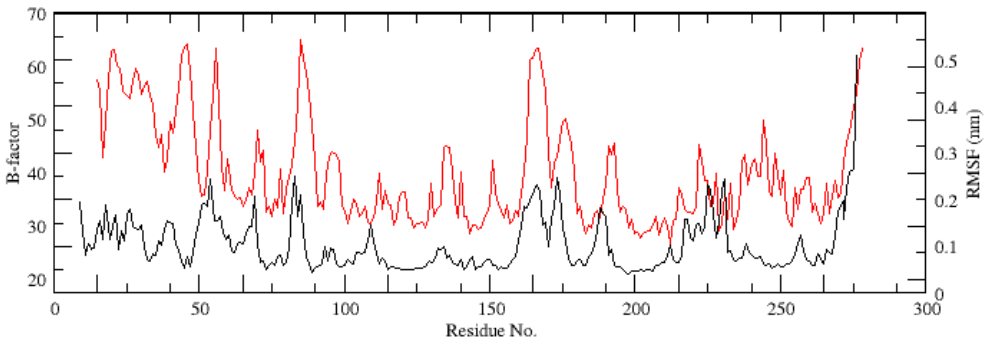


Figure 6.4.7: RMSF (black) vs B-factors (red) for 3CS9.pdb

Principal Component Analysis

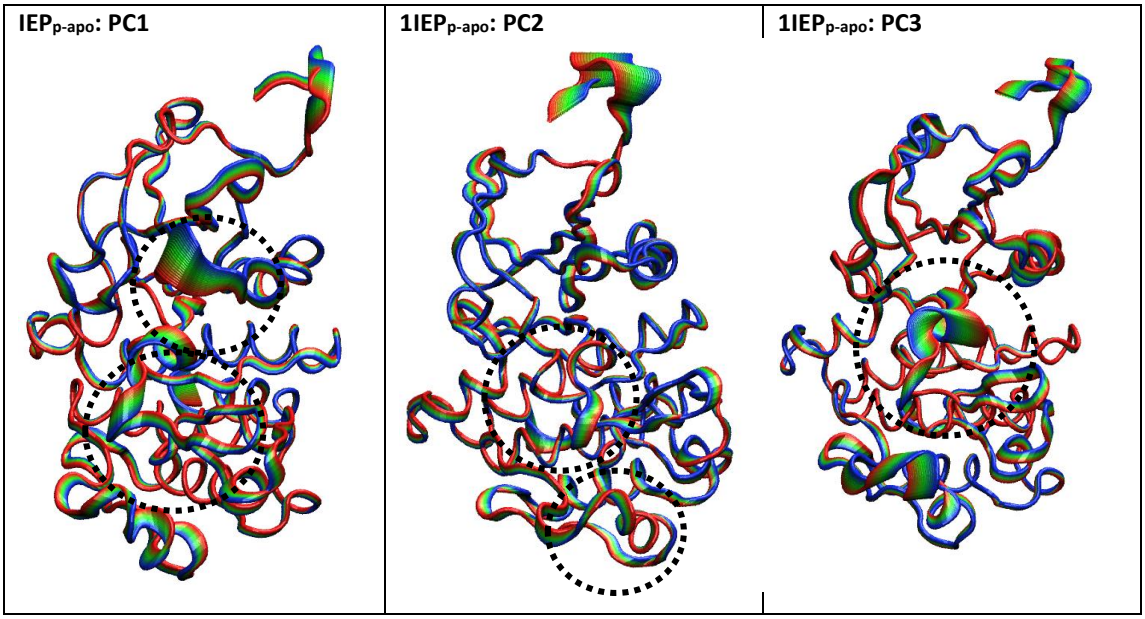


Figure 6.4.8: PCA projections for the first three principal components for inactive BCR-Abl apo structure (1IEP$_{p\text{-}apo}$).

The PCA shows that the most significant motions in the structure involve the residues prior to the C-Helix, as well as the helix itself. In the 1IEP simulation, the loop section preceding the C-helix moved closer into the binding site; however in the pseudo apo structure the loop folds upwards away from the pocket, due to a loss of interactions with the ligand.

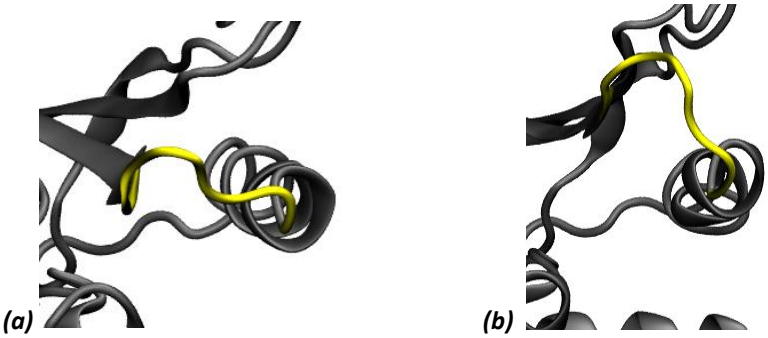


Figure 6.4.9: Conformation of loop section preceding C-helix (1IEPp-apo); a) Initial structure b) Final structure post MD simulation

The structural analysis results showed that the C-helix itself appeared to have greater flexibility in the simulation of 1IEP$_{p\text{-}apo}$ compared to the 1IEP simulation; a review of this region showed that the helix is partly unfolded in the pseudo apo structure and is therefore less rigid.

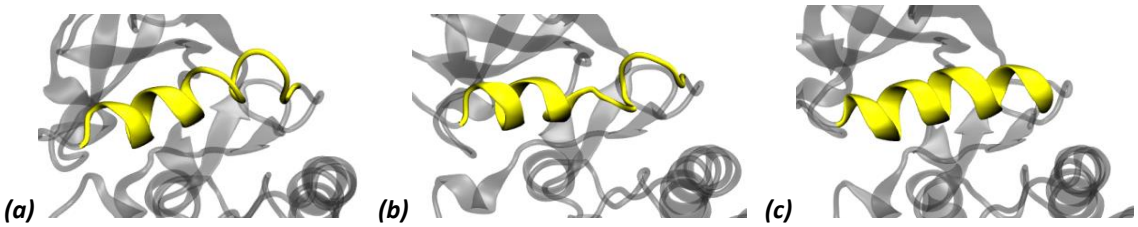


Figure 6.4.10: The change in C-helix conformation: a) Initial structure (IEPp-apo); b) Final structure post MD simulation (1IEPp-apo); c) Final structure post MD simulation (1IEP)

Analysis of catalytic and regulatory spines

The hydrophobic spines remained misaligned in the pseudo-apo structure and the mean distance between the residues of was similar to that of the holo structure. This suggests that the spine residues are stable in this conformation despite the loss of interactions with the ligand and overall higher RMS.

R-spine residues	Mean distance (CA-CA, Å)	
	$1IEP_{p\text{-}apo}$	**1IEP**
Leu301 – Met290	6.054	6.487
Met290 – Phe382	12.860	12.284
Phe382 – His361	11.096	12.302

Table 6.4.1: Mean distance between R-spine residues

C-spine residues	Mean distance (CA-CA, Å)	
	1IEPp-apo	**1IEP**
Ala269 - Val256	7.323	7.215
Val256 – Val371	17.889	17.273
Val256 – Leu370	14.351	14.756
Val371 – Leu370	4.676	3.879
Leu370 – Cys369	3.865	3.862
Val371 – Leu324	9.661	9.589
Leu324 – Cys369	7.826	7.905
Leu324 – Ile432	8.195	8.270
Cys369 – Leu428	9.984	9.976
Ile432 – Leu428	6.164	6.228

Table 6.4.2: Mean distance between C-spine residues

Critical hydrogen-bond analysis

Compared to the holo structure the interactions between Tyr393 and Asp363 increased slightly in the pseudo-apo structure, likely due to the increased RMS in this region, whereas the interactions of the salt bridge between Lys271 and Glu286 were diminished; the hydrogen bond was stable for over 90% of the simulation in the 1IEP holo structure but only 30% in the pseudo apo simulation. This is due to Glu286 rotating during the simulation, breaking the interaction with Lys271 (Figure 6.4.11 and Figure 6.4.12).

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{p\text{-}apo}$	1IEP
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	98.60	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	33.77	91.21
		Lys271 (S)	Glu286 (M)	0.10	-
		Glu286 (S)	Lys271 (S)	0.10	-
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	23.38	45.25
		Gly383(M)	Asp381(M)	-	0.50
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.00	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	88.71	80.62
		His361(S)	Asp421(S)	20.48	33.47

Table 6.4.3: Hydrogen bond analysis: $1IEP_{p\text{-}apo}$ vs. 1IEP (4 Å, 30°); S = sidechain, M = main chain.

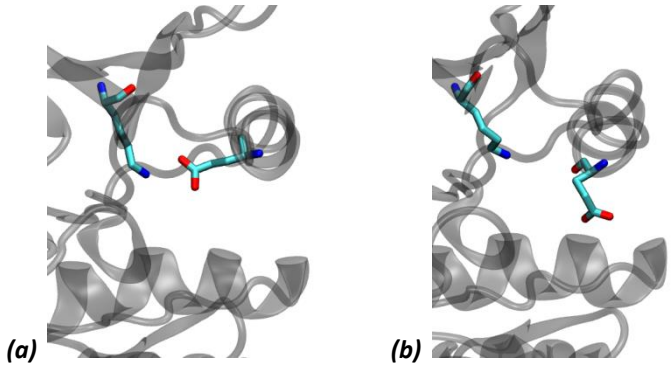


Figure 6.4.11: Loss of interaction between Lys271 and Glu286 in the 1IEP$_{p\text{-}apo}$ simulation; a) initial position b) rotated position

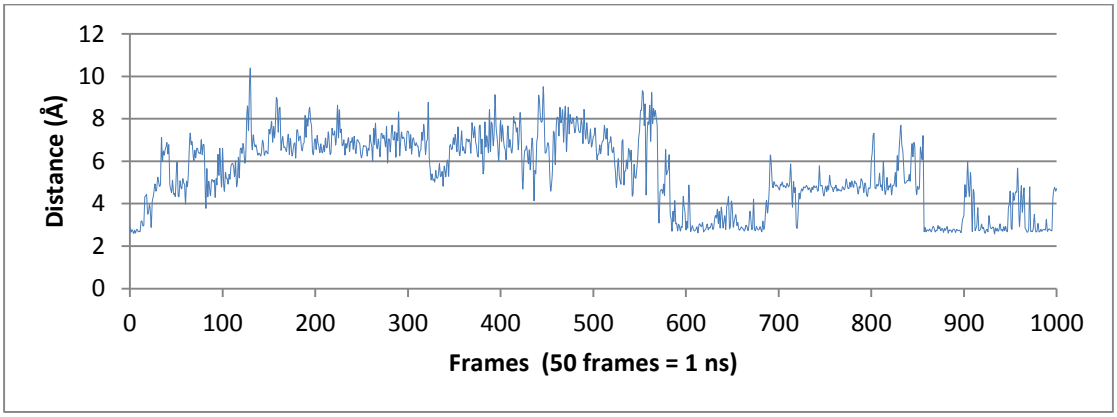


Figure 6.4.12: Distance between Lys271 and Glu286 over the 20 ns simulation for 1IEP$_{p\text{-}apo}$

Binding site analysis

Fluctuations in the size of the binding pocket were measured using PocketAnalyzer. As in the holo structure, the two end portions of the binding site were found to the most stable sections of the pocket throughout the simulation. In the simulation of the holo structure the middle section, around the methyl benzene and benzamide rings was identified for 20% of the simulation, however in the simulation of the pseudo apo structure this section is even less stable. This is to be expected given that the ligand has been removed from the structure and the structural analysis highlighted an increased in RMS in the regions surrounding the binding site.

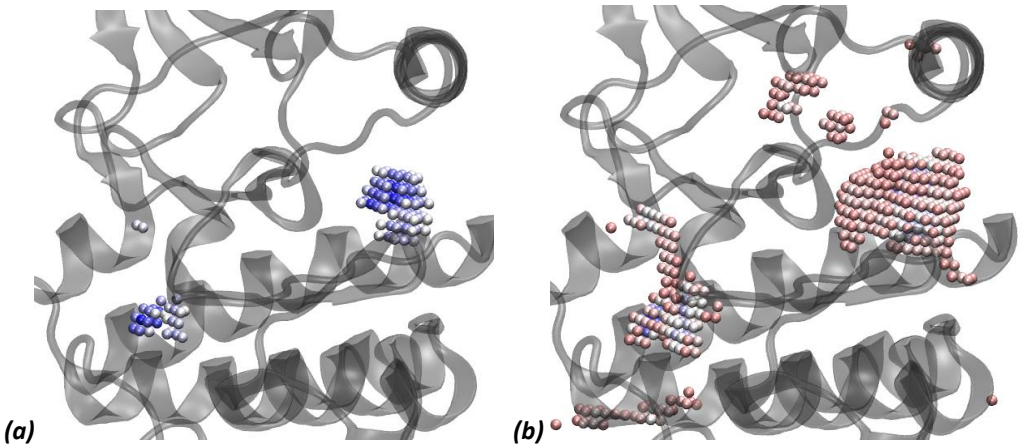


Figure 6.4.13: Binding site analysis, average pocket overlaid on the starting structure (1IEP$_{p\text{-}apo)}$ a) Most conserved region of pocket (pocket present 40-100% of simulation). B) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown).

6.5 Results: WT Abl kinase domain in complex with nilotinib (3CS9.pdb)

6.5.1 Structural stability and flexibility

RMSD and RMSF of major structural motif

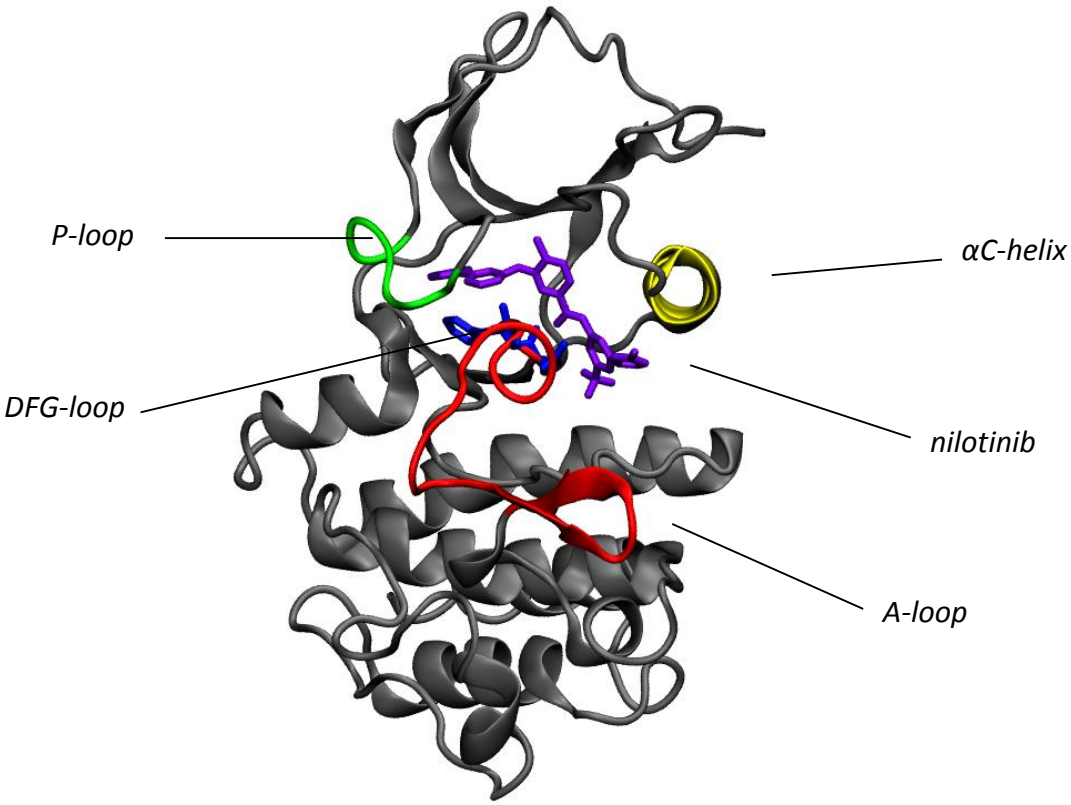


Figure 6.5.1: The key motifs of the inactive structure (3CS9.pdb)

As with the previous simulations the RMSD for the inactive structure in complex with nilotinib was analysed for the αC-helix, DFG-loop, A-loop and the residues around the P-loop in order to capture any differences in movement in these key motifs compared to the imatinib complex.

As expected the results were generally well aligned with that of the 1IEP simulation, with similar RMS of the protein structure, A-loop and P-loop (Figure 6.5.2). There were some differences seen in the region of the C-Helix and also the DFG-loop; the results showed that these sections had increased RMS in the nilotinib complex (Figure 6.5.3).

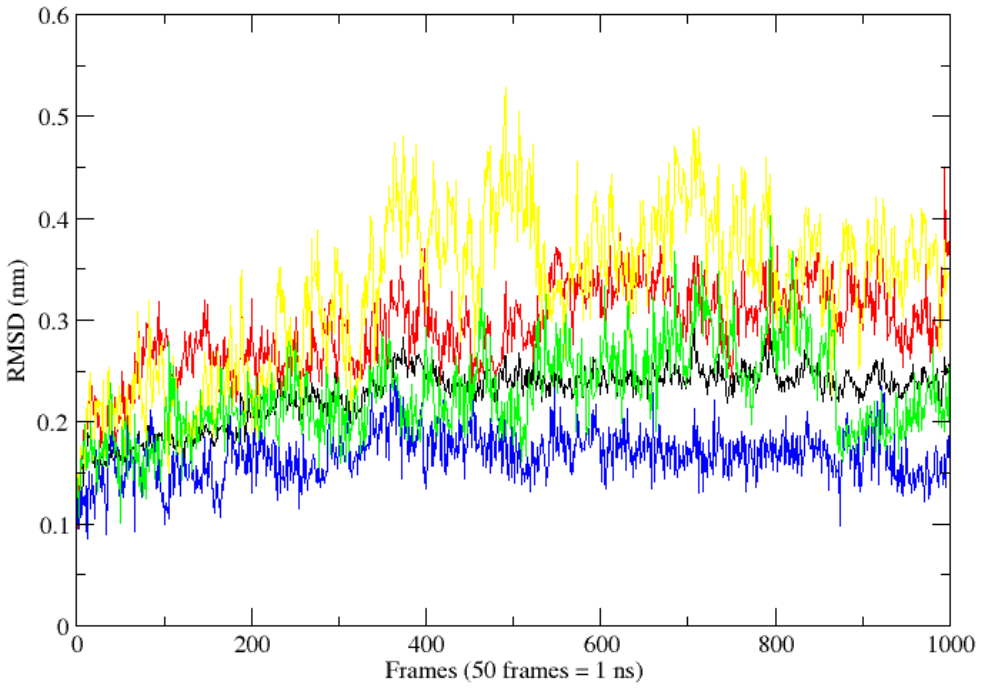


Figure 6.5.2: RSMD results for 3CS9.pdb; protein shown in black, A-loop in red, DFG-loop in blue, αC-helix in yellow and the P-loop in green.

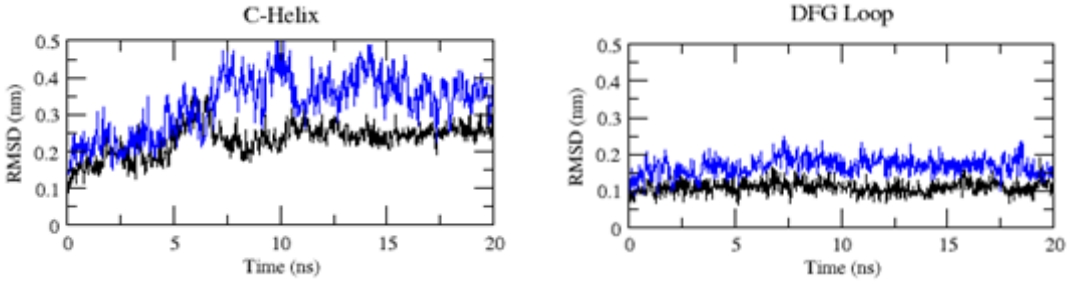


Figure 6.5.3: RMSD results for inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl with nilotinib (3CS9, blue)

On reviewing the structure at the beginning and end of the simulation, the C-helix can be seen to curl upwards away from the binding site.

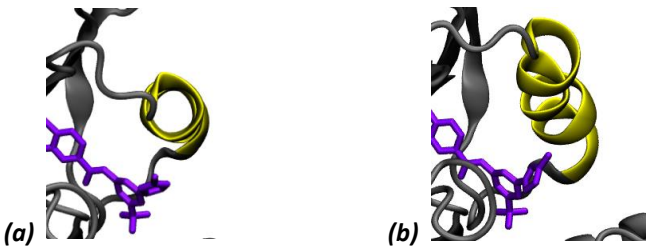


Figure 6.5.4: Conformation of the C-helix (3CS9); a) Initial structure b) Final structure post MD simulation

The RMSF results were also closely aligned with those of the 1IEP structure but also identified three regions where the RMSF was increased compared to the imatinib simulation; Resides 80-90 (β5-sheet), 180-190 (αF-helix), 220-230 (αH-helix) (Figure 6.5.5 and Figure 6.5.6)

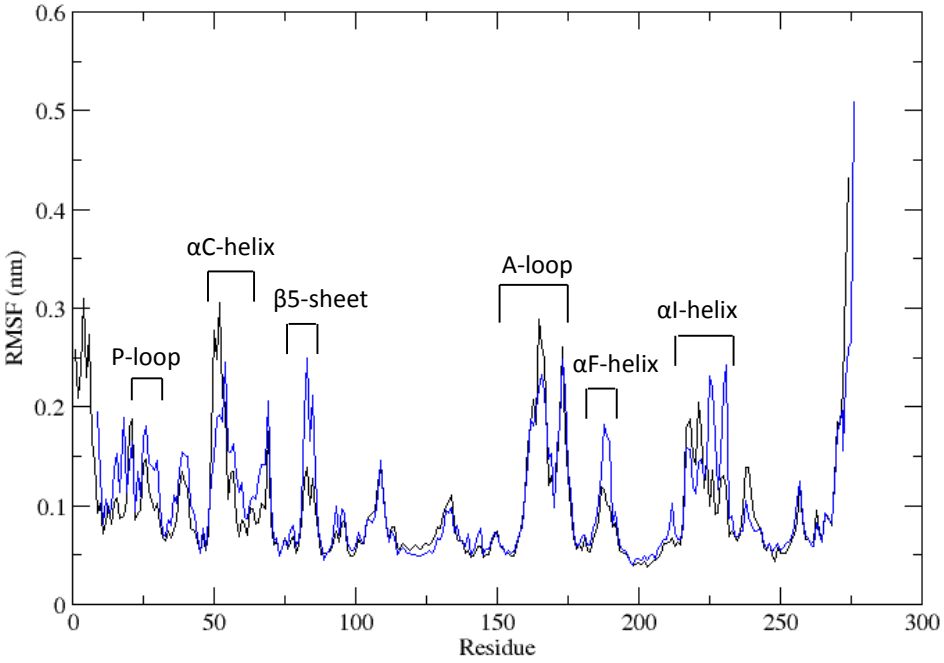


Figure 6.5.5: Comparison of RMSF plots of inactive BCR-Abl with imatinib (1IEP, black) and inactive BCR-Abl with nilotinib (3CS9, blue).

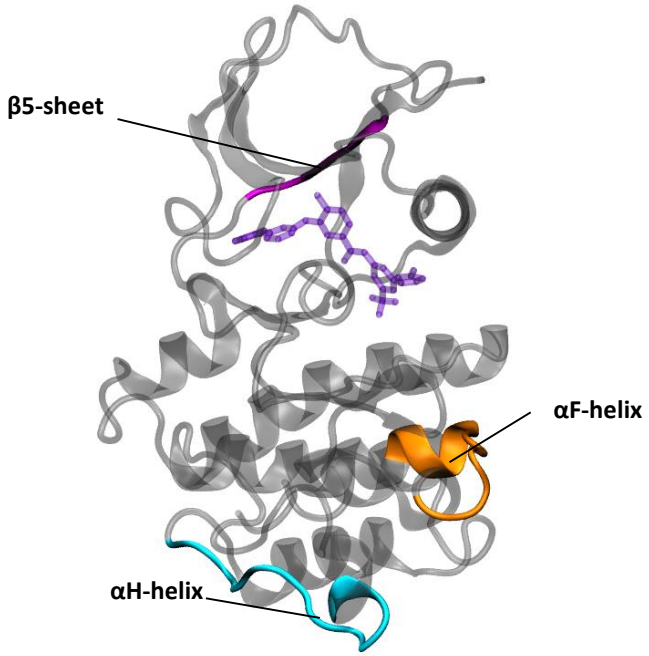


Figure 6.5.6: Additional regions of the kinase with increased RMSF compared to 1IEP.pdb (3CS9.pdb): β5-sheet, residues 80-85, magenta; αF-helix, residues 180-190, orange; αH-helix, residues 220-230, cyan.

Principal Component Analysis

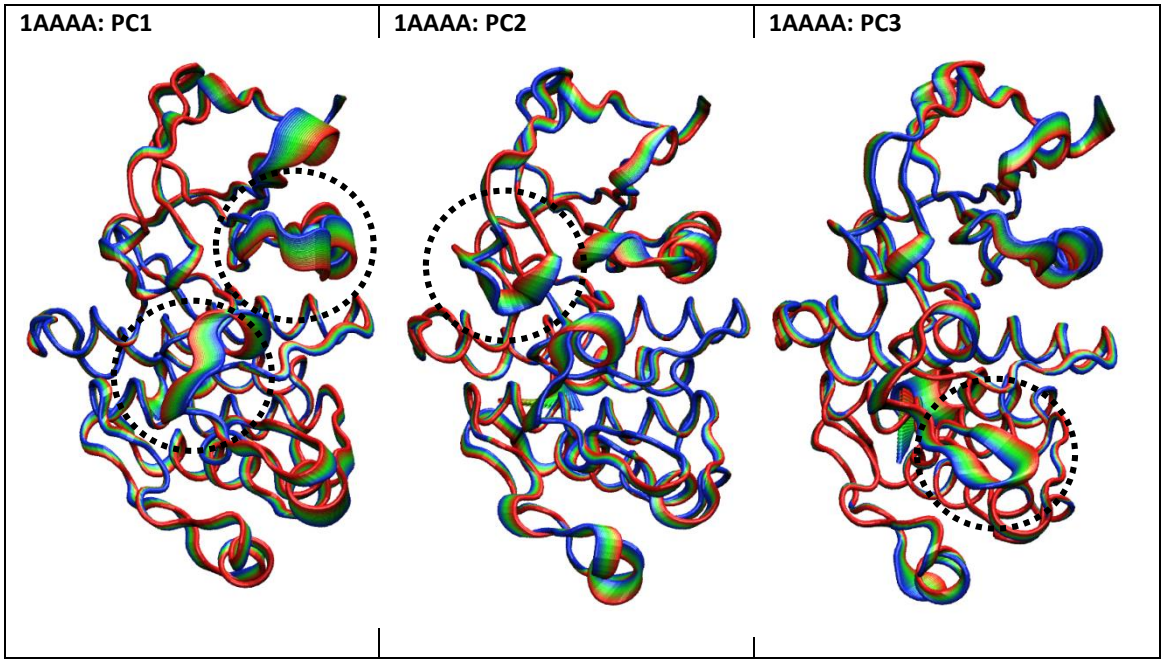


Figure 6.5.7: PCA projections for the first three principal components for inactive BCR-Abl with nilotinib (3CS9).

The PCA results were aligned with the results of the RMSF and RMSD analysis and showed that the predominant motions (PC1) as well as those for the second and third principal components were located around the C-Helix and the A-loop, which is consistent with what was seen in the imatinib simulation.

Analysis of catalytic and regulatory spines

The distances between residues in the two hydrophobic spines, the regulatory spine (R-spine) and the catalytic spine (C-spine) were analysed (Figure 6.5.8).

6: Inactive BCR-Abl

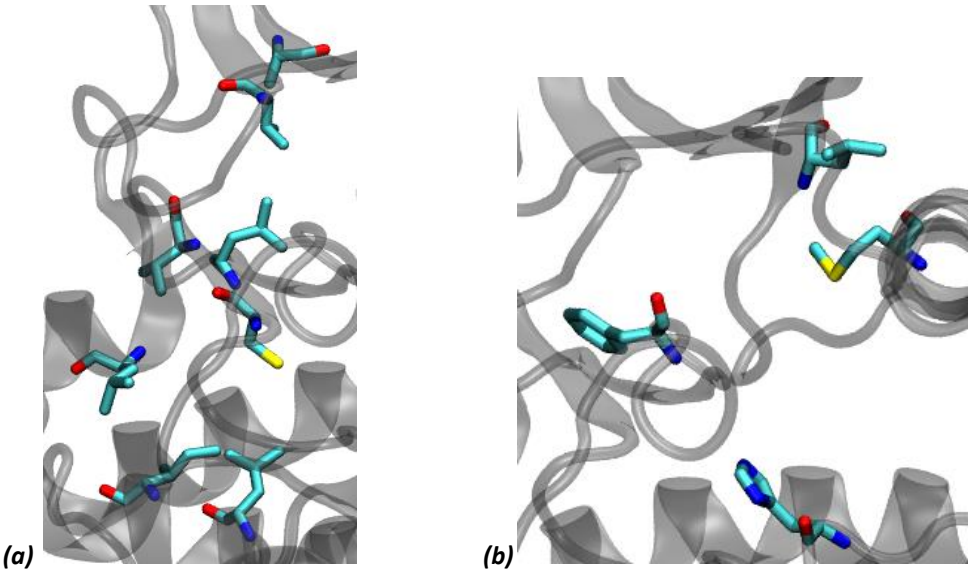


Figure 6.5.8: Catalytic (a) and regulatory (b) spines in the inactive conformation (3CS9.pdb)

The alignment of both the regulatory (R-) and catalytic (C-) spines was stable throughout the simulation and the average distance between the residues remained similar to those of measured for the crystal structure (Table 6.5.1 and Table 6.5.2).

R-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Leu301 – Met290	6.500	6.460
Met290 – Phe382	12.874	12.259
Phe382 – His361	11.514	11.341

Table 6.5.1: Mean distance between R-spine residues

C-spine residues	Mean distance (CA-CA, Å)	Distance, crystal structure (CA-CA, Å)
Ala269 - Val256	7.381	6.590
Val256 – Val371	17.485	16.864
Val256 – Leu370	14.737	14.270
Val371 – Leu370	3.843	3.754
Leu370 – Cys369	3.865	3.817
Val371 – Leu324	9.369	9.511
Leu324 – Cys369	7.781	7.693
Leu324 – Ile432	7.873	8.151
Cys369 – Leu428	9.812	9.929
Ile432 – Leu428	6.269	6.207

Table 6.5.2: Mean distance between C-spine residues

This is in close alignment with the results from the simulations of the imatinib and pseudo-apo structures and reinforces the concept that the spines are stable in the WT structure.

Critical hydrogen-bond analysis

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				3CS9	1IEP
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	97.70	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	82.32	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	19.08	45.25
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.60	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	85.71	80.62
		His361(S)	Asp421(S)	32.97	33.47

Table 6.5.3: Hydrogen bond analysis: 3CS9 vs 1IEP (4 Å, 30°); S = sidechain, M = main chain.

Overall the results of the hydrogen bond analysis were well aligned with the imatinib simulation, with only minor differences; the hydrogen bond between D and F of DFG was not recorded and the hydrogen bond between Tyr393 and Asp363 was extremely stable throughout the simulation, more so than in the 1IEP simulation (98% vs 90%). The salt bridge between Lys271 and Glu286 and the hydrogen bond between the Asp and Gly residues of the DFG helix were slightly less stable in the imatinib complex, however the interactions between the αG-helix and the R of the HRD motif, as well as the hydrogen bond between His361 and Asp421 were more stable compared to the imatinib simulation.

Binding site analysis

As with the imatinib and pseudo apo structures, changes in the size of the binding pocket during the simulation were measured using PocketAnalyzer.

The pocket appeared to be more stable in the nilotinib complex, with an increased area of the binding site being identified for a higher percentage of the simulation. However, as with the imatinib complex the central section of the pocket was found to be the least stable.

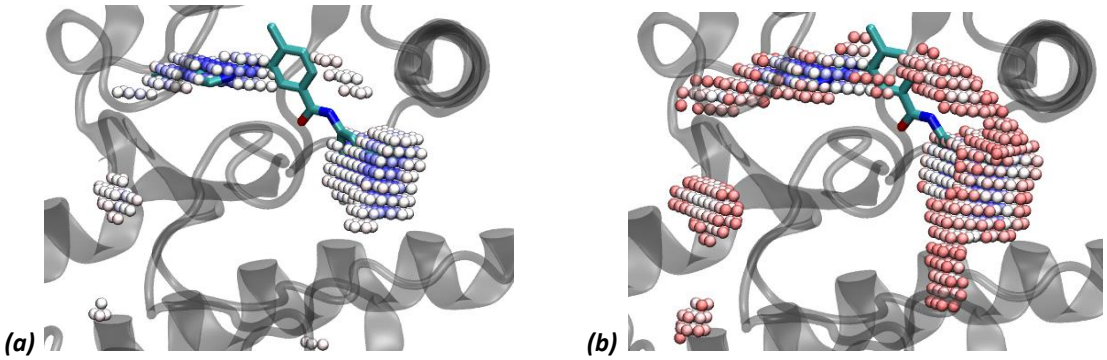


Figure 6.5.9: Binding site analysis, average pocket overlaid on the starting structure (3CS9.pdb) a) Most conserved region of pocket (pocket present 40-100% of simulation). B) Less conserved regions of pocket (pocket present 20-100% of the simulation); Blue spheres = 80-100%, Pale blue 60-80%, white, 40-60%, pale red 20-40% (0-20% not shown).

6.5.2 Protein-ligand binding

RMSD and RMSF of nilotinib

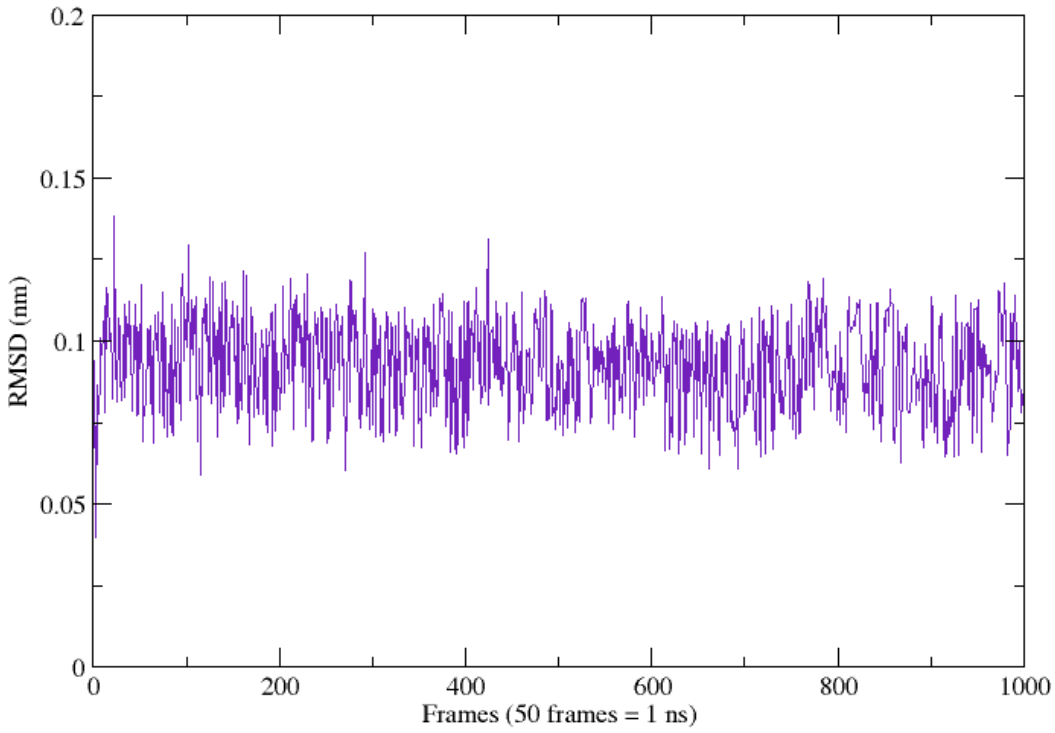


Figure 6.5.10: RMSD of nilotinib (3CS9.pdb)

The RMSD and RMSF analysis showed that the ligand was stable in the binding pocket, with an average RMSD of approximately 0.1 nm for the duration of the simulation. The RSMF analysis showed that this was in line with the fluctuations of the overall protein structure (Figure 6.5.11 and Figure 6.5.12).

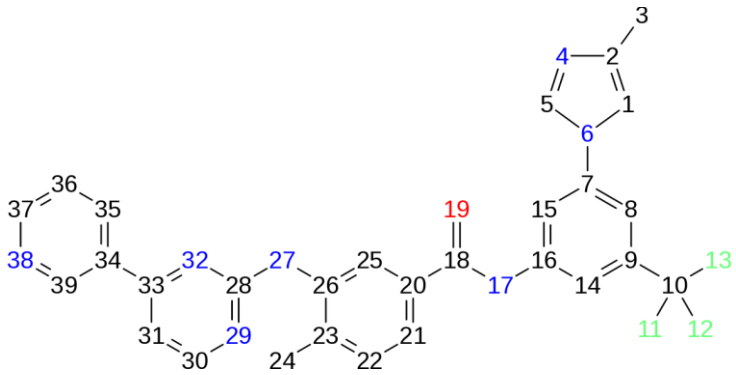


Figure 6.5.11: Index of atoms in nilotinib (3CS9.pdb)

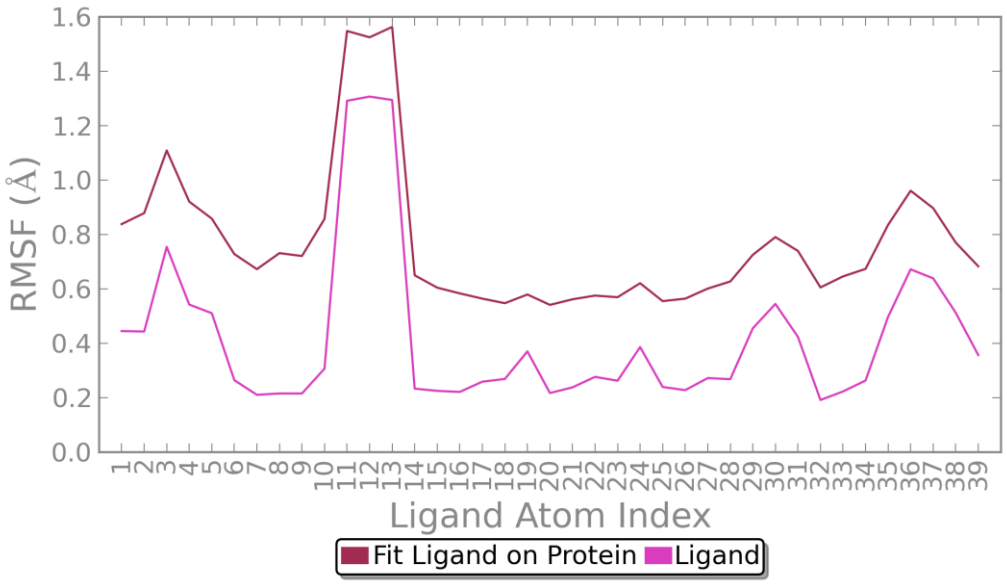


Figure 6.5.12: RMSF of ligand nilotinib (3CS9.pdb)

Protein-ligand interactions

In the crystal structure nilotinib made four hydrogen bonds with the residues of the binding site; the amino group of pyrimidine and Thr315, the amide NH and Glu286, the amide CO and Asp381 and the pyridine N and Met318 (Figure 6.5.13)

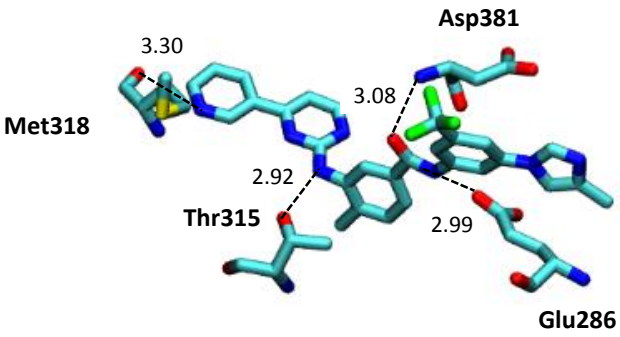


Figure 6.5.13: The nilotinib binding site in 3CS9.pdb. Hydrogen bond distances (Å) labelled. Figure adapted from reference [262]

Protein-ligand interactions during the simulation were analysed using the Simulation Interactions Diagram; the results confirmed that residues Thr315, Glu286, Asp381 and Met318 maintained bonding with nilotinib throughout the simulation (Figure 6.5.14 and Figure 6.5.15; residues Glu54, Thr83, Met86 and Asp149).

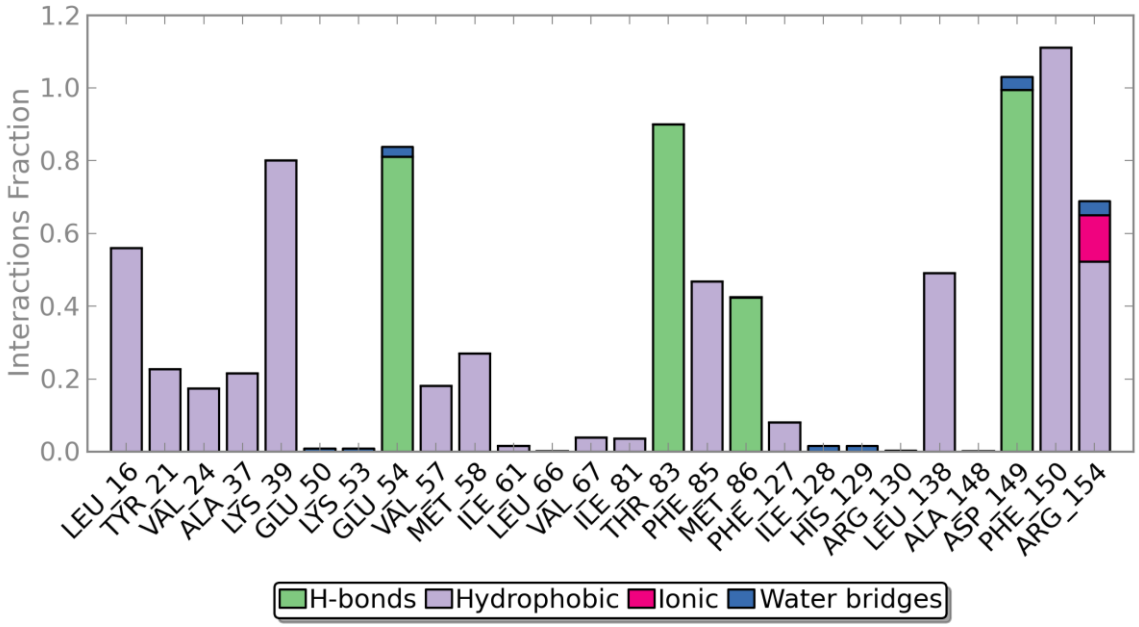


Figure 6.5.14: Protein-ligand interactions by residue (3CS9.pdb)

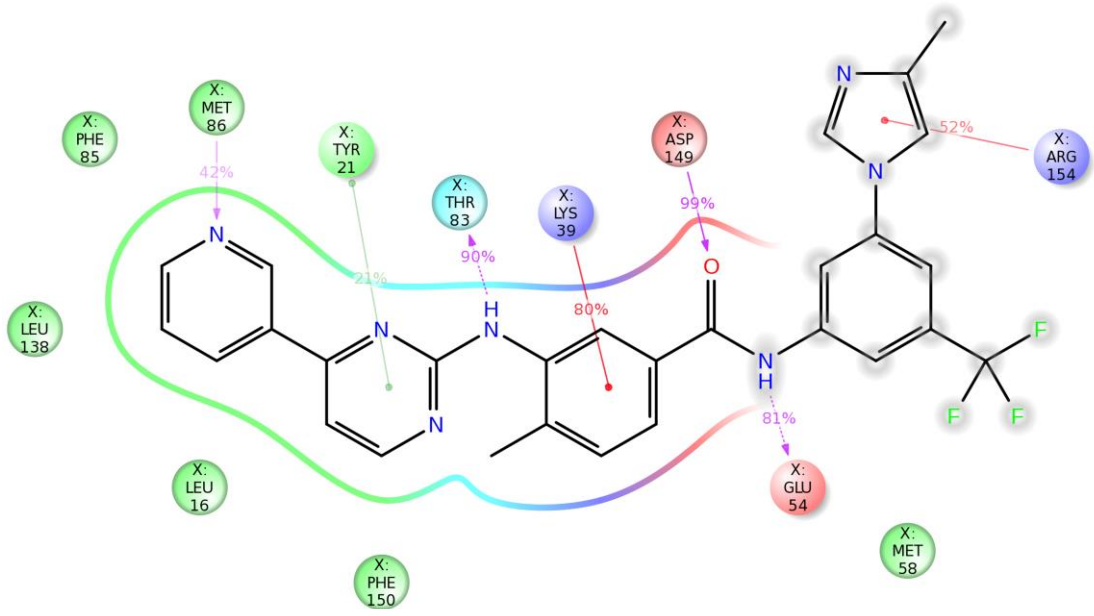


Figure 6.5.15: Protein-ligand interaction diagram for nilotinib (3CS9.pdb)

The analysis also confirmed stable interactions with Phe382, Lys271, Leu248 and Arg386 (Figure 6.5.14 and Figure 6.5.15; Phe150, Lys39, Leu16, Arg154).

A comparison of the interactions with imatinib and nilotinib in the 1IEP and 3CS9 simulations shows that of the four shared hydrogen bonds, three are more stable in the nilotinib complex (Thr315, Glu286, Asp381). This supports what is known regarding nilotinib being a more potent inhibitor of the Abl kinase. Interestingly, the interactions between the ligand and residues Tyr253 and Phe317 were stronger in the 1IEP structure than 3CS9. This supports previous research that these residues are more important for the binding of imatinib than nilotinib [32].

6.6 Summary of molecular dynamics study

The structure and dynamics of the inactive conformation have been investigated using molecular dynamics simulations on three inactive structures; two complexes (1IEP with imatinib and 3CS9 with nilotinib) and one with the inhibitor removed ($1IEP_{p\text{-}apo}$). This work has been completed to confirm the stable characteristics of the inactive state and compare the binding modes of imatinib and nilotinib.

The C-helix region and the first half of the A-loop appear to be more flexible in the inactive structure than the active whereas the P-loop is more rigid; although there were fluctuations in the P-loop, it remained in a kinked conformation in all three simulations. The complex structures demonstrated stability in the simulation; the structural analysis results were well aligned

although the nilotinib complex demonstrated slightly elevated RMSF around the C-, F- and H-helices compared to the imatinib complex.

Removing imatinib from the 1IEP structure to create a pseudo-apo inactive structure caused a significant increase in the overall RMSF and the unfolding of the C- and H- helices. When VX-680 was removed from the 2F4J structure the increase in RMSF was much less and structural motifs remained well aligned with that of the original 2F4J holo structure. This suggests that sections of the protein are more flexible in the inactive conformation than the active.

The positions of residues in the regulatory (R-) and catalytic (C-) spines were stable in all three simulations. The misalignment of the R-spine is driven by the positioning of the DFG Phe residue, therefore the realignment of the spine may only be realised via a change in conformation of the DFG loop. The DFG loop was extremely stable in the simulations of 1IEP and 3CS9, with a lower RMSD than the protein overall. In the simulation of 1IEPp-apo the RMSD of the DFG-loop increased but the Asp-out conformation remained stable. This suggests that although it might destabilise the protein structure, the loss of interaction with the ligand may not facilitate a change in the DFG conformation within a 20 ns timescale which is consistent with published literature. The position of the C-spine residues remained well aligned in both the active and inactive simulations suggesting that the C-spine is aligned and stable in both the active and inactive conformations.

A review of the critical hydrogen bonds associated with the inactive conformation confirmed that the hydrogen bond between Tyr393 and Asp363 was extremely stable in all three simulations; however several of the hydrogen bonds identified as critical to the active structure, were also stable in the inactive structure and are therefore unsuitable to be used as markers of conformation change: Lys271 – Glu286 (salt bridge between β sheet and αC-helix), Asp381 – Gly383 (hydrogen bond between D and G of DFG), Arg362 – Asp421 (hydrogen bond between αG-helix and HRD motif (R)), His361 – Asp421 (hydrogen bond securing R spine to αG-helix). The hydrogen bond between the magnesium binding loop and the HRD motif (Leu384 and Arg362), was absent in the three simulations of the inactive structure but stable in the 2F4J and $2F4J_{p\text{-}apo}$ simulations suggesting that this hydrogen bond could be used as a marker of the active conformation.

Based on these analyses the following characteristics appear to be stable markers of the inactive conformation.

- A folded A-loop conformation
- The DFG-out conformation (Asp381-Phe382-Gly383)

- Strong hydrogen bond interactions between Tyr393 and Asp363
- A misaligned regulatory spine (R-spine), due to the position of Phe382
- Kinked and intermediate conformations of the P-loop

Both nilotinib and imatinib were stable during the simulations, despite the central sections of the ligands not being consistently identified by pocket analysis, and analysis of the protein-ligand interactions confirmed that the hydrogen bonds in the crystal structures were maintained throughout the simulations. A comparison of the interactions with imatinib and nilotinib in the 1IEP and 3CS9 simulations shows that of the four shared hydrogen bonds, three are more stable in the nilotinib complex (Thr315, Glu286, and Asp381). This supports what is known regarding nilotinib being a more potent inhibitor of the Abl kinase.

In the remainder of this chapter the differences in protein-ligand binding will be further discussed using the results from ensemble docking and approximate free energy calculations (MD-GBSA) studies.

6.7 Ensemble docking

6.7.1 Methodology

Docking simulations were completed using the Schrödinger suite of programs (Suite2008; Glide 5.0, LigPrep 2.2, Maestro 8.5, Schrodinger LLC, New York NY 2008). 200 conformations were extracted from the 20 ns MD simulations as docking receptors and prepared in Maestro using the Protein Preparation Wizard. The ligands imatinib and nilotinib were prepared using the LigPrep module. A scoring grid was generated, centred on the imatinib/nilotinib binding site. Docking simulations and docking scores were performed using the Glide module using both sets of receptors and both ligands (Figure 6.7.1).

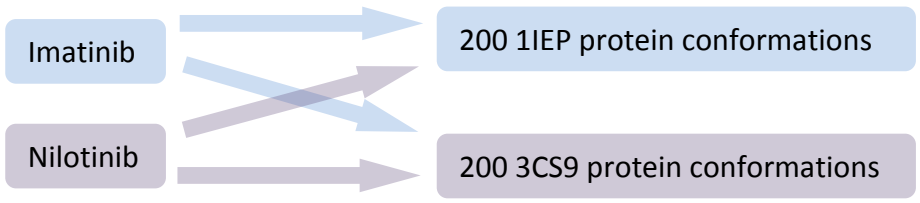


Figure 6.7.1: Docking study methodology

6.7.2 Results

The docking studies showed that both imatinib and nilotinib bound favourably to the 1IEP and 3CS9 receptor conformations, and predicted that nilotinib would have improved binding affinity compared to imatinib in line with experimental results.

	Imatinib	Nilotinib
IC50 (nM)	768.3	57
GlideScore (1IEP as receptor, kcal/mol)	-13.40	-13.78
GlideScore (3CS9 as receptor, kcal/mol)	-14.78	-15.41

Table 6.7.1: GlideScore for imatinib and nilotinib docked to $1IEP_{WT}$ and $3CS9_{WT}$ (best achieved) vs IC_{50} values. IC_{50} values from Zabriskie et al [331].

As a check against the docking scores, approximate free energy calculations were completed for the two complexes using the MM-GBSA method.

6.8 MM-GBSA calculations

6.8.1 Methodology

MM-GBSA calculations were performed on the 1IEP and 3CS9 trajectories using the pymdbsa script by Romain Wolf at Novartis, which uses the pbsa routine in Amber Tools [161-163]. Each trajectory frame was split into receptor and ligand. The free energy for each component (ligand/receptor/complex) can be defined in terms of a gas phase energy, a solvation free energy and an entropy term (Equation 6.1 and 6.2).

$$G = E_{gas} + G_{solv} - T \cdot S \quad \textbf{(6.1)}$$

$$G = E_{bat} + E_{vdw} + E_{coul} + G_{solv,polar} + G_{solv,nonpolar} - T \cdot S \quad \textbf{(6.2)}$$

Where E_{bat} is the sum of bond, angle and torsion terms in the force field, E_{vdw} and E_{coul} are the van der Waals and Coulomb energy terms, $G_{solv,polar}$ is the polar contribution to the solvation free

energy and $G_{solv,nonpolar}$ in the non-polar solvation free energy. The $G_{solv,polar}$ term is derived using the Generalised-Born (GB) model and the $G_{solv,nonpolar}$ is calculated via Equation 6.3.

$$G_{solv,nonpolar} = \gamma \cdot SASA + const. \quad \textit{(6.3)}$$

Where SASA is the solvent accessible surface area and γ is the surface tension (0.0072 for AMBER). Given the structural similarity between the two ligands the entropy terms were not calculated.

The free energy of binding can then be evaluated as:

$$\Delta G_{bind} = G_{com} - G_{rec} - G_{lig} \quad \textit{(6.4)}$$

Where com, rec and lig stand for complex, receptor and ligand, respectively.

The final interaction energy is an average of all the ΔG values for the individual frames [131, 161].

6.8.2 Results

The MM-GBSA calculations were consistent with the docking results, correctly predicting that nilotinib has improved binding affinity compared to imatinib.

Components	1IEP + imatinib	3CS9 + nilotinib
Total ligand energy (kcal/mol)	-144.95	-223.99
Total receptor energy	-6368.80	-6413.74
Total complex energy	-6586.41	-6712.63
Difference (kcal/mol)	**-72.66**	**-74.90**

Table 6.8.1: MM-GBSA results for imatinib and nilotinib ($1IEP_{WT}$ and $3CS9_{WT}$)

6.9 Chapter summary

In this chapter the structure and dynamics of the inactive conformation have been investigated using molecular dynamics simulations on three inactive structures; two complexes (1IEP with imatinib and 3CS9 with nilotinib) and one with the inhibitor removed ($1IEP_{p\text{-}apo}$). This work has been completed to confirm the stable characteristics of the inactive state and compare the binding modes of imatinib and nilotinib. The differences in protein-ligand binding were then

further investigated using ensemble docking and approximate free energy calculations (MM-GBSA).

Based on the results of this chapter the following characteristics appear to be stable markers of the active and inactive conformations.

The active conformation:

- An extended A-loop conformation
- The DFG-in conformation (Asp381-Phe382-Gly383), but not necessarily the hydrogen bond between Asp381 and Gly383.
- Strong hydrogen bond interactions between Leu324 and Arg362
- An aligned regulatory spine (R-spine, based on the distances between residues)
- Elongated and intermediate conformations of the P-loop

The inactive conformation

- A folded A-loop conformation
- The DFG-out conformation (Asp381-Phe382-Gly383), but not necessarily the hydrogen bond between Asp381 and Phe382.
- Strong hydrogen bond interactions between Tyr393 and Asp363
- A misaligned regulatory spine (R-spine, based on the distances between residues)
- Kinked and intermediate conformations of the P-loop

Several of the hydrogen bonds identified as critical to the active structure, were also stable in the inactive structure and are therefore unsuitable to be used as markers of activity.

Both nilotinib and imatinib were stable during the simulations despite the central sections of the ligands not being consistently identified by pocket analysis. The results of the interaction analysis showed that nilotinib is the more stable of the two ligands, supporting what is known experimentally regarding nilotinib being a superior inhibitor compared to imatinib.

The behaviour of pseudo apo structures suggests that the inactive conformation may be less stable than the active conformation, given the unfolding of several helices and the increase in RMSF when the ligand was removed. This suggests that a significant loss of interaction between the kinase and the ligand could destabilise the inactive conformation.

6: Inactive BCR-Abl

The docking and MMGBSA results correctly predicted that nilotinib has improved binding affinity over imatinib, in line with what is known experimentally.

Chapter 7

7 Single mutation study

7.1 Aim

The aim of this chapter is to investigate the individual effects of three common mutations, F317L, Y253F and T315I, on the conformation of Abl and the binding modes of imatinib and nilotinib. Docking and MMGBSA calculations will be utilised to investigate whether these methods can correctly rank the mutations according to what is known experimentally. The F317L, Y253F and T315I mutations were selected based on a mixed resistance profile to imatinib and nilotinib and differing mechanisms of resistance. Y253F was chosen over the more common mutation T253H to avoid the challenges associated with histidine protonation states.

Published in vitro data was used to select the mutations for study. The data in Table 7.1.1 supports two important facts; imatinib and nilotinib are both insensitive to the gatekeeper mutation, T315I, and nilotinib shows greater sensitivity to the F317L and Y253F mutations than imatinib. Although imatinib and nilotinib both demonstrate intermediate sensitivity to the F317L mutation, this mutation appears to be less resistant to nilotinib. The IC_{50} data for nilotinib suggests intermediate sensitivity to Y253F, whereas imatinib is insensitive to this mutation.

	Imatinib (IC_{50} in nM)				Nilotinib (IC_{50} in nM)		
	Azam	**O'Hare**	**Weisberg**	**Zabriskie**	**O'Hare**	**Weisberg**	**Zabriskie**
WT	600	260	649 ± 52	768 ± 143	13	25 ± 2	57.0 ± 11.1
F317L	2,300	1,050	1,583 ± 236	2459 ± 267	50	80 ± 6	137 ± 19
Y253F	NR	3,475	NR	NR	125	NR	NR
T315I	>20,000	>6,400	7,393 ± 157	>10,240	>2,000	>10,000	>10,240

Table 7.1.1. Sensitivity of Bcr-Abl kinase domain mutants to imatinib and nilotinib (based on assays using Ba/F3 cells. The IC_{50} value is the concentration of inhibitor resulting in a 50% reduction in cell viability; NR, not recorded. Adapted from references: Azam et al, 2003 [44], O'Hare et al, 2009 [333] Weisberg et al, 2005 [32] and Zabriskie et al, 2014 [331])

Sensitivity	Imatinib (IC_{50} in nM)	Nilotinib (IC_{50} in nM)
Sensitive	< 1000	< 200
Moderately resistant	1000 – 4000	200 - 1000
Highly resistant	> 4000	> 1000

Table 7.1.2. Sensitivity scale for imatinib and nilotinib. Adapted from reference [331].

According to the experimental data in Table 7.1.1, the mutations are expected to be ranked as follows for both imatinib and nilotinib: WT > F317L > Y253F > T315I.

The F317L mutation causes the loss of interaction with the ligand whereas the Y253F mutation hinders sensitivity by the distorting the conformation of the P-loop away from the kinked conformation associated with imatinib induced fit binding. The gatekeeper mutation T315I disrupts binding directly through the loss of a hydrogen bond with the ligand and blocking access to the binding site.

One of the objectives of this chapter will be to investigate if the any of the three mutations studied destabilise the inactive structure. Since it is unlikely that a large conformational change will be seen within these timescales, the simulations will be reviewed to assess if there has been any movement towards an intermediate state. Previous studies have discussed that one pathway from the inactive state involves the rotation of the C-helix away from the binding site, breaking the Glu-Lys salt bridge and Glu286 forming a hydrogen bond with Arg386 on the A-loop. This gives the Phe residue on the DFG loop enough space to flip 180 degrees, stabilising the R-spine and triggering the formation of the active state. Therefore the position of the Glu286 on the C-helix and its proximity to Arg386 is considered to be an important marker of conformational shift [12, 81].

7.2 Molecular dynamics study of mutations F317L, Y253F and T315I

7.2.1 Review of the mutations

F317L: phenylalanine mutated to leucine

The side chain of Phe317 (F317), which is part of the hinge region, effectively forms the ceiling of the ATP binding site. This residue stacks partially on the pyrimidine ring of imatinib and nilotinib, and the side chain shields the hinge region from the solvent forming a hydrophobic enclosure. In Chapter 5 it was confirmed that Phe317 makes pi-stacking interactions with imatinib and nilotinib in the binding site, therefore a mutation at this residue is likely to directly affect ligand binding. Mutation F317L is known to be moderately resistant to imatinib but not to nilotinib; this is due to nilotinib's increased potency as an inhibitor. The trifluoromethyl and imidazole groups contribute greatly to this potency, therefore the interactions involving the hinge region residues, including Phe137, and the pyridinyl and pyrimidinyl rings of nilotinib contribute far less to the overall binding affinity [32, 262, 300, 333].

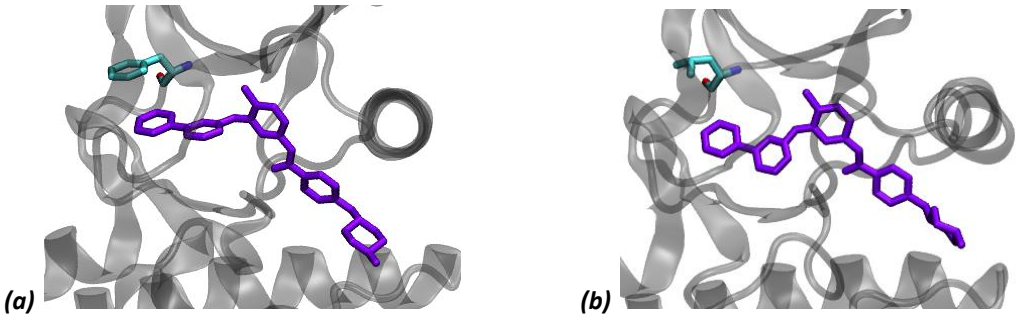


Figure 7.2.1: (a) The position of Phe317 in WT Abl (1IEP.pdb); (b) The position of Leu317 in mutant Abl (1IEP$_{F317L}$)

Y253F: tyrosine mutated to phenylalanine

Tyr253 (Y253) in WT Abl forms one side of the hydrophobic cage around the imatinib binding site; the tyrosine aromatic ring is approximately 4 Å away from the pyridine-pyrimidine rings of imatinib, possibly forming π-π interactions. The positioning of Tyr 253 allows it to form a water bridge interaction with Asp322 that folds the P-loop into the kinked conformation for preferential binding with imatinib/nilotinib. The mutation of Tyr253 to Phe253, would be unlikely to have a significant effect on ligand binding directly but would disrupt the conformation of the P-loop and therefore the preferential conformation of the binding site. This is in agreement with clinical data, where the Y253F mutation is associated with resistance to imatinib and intermediate sensitivity to nilotinib [32, 300, 333, 334].

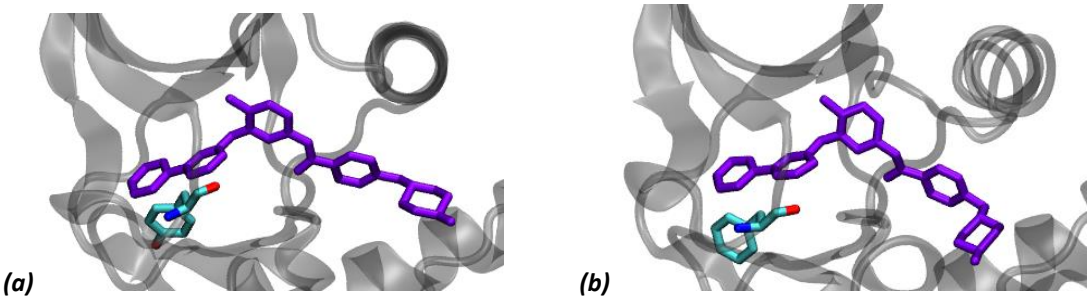


Figure 7.2.2: (a) The position of Tyr 253 in WT Abl (1IEP.pdb); (b) The position of Phe 253 in mutant Abl (1IEP$_{Y253F}$)

T315I: threonine mutated to isoleucine

The threonine residue at position 315 in Abl is the so-called gatekeeper residue. The size of the residue at this position determines how easily the hydrophobic pocket can be accessed especially in the (DFG-in) active form. In the case of imatinib binding, the hydroxyl group of T315 forms a hydrogen bond to the amine linker between the pyrimidine and phenyl rings of imatinib. Mutation of the threonine therefore results in a loss of this hydrogen bonding interaction, while the larger sizes of the residues at this position (such as isoleucine in the case of T315I) also result in steric clashes, preventing imatinib from binding to mutant Abl structures [4].

Mutagenesis studies have also shown that the gatekeeper mutations in Abl and other kinases (PDGFR, EGFR and Src) have increased activity in cells expressing this mutation. The increased activity is due to the strengthening of the R-spine, which is a conserved feature of active kinases and is disrupted in inactive kinases. The increase in activity can further contribute to drug resistance [260, 300, 317, 333].

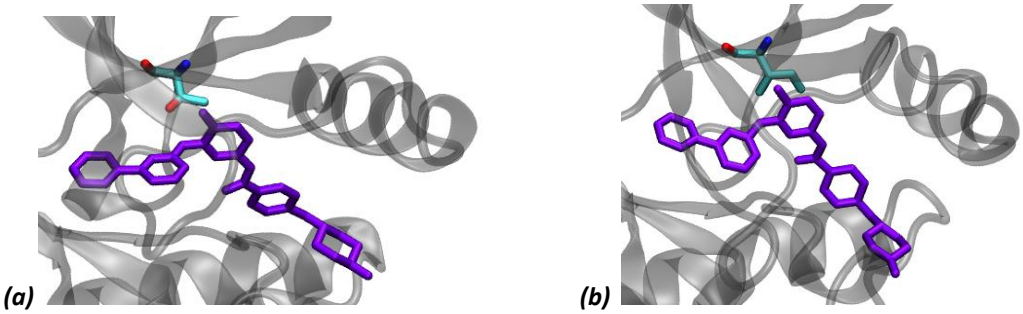


Figure 7.2.3: (a) The position of Thr315 in WT Abl (1IEP.pdb); (b) The position of Ile315 in mutant Abl (1IEP$_{T315I}$)

In contrast to imatinib and nilotinib, ponatinib maintains inhibition to BCR-Abl by forming favourable interactions with the mutated isoleucine residue in T315I.

It would be valuable to repeat this study with ponatinib in order to diversify the results and include an example of a TKI that is capable of inhibiting all three single mutants (F317L, Y253F, T315I).

7.2.2 Protein structure preparation and MD parameters

The SCAP program was used to mutate residues Y253, F317 and T315 [71]. Protein structure preparation and molecular dynamics simulations were then completed as described in 5.2.4.

7.3 Results: F317L

According to experimental data the F317L mutation confers resistance to imatinib and diminished binding affinity with nilotinib [300, 333]. Analysis of the WT structures confirmed that Phe317 interacts with imatinib and nilotinib; this interaction is expected to be diminished or lost with the substitution of a leucine residue at this position. Nilotinib is a more potent inhibitor than imatinib and less susceptible to hinge region mutations, therefore can maintain binding with Abl$_{F317L}$ despite the loss of interactions. Nilotinib is expected to be more stable than imatinib in the F317L simulations; any changes in the conformation of the binding site are likely to have a greater destabilising impact on imatinib given it has a more stringent induced-fit

binding mode [32, 262]. Given the location on the hinge region, the F317L mutation is not expected to cause a significant change in the kinase conformation in either the imatinib or nilotinib complexes.

7.3.1 Imatinib

Structural stability and flexibility of the protein

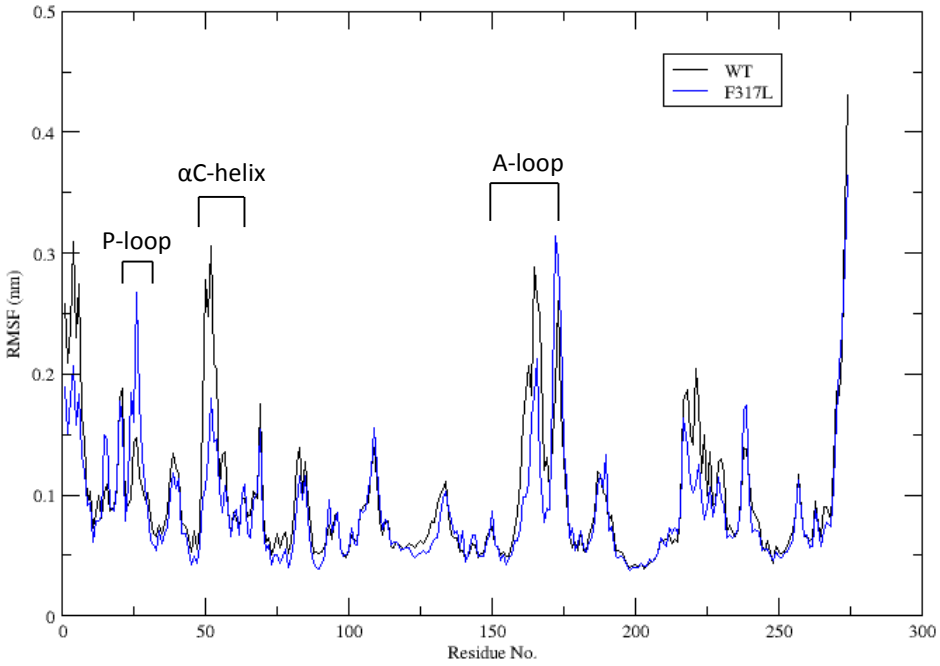
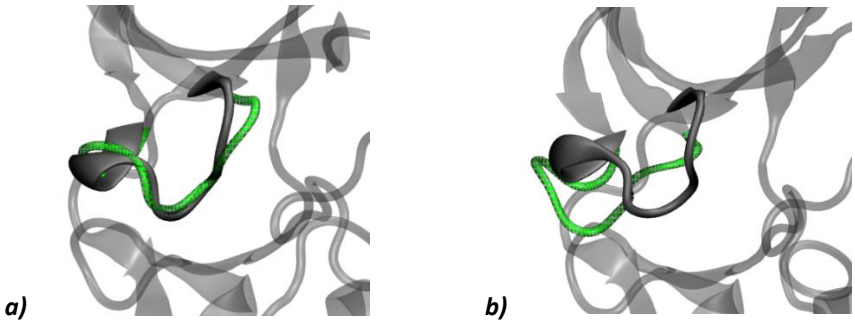


Figure 7.3.1: RMSF results for 1IEP$_{F317L}$

The RMSF profile for 1IEP$_{F317L}$ displayed increased mobility in the P-loop and second half of the A-loop compared to the wild-type structure; fluctuations in the other regions in the kinase were comparable to that of the wild-type.

As expected the increased fluctuation in the P-loop did not translate to a significant change in conformation; the loop remained in the inactive kinked conformation for the duration of the simulation.



a) *b)*

Figure 7.3.2: P-loop conformation: initial (a) and during simulation (b). WT shown in grey, F317L mutated structure in green

PCA confirmed that the most significant motions were located at the P- and A-loop but that the inactive conformation of these motifs was maintained.

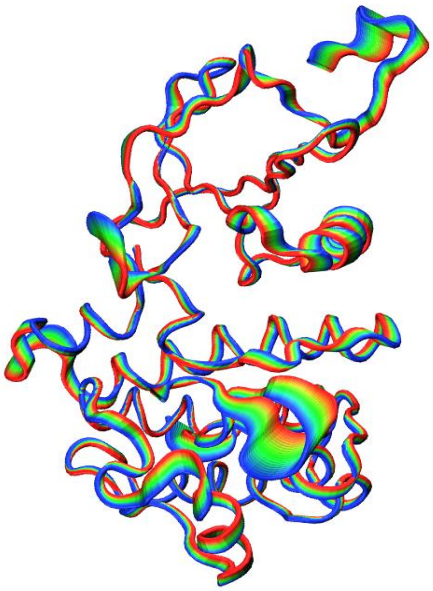


Figure 7.3.3: Principal component analysis for 1IEPF317L (PC1, projected on structure)

Fluctuations located around the HRD motif caused the distance between Phe382 and His361 to decrease by approximately 2.4Å, however the Phe residue in the R-spine remained misaligned with respect to the rest of the stack (Table 7.3.1).

R-spine residues	Distance (Å)	
	1IEP_WT (Å)	1IEP_F317L (Å)
Leu301 – Met290	6.487	6.746
Met290 – Phe382	12.284	12.927
Phe382 – His361	12.302	9.859

Table 7.3.1: Mean distance between R-spine residues

The mobility of the HRD motif was also captured in the hydrogen bond analysis; the well conserved hydrogen bond between Tyr393 and Asp363 was less stable in the F317L mutated structure. The Tyr393 – Asp363 hydrogen bond aids the folding of the A-loop and the reduced stability of this bond would likely increase the flexibility of the A-loop, which is aligned with the RMSF analysis. The Lys271-Glu286 salt bridge, although not as stable as in the WT simulation, was well formed throughout the simulation and there was no obvious shift towards the intermediate pathway via the rotation of the C-helix (Table 7.3.2).

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{F317L}$	$1IEP_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	16.28	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	84.62	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	50.15	45.25
		Gly383(M)	Asp381(M)	0.10	0.50
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	98.50	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	95.20	80.62
		His361(S)	Asp421(S)	12.49	33.47

Table 7.3.2: Hydrogen bond analysis: $1IEP_{F317L}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Stability and flexibility of imatinib

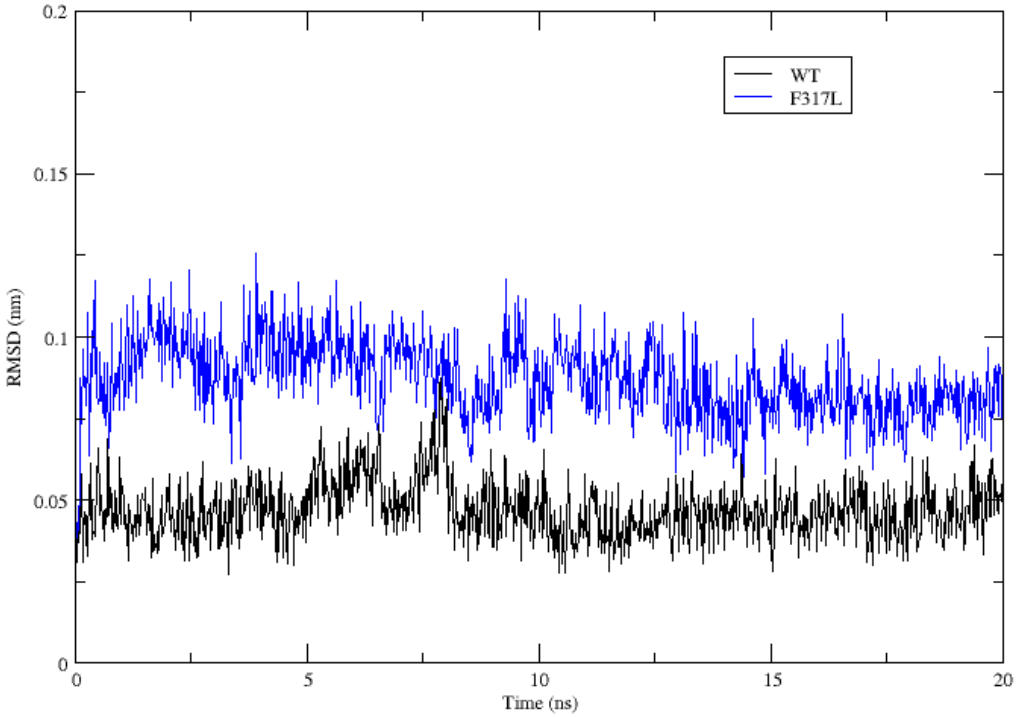


Figure 7.3.4: RMSD of imatinib (1IEP$_{F317L}$)

The RMSD for imatinib was approximately 0.05 nm higher in 1IEP$_{F317L}$ compared to the 1IEP$_{WT}$. The RSMF analysis (Figure 7.3.5 and Figure 7.3.6) showed that this was in line with the fluctuations of the overall protein structure and that the piperazine ring (at the mouth of the binding site) was the most mobile part of the ligand. The movement of the ligand in the binding site is further illustrated in Figure 7.3.7, which shows the conformation of the ligand in 1IEP$_{F317L}$ compared to 1IEP$_{WT.}$

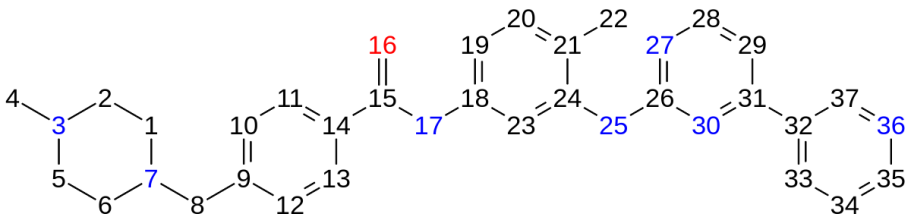


Figure 7.3.5: Index of atoms in imatinib (1IEP$_{F317L}$)

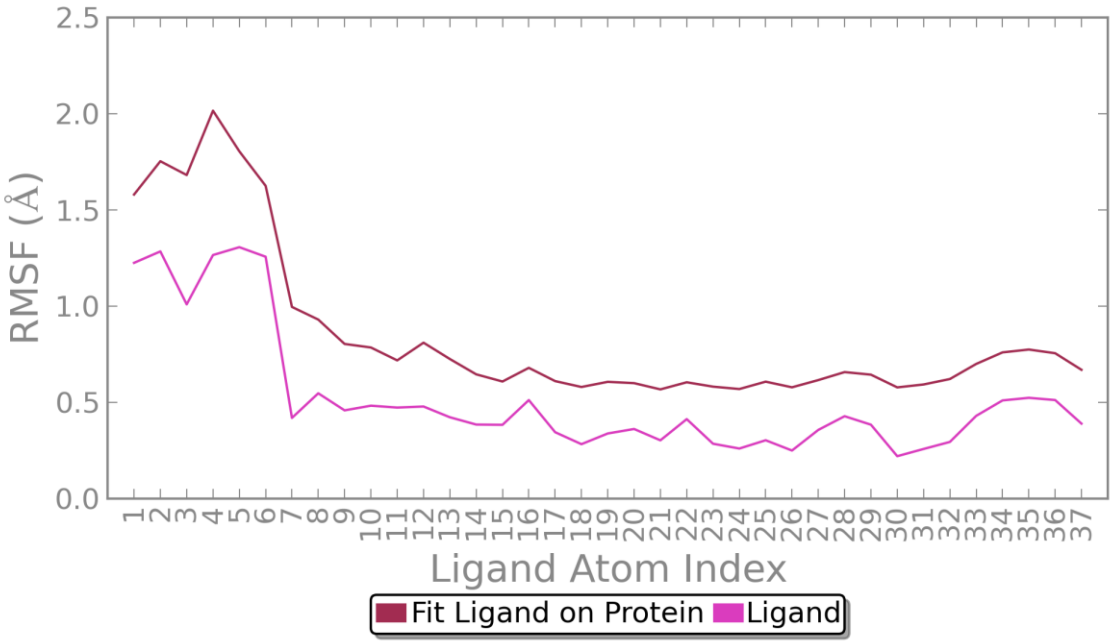


Figure 7.3.6: RMSF of imatinib (1IEP$_{F317L}$)

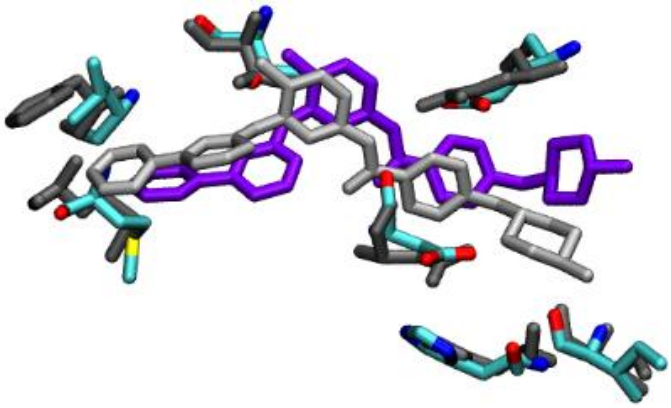


Figure 7.3.7: Comparison of imatinib binding site in the WT and F317L mutant structures; grey/silver, WT; purple/cyan, F317L (5 ns)

Protein-ligand interactions

As expected, the substitution of a leucine residue at position 317 (Leu93 in Figure 7.3.8) resulted in a loss of pi-stacking interactions between the protein and the ligand.

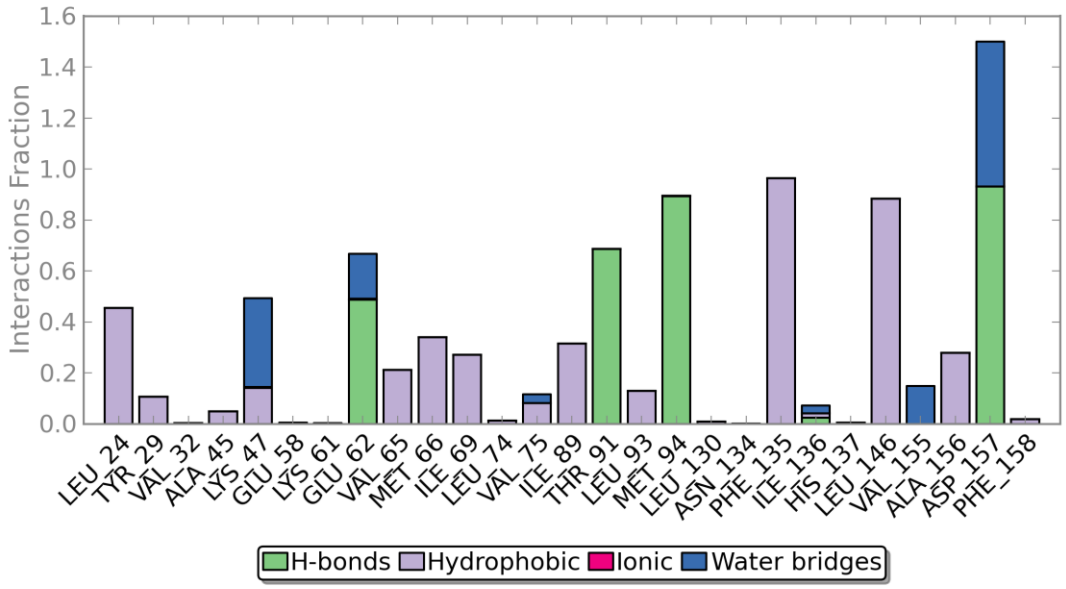


Figure 7.3.8: Protein-ligand interactions by residue (1IEP$_{F317L}$)

The increased fluctuation and change in conformation of the ligand also resulted in the loss of the two hydrogen bonds between the piperazine ring of imatinib and Ile360 and His361 (Ile136 and His137 in Figure 7.3.8). The loss of the hydrogen bond with imatinib means that the His361 (His137 in Figure 7.3.8) is no longer tightly bound; this supports the structural analysis which showed increased fluctuation on the HRD motif.

In summary, the analysis of the 1IEP$_{F317L}$ structure suggests that the loss of pi-stacking and change in shape caused by the substitution of phenylalanine with leucine at position 317 disrupts the preferred induced fit of imatinib, causing the ligand to shift to a suboptimal position in the pocket, and the loss of two critical hydrogen bonds. The decrease in protein-ligand interactions at the hinge region and entrance to the binding site increased the mobility of the residues at these points, which had causal effects on the fluctuations of the P-loop and A-loop.

7.3.2 Nilotinib

Stability and flexibility of protein

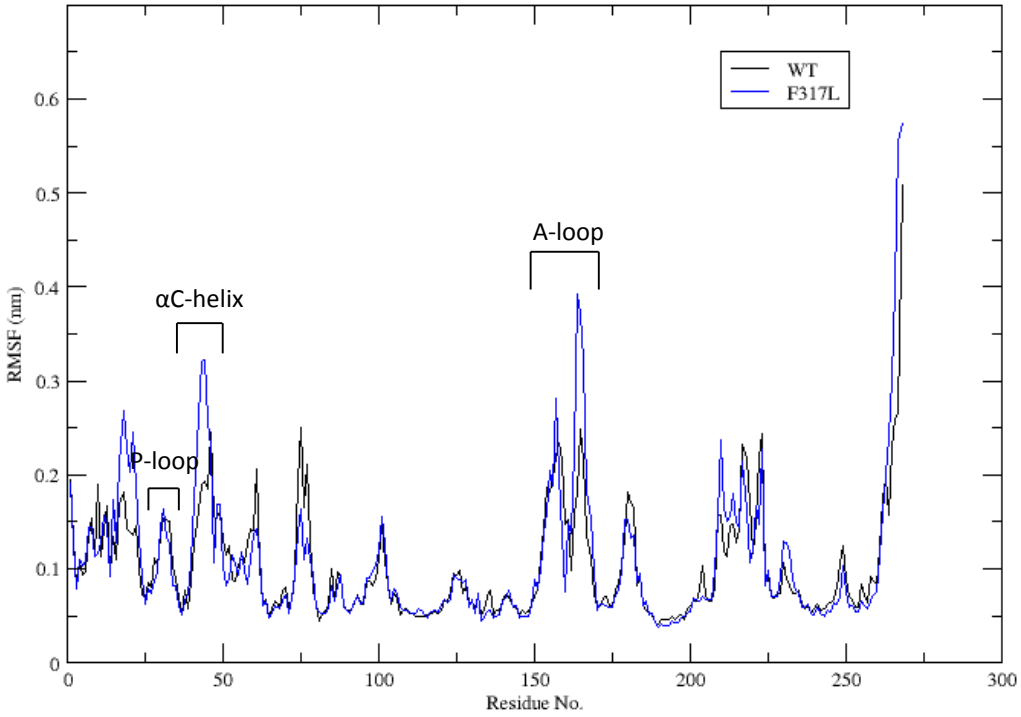


Figure 7.3.9: RMSF results for $3CS9_{WT}$ and $3CS9_{F317L}$

The RMSF profile for $3CS9_{F317L}$ displayed increased mobility in the P-loop, αC-helix and second half of the A-loop compared to the wild-type structure; fluctuations in the other regions in the kinase were comparable to that of the wild-type. As with the $1IEP_{F317L}$ simulation, the increased fluctuation in the P-loop did not translate to a significant change in conformation and maintained the inactive kinked conformation. In the WT simulation the αC-helix was seen to move away from the binding site; in the F317L mutated structure these motions were more pronounced as demonstrated in the PCA projection in Figure 7.3.10.

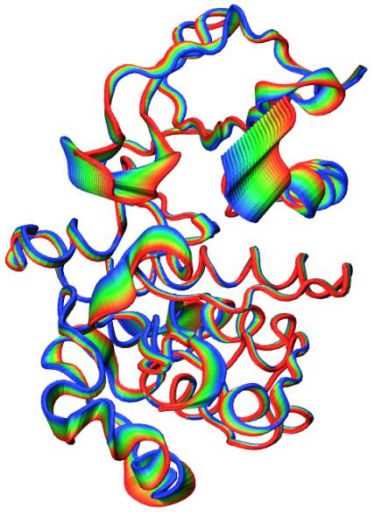


Figure 7.3.10: Principal component analysis for 3CS9$_{F317L}$ (PC1, projected on structure)

A review of the R-spine showed that the residues maintained similar positions to that of the WT structure, suggesting that the F317L had minimal impact to the position of these residues (Table 7.3.3).

R-spine residues	Mean distance (CA-CA, Å)	
	3CS9_WT	3CS9_F317L
Leu301 – Met290	6.500	6.632
Met290 – Phe382	12.874	12.066
Phe382 – His361	11.514	11.087

Table 7.3.3: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was well conserved throughout the simulation, and appeared to be more stable in the F317L simulation than the WT. The Lys271- Glu286 salt bridge was less stable than in the WT simulation, this is due to the increased movement of the αC-helix in 3CS9$_{F317L}$ which results in the breaking of the hydrogen bond between Lys271 and Glu286.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$3CS9_{F317L}$	$3CS9_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	97.90	97.70
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	29.77	82.32
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	77.82	19.08
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	97.40	99.60
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	75.62	85.71
		His361(S)	Asp421(S)	9.89	32.97

Table 7.3.4: Hydrogen bond analysis: $3CS9_{F317L}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Since the breaking of the salt bridge can mark the initiation of a shift towards the intermediate conformation, the structure was reviewed for the formation of the Glu286-Arg386 hydrogen bond following the movement of the αC-helix away from the binding site; Figure 7.3.11 shows that Glu286 does form a hydrogen bond with Arg386, however the C-helix is only partially rotated and due to the position of the ligand, and the hydrogen bond between Asp381 and nilotinib, the Phe382 residue cannot facilitate the 180° flip. This is supported by the R-spine analysis, which demonstrated that the misaligned conformation of the R-spine is stable in $3CS9_{F317L}$. Glu286 fluctuates between forming a hydrogen bond with Lys271 and Arg386 throughout the simulation.

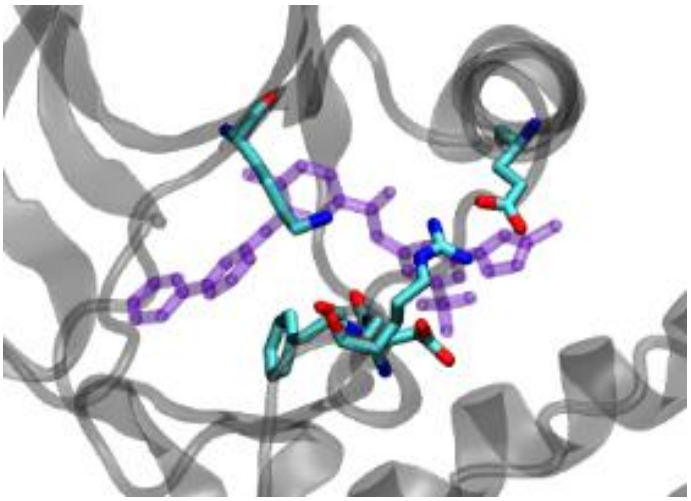


Figure 7.3.11: Glu286-Arg386 bond formation ($3CS9_{F317L}$)

Stability and flexibility of nilotinib

Nilotinib demonstrated stable motions within the pocket, with a marginally lower RMSD compared to the WT structure.

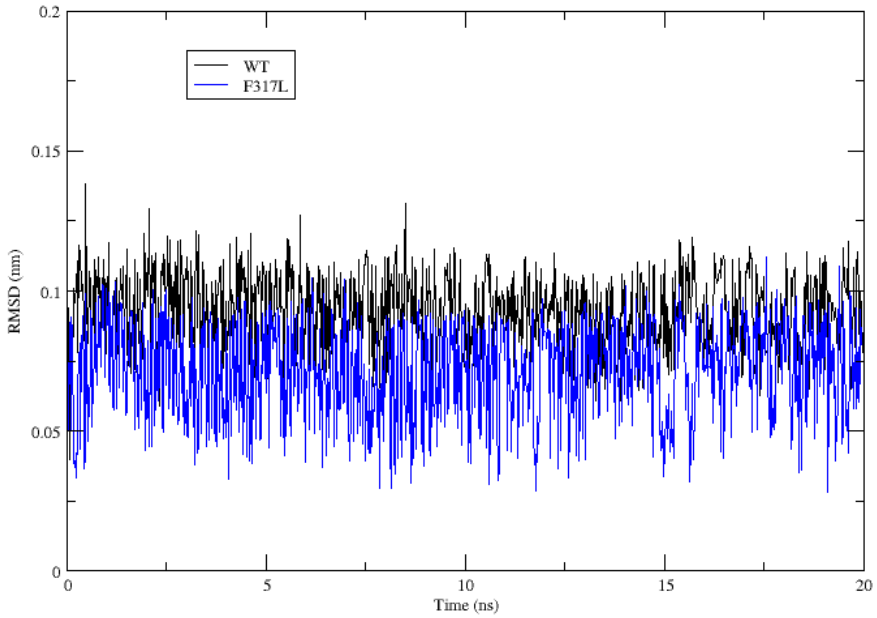


Figure 7.3.12: RMSD of nilotinib (3CS9$_{F317L}$)

RMSF analysis confirmed that fluctuations of the ligand were in line with that of the overall protein structure (Figure 7.3.13 and Figure 7.3.14) and that the largest fluctuations were located at the trifluoromethyl and imidazole groups at the entrance to the binding site.

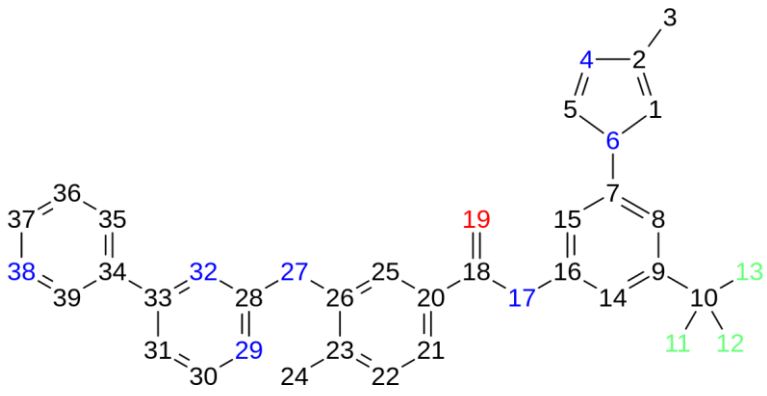


Figure 7.3.13: Index of atoms in nilotinib (3CS9$_{F317L}$)

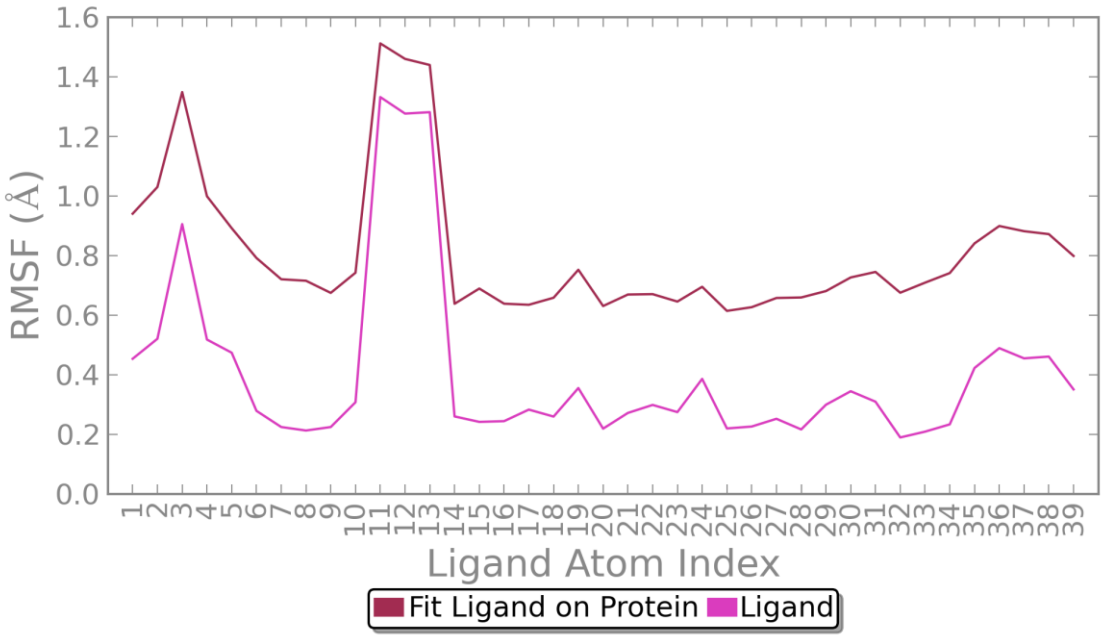


Figure 7.3.14: RMSF of nilotinib (3CS9$_{F317L}$)

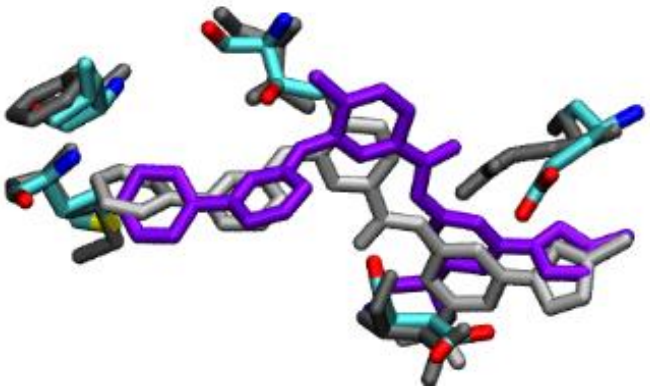


Figure 7.3.15: Comparison of nilotinib binding site in the WT and F317L mutant structures; grey/silver, WT; purple/cyan, F317L (5 ns)

Protein-ligand interactions

Figure 7.3.15 shows the conformation of the ligand in 3CS9$_{F317L}$ compared to 3CS9$_{WT.}$ The nilotinib amide group that binds to the Glu286 residue has rotated, and Glu286 occupies a slightly more distant position in 3CS9$_{F317L}$ compared to the WT. This is in contrast to the positions of other binding site residues which are well aligned with the WT simulation. The protein ligand interactions analysis confirmed that the replacement of Phe317 with a leucine (Leu85 in Figure 7.3.16) results in the loss of pi-stacking interactions with nilotinib. The hydrogen bonds between

nilotinib Thr315, Met318 and Asp381 (Thr83, Met86 and Asp149 in Figure 7.3.16) are maintained, though diminished with respect to the WT, however due the increase in distance between Glu286 and nilotinib, this hydrogen bond is mediated by a water molecule in the F317L simulation (Glu286 is Glu84 in Figure 7.3.16).

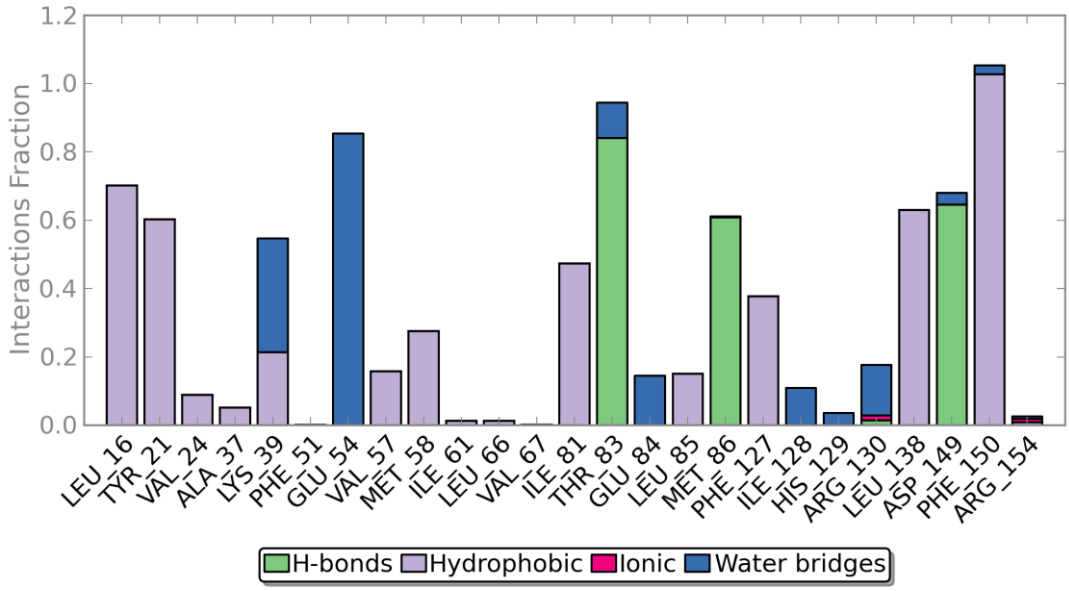


Figure 7.3.16: Protein-ligand interactions by residue (3CS9$_{F317L}$)

In summary, the analysis of the 3CS9$_{F317L}$ structure suggests that the loss of pi-stacking interactions disrupts the preferred binding conformation of nilotinib, causing the ligand to shift in conformation and the replacement of the Glu286 hydrogen bond with a water bridge.

Interactions involving the hinge region residues, including Phe137, and the pyridinyl and pyrimidinyl rings of nilotinib contribute far less to the overall binding affinity compared in imatinib, which supports data that F317L is sensitive to nilotinib inhibition [32, 262]. The decrease in protein-ligand interactions at the hinge region increased the mobility of the residues in this region and the fluctuations of the P-loop. The diminishment of the hydrogen bond between Glu286 and nilotinib enabled the rotation of the αC-helix and subsequent Glu286-Arg386 bond. This likely has a causal effect on the residues surrounding the HRD motif and consequential effects on the flexibility of the A-loop.

7.4 Results: Y253F

The Y253F mutation is associated with resistance to imatinib and intermediate sensitivity to nilotinib [32, 300, 333, 334]. Tyr253 enables the kinked conformation of the P-loop for preferential binding with imatinib/nilotinib. The Y253F mutation would be unlikely to have a

significant effect on ligand binding interactions directly but would disrupt the conformation of the P-loop and therefore the preferential conformation of the binding site.

7.4.1 Imatinib

Stability and flexibility of the protein

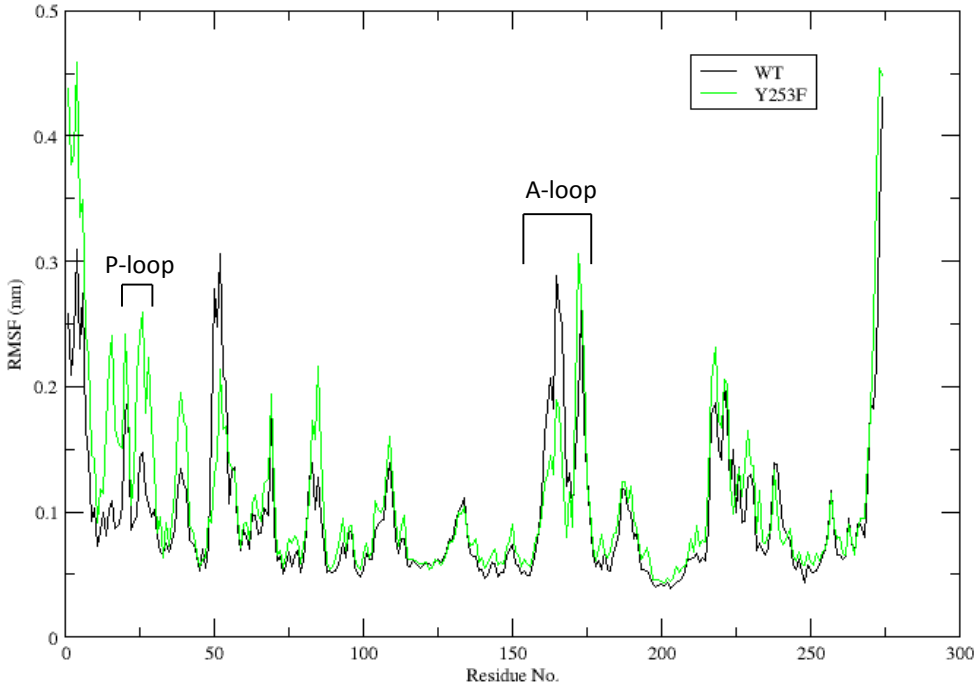


Figure 7.4.1: RMSF results for $1IEP_{WT}$ and $1IEP_{Y253F}$

The RMSF profile for $1IEP_{Y253F}$ showed a significant increase in the mobility of the P-loop (residues 25-30 in Figure 7.4.1) compared to the wild-type structure, and a moderate increase in RMS across the protein structure overall.

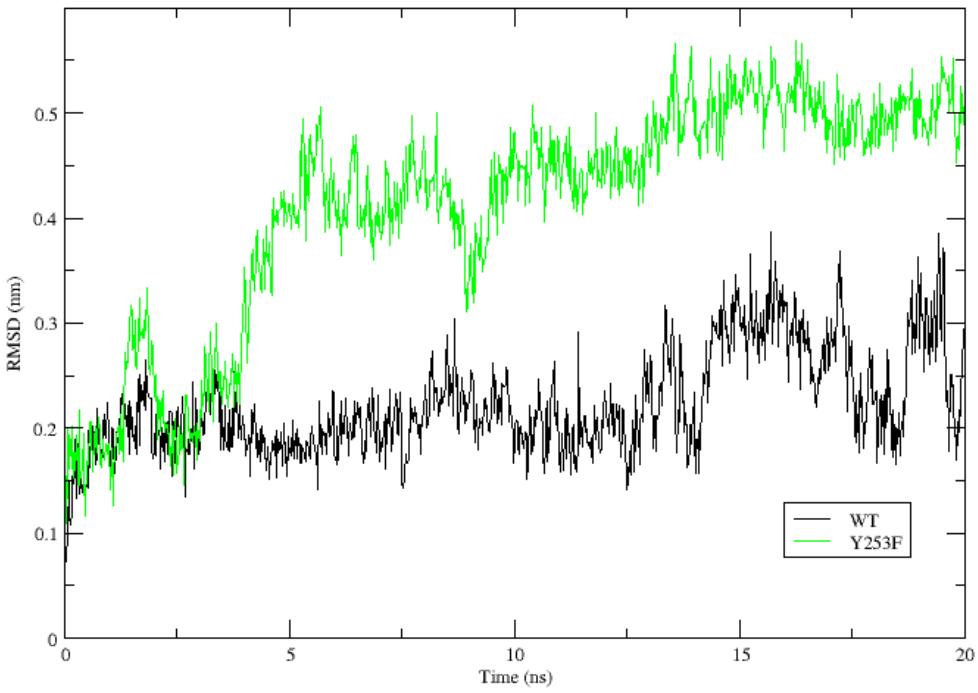


Figure 7.4.2: RMSD of the P-loop (1IEP$_{WT}$ and 1IEP$_{Y253F}$)

Although there was a substantial increase in P-loop flexibility and distortion in the loop itself, the P-loop remained in the inactive kinked conformation for the duration of the simulation.

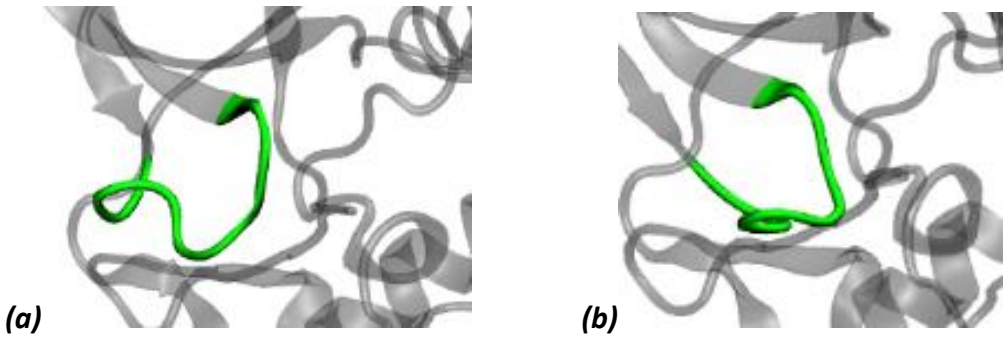


Figure 7.4.3: Fluctuations of the P-loop: initial (a) and during simulation (b) (1IEP$_{Y253F}$)

The RMSD analysis highlighted a large fluctuation in the DFG loop during the 1IEP$_{Y253F}$ simulation (Figure 7.4.4); the DFG-loop was highly stable WT structure and increased motion in this region can signify a shift in the inactive state. Observational analysis (Figure 7.4.5) showed that Phe383 residue had shifted away from the WT position but had not flipped away from the active site; this was supported by the R-spine analysis which identified that the distance between Phe382 and Met290 has increased (Table 7.4.1).

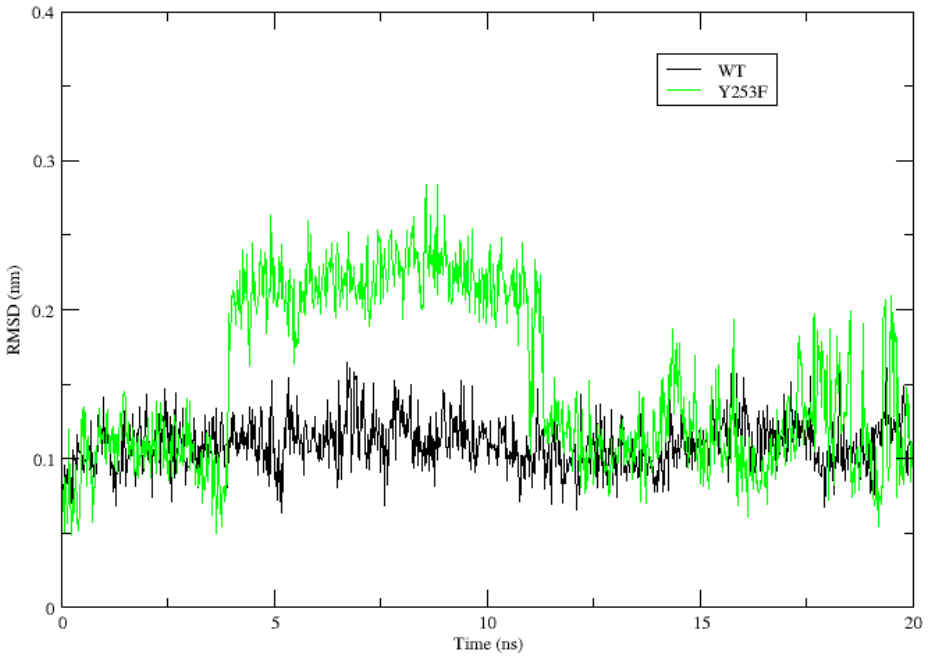


Figure 7.4.4: RMSD of the DFG-loop ($1IEP_{WT}$ and $1IEP_{Y253F}$)

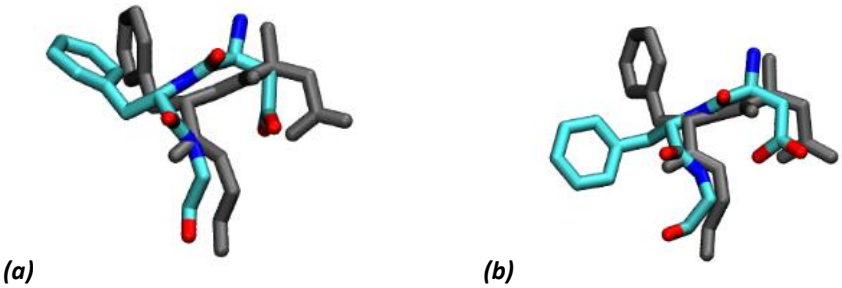


Figure 7.4.5: (a) Initial conformation of DFG loop (b) Conformation shift of DFG loop during simulation (between ~4 ns and 11 ns)

R-spine residues	Mean distance (CA-CA, Å)	
	1IEP_WT (Å)	1IEP_F317L (Å)
Leu301 – Met290	6.487	6.446
Met290 – Phe382	12.284	13.493
Phe382 – His361	12.302	11.584

Table 7.4.1: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the Y253F mutated structure; reduced stability was observed in the Lys271-Glu286 salt bridge, but the hydrogen bond was conserved throughout the simulation (Table 7.4.2)

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{Y253F}$	$1IEP_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	99.70	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	79.82	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	38.46	45.25
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.90	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	88.61	80.62
		His361(S)	Asp421(S)	23.38	33.47

Table 7.4.2: Hydrogen bond analysis: $1IEP_{Y253F}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Stability and flexibility of imatinib

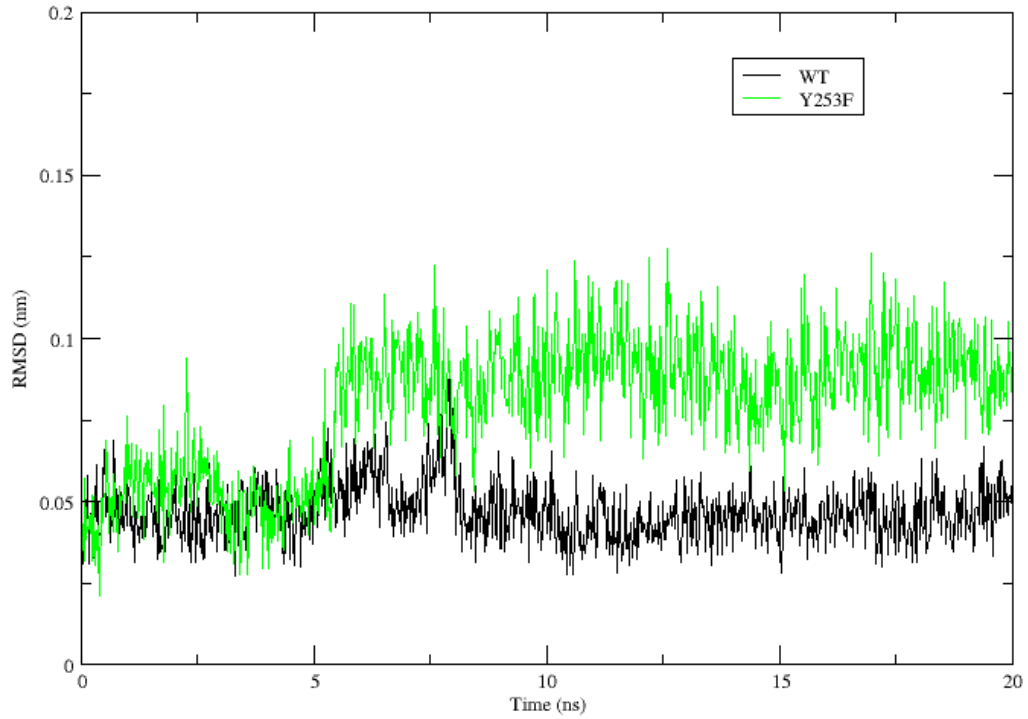


Figure 7.4.6: RMSD of imatinib ($1IEP_{Y253F}$)

The mobility of the ligand in the $1IEP_{Y253F}$ simulation was initially well aligned with the WT but increased by approximately 1 Å during the simulation; the RSMF analysis showed that this was in line with the fluctuations of the overall protein structure and that the piperazine ring (at the mouth of the binding site) was responsible for the most of the motions The RSMF analysis (Figure 7.4.7 and Figure 7.4.8) showed that that the piperazine ring (at the mouth of the binding site) the amide CO were the most flexible regions of the ligand. The movement of the ligand in the binding site is further illustrated in Figure 7.4.9, which shows the conformation of the ligand in $1IEP_{Y253F}$ compared to $1IEP_{WT.}$

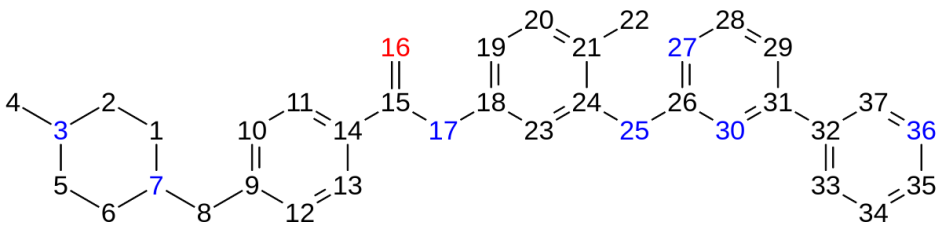


Figure 7.4.7: Index of atoms in imatinib ($1IEP_{Y253F}$)

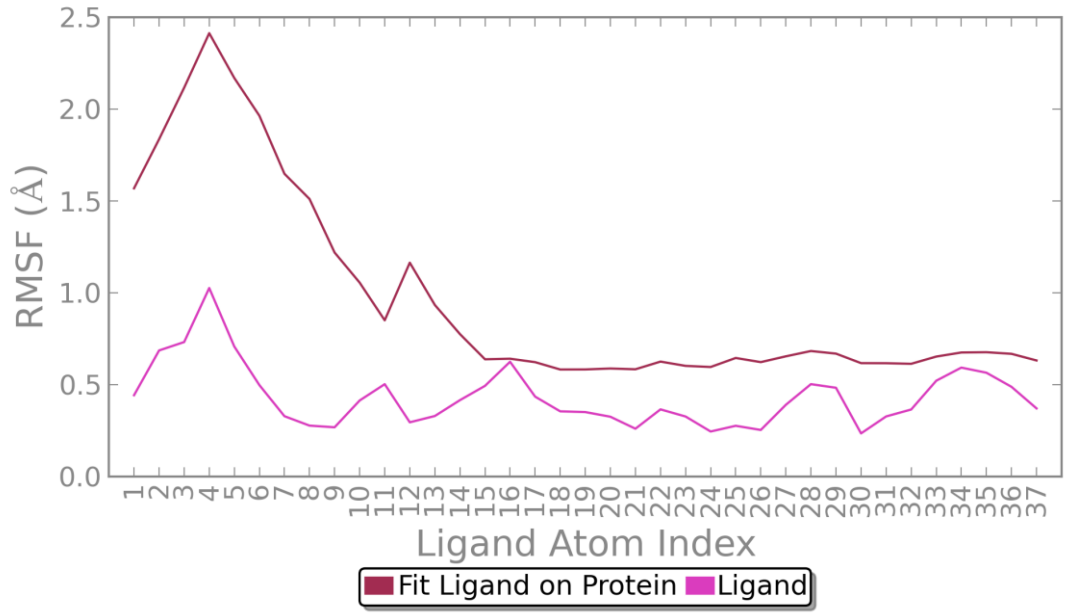


Figure 7.4.8: RMSF of imatinib (1IEP$_{Y253F}$)

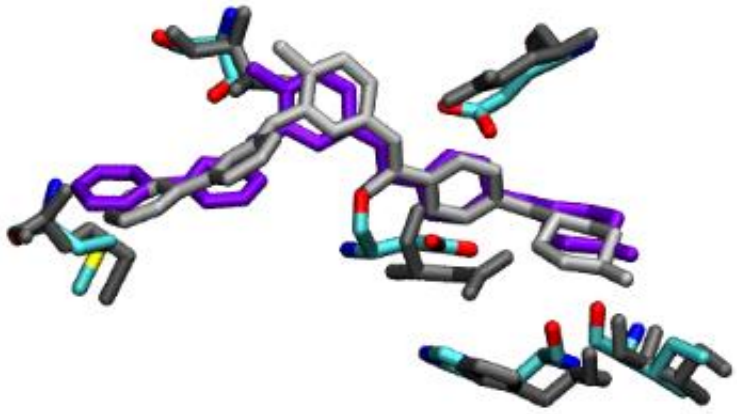


Figure 7.4.9: Comparison of imatinib binding site in the WT and Y253F mutant structures; grey/silver, WT; purple/cyan, Y253F (10 ns)

Protein-ligand interactions

As expected, the substitution of a phenylalanine residue at position 253 resulted in a loss of interaction between the protein and the ligand at this position (Phe253 is Phe29 in Figure 7.4.10).

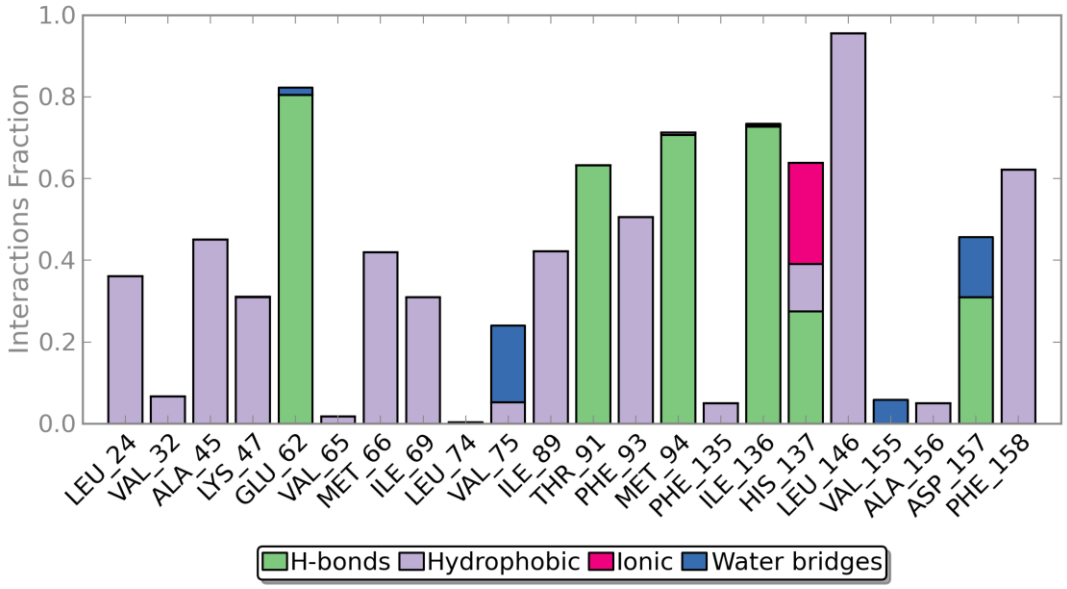


Figure 7.4.10: Protein -ligand interactions by residue (1IEP$_{Y253F}$)

The disruption to the preferred binding mode of imatinib resulted in the reduced hydrogen bond interactions with Asp 381 on the DFG-loop (Asp 157 in Figure 7.4.10).

In summary, the analysis of the 1IEP$_{Y253F}$ simulation suggests that the loss of interaction disrupts the preferred induced fit of imatinib, causing the ligand to shift to a suboptimal position in the pocket. The decrease in protein-ligand interactions at the P-loop and central section (near the amide CO) of the binding site increased the mobility of the residues at these points, resulting in different conformations of the P-loop and DFG-loop.

7.4.2 Nilotinib

Structural stability and flexibility of the protein

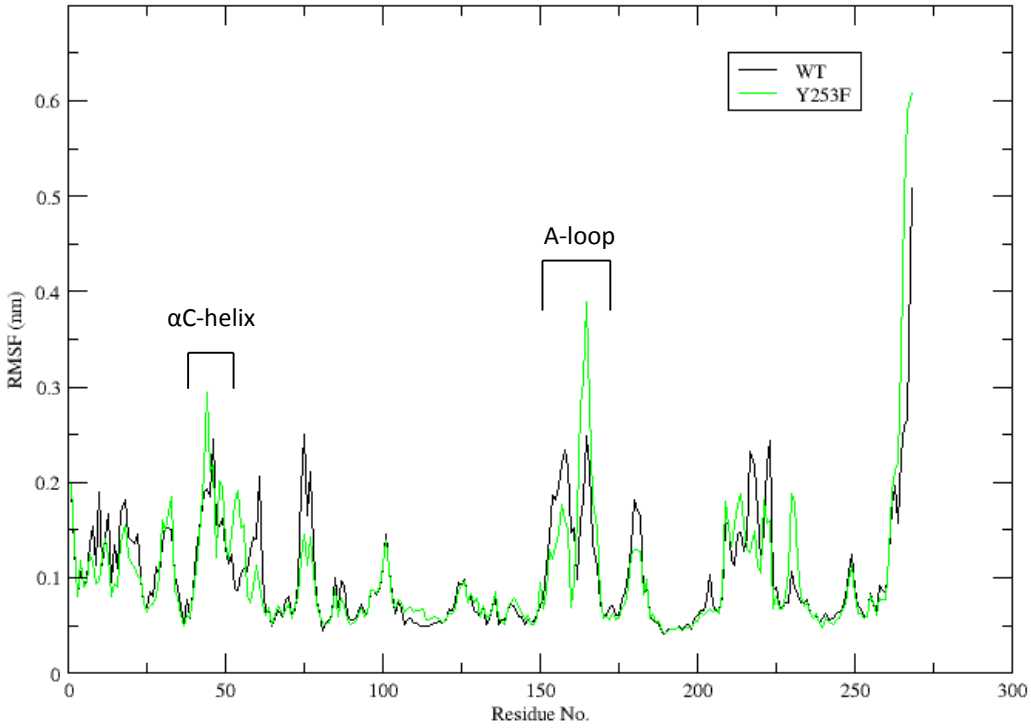


Figure 7.4.11: RMSF results for 3CS9$_{WT}$ and 3CS9$_{Y253F}$

The RMSF profile for 3CS9$_{Y253F}$ displayed increased mobility in the region of the αC-helix (residues 40-58 in Figure 7.4.11) and the A-loop (residues 149-170 in Figure 7.4.11) compared to the wild-type structure, which was supported by the results of the PCA (Figure 7.4.12).

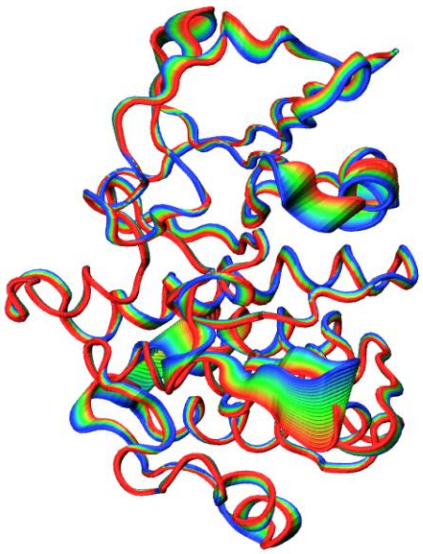


Figure 7.4.12: Principal component analysis for 3CS9$_{Y253F}$ (PC1, projected on structure)

Fluctuations in the other regions in the kinase, including the P-loop, were comparable to that of the wild-type; RMSD analysis of the P-loop (Figure 7.4.13) showed that although there was an initial increase in mobility, the fluctuations for the latter half of the simulation were comparable to the simulation of $3CS9_{WT.}$

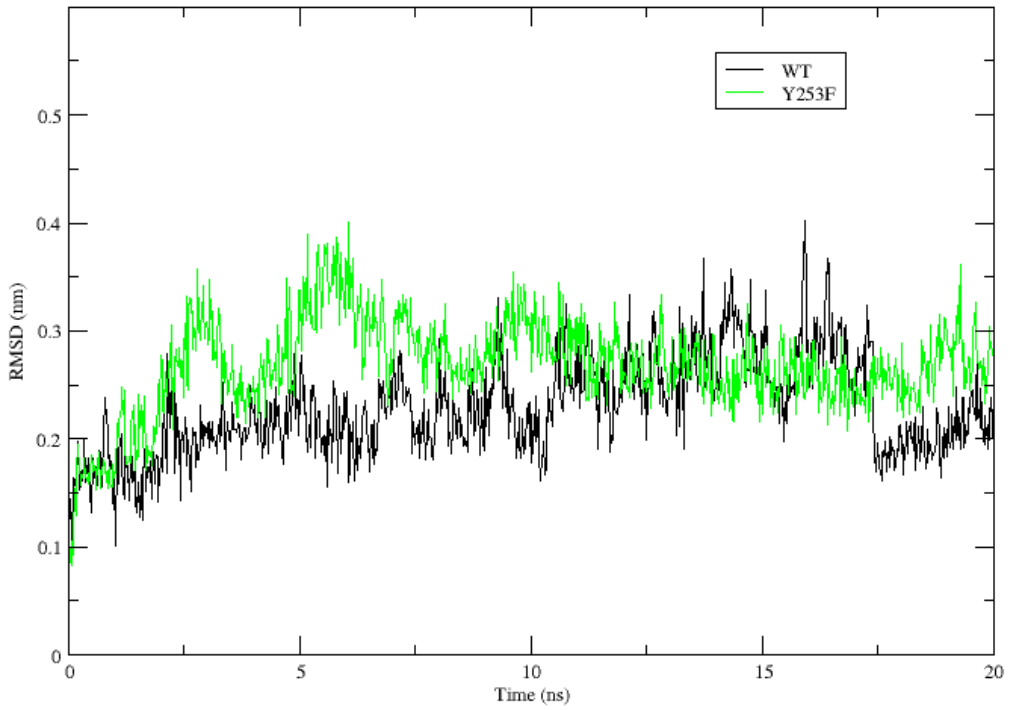


Figure 7.4.13: RMSD of the P-loop ($3CS9_{WT}$ and $3CS9_{Y253F}$)

The average distance between Phe382 and neighbouring R-spine residues Met290 and His361 decreased was slightly shorter in the $3CS9_{Y253F}$ simulation; however the spine maintained the inactive misaligned conformation throughout the simulation.

R-spine residues	Mean distance (CA-CA, Å)	
	3CS9_WT (Å)	3CS9_Y253F (Å)
Leu301 – Met290	6.500	6.871
Met290 – Phe382	12.874	11.282
Phe382 – His361	11.514	10.807

Table 7.4.3: Mean distance between R-spine residues

Similar to the $3CS9_{F317L}$ simulation, the hydrogen bond between Tyr393 and Asp363 was well conserved throughout the simulation, and appeared to be more stable in the Y253F simulation than the WT. The Lys271- Glu286 salt bridge was less stable than in the WT simulation, this is

due to the increased movement of the αC-helix in $3CS9_{Y253F}$ which results in the breaking of the hydrogen bond between Lys271 and Glu286.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$3CS9_{Y253F}$	$3CS9_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	98.00	97.70
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	34.77	82.32
		Glu286 (S)	Lys271 (S)	0.80	-
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	10.39	19.08
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	98.40	99.60
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	79.62	85.71
		His361(S)	Asp421(S)	17.68	32.97

Table 7.4.4: Hydrogen bond analysis: $3CS9_{Y253F}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Review of the simulation showed that Glu286 does form a hydrogen bond with Arg386 when the αC-helix is in partially rotated conformation; however the helical conformation the helix becomes destabilised during the simulation and the αC-helix further rotates out of the binding site, in a shift towards the intermediate Src conformation (Figure 7.4.14). Arg386 cannot maintain the hydrogen bond with Glu286 beyond the partially rotated conformation of the αC-helix (Figure 7.4.15).

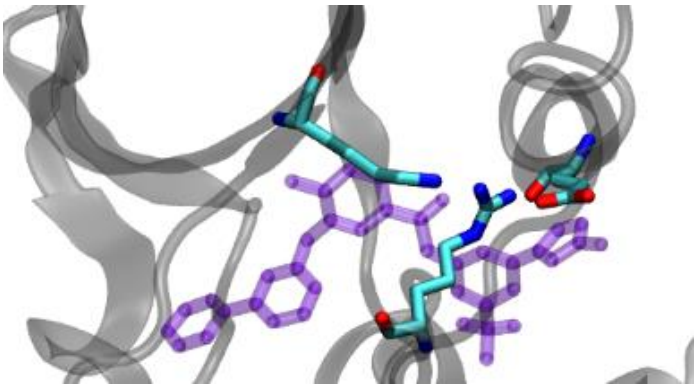


Figure 7.4.14: Partial rotation of the C-helix in 3CS9$_{Y253F}$

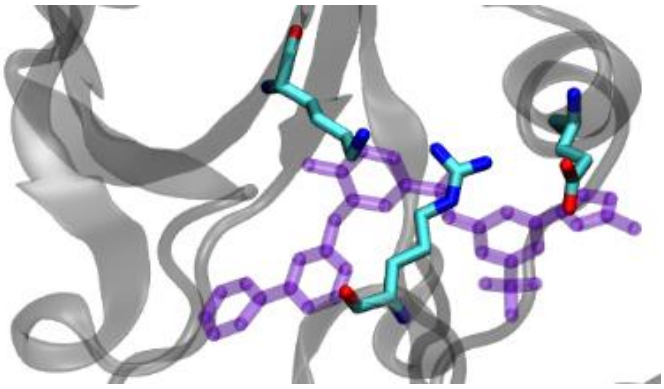


Figure 7.4.15: Full rotation of the C-helix in 3CS9$_{Y253F}$

Stability and flexibility of nilotinib

Nilotinib demonstrated stable motions within the pocket, with a marginally higher RMSD compared to the WT structure (Figure 7.4.16).

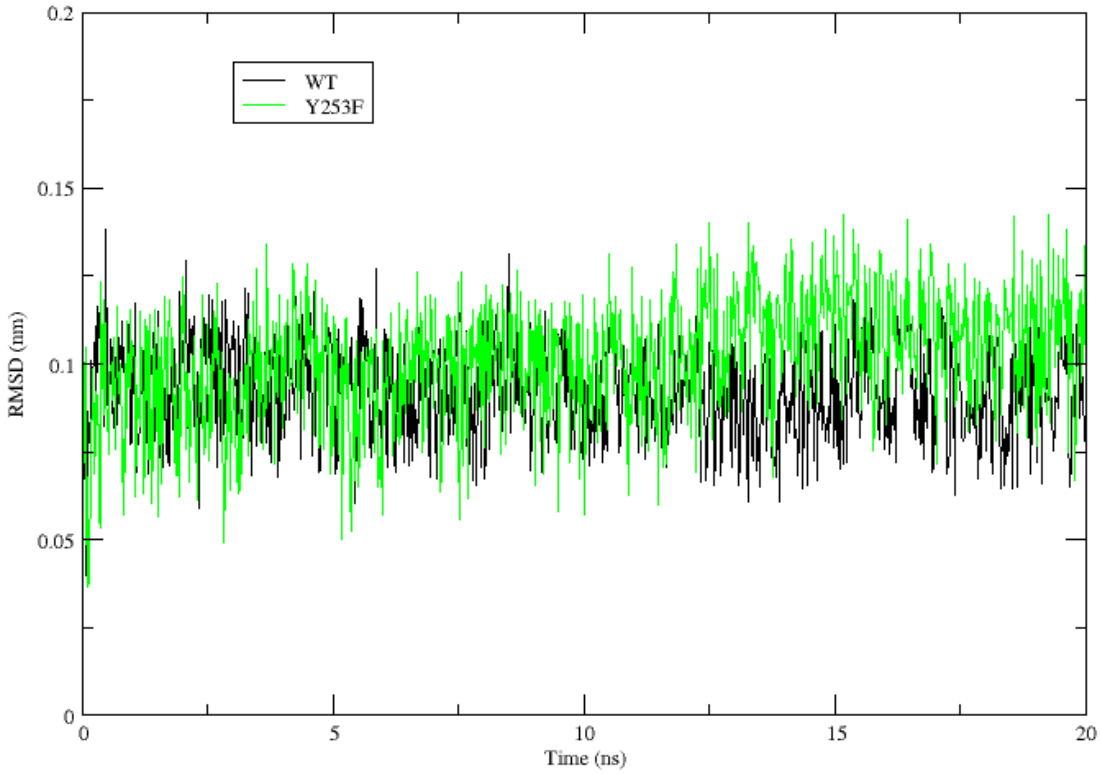


Figure 7.4.16: RMSD for nilotinib (3CS9$_{Y253F}$)

RMSF analysis confirmed that fluctuations of the ligand were in line with that of the overall protein structure (Figure 7.4.17 and Figure 7.4.18) and that the largest fluctuations were located at the trifluoromethyl and imidazole groups at the entrance to the binding site.

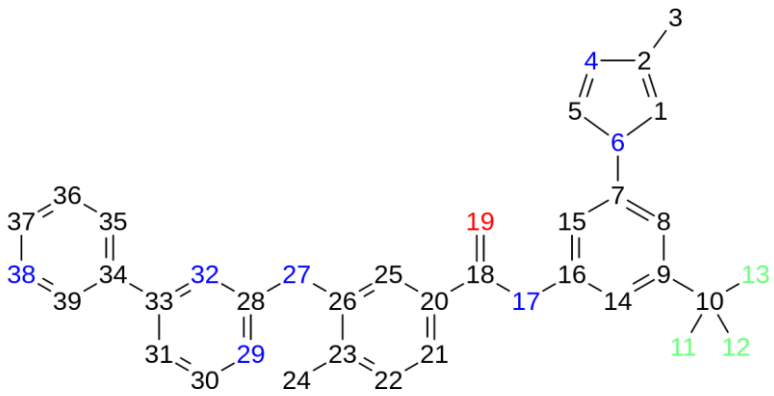


Figure 7.4.17: Index of atoms for nilotinib (3CS9$_{Y253F}$)

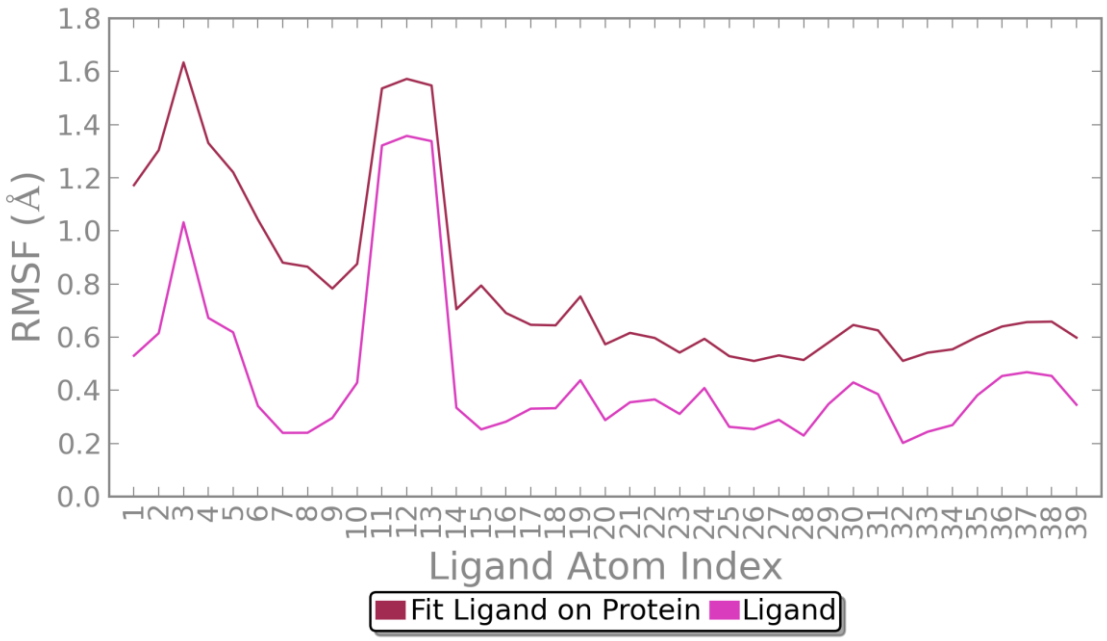


Figure 7.4.18: RMSF of nilotinib (3CS9$_{Y253F}$)

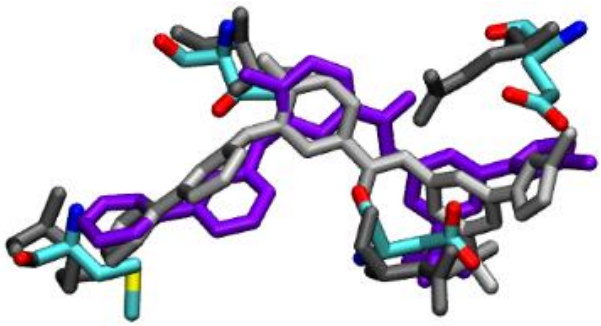


Figure 7.4.19: Comparison of nilotinib binding site in the WT and Y253F mutant structures; grey/silver, WT; purple/cyan, Y253F (15 ns)

Protein-ligand interactions

Figure 7.4.19 shows the conformation of the ligand in $3CS9_{Y253F}$ compared to $3CS9_{WT.}$ Akin to the F317L simulation, the nilotinib amide group that binds to the Glu286 residue has rotated, and Glu286 occupies a slightly more distant position in $3CS9_{F317L}$ compared to the WT. This is in contrast to the positions of other binding site residues which are well aligned with the WT simulation. The protein ligand interactions analysis (Figure 7.4.20) confirmed that the replacement of Tyr253 (Tyr21 Figure 7.4.20) in with a phenylalanine results in the loss of interaction with nilotinib. The hydrogen bonds between nilotinib Thr315, Met318 and Asp381 are maintained (displayed as Thr83, Met86, Asp149 in Figure 7.4.20), though diminished with respect to the WT, however due the increase in distance between Glu286 (Glu54 in Figure 7.4.20) and nilotinib, this hydrogen bond is mediated by a water molecule in the Y253F simulation.

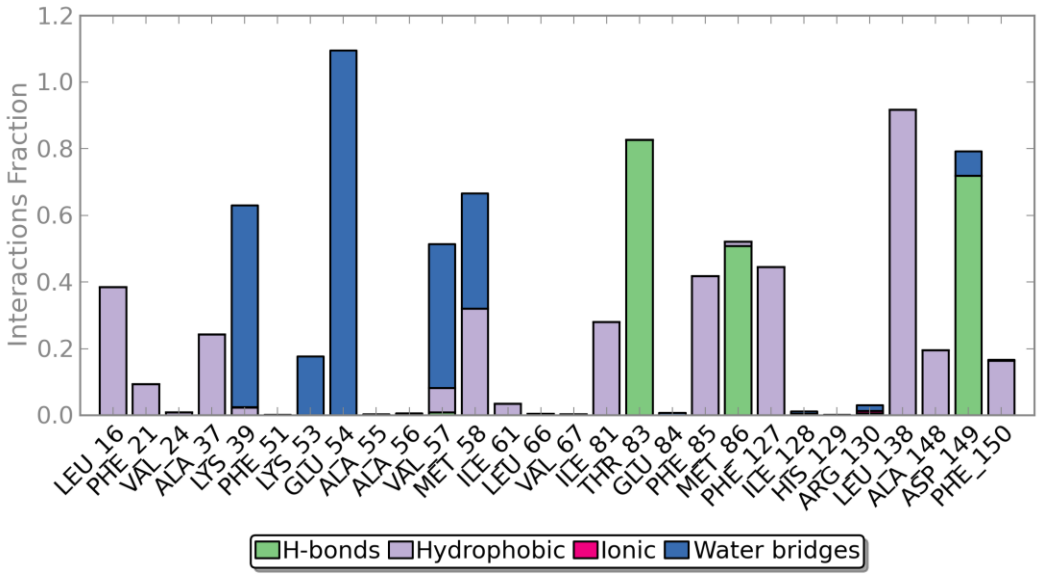


Figure 7.4.20: Protein-ligand interactions by residue for nilotinib ($3CS9_{Y253F}$)

In summary, the analysis of the $3CS9_{Y253F}$ simulation shares striking similarities with the $3CS9_{F317L}$ simulation. The results of both simulations suggest that the loss of interaction at residue position 253 disrupts the preferred binding conformation of nilotinib, causing the ligand to shift to change conformation and the replacement of the Glu286 hydrogen bond with a water bridge. The diminishment of the hydrogen bond between Glu286 and nilotinib enabled the rotation of the αC-helix and subsequent Glu286-Arg386 bond. This likely has a causal effect on the residues surrounding the HRD motif and consequential effects on the flexibility of the A-loop.

As previously stated in section 7.3.2, the interactions involving the pyridinyl and pyrimidinyl rings of nilotinib contribute far less to the overall binding affinity compared in imatinib, which supports data that Y253F and F317L are sensitive to nilotinib inhibition [32, 262].

7.5 Results: T315I

The residue Thr315 forms a hydrogen bond with the amine linker between the pyrimidine and phenyl rings of imatinib and nilotinib. Mutation of the threonine to isoleucine results in the loss of this hydrogen bonding interaction in addition to steric clashes between the isoleucine residue and ligand [4]. The T315I mutation is strongly resistant to both nilotinib and imatinib.

7.5.1 Imatinib

Stability and flexibility of the protein

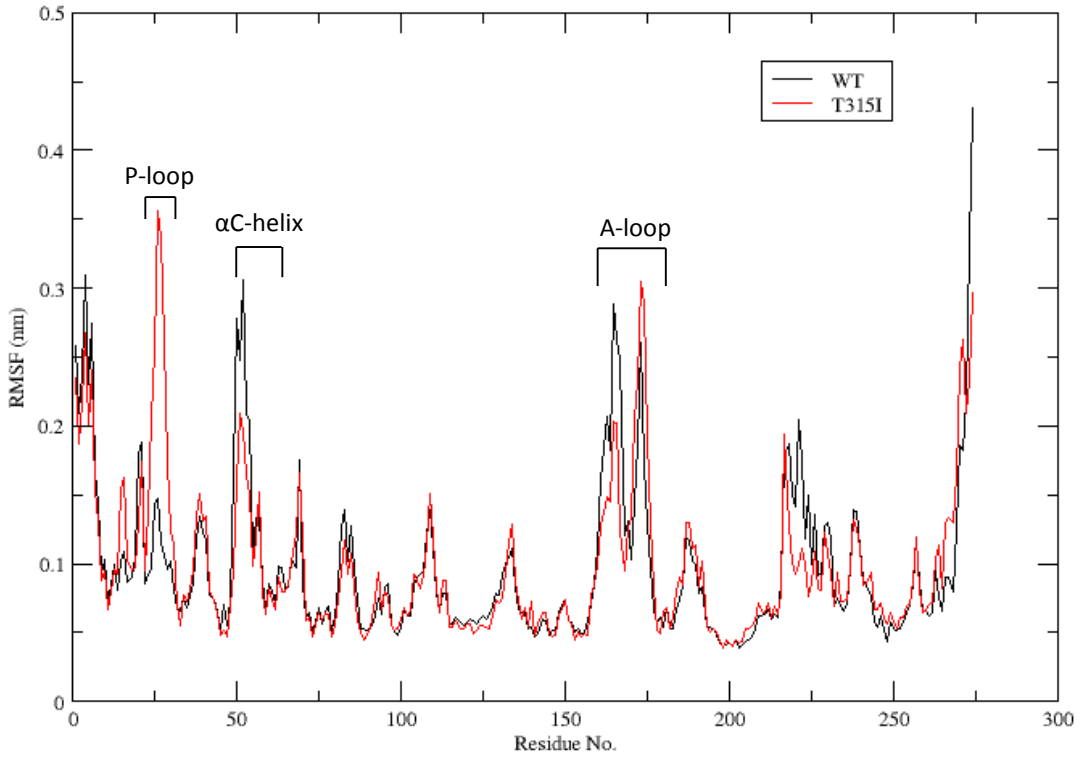


Figure 7.5.1: RMSF results for $1IEP_{WT}$ and $1IEP_{T315I}$

The RMSF profile for $1IEP_{T315I}$ showed a significant increase in the mobility of the P-loop compared to the wild-type structure, however fluctuations across the rest of the protein structure was comparable to that of $1IEP_{WT.}$ The RMSD analysis of the P-loop showed large fluctuations, and on review of the simulation the P-loop could be seen to shift towards the active elongated conformation (Figure 7.5.2 and Figure 7.5.3).

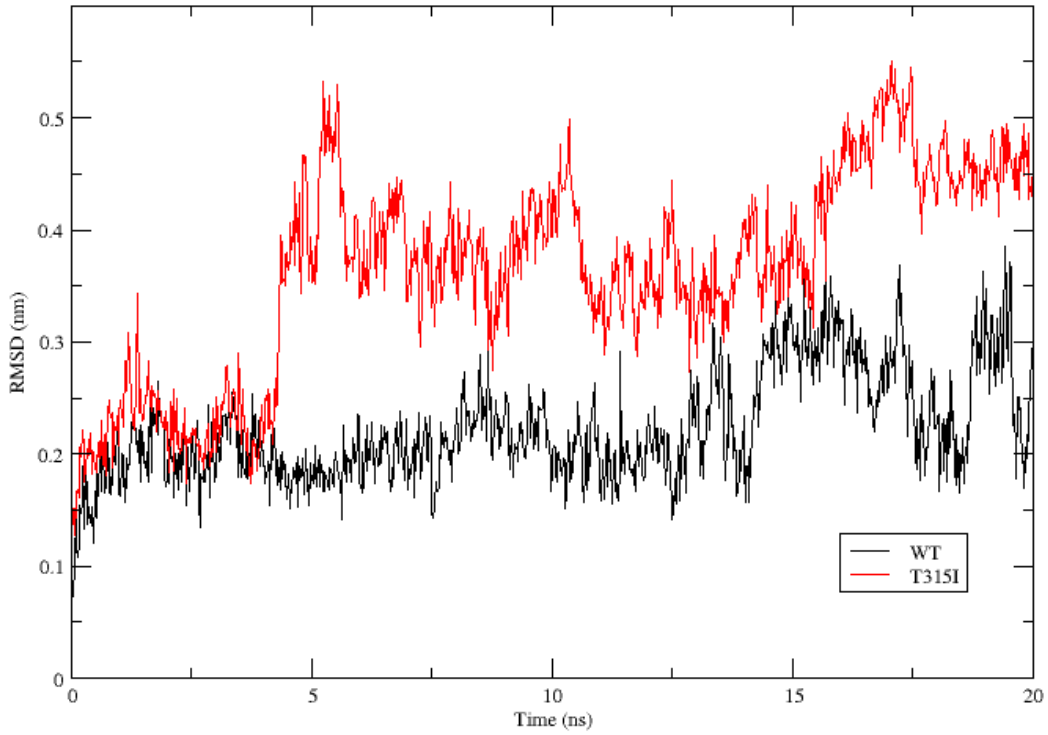


Figure 7.5.2: RMSD of the P-loop ($1IEP_{WT}$ and $1IEP_{T315I}$)

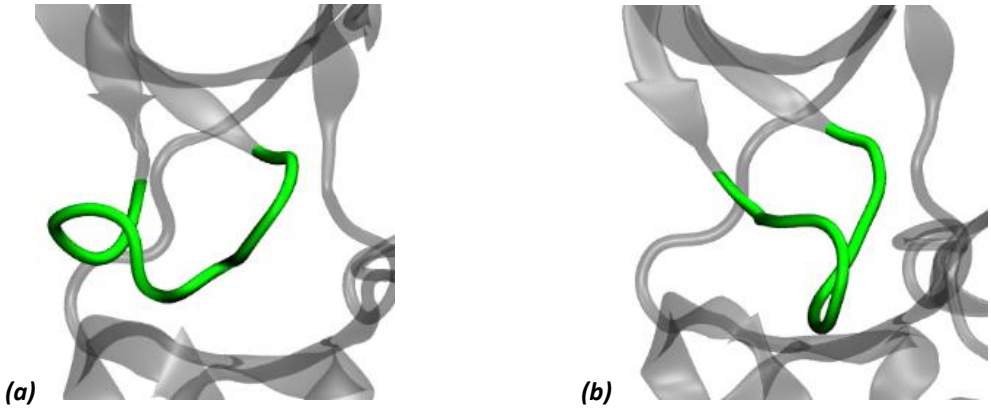


Figure 7.5.3: P-loop conformation: initial (a) and during simulation (b). ($1IEP_{T315I}$)

Fluctuations around the Phe382 residue resulted in small changes in the distance between Ph382 and neighbouring residues of the R-spine, however the Phe382 residue did not experience a significant change in conformation and the R-spine remained misaligned for the duration of the simulation (Table 7.5.1: Mean distance between R-spine residues).

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the T315I mutated structure. Reduced stability was observed in the Lys271-Glu286 salt bridge, but the hydrogen bond was conserved throughout the simulation (Table 7.5.2).

R-spine residues	Mean distance (CA-CA, Å)	
	1IEP_WT (Å)	1IEP_T315I (Å)
Leu301 – Met290	6.487	6.481
Met290 – Phe382	12.284	13.886
Phe382 – His361	12.302	11.891

Table 7.5.1: Mean distance between R-spine residues

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{T315I}$	$1IEP_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	97.03	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	67.53	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	64.64	45.25
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.80	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	90.41	80.62
		His361(S)	Asp421(S)	14.69	33.47

Table 7.5.2: Hydrogen bond analysis: $1IEP_{T315I}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Stability and flexibility of imatinib

The RMSD of imatinib in the T315I mutated structure was slightly elevated compared to the WT simulation (Figure 7.5.4). The RSMF analysis (Figure 7.5.5 and Figure 7.5.6) showed that the fluctuations of the ligand were in line with the fluctuations of the overall protein structure and that the piperazine ring (at the mouth of the binding site) was the most mobile part of the ligand. The movement of the ligand in the binding site is further illustrated in Figure 7.5.7, which shows the conformation of the ligand in $1IEP_{T315I}$ compared to $1IEP_{WT}$.

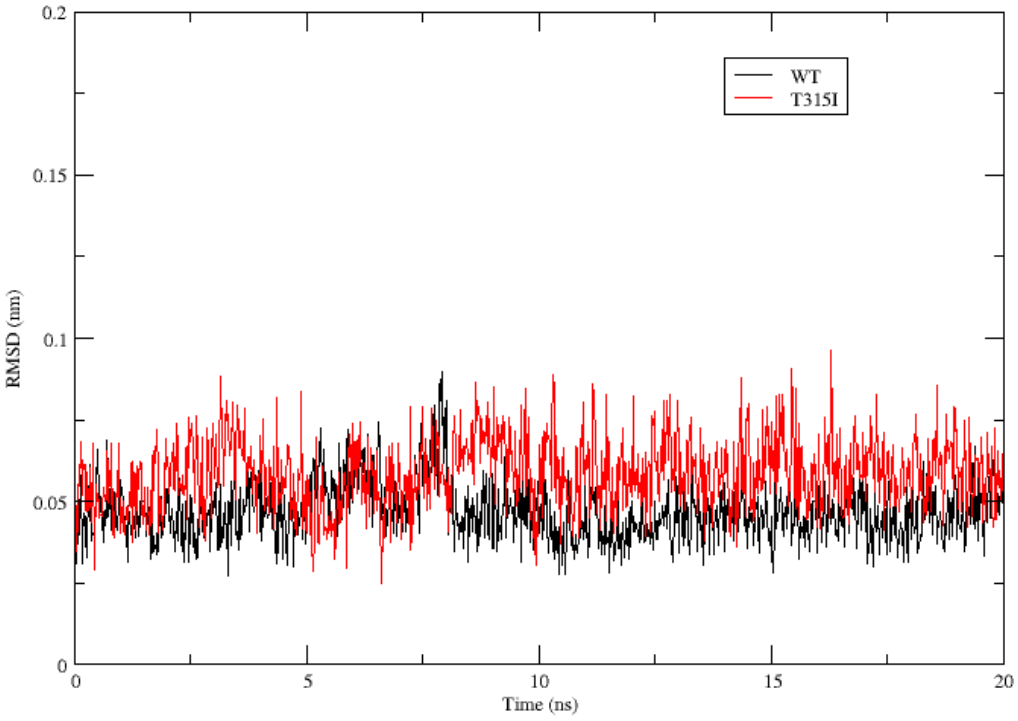


Figure 7.5.4: RMSD of imatinib (1IEP$_{T315I}$)

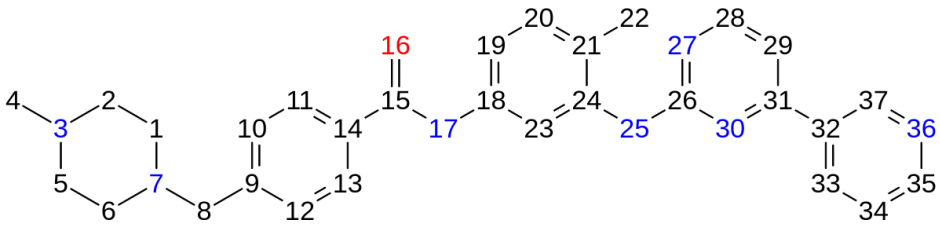


Figure 7.5.5: Index of atoms in imatinib (1IEP$_{T315I}$)

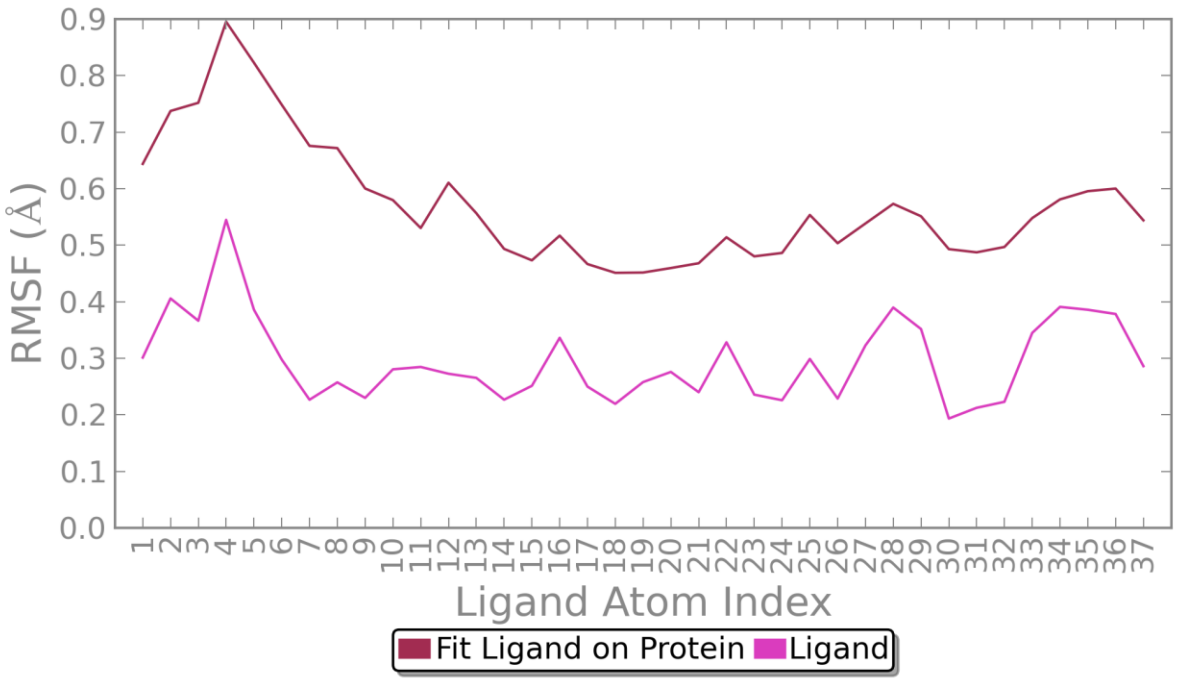


Figure 7.5.6: RMSF of imatinib (1IEP$_{T315I}$)

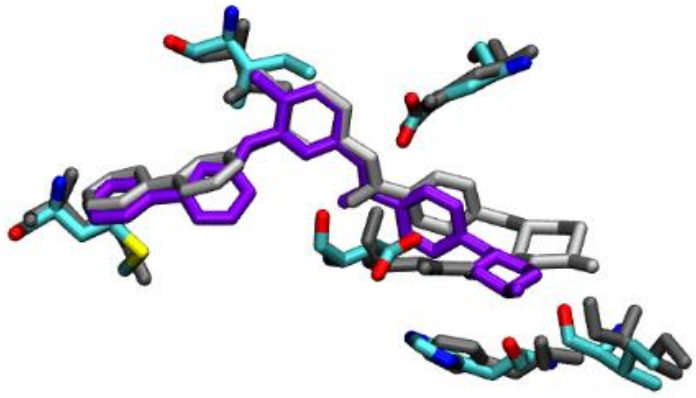


Figure 7.5.7: Comparison of imatinib binding site in the WT and T315I mutant structures; grey/silver, WT; purple/cyan, T315I (10 ns)

Protein-ligand interactions

As expected, the substitution of an isoleucine residue at position 315 resulted in a loss of a hydrogen bond between the protein and the ligand at this position. The disruption to the preferred binding mode of imatinib further resulted in the loss of the hydrogen bond with Ile360

(Ile136 in Figure 7.5.8) at the entrance to the binding site, a loss of interactions with Tyr253 (Tyr29 in Figure 7.5.8), and reduced hydrogen bond interactions with Asp381 (Asp157 in Figure 7.5.8) on the DFG-loop.

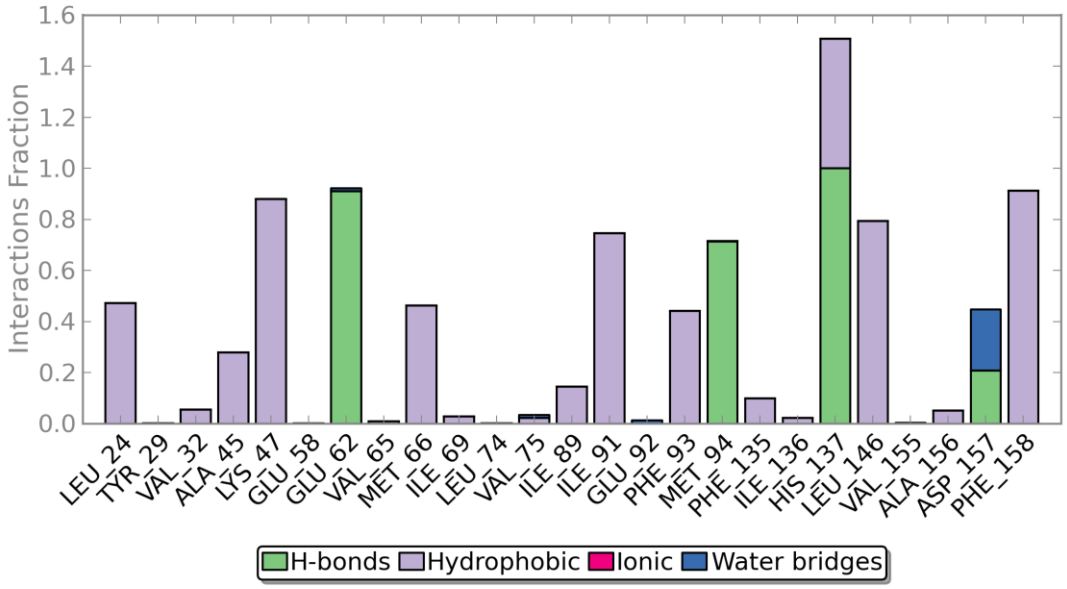


Figure 7.5.8: Protein-ligand interactions by residue (1IEP$_{T315I}$)

In summary, the 1IEP$_{T315}$ simulation suggests that the loss of the hydrogen bond with residue 315 and slightly larger size of isoleucine compared to threonine, disrupts the preferred induced fit of imatinib, causing the ligand to shift to a suboptimal position in the pocket. The decrease in protein-ligand interactions at residue 253 (Tyr29 in Figure 7.5.8) on the P-loop and central section (near the amide CO) of the binding site increased the mobility of the residues at these points, resulting in the intermediate elongated conformation of the P-loop.

7.5.2 Nilotinib

Stability and flexibility of protein

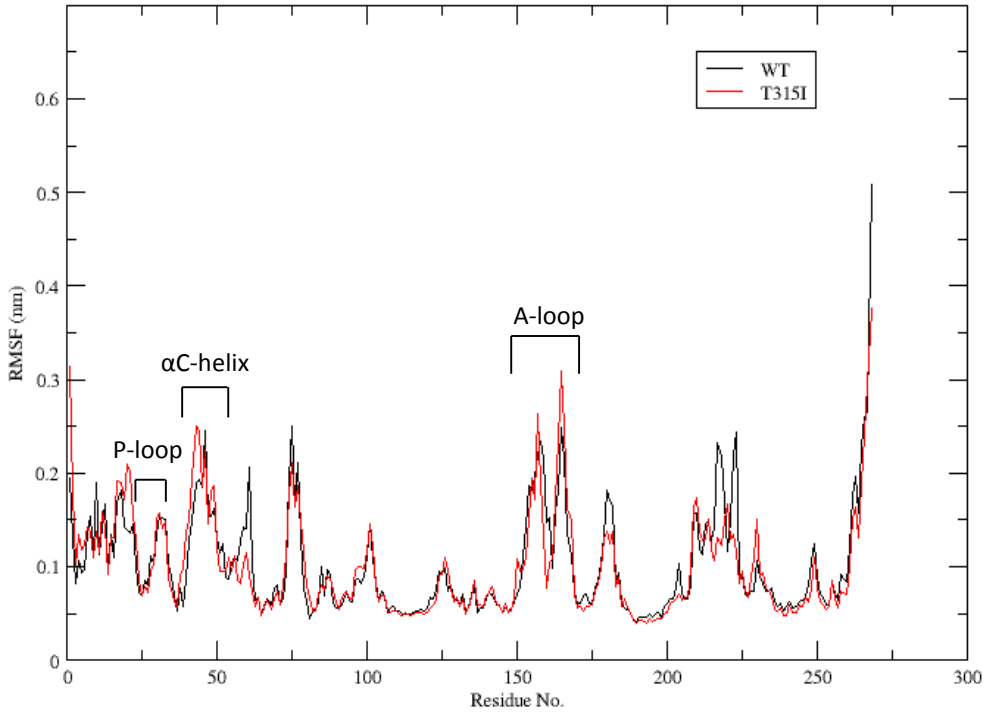


Figure 7.5.9: RMSF results for $3CS9_{WT}$ and $3CS9_{T315I}$

The RMSF results for $3CS9_{T315I}$ showed a small increase in the mobility of the P-loop, αC-helix and second half of the A-loop; fluctuations in the other regions in the kinase were comparable to that of the wild-type (Figure 7.5.9). The PCA (Figure 7.5.11) supported these results, showing major motions in the A-loop and αC-helix.

Regional RMSD analysis revealed fluctuations in the conformation of the DFG loop (Figure 7.5.10); the DFG-loop was highly stable in the other nilotinib simulations and increased motion in this region can signify a shift in the inactive state. Observational analysis revealed that although less stable, there was no distinct change in conformation.

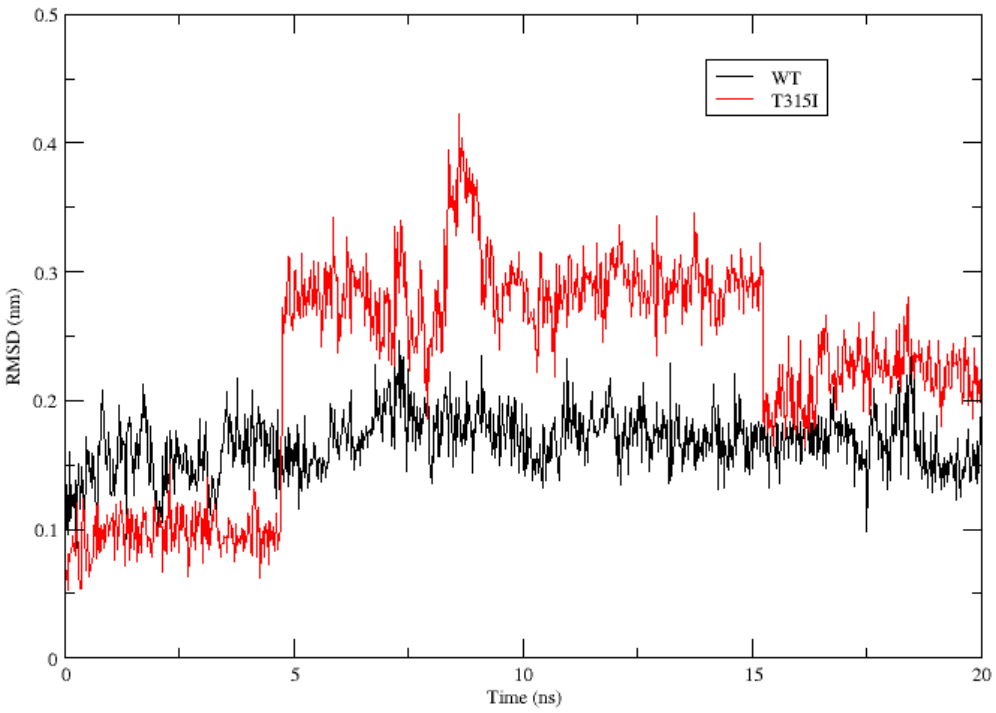


Figure 7.5.10: RMSD of the DFG-loop ($3CS9_{WT}$ and $3CS9_{T315I}$)

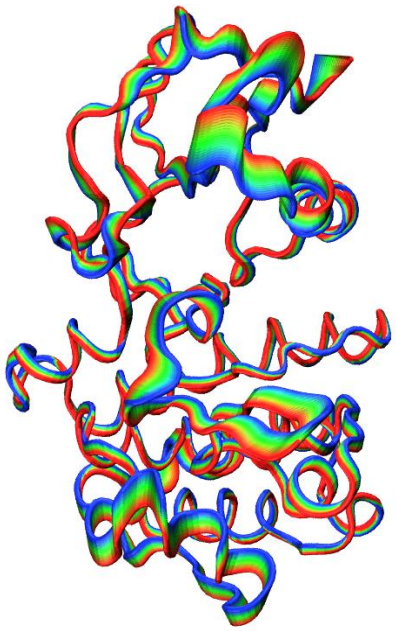


Figure 7.5.11: Principal component analysis for $3CS9_{Y253F}$ (PC1, projected on structure)

The observational analysis of the DFG loop was supported by the R-spine analysis, which showed that the residues maintained similar positions to that of the WT structure. This suggesting that despite the fluctuations in the DFG loop the T315I had minimal impact to the position of these residues (Table 7.5.3).

R-spine residues	Mean distance (CA-CA, Å)	
	3CS9_WT	3CS9_T315I
Leu301 – Met290	6.500	6.258
Met290 – Phe382	12.874	12.966
Phe382 – His361	11.514	10.965

Table 7.5.3: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the T315I simulation; reduced stability was observed in the Lys271-Glu286 salt bridge, but the hydrogen bond was conserved throughout the simulation (Table 7.5.4)

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$3CS9_{T315I}$	$3CS9_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	98.90	97.70
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	64.14	82.32
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	71.13	19.08
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.40	99.60
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	83.42	85.71
		His361(S)	Asp421(S)	20.38	32.97

Table 7.5.4: Hydrogen bond analysis: $3CS9_{T315I}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Review of the simulation showed that Glu286 does form a hydrogen bond with Arg386 when the αC-helix is in partially rotated conformation, but only for a small percentage of the simulation.

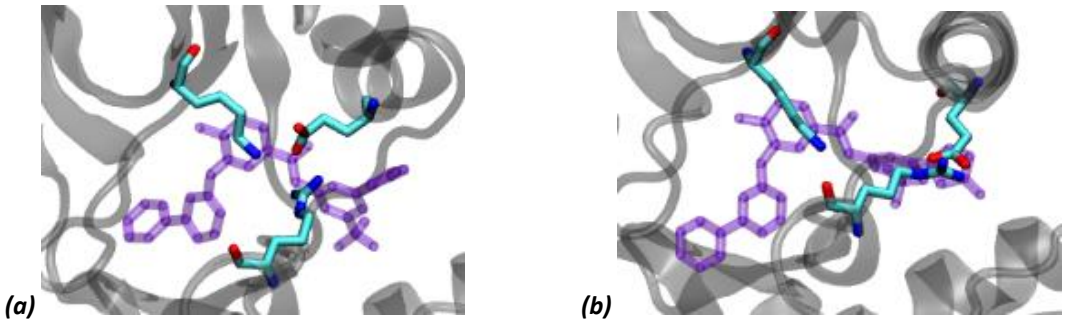


Figure 7.5.12: (a) interactions between Lys271, Glu286 and Arg386; (b) Glu286-Arg386 bond formation (3CS9$_{T315I}$)

Stability and flexibility of nilotinib

The RMSD of the ligand was initially lower than in the WT simulation but rapidly increased by approximately 1 Å halfway through the simulation (Figure 7.5.13).

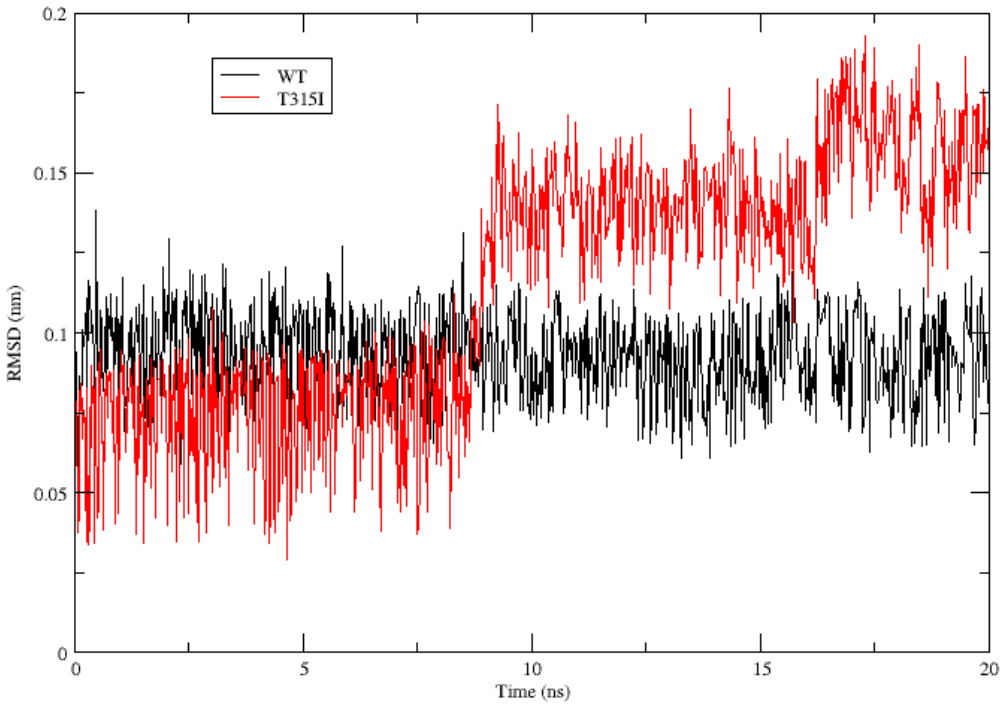


Figure 7.5.13: RMSD of nilotinib (3CS9$_{T315I}$)

The RSMF analysis shows that nilotinib has a poor fit in the binding site; the fluctuations of the trifluoromethyl and imidazole groups appear to be clashing against the protein and the pyridine ring is distal from the hinge region.

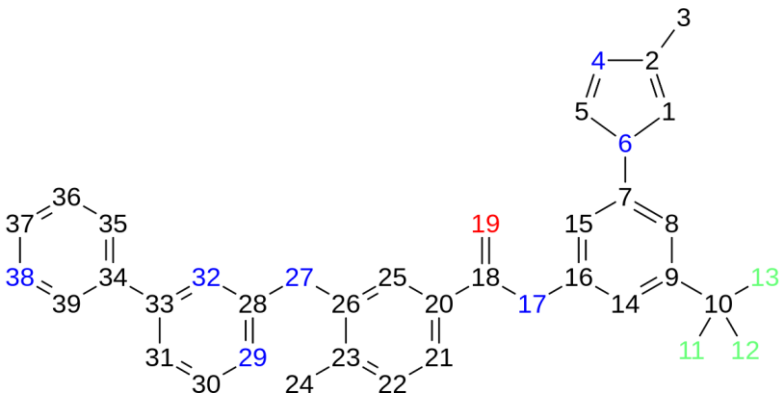


Figure 7.5.14: Index of atoms for nilotinib (3CS9$_{T315I}$)

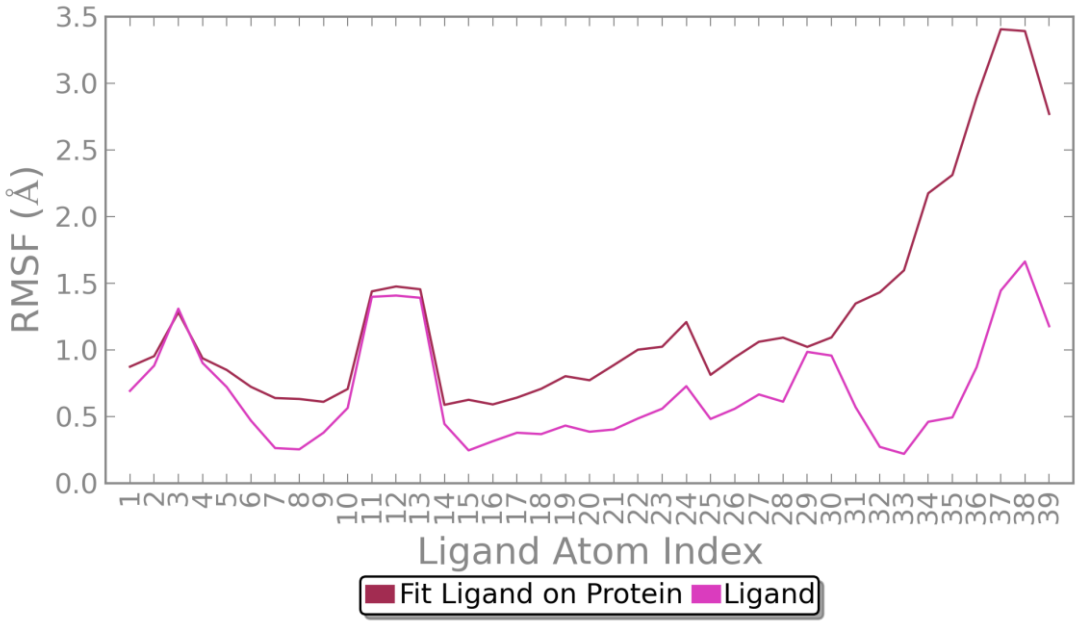


Figure 7.5.15: RMSF of nilotinib (3CS9$_{T315I}$)

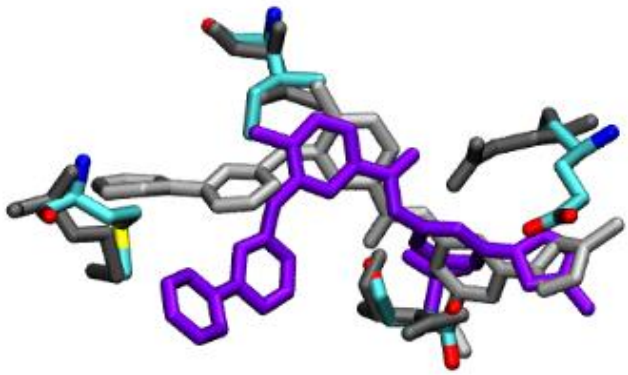


Figure 7.5.16: Comparison of nilotinib binding site in the WT and T315I mutant structures; grey/silver, WT; purple/cyan, T315I (17 ns)

Protein-ligand interactions

As expected, the substitution of an isoleucine residue at position 315 resulted in a loss of a hydrogen bond between the protein and the ligand at this position. The loss of this hydrogen bond and the larger size of isoleucine resulted in nilotinib shifting position in the binding site, which further resulted in the loss of the hydrogen bond with Glu286 (replaced by a water bridge, Glu54 in Figure 7.5.17) and a less stable hydrogen bond with Met318 (Met86 in Figure 7.5.17).

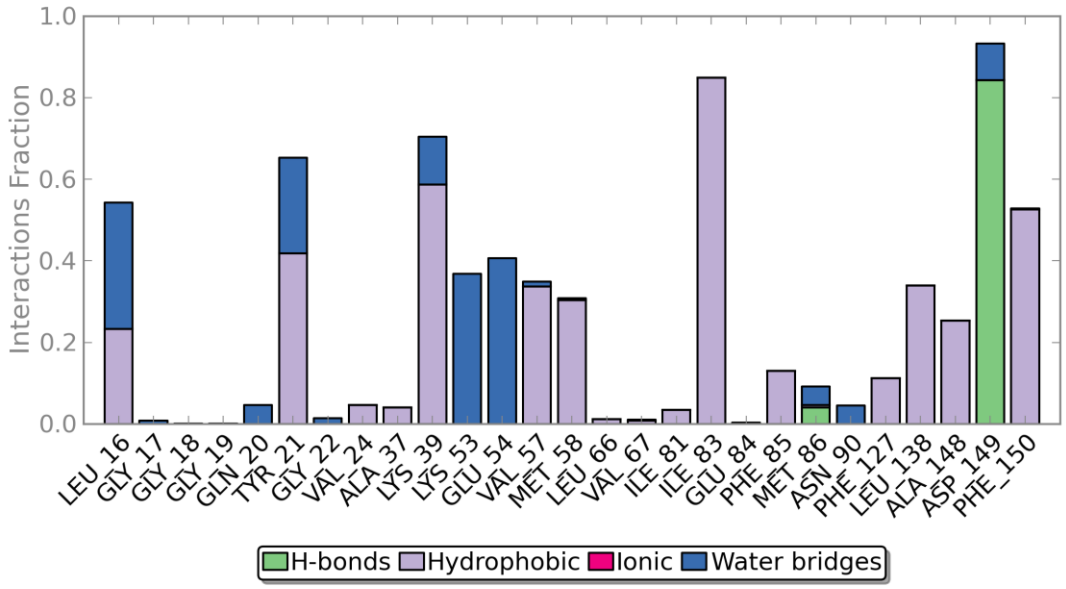


Figure 7.5.17: Protein-ligand interactions by residue for nilotinib (3CS9$_{T315I}$)

In summary, the analysis of the 3CS9$_{T315}$ simulation suggests that the loss of the hydrogen bond with residue 315 and slightly larger size of isoleucine compared to threonine, causes nilotinib to change position in the pocket in order to accommodate the different conformation of the binding site. In the new position nilotinib can only maintain stable hydrogen bonding with Asp381 and hydrogen bonds with the other binding site residues are either lost or diminished. The new binding position of nilotinib allows for interactions with the P-loop and residues surrounding the αC-helix, stabilising these motifs; there were fewer fluctuations in these regions in the T315I simulation compared to the Y253F and F317L simulation.

7.6 Summary of molecular dynamics study

The MD simulations enabled insight into the internal dynamics of the mutated kinase structures; both within the binding site and in conformationally important regions such as the P-loop and αC-helix.

The analysis of the $1IEP_{F317L}$ and $1IEP_{Y253F}$ simulations suggest that the loss of interaction disrupts the preferred induced fit of imatinib, causing the ligand to shift to a suboptimal position in the pocket, and the further loss of interactions and translates to an increase in the flexibility of the P-loop and A-loop. These results support the concept that imatinib resistance is conferred due to the inability to achieve the preferred induced fit.

The analysis of the $3CS9_{F317L}$ and $3CS9_{F317L}$ simulations also suggest that the loss of interaction disrupts the preferred binding of nilotinib, causing nilotinib to shift within the pocket and the diminishment of the hydrogen bond between Glu286 and nilotinib. Without stable interactions with the ligand the αC-helix was able to rotate and form the Glu286-Arg386 bond associated with the intermediate conformation. The formation of the Glu-Arg bond and the loss of interaction at the hinge region had a causal effect on the flexibility of the HRD motif, A-loop and P-loop. Given that nilotinib maintains sensitivity to F317L and Y253F, the diminishment of the hydrogen bond between nilotinib and Glu286 supports the concept that that interactions involving the pyridinyl and pyrimidinyl rings of nilotinib contribute far less to the overall binding affinity.

The analysis of the $1IEP_{T315}$ and $3CS9_{T315}$ simulations indicate that the loss of the hydrogen bond with residue 315 and slightly larger size of isoleucine compared to threonine, causing the ligand to shift to a suboptimal position in the pocket. In the case $1IEP_{T315I}$ imatinib moved away from the hinge region, increasing the mobility of the P-loop, resulting in a shift towards the more elongated conformation of the P-loop associated with active Abl. In the $3CS9_{T315}$ simulation, the new position nilotinib allowed stable hydrogen bonding with Asp381 and interactions with the P-loop and residues surrounding the αC-helix, stabilising these motifs; however hydrogen bonds with the other binding site residues were either lost or diminished. This supports the expectation that the T315I gatekeeper causes a significant change to the conformation of the imatinib/nilotinib binding site which impedes binding of both imatinib and nilotinib. As mentioned in the introduction to this chapter, it would be valuable to repeat this study with ponatinib in order to diversify the results and include an example of a TKI that is capable of inhibiting the gatekeeper mutation.

7.7 Ensemble docking

7.7.1 Methodology

Imatinib and nilotinib were docked to the snapshot conformations of $1IEP_{F317L}$, $1IEP_{Y253F}$, $1IEP_{T315I}$, $3CS9_{F317L}$, $3CS9_{Y253F}$, and $3CS9_{T315I}$.

Docking simulations were completed as described in section 6.7.1.

7.7.2 Results

Table 7.7.1 and Table 7.7.2 compare the best GlideScores for imatinib docked to 1IEP and 3CS9 (WT, F317L, Y253F and T315I) to what is known experimentally. Experimental data ranks the mutants as follows: WT < F317L < Y253F < T315I.

The results of the ensemble docking study using the 3CS9 snapshots as receptors correctly ranked the mutations according to experimental data (Table 7.7.2); however the docking scores from the 1IEP receptor study were not aligned with experimental data and the mutations were not correctly ranked in order of binding affinity as expected (Table 7.7.1).

	Experimental sensitivity	Experimental rank	GlideScore (1IEP)	GlideScore rank
WT		1	-13.40	3
F317L		2	-14.93	1
Y253F		3	-12.12	4
T315I		4	-13.96	2

Table 7.7.1: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental sensitivity	Experimental rank	GlideScore (3CS9)	GlideScore rank
WT		1	-14.78	1
F317L		2	-13.07	2
Y253F		3	-12.86	3
T315I		4	-11.71	4

Table 7.7.2: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Table 7.7.3 and Table 7.7.4 compare the best GlideScores for nilotinib docked to the 1IEP and 3CS9 structures to what is known experimentally. As with imatinib, the results should have correctly ranked the structures: WT < F317L < Y253F < T315I. Furthermore the docking scores for nilotinib should have been better (more negative) than those for imatinib.

The results of the ensemble docking study using the 3CS9 snapshots as receptors correctly ranked the mutations according to experimental data and ranked nilotinib as a more potent inhibitor compared to imatinib (Table 7.7.4); however the docking scores from the 1IEP receptor study were not aligned with experimental data and the mutations were not correctly ranked in order of binding affinity as expected (Table 7.7.3).

	Experimental sensitivity	Experimental rank	GlideScore	GlideScore rank
WT		1	-13.78	3
F317L		2	-14.50	1
Y253F		3	-13.68	4
T315I		4	-14.17	2

Table 7.7.3: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental sensitivity	Experimental rank	GlideScore	GlideScore Rank
WT		1	-15.41	1
F317L		2	-14.22	2
Y253F		3	-14.19	3
T315I		4	-12.97	4

Table 7.7.4: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

7.7.3 Discussion

Given the MD simulations of 1IEP and 3CS9 had shown the two protein structures to be very similar, the inconsistency in the 1IEP docking results was unexpected. The reasons behind this could be either due to poor performance from the docking software Glide or an error in the simulation. Unfortunately it is difficult to draw conclusions based on only two receptors and ligands; if more time had been available the protocol would ideally have been tested using a large dataset of TKIs and their receptors, and then compared to available experimental binding affinities.

Docking software is often more accurate at ranking and predicting the binding affinity of ligands when there is high correlation between the logarithm of ligand molecular weight and experimental binding affinity [335]. Table 7.7.5 shows that this is not the case for the Abl tyrosine kinase inhibitors; the experimental IC50 values (Zabriskie et al., 2014) and the logarithm of the molecular weight are poorly correlated. This suggests that Glide might not be as efficient at predicting binding affinity for the Abl tyrosine kinase inhibitors as other data sets.

Inhibitor	IC_{50} (nM)	MW (g/mol)	Log MW
Imatinib	768.3	493.60	2.69
Nilotinib	57.0	529.52	2.72
Bosutinib	111.6	530.45	2.72
Dasatinib	2.8	488.01	2.69
Ponatinib	3.9	532.56	2.73

Table 7.7.5: IC_{50} values of Abl TKIs vs. molecular weight. IC_{50} values from Zabriskie et al [331].

7.8 MDGBSA calculations

7.8.1 Methodology

MM-GBSA calculations were performed on the 1IEP and 3CS9 trajectories were completed as described in section 6.8.1.

7.8.2 Results

As with the GlideScores in the docking study, the more negative the MM-GBSA result, the higher the binding affinity of the ligand.

As seen in the docking study, the MM-GBSA results for the 3CS9 structures were reasonably well aligned with experimental data, whereas those for the 1IEP structures were not.

Energy components (kcal/mol)	WT	F317L	Y253F	T315I
Total ligand energy	-144.95	-145.99	-147.91	-147.35
Total receptor energy	-6368.80	-6076.51	-6357.99	-6341.49
Total complex energy	-6586.41	-6317.04	-6580.38	-6563.31
Difference	-72.66	-94.53	-74.48	-74.48

Table 7.8.1: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I) with imatinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-72.66	3
F317L		2	-94.53	1
Y253F		3	-74.48	=2
T315I		4	-74.48	=2

Table 7.8.2: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Energy components (kcal/mol)	WT	F317L	Y253F	T315I
Total ligand energy	-223.99	-221.62	-223.40	-224.63
Total receptor energy	-6413.74	-6422.64	-6380.38	-6372.00
Total complex energy	-6712.63	-6706.84	-6666.73	-6650.87
Difference	-74.90	-62.58	-62.95	-54.25

Table 7.8.3: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I) with nilotinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-74.90	1
F317L		2	-62.58	3
Y253F		3	-62.95	2
T315I		4	-54.25	4

Table 7.8.4: MM-GBSA results for 3CS9 with nilotinib (WT, F317L, Y253F, T315I) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

One notable difference in the 3CS9 MM-GBSA results compared to the ensemble docking study and experimental data is the rankings of F317L and Y253F. These mutations had similar MM-GBSA results but the Y253F mutation was ranked as slightly more sensitive to nilotinib inhibition.

7.8.3 Discussion

Both the docking scores and MM-GBSA results suggested there was an inaccuracy in the 1IEP simulation. As previously mentioned, is it difficult to identify the source of this error without repeating the simulations for a larger dataset; however, one common reason for inaccuracies or unexpected results in docking scores is the exclusion of conserved water molecules [221, 336].

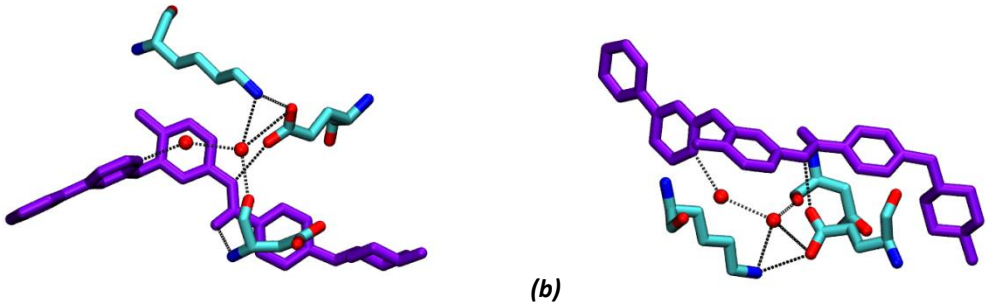
Since the crystal structure of 1IEP contained water molecules within the binding site (whereas the 3CS9 structure did not) this warranted further investigation.

7.9 Water molecule prediction

When preparing a PDB structure it is difficult to decide which, if any, of the crystallographic water molecules should be retained. In the case of 1IEP, previous research was contradictory; in some cases the water molecules were kept and in others they were removed. In this research all the waters in both the 2F4J, 1IEP and 3CS9 crystal structures were removed prior to MD production. At the time of preparation there were few methods of evaluating this decision; however, recently methods have been developed that can reliably determine the positioning of water molecules in the binding site of a protein or protein-ligand complex. The position and interactions of the water molecules in the 1IEP binding site were retrospectively evaluated via a) observational analysis of the simulation b) using the water grid function in PTRAJ (AMBER) to determine the average positions of the water molecules in the binding site and c) completing robust free energy calculations via the statistical thermodynamics approaches FDTI and JAWS [226].

7.9.1 Observational analysis

A review of the crystal structure showed that there were two waters in the binding site; one of the waters, Wat14, was held in place by a tetrahedral hydrogen bonding network between Wat75, Glu286, Lys271 and Asp381.



(a) ***(b)***

Figure 7.9.1: Hydrogen bonding network with W_1 and W_2 in the 1IEP crystal structure; a) underside of binding site b) parallel to binding site

On reviewing the 1IEP simulation, two water molecules (W_1 and W_2) were observed entering the pocket and occupying the binding site for a large part of the simulation. W_1 entered the pocket within the first 2ns of the simulation and formed hydrogen bonds with Glu286, Lys271 and Asp381.

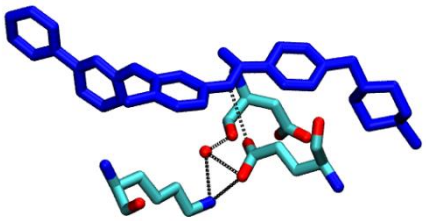
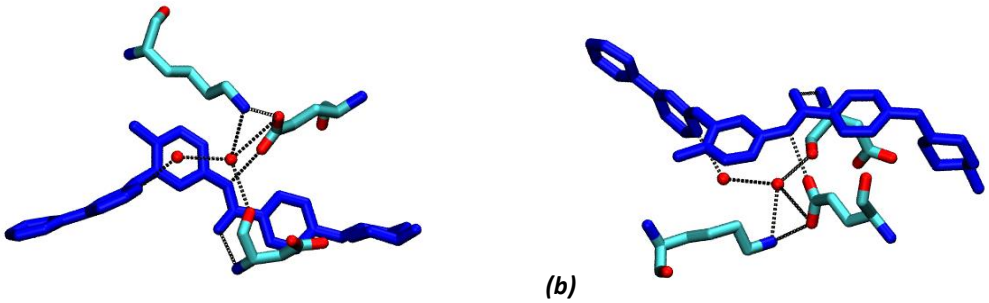


Figure 7.9.2: Hydrogen bonding network with W_1 in the 1IEP WT simulation; underside of binding site

The second water molecule, W_2, entered the binding site approximately halfway through the simulation, displacing W_1 and forming hydrogen bonds with W_1, Glu286, Lys271 and Asp381, as in the crystal structure.



(a) ***(b)***

Figure 7.9.3: Hydrogen bonding network with W1 and W2 in the 1IEP WT simulation; a) underside of binding site b) parallel to binding site

Hydrogen bond analysis (Table 7.9.1) confirmed that this hydrogen bonding network was very stable, especially between the waters, the Glu286 and Asp381 residues and imatinib.

Residues	Stability (% of simulation)
Lys271 – W_1	35.7
Glu286 – W_1	62.2
Asp381 - W_1	67.2
Imatinib - W_1	85.7
Lys271 – W_2	10.4
Glu286 – W_2	65.6
Asp381 – W_2	70.6
W_1 - W_2	22.5

Table 7.9.1: Stability of hydrogen bonding network with W_1 and W_2. 1IEP simulation.

The interactions with the two water molecules likely stabilise the positions of Glu286 and Asp381 for optimal ligand binding.

7.9.2 Water molecule prediction using Amber (PTRAJ)

The grid module in ptraj (part of the AMBER toolkit [162]) creates a grid in X-Plor density format with a count of the specified atoms (e.g. water molecules) in each grid cell for an entire MD simulation. The "average" output shows the most common position of the water molecules throughout the simulation, which gives an indication of the position of hydration sites within the pocket [234]. The figure below shows the p*traj* output for 1IEP, it shows water molecules (in green) clustered around the ligand within the binding site.

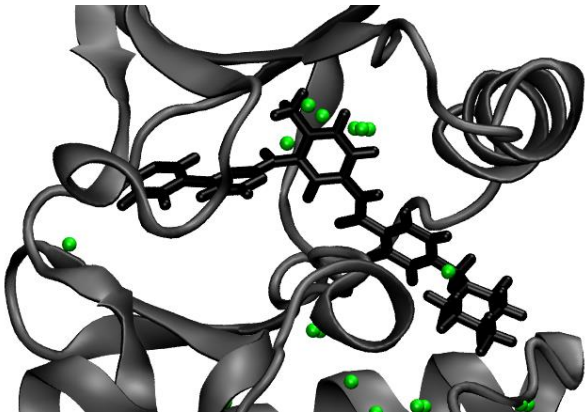


Table 7.9.2: Ptraj demonstrates how waters cluster around imatinib in the 1IEP simulation

In the 3CS9 simulation, although there were water molecules local to the binding site, none appeared to directly interact with the ligand.

7.9.3 Free energy calculations to determine the stability of W282

In order to confirm that W_2 is a conserved and tightly bound water molecule, two free energy simulations were run by another member of the group, Michael Bodnarchuk, to estimate the binding free energy of W_2; double-decoupling FDTI and JAWS.

Double-Decoupling FDTI

Calculations were performed on $W_2$by gradually switching off firstly the electrostatic interactions, followed by perturbing the Lennard-Jones terms to zero. Calculations were performed using 16 equally spaced λ windows, with each λ state firstly equilibrated for 25 M MC moves, followed by a further 30 M data collection moves. Decoupling the electrostatic interactions resulted in a free energy change of +18.824 kcal/mol, whilst perturbing the LJ terms gave a free energy change of -0.773 kcal/mol. This gave a decoupling energy of 18.051 kcal/mol. The simulations were performed using a hardwall potential of radius 1.8 Å which prevented

other molecules from entering the decoupling region. This restraint needs to be accounted for, with its correction term equal to -0.12 kcal/mol. Accounting for the symmetry of the substrate (a term equal to +0.40 kcal/mol), we can find the binding free energy of this molecule to be -11.931 kcal/mol (6.4 - 18.051 + 0.12 - 0.40).

JAWS

The same input structures were used for JAWS as for the double decoupling simulations. Simulations were run at a range of different biasing potentials in order to sample both the on and off states of W_2.

Figure 7.9.4 shows that the JAWS binding free energy fluctuates around the FDTI result. The associated error for the JAWS results is around 1.1 kcal/mol, with the FDTI error around 0.60 kcal/mol. This suggests that the two methods are giving good agreement and that W_2 is a tightly bound to imatinib in the 1IEP structure.

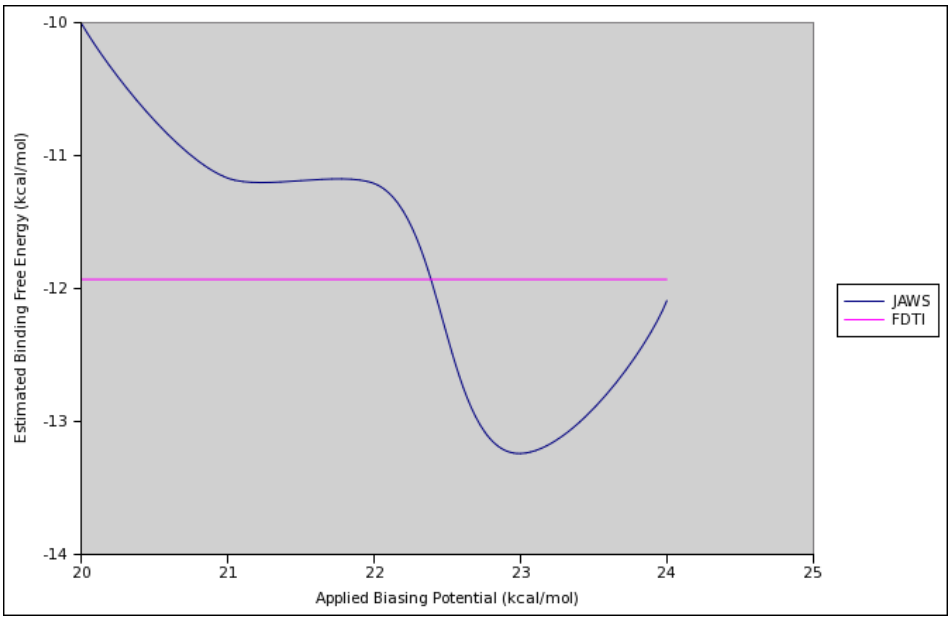


Figure 7.9.4: Binding free energy of W_2 determined by FDTI (pink) and JAWS (blue).

The docking results for imatinib docked to the 3CS9 structure were in line with the expected results; this may be because the binding site residues, particularly Glu286 and Asp381, are in the correct position for imatinib binding despite the absence of the water molecules.

7.10 Chapter summary

The aim of this chapter was to investigate the individual effects of three common mutations, F317L, Y253F and T315I, on the conformation of Abl and binding modes of imatinib and nilotinib, and to test whether the ensemble docking method and MMGBSA calculations could correctly rank the mutations according to experimental data.

The analysis of the F317L and Y253F simulations suggests that the loss of interaction does disrupt the preferred binding of imatinib and nilotinib, causing the ligand to shift to a suboptimal position in the pocket and the further loss of interactions. In 1IEP this translates to diminished hydrogen bonding with His361 (F317L) and Ile360 (F317L) and Asp381 (Y253F) and a subsequent increase in the flexibility of the P-loop and A-loop. In the 3CS9 simulations ligand shift resulted in the diminishment of the hydrogen bond between Glu286 and nilotinib. Without stable interactions with the ligand the αC-helix was able to rotate and form the Glu286-Arg386 bond associated with the intermediate conformation. The formation of the Glu-Arg bond and the loss of interaction at the hinge region had a causal effect on the flexibility of the HRD motif, A-loop and P-loop. These results support the concept that imatinib resistance to F317L and Y253F is conferred due to the inability to achieve the preferred induced fit and that interactions involving the pyridinyl and pyrimidinyl rings of nilotinib contribute far less to the overall binding affinity, allowing nilotinib to maintain sensitivity to F317L and Y253F.

The analysis of the $1IEP_{T315}$ and $3CS9_{T315}$ simulations indicate that the loss of the hydrogen bond with residue 315 and slightly larger size of isoleucine compared to threonine, causes a significant shift in the position of the ligand. In the $1IEP_{T315I}$ simulation imatinib moved away from the hinge region, increasing the mobility of the P-loop, resulting in a shift towards the more elongated conformation of the P-loop associated with active Abl. The disruption to the preferred binding mode of imatinib further resulted in the loss of the hydrogen bond with Ile361, depletion in interactions with Tyr253, and diminished hydrogen bond interactions with Asp381. In contrast, the new position of nilotinib in the $3CS9_{T315}$ simulation allowed stable hydrogen bonding with Asp381 and interactions with the P-loop and residues surrounding the αC-helix, stabilising these motifs; however hydrogen bonds with the other binding site residues were either lost or diminished. This supports the expectation that the T315I gatekeeper causes a significant change to the conformation of the imatinib/nilotinib binding site which impedes binding of both imatinib and nilotinib.

The results of the docking and MM-GBSA studies were mixed; Table 7.10.1 and Table 7.10.2 compare the results for each ligand to experimental data.

	Experimental		Ensemble docking (GlideScore)				MM-GBSA	
	Sensitivity	Rank	1IEP	Rank	3CS9	Rank	Difference (kcal/mol)	Rank
WT		1	-13.40	3	-14.78	1	-72.66	3
F317L		2	-14.93	1	-13.07	2	-94.53	1
Y253F		3	-12.12	4	-12.86	3	-74.48	=2
T315I		4	-13.96	2	-11.71	4	-74.48	=2

Table 7.10.1: Results summary – Imatinib, single mutations study. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental		Ensemble docking (GlideScore, kcal/mol)				MM-GBSA (kcal/mol)	
	Sensitivity	Rank	1IEP	Rank	3CS9	Rank	Difference	Rank
WT		1	-13.78	3	-15.41	1	-74.90	1
F317L		2	-14.50	1	-14.22	2	-62.58	3
Y253F		3	-13.68	4	-14.19	3	-62.95	2
T315I		4	-14.17	2	-12.97	4	-54.25	4

Table 7.10.2: Results summary – Nilotinib, single mutations study. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Using the 3CS9 snapshots as receptors, the docking study correctly ranked the mutations according to experimental data (WT > F317L > Y253F > T315I), and ranked nilotinib as a more potent inhibitor compared to imatinib. The results of the MM-GBSA study were also aligned with the experimental data, however did not correctly the F317L and Y253F structures; Y253F had a slightly more negative result compared to F317L. The overall rankings from the MM-GBSA study were: WT > Y253F > F317L > T315I.

The docking and MM-GBSA results using the 1IEP structures were poorly aligned with experimental data and the mutations were not correctly ranked.

The reasons behind this inconsistency could be either due to poor performance from the docking/MM-GBSA software or an error in the simulation. On review of the 1IEP MD simulation

it was found that two water molecules interacted consistently with imatinib in the simulation; in particular W_2 formed a tetrahedral hydrogen bonding network between W_1, Glu62, Lys47 and Asp157 as seen in the crystal structure of 1IEP and likely stabilises the positions of Glu286 and Asp381. Free energy calculations confirmed that this water is tightly bound and should have been considered explicitly in the docking and MMGBSA studies; the exclusion of conserved waters is known to cause inaccuracies in the prediction of binding affinity values [214-221]. Although the evidence from the water molecule analysis suggests that including the W_1 and W_2 water molecules would have been appropriate for the 1IEP simulations, there is no conclusive evidence to confirm that this is and not some other error source is the driver behind the unexpected docking and MM-GBSA results. To confirm this, multiple simulations using the 1IEP structure, with and without water molecules W_1 and W_2, as well as alternative structures of the Abl kinase domain, such as 1OPJ, would need to be completed. This was unfortunately not possible to complete as part of this research due to time constraints.

Chapter 8

8 Compound mutation study

8.1 Aim

The aim of this chapter is to investigate the effect of two compound mutations, Y253F/F317L and T315I/Y253F/F317L, on the conformation of Abl and the binding modes of imatinib and nilotinib. At the time of this research there was no clinical data available regarding the prevalence of compound mutations in patients and no in-vitro studies had been completed. In the absence of experimental data, the selection of the compound mutations was based on combining the single mutations already studied. Since the completion of this research, there have been several studies evaluating the how well the TKIs used in current clinical practice inhibit compound mutations. Although there are no experimental data available for the compound mutations studied in this work (Y253F/F317L and T315I/Y253F/F317L) previous studies have shown that compound mutations that include T315I as a component typically demonstrate a high level of resistance to imatinib and nilotinib [330, 331].

Having validated the methods by reproducing experimental rankings in Chapter 7, docking and MMGBSA calculations were utilised to predict the level of resistance conferred by the compound mutations. In the previous chapter it was shown that a single occurrence of these mutations has dramatic effects on the dynamics of the kinase as well as protein-ligand interactions.

Both compound mutations are expected to confer resistance to imatinib and nilotinib. The double mutation is expected to be ranked as more resistant than the single mutants Y253F and F317L; the triple mutation is predicted to be ranked the most resistant, given the inclusion of the gatekeeper. The overall rankings are expected to be: WT > F317L > Y253F > YF > T315I > TYF.

	Imatinib (IC_{50} in nM)	Nilotinib (IC_{50} in nM)	Ponatinib (IC_{50} in nM)
WT	768.3 ± 143	57.0 ± 11.1	111.6 ± 12.3
G250E/V299L	6321.3 ± 1411.9	717.8 ± 177.1	17.5 ± 3.1
Y253H/E255V	>10240.0	>10240.0	203.5 ± 41.7
Y253H/F317L	>10240.0	>10240.0	27.3 ± 2.9
E255V/V299L	>10240.0	3890.3 ± 2041.4	42.8 ± 13.0
V299L/F317L	4276.7 ± 494.7	287.8 ± 26.8	9.5 ± 1.9
V299L/M351T	2599.7 ± 378.6	95.7 ± 17.8	8.3 ± 1.5
V299L/F359V	2760.7 ± 462.4	690.0 ± 130.6	18.8 ± 3.6
F317L/F359V	8397.3 ± 1395.0	1884.7 ± 344.2	53.2 ± 17.7
M244V/T315I	9865.0 ± 212.3	9914.0 ± 326.2	29.1 ± 8.6
G250E/T315I	>10240.0	>10240.0	152.4 ± 29.7
Q252H/T315I	>10240.0	>10240.0	114.3 ± 23.2
Y253H/T315I	>10240.0	>10240.0	357.9 ± 77.4
E255V/T315I	>10240.0	>10240.0	659.5 ± 9.4
F311I/T315I	>10240.0	>10240.0	215.7 ± 23.4
T315I/M351T	>10240.0	>10240.0	84.8 ± 19.7
T315I/F359V	>10240.0	>10240.0	103.5 ± 17.1
T315I/H396R	>10240.0	>10240.0	90.8 ± 24.7
T315I/E453K	>10240.0	>10240.0	93.4 ± 1.7

Table 8.1.1. Sensitivity of Bcr-Abl kinase domain mutants to imatinib and nilotinib (based on assays using Ba/F3 cells. The IC_{50} value is the concentration of inhibitor resulting in a 50% reduction in cell viability; NR, not recorded. Adapted from references: Zabriskie et al, 2014 [331])

Sensitivity	Imatinib (IC_{50} in nM)	Nilotinib (IC_{50} in nM)	Ponatinib (IC_{50} in nM)
Sensitive	< 1000	< 200	< 25
Moderately resistant	1000 – 4000	200 - 1000	25 – 150
Highly resistant	> 4000	> 1000	> 150

Table 8.1.2. Sensitivity scale for imatinib, nilotinib and ponatinib. Adapted from reference [331]

8.2 Molecular dynamics study of compound mutations

8.2.1 Review of the mutations

Y253F/F317L compound mutation

The Y253F/F317L double mutation combines two of the mutations studied in Chapter 7; Y253F and F317L. Tyr253 (Y253) in WT Abl forms one side of the hydrophobic cage around the imatinib binding site and aids the folding of the P-loop into the kinked conformation for preferential binding with imatinib/nilotinib. The side chain of Phe317 (F317), which is part of the hinge region, effectively forms the roof of the ATP binding site. This residue stacks partially on the pyrimidine ring of imatinib and nilotinib, and the side chain shields the hinge region from the solvent forming a hydrophobic enclosure.

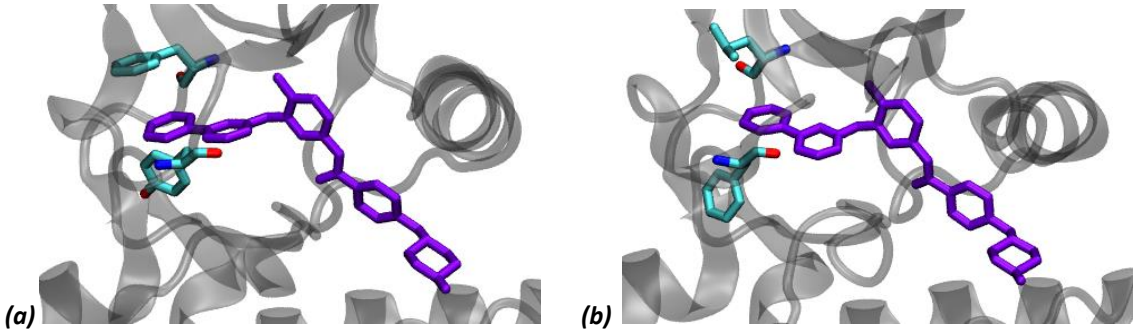


Figure 8.2.1: Location of the residues in the double mutant structure (a) The position of Tyr253 and Phe317 in WT Abl (1IEP.pdb); (b) The position of Phe253 and Leu317 in mutant Abl (1IEP$_{YF}$)

The analysis of the single F317L and Y253F simulations suggests that the loss of interaction disrupts the preferred binding of imatinib and nilotinib, causing the ligand to shift to a suboptimal position in the pocket and the further loss of interactions. In 1IEP this translates to diminished hydrogen bonding with His361 (F317L) and Ile360 (F317L) and Asp381 (Y253F) and a subsequent increase in the flexibility of the P-loop and A-loop. In the 3CS9 simulations ligand shift resulted in the diminishment of the hydrogen bond between Glu286 and nilotinib. Without stable interactions with the ligand the αC-helix was able to rotate and form the Glu286-Arg386 bond associated with the intermediate conformation.

The combination of the two mutations is therefore likely to cause significant disruption to the binding interactions of both ligands.

T315I/Y253F/F317L compound mutation

The T315I/Y253F/F317L combines the double mutation with the T315I gatekeeper mutation and therefore all three of the single mutations studied in Chapter 7.

The threonine residue at position 315 in Abl is known as the gatekeeper residue and the size of the residue at this position determines how easily the hydrophobic pocket can be accessed especially in the (DFG-in) active form. In the case of imatinib and nilotinib binding, the hydroxyl group of T315 forms a hydrogen bond to the amine linker between the pyrimidine and phenyl rings of the ligand..

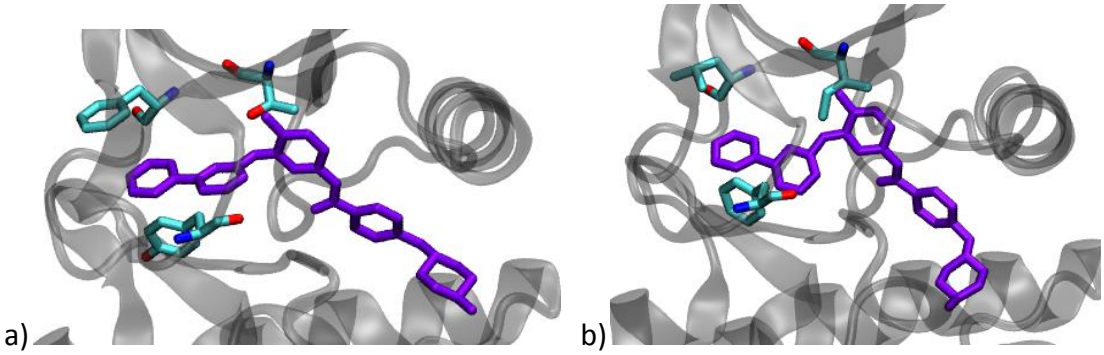


Figure 8.2.2: (a) The position of Tyr253, Phe317 and Thr315 in WT Abl (1IEP.pdb); (b) The position of Phe253, Leu317 and Ile 315 in mutant Abl ($1IEP_{TYF}$)

The protein-ligand interactions analysis of the T315I simulations in Section 7.5 indicates that the loss of the hydrogen bond with residue 315 and slightly larger size of isoleucine compared to threonine, causes a significant shift in the position of the ligand. In the $1IEP_{T315I}$ simulation imatinib moved away from the hinge region, increasing the mobility of the P-loop, resulting in a shift towards the more elongated conformation of the P-loop associated with active Abl. The disruption to the preferred binding mode of imatinib further resulted in the loss of the hydrogen bond with Ile361, depletion in interactions with Tyr253, and diminished hydrogen bond interactions with Asp381. In contrast, the new position of nilotinib in the $3CS9_{T315}$ simulation allowed stable hydrogen bonding with Asp381 and interactions with the P-loop and residues surrounding the αC-helix, stabilising these motifs; however hydrogen bonds with the other binding site residues were either lost or diminished.

Given the disruptive effects of each singular mutation, the combination of Y253F, F317L and T315I are expected to cause substantial loss of binding and a significant change in the position of imatinib and nilotinib in the pocket.

8.2.2 Protein structure preparation and MD parameters

The SCAP program was used to mutate multiple residues in the 1IEP and 3CS9 structures; Y253 and F317 for the double mutation and Y253, F317L and T315 for the triple mutation [71]. Protein structure preparation and molecular dynamics simulations were then completed as described in 5.2.4.

8.3 Results: Double compound mutation

8.3.1 Imatinib

Stability and flexibility of the protein

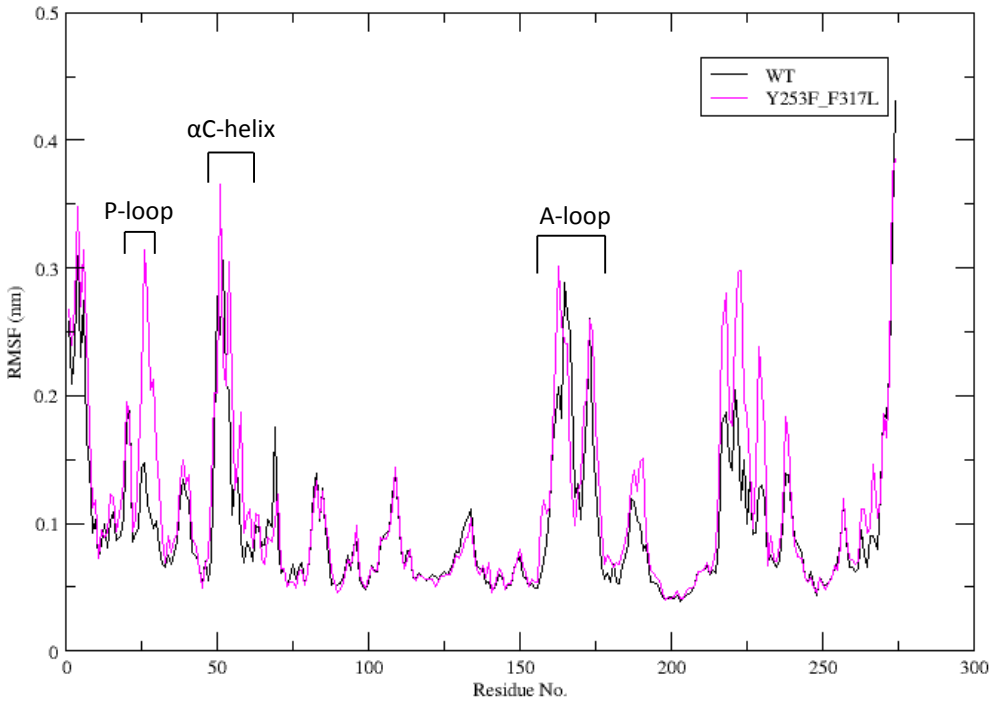


Figure 8.3.1: RMSF results for $1IEP_{WT}$ and $1IEP_{YF}$

The RMSF profile for $1IEP_{YF}$ displayed increased mobility in the P-loop, αF-Helix and αH-Helix; fluctuations in the other regions in the kinase were comparable to that of the wild-type. Fluctuations in the P-loop translated to a change in conformation; the loop unfolded from the kinked conformation towards the elongated active conformation.

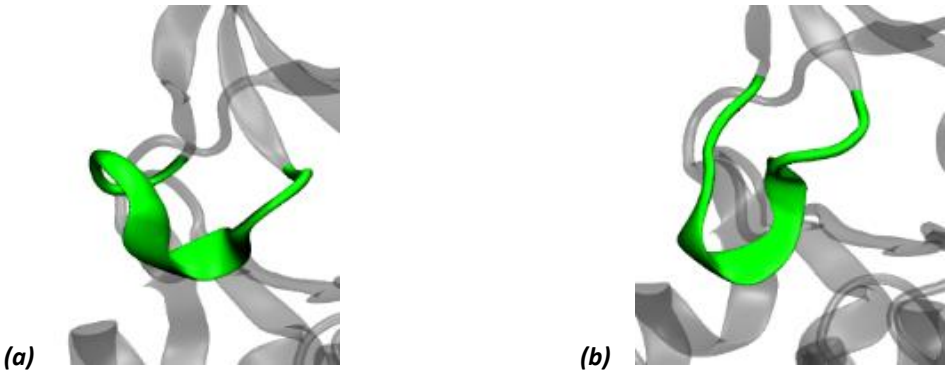


Figure 8.3.2: P-loop conformation: (a) initial and (b) during simulation

The P-loop did retain some secondary structure however and did not encroach on the binding site. The RMSD analysis highlighted a large fluctuation in the DFG loop during the simulation; observational analysis showed that Phe383 residue was extremely mobile and fluctuated between several different DFG-in conformations during the simulation. This was supported by the R-spine analysis which identified that the distance between Phe382 and His361 had decreased but that the spine remained misaligned.

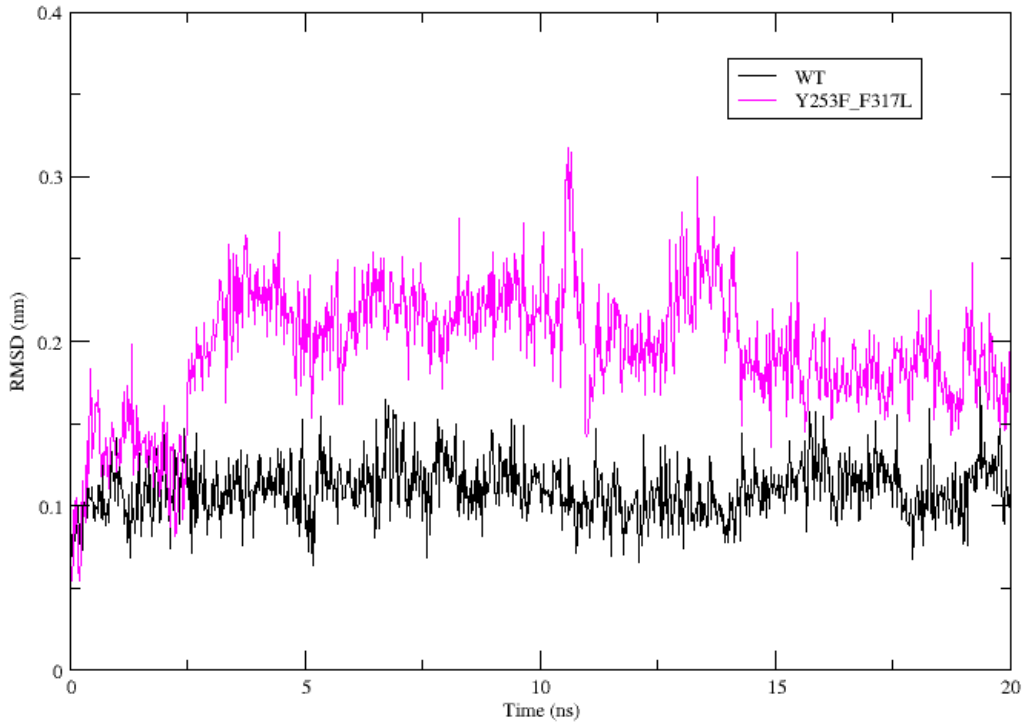


Figure 8.3.3: RMSD of the DFG-loop ($1IEP_{YF}$)

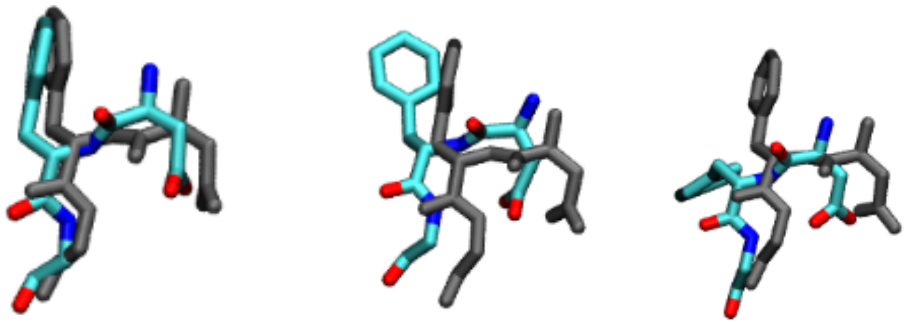


Figure 8.3.4: Different conformations of the DFG loop in 1IEP$_{YF}$; Double mutant in cyan, WT in grey

R-spine residues	Mean distance (CA-CA, Å)	
	1IEP_WT	1IEP_YF
Leu301 – Met290	6.487	7.225
Met290 – Phe382	12.284	12.461
Phe382 – His361	12.302	10.907

Table 8.3.1: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the Y253F/F317L simulation, however the Lys271-Glu286 salt bridge, was not well conserved and was only stable for approximately 59% of the simulation (Table 8.3.2). Observational analysis showed that that Glu286 is partially rotated and forms a hydrogen bond with Arg386, however the bond is not stable given the position of the C-Helix is turned upwards away from the binding site. The breakage of the salt bridge is not driven by the rotation of the C-helix in this case but by the helix's upward motion and fluctuations in the Lys271 residue.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{YF}$	$1IEP_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	97.30	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	44.16	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	52.85	45.25
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.50	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	81.72	80.62
		His361(S)	Asp421(S)	27.27	33.47

Table 8.3.2: Hydrogen bond analysis: $1IEP_{YF}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

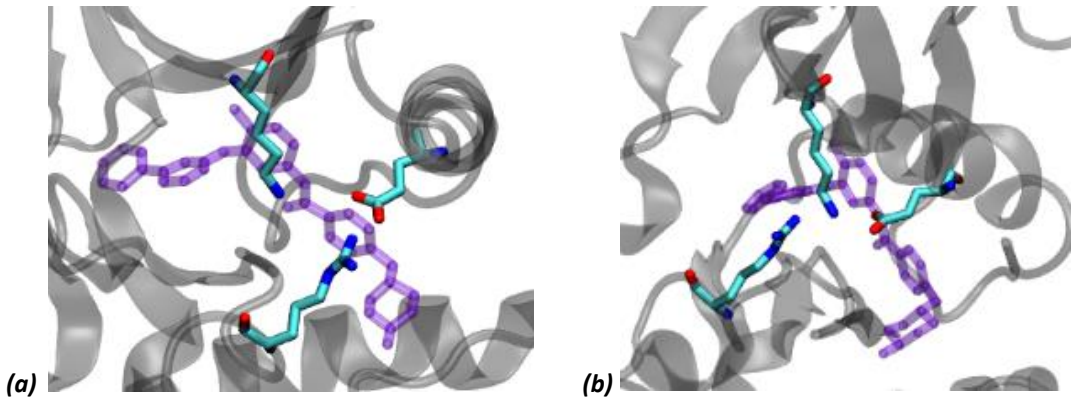


Figure 8.3.5: The positions of Glu286, Arg386 and Lys271; (a) facing the αC-Helix (b) rotated 90°

Stability and flexibility of imatinib

The RMSD analysis showed that imatinib was less stable in the binding site in the Y253F/F317L simulation compared to the WT and RMSF analysis highlighted fluctuations at the pyrimidine and piperazine rings (Figure 8.3.6, Figure 8.3.7 and Figure 8.3.8).

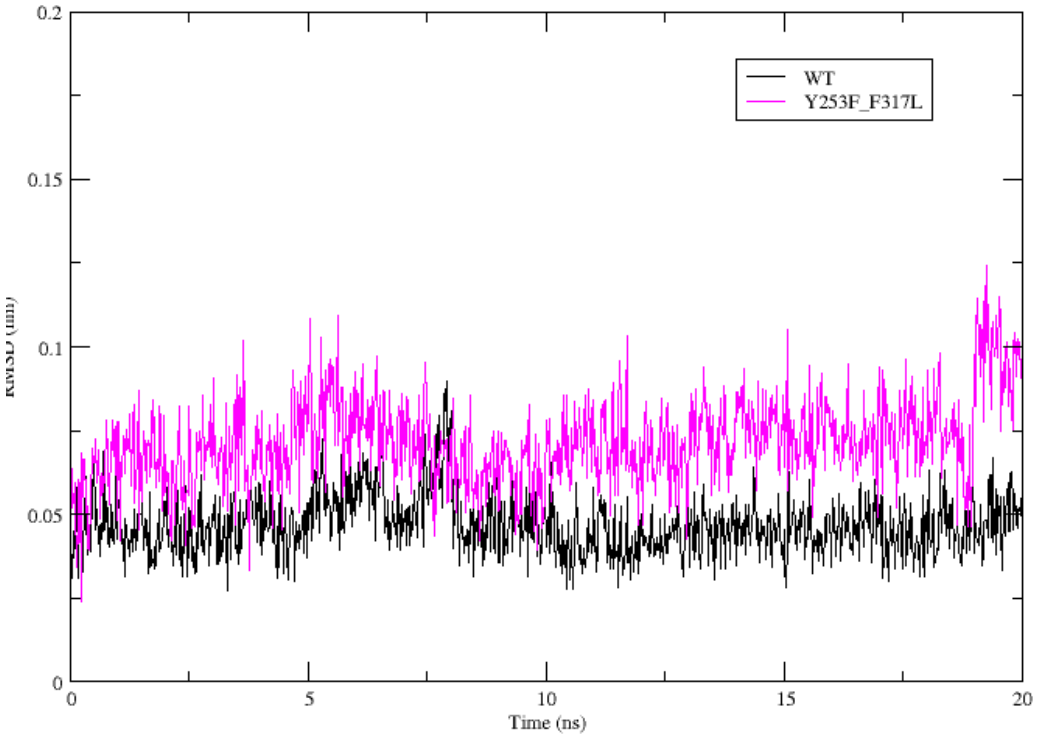


Figure 8.3.6: RMSD of imatinib ($1IEP_{YF}$)

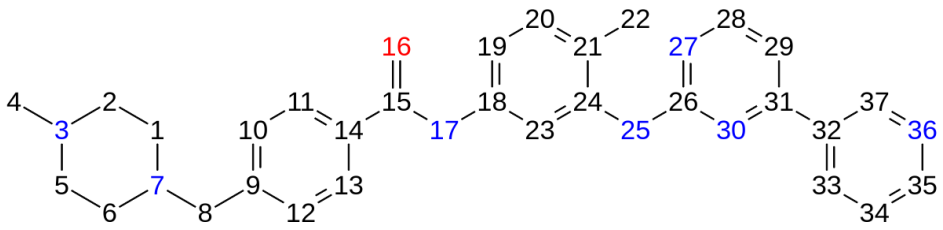


Figure 8.3.7: Index of atoms in imatinib ($1IEP_{YF}$)

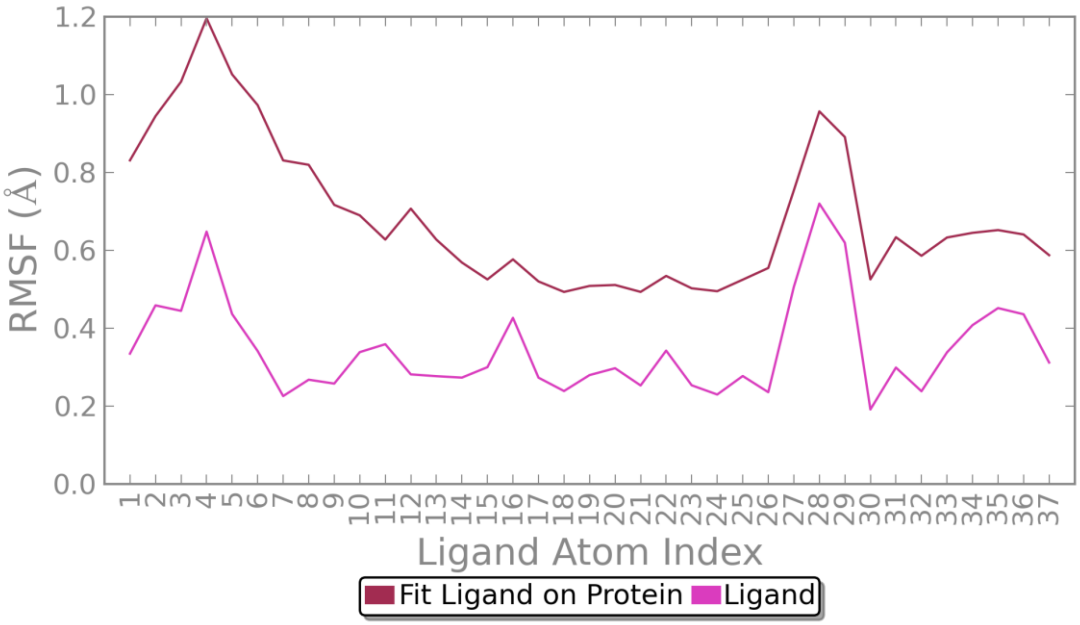


Figure 8.3.8: RMSF of imatinib (1IEP$_{YF}$)

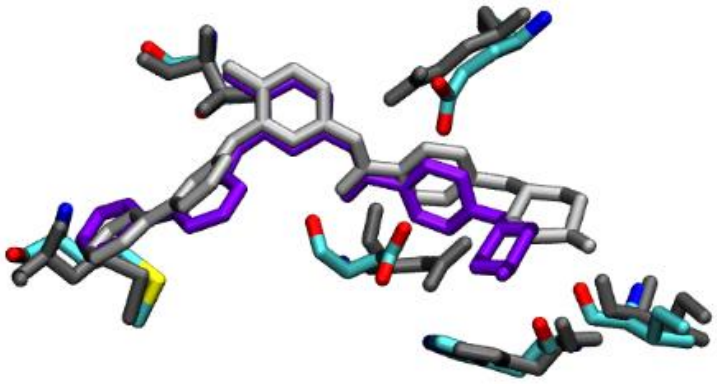


Figure 8.3.9: Comparison of imatinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (15 ns)

Protein-ligand interactions

As expected, the substitution of a phenylalanine residue at position 253 and a leucine at 317 resulted in a loss of interactions between the protein and the ligand (residues 253 and 317 are renumbered in Figure 8.3.10 to 29 and 93 respectively).

Unexpectedly the hydrogen bonding interactions with Asp381 (Asp157 in Figure 8.3.10), Thr315 (Thr91 in Figure 8.3.10), Glu286 (Glu62 in Figure 8.3.10), Met290 (Met66 in Figure 8.3.10), Ile360

(Ile136 in Figure 8.3.10) and His361 (His137 in Figure 8.3.10) appear to be well maintained; however given the increased RMSD of the ligand it is unclear how stable the ligand is in this conformation given the flexibility if the P- and DFG-loops.

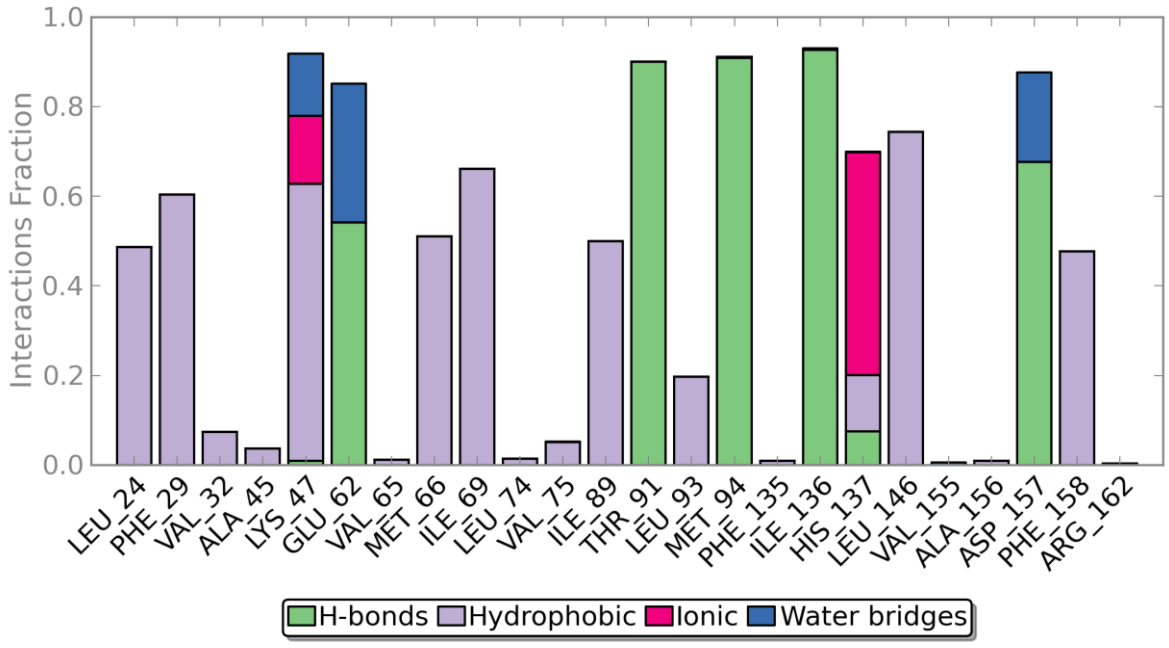


Figure 8.3.10: Protein-ligand interactions by residue (1IEP$_{YF}$)

8.3.2 Nilotinib

Structural stability and flexibility

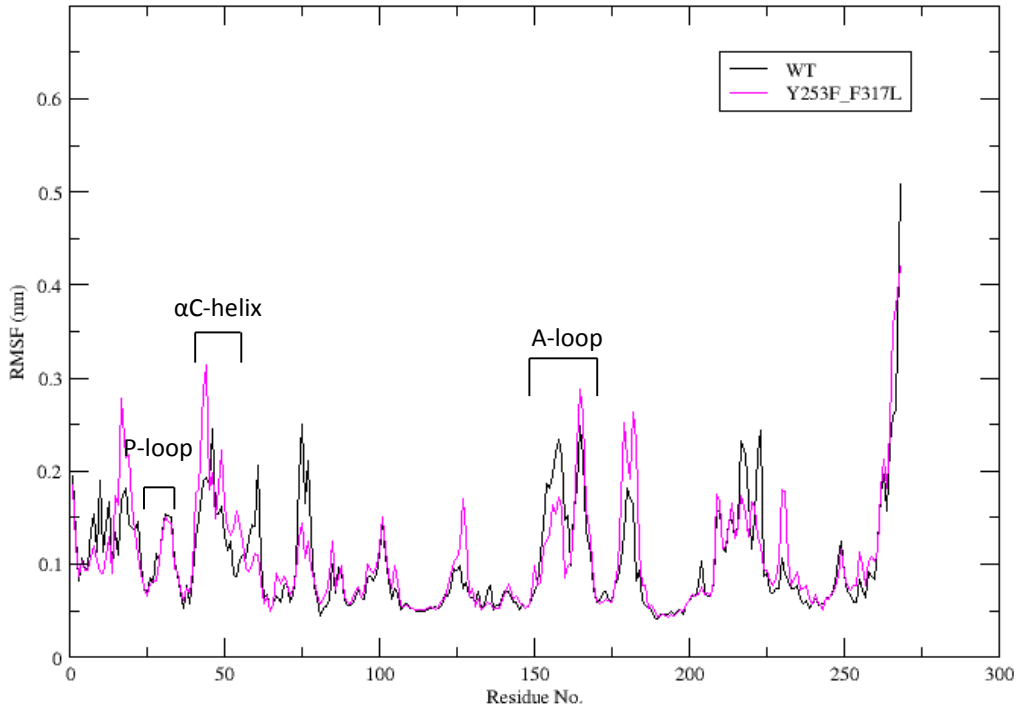


Figure 8.3.11: RMSF results for $3CS9_{WT}$ and $3CS9_{YF}$

The RMSF profile for $3CS9_{YF}$ displayed increased mobility in the P-loop, αC-helix and second half of the A-loop; fluctuations in the other regions of the kinase were comparable to that of the wild-type simulation. The increased flexibility in the P-loop translated in a change in the kinked conformation towards the more elongated active conformation (Figure 8.3.12).

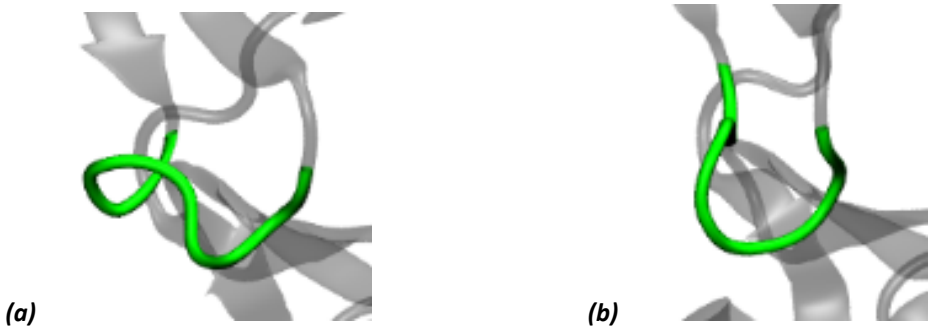


Figure 8.3.12: Conformation of the P-loop (a) initial (b) final

A review of the R-spine showed that the residues maintained similar positions to that of the WT structure, suggesting that the double mutation had minimal impact on the position of these residues or the conformation of the DFG-loop.

R-spine residues	Mean distance (CA-CA, Å)	
	3CS9_WT	3CS9_YF
Leu301 – Met290	6.500	6.418
Met290 – Phe382	12.874	11.665
Phe382 – His361	11.514	11.279

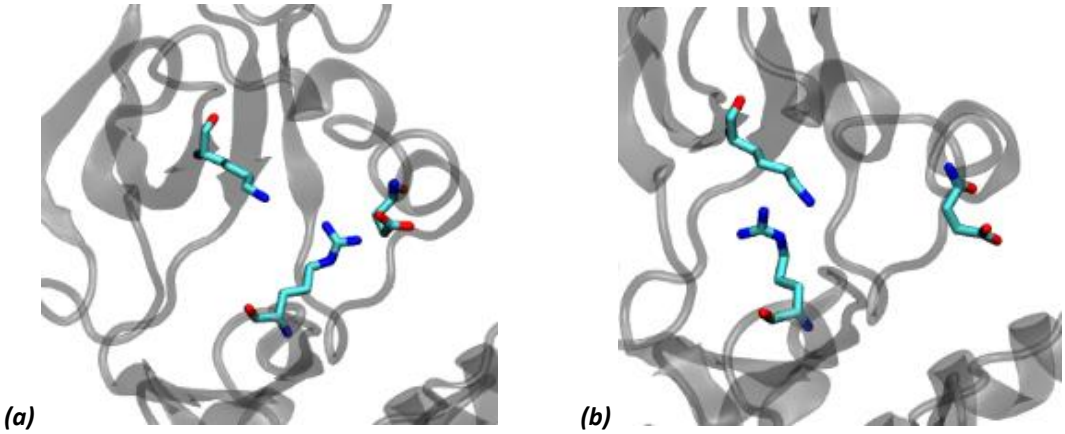
Table 8.3.3: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the $3CS9_{YF}$ simulation, however the Lys271-Glu286 salt bridge was all but absent.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$3CS9_{YF}$	$3CS9_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	96.39	97.70
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	0.20	82.32
		Glu286 (S)	Lys271 (S)	0.30	-
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	75.95	19.08
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	98.50	99.60
		Arg362(M)	Asp421(S)	0.90	-
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	67.23	85.71
		His361(S)	Asp421(S)	12.73	32.97

Table 8.3.4: Hydrogen bond analysis: $3CS9_{YF}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Observational analysis revealed that the Lys271-Glu286 bond was not present at the start of the simulation due to the denaturation of the αC-helix, and during the simulation Glu286 continues to rotate out of the binding site.



(a) ***(b)***

Figure 8.3.13: Conformation of Glu286 (a) initial (b) fully rotated

Glu286 does form a hydrogen bond with Arg386 but the upwards motion of the C-helix means that this bond is not well conserved. The denaturation of the C-helix suggests that the secondary structure of the $3CS9_{YF}$ simulation may be unstable.

Stability and flexibility of nilotinib

Although initially well aligned with the WT simulation, the RMSD of the ligand gradually increased during the simulation (Figure 8.3.14). The RMSF analysis indicated a poor fit in the binding site; the fluctuations of the trifluoromethyl were tightly aligned with those of the protein, however the pyridine ring appeared to be distal from the hinge region (Figure 8.3.15, Figure 8.3.16 and Figure 8.3.17).

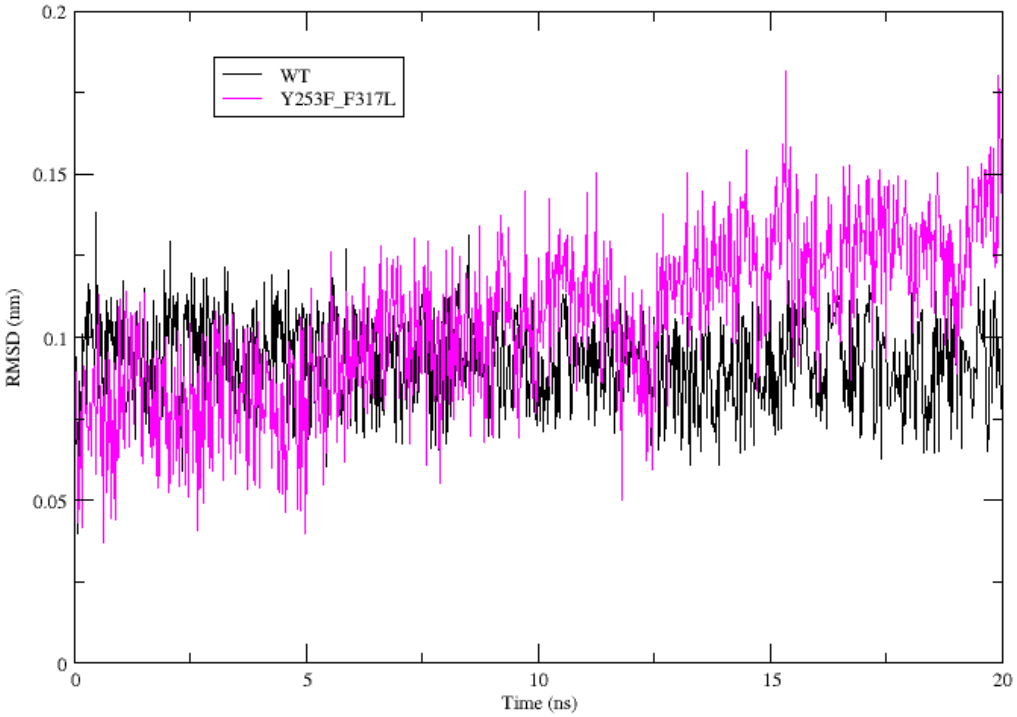


Figure 8.3.14: RMSD for nilotinib (3CS9$_{YF}$)

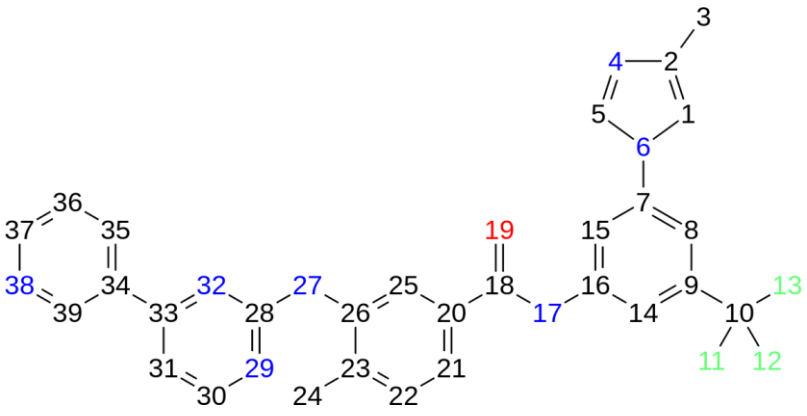


Figure 8.3.15: Index of atoms for nilotinib (3CS9$_{YF}$)

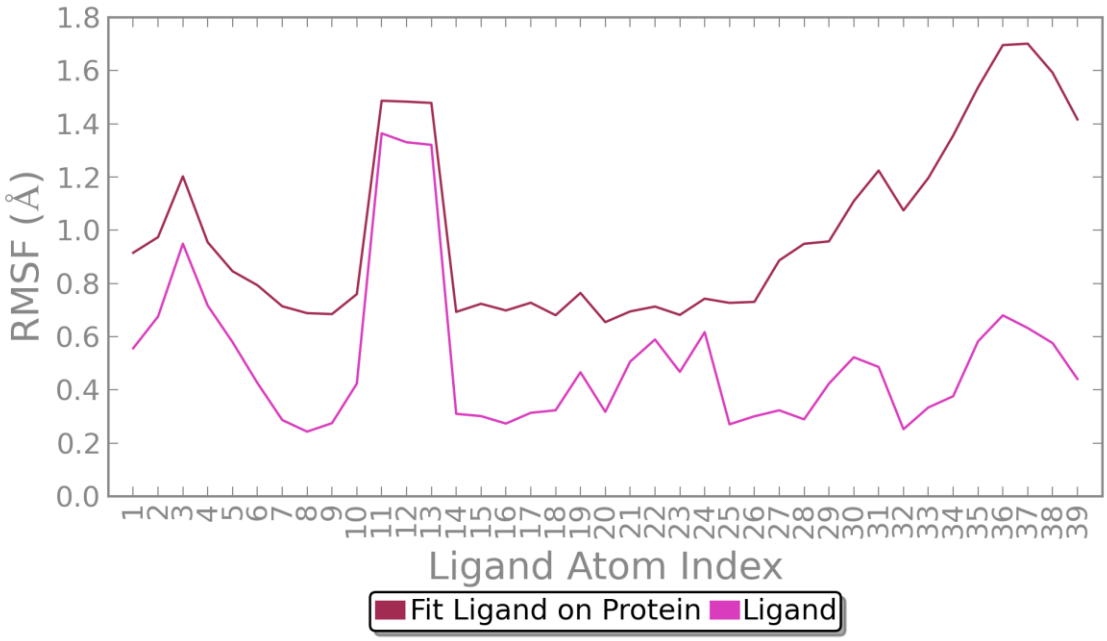


Figure 8.3.16: RMSF for nilotinib (3CS9$_{YF}$)

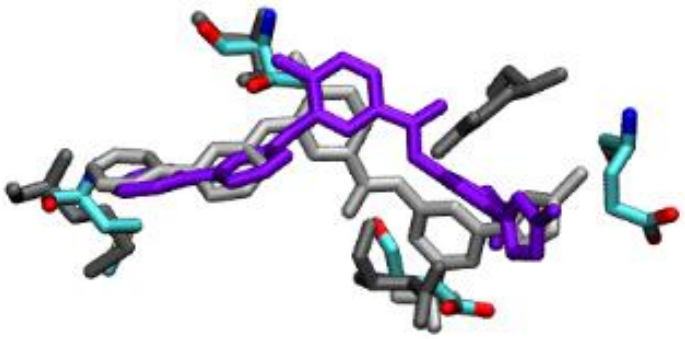


Figure 8.3.17: Comparison of nilotinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (15 ns)

Protein-ligand interactions

As in the imatinib simulation the substitution of a phenylalanine residue at position 253 and a leucine at 317 resulted in a loss of interaction between the protein and the ligand; the shift in conformation also resulted in the hydrogen bonds between nilotinib and the protein becoming diminished, particularly Glu286 and Met290 (Glu54 and Met58 in (Figure 8.3.18); interestingly the new conformation of nilotinib does facilitate a strong water mediated hydrogen bond with Lys271 (Lys39 in Figure 8.3.18).

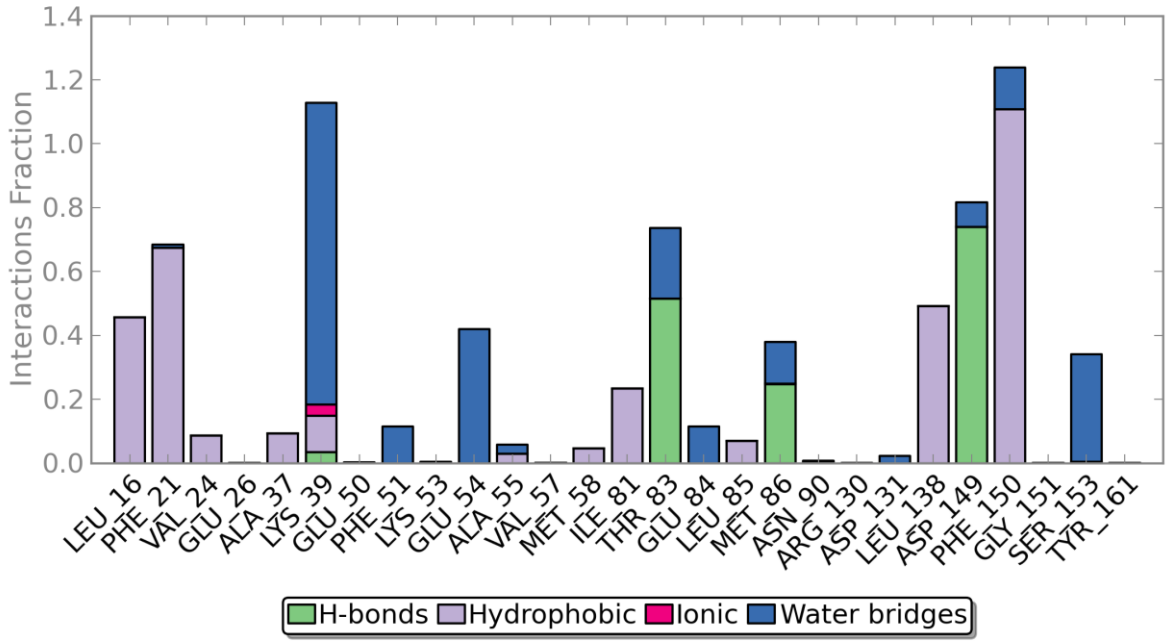


Figure 8.3.18: Protein-ligand interactions by residue (nilotinib, 3CS9$_{YF}$)

8.4 Results: Triple compound mutation

8.4.1 Imatinib

Stability and flexibility of the protein

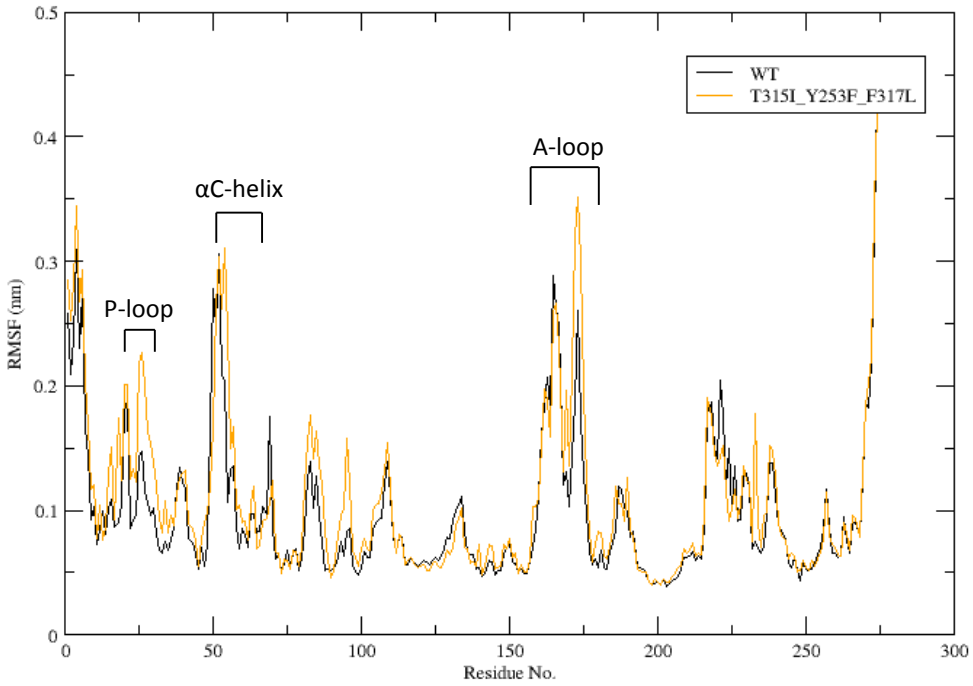


Figure 8.4.1: RMSF results for $1IEP_{WT}$ and $1IEP_{TYF}$

The RMSF profile for $1IEP_{TYF}$ showed an increase in the mobility of the P-loop and second half of the A-loop compared to the wild-type structure, and a moderate increase in RMS across the protein structure overall. Similar to the $1IEP_{YF}$ simulation, the fluctuations in the P-loop translated to a change in conformation; the loop unfolded from the kinked conformation towards the elongated active conformation.

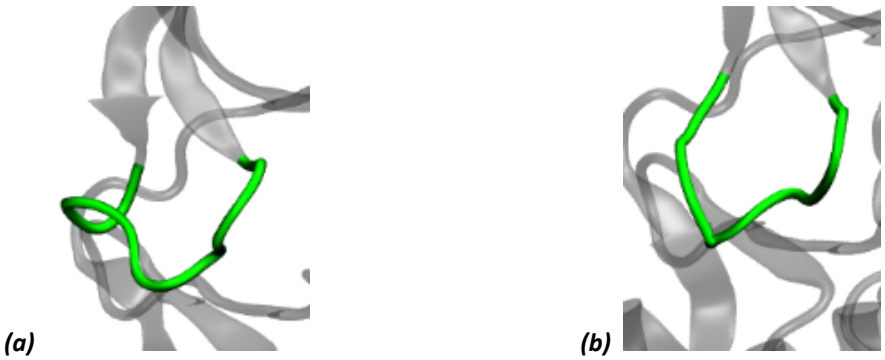


Figure 8.4.2: Different conformations of the P-loop in $1IEP_{TYF}$ (a) initial (b) final

A review of the R-spine showed that the residues maintained similar positions to that of the WT structure, suggesting that the double mutation had minimal impact on the position of these residues or the conformation of the DFG-loop.

R-spine residues	Mean distance (CA-CA, Å)	
	1IEP_WT	1IEP_TYF
Leu301 – Met290	6.487	6.983
Met290 – Phe382	12.284	12.769
Phe382 – His361	12.302	11.843

Table 8.4.1: Mean distance between R-spine residues

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the Y253F/F317L mutated structure; reduced stability was observed in the Lys271-Glu286 salt bridge, but the hydrogen bond was conserved throughout the simulation

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$1IEP_{TYF}$	$1IEP_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	97.00	90.21
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	79.02	91.21
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	34.57	45.25
		Gly383(M)	Asp381(M)	0.10	0.50
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	99.30	99.90
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	90.31	80.62
		His361(S)	Asp421(S)	19.48	33.47

Table 8.4.2: Hydrogen bond analysis: $1IEP_{TYF}$ vs. $1IEP_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

Stability and flexibility of imatinib

Imatinib demonstrated stable motions within the pocket, with a marginally lower RMSD compared to the WT structure. RMSF analysis confirmed that fluctuations of the ligand were in line with that of the overall protein structure (Figure 8.4.3) and that the largest fluctuations were located at the trifluoromethyl and imidazole groups at the entrance to the binding site (Figure 8.4.4 and Figure 8.4.5).

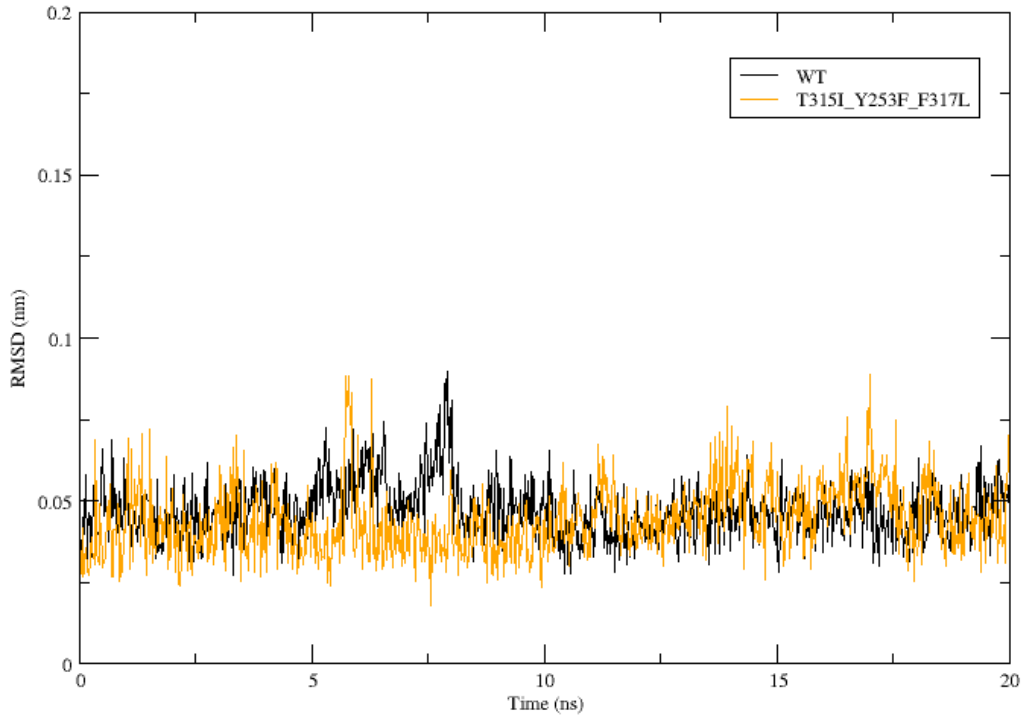


Figure 8.4.3: RMSD of imatinib (1IEP$_{TYF}$)

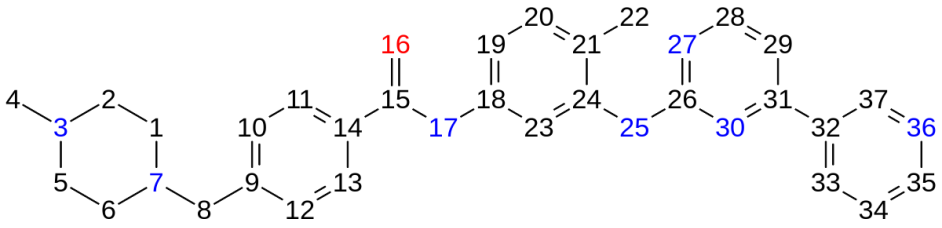


Figure 8.4.4: Index of atoms for imatinib

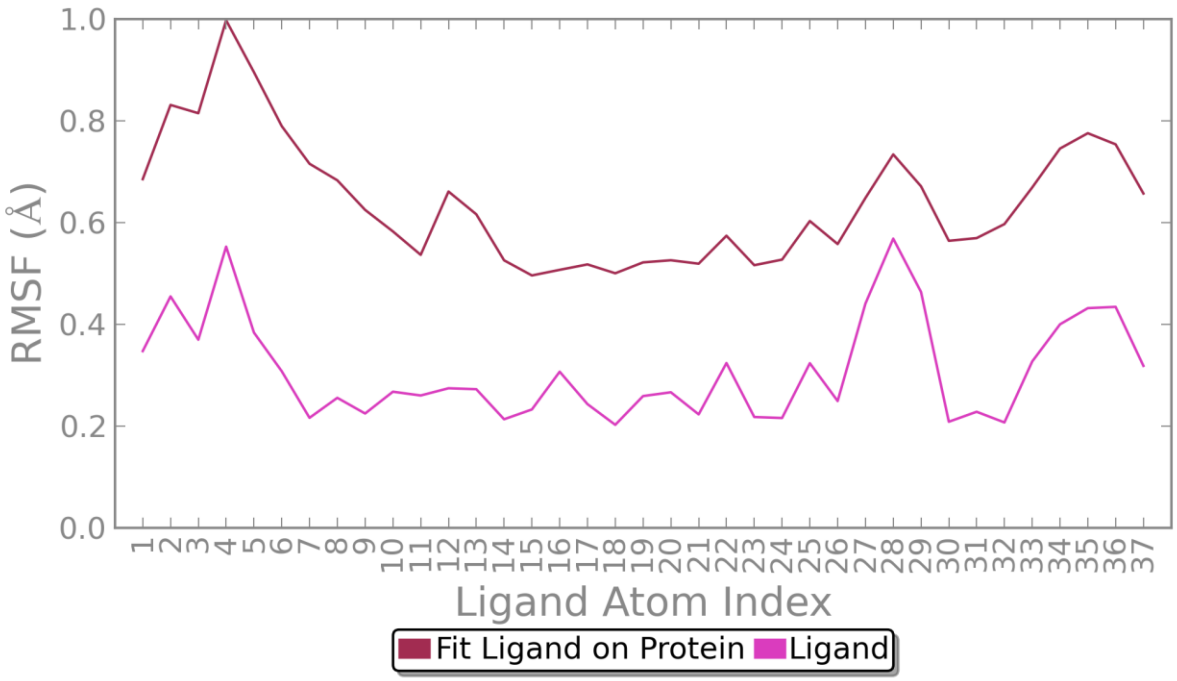


Figure 8.4.5: RMSF of imatinib ($1IEP_{TYF}$)

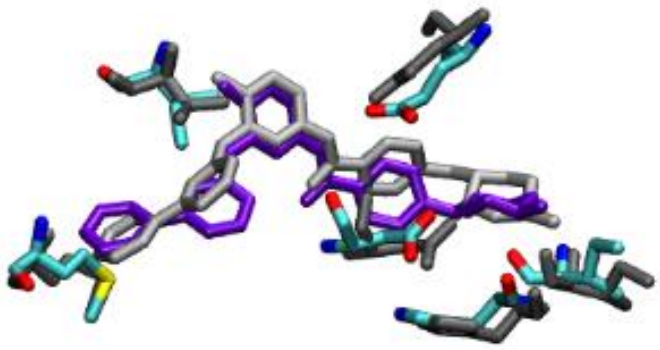


Figure 8.4.6: Comparison of imatinib binding site in the WT and YF mutant structures; grey/silver, WT; purple/cyan, YF (6 ns)

The movement of the ligand in the binding site is further illustrated in Figure 8.4.6, which shows the conformation of the ligand in $1IEP_{TYF}$ compared to $1IEP_{WT}$.

Protein-ligand interactions

The triple mutation resulted in the loss of a hydrogen bond between imatinib and Ile315, and the loss of interaction with Leu317 (Ile91 and Leu93 in Figure 8.4.7).

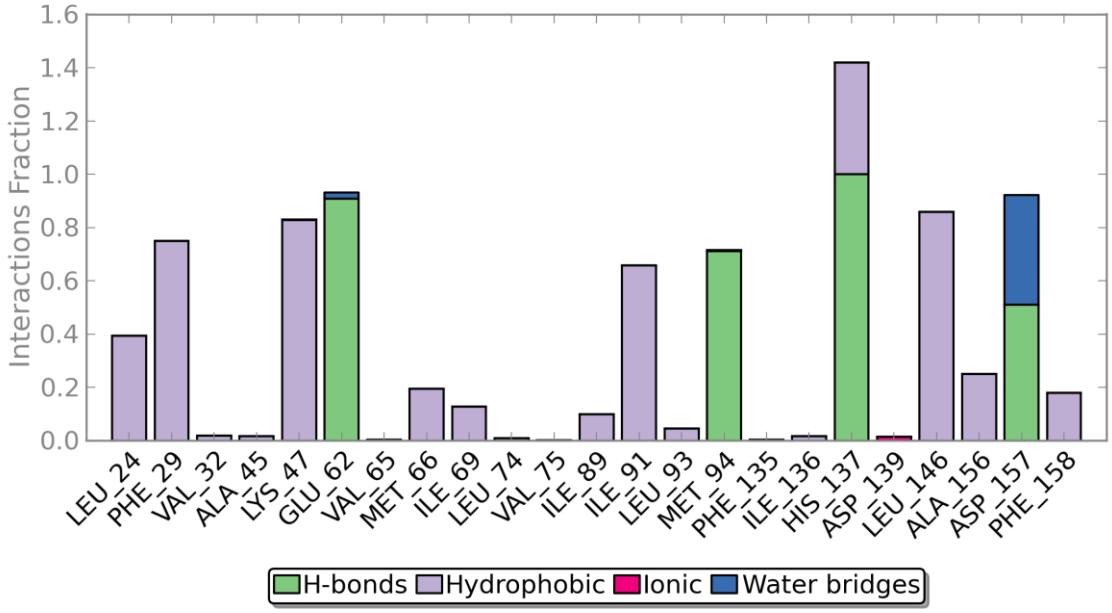


Figure 8.4.7: Protein-ligand interactions by residue ($1IEP_{TYF}$)

The increased fluctuation and change in conformation of the ligand also resulted in the loss of the hydrogen bond between the piperazine ring of imatinib and Ile360 (Ile136 in Figure 8.4.7), however the hydrogen bond with His361 (His137 Figure 8.4.7) was maintained, as were the interactions with Phe253 (Phe29 in Figure 8.4.7). Hydrogen bond interactions with Asp381 (Asp157 in Figure 8.4.7) were diminished compared to the WT simulation.

8.4.2 Nilotinib

Stability and flexibility of the protein

The RMSF profile for $3CS9_{TYF}$ demonstrated increased mobility in the P-loop and αC-helix compared to the wild-type structure; fluctuations in the other regions of the kinase were comparable to that of the wild-type (Figure 8.4.8). The fluctuations in the P-loop translated to a change in conformation; the loop unfolded from the kinked conformation towards the elongated active conformation.

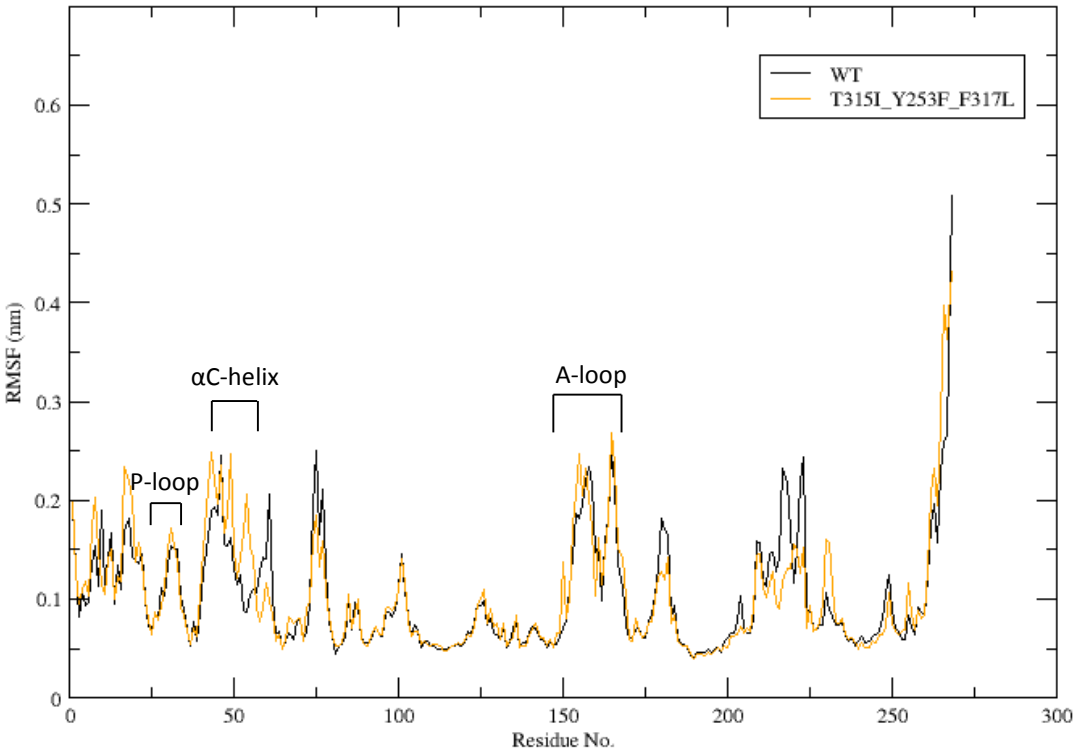


Figure 8.4.8: RMSF results for $3CS9_{WT}$ and $3CS9_{TYF}$

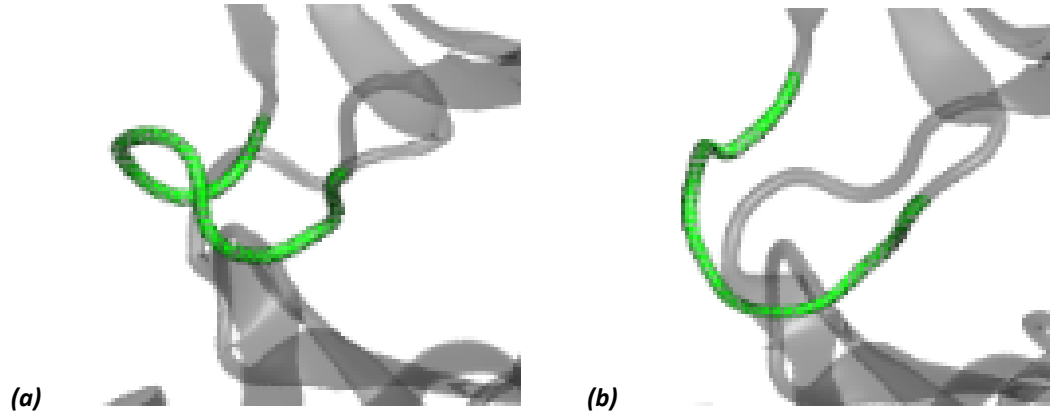


Figure 8.4.9: Different conformations of the P-loop in $1IEP_{TYF}$ (a) initial (b) final

The average distance between Phe382 and neighbouring R-spine residues Met290 and His361 decreased was slightly shorter in the $3CS9_{Y253F}$ simulation; however the spine maintained the inactive misaligned conformation throughout the simulation. Observational analysis revealed that the DFG-loop had a subtle change in conformation during the simulation as supported by these results.

R-spine residues	Mean distance (CA-CA, Å)	
	3CS9_WT	3CS9_TYF
Leu301 – Met290	6.500	6.199
Met290 – Phe382	12.874	11.721
Phe382 – His361	11.514	10.970

Table 8.4.3: Mean distance between R-spine residues

Figure 8.4.10: Different conformations of the DFG-loop ($1IEP_{TYF}$)

The hydrogen bond between Tyr393 and Asp363 was extremely stable in the triple mutated structure (Table 8.4.4).

Reduced stability was observed in the Lys271-Glu286 salt bridge; observational analysis revealed that this was due to the denaturation of the αC-helix and fluctuations locally to Lys271; Glu286 residue also formed a hydrogen bond with Arg386 but did not fully rotate out of the binding site.

Interaction marker	Description of interaction	Donor	Acceptor	Stability (% present in simulation)	
				$3CS9_{TYF}$	$3CS9_{WT}$
Hydrogen bonds associated with the inactive conformation					
Tyr393-Asp363	Hydrogen bond between Tyr393 and D of the HRD motif	Tyr393 (S)	Asp363 (S)	98.40	97.70
Hydrogen bonds associated with the active conformation					
Lys271 – Glu286	Salt bridge between β sheet and αC-helix	Lys271 (S)	Glu286 (S)	38.88	82.32
		Glu286 (S)	Lys271 (S)	0.50	-
Asp381 – Gly383	Hydrogen bond between D and G of DFG	Gly383(M)	Asp381(S)	21.74	19.08
Arg362 – Asp421	Hydrogen bond between αG-helix and HRD motif (R)	Arg362(M)	Asp421(S)	98.70	99.60
His361 – Asp421	Hydrogen bond securing R spine to αG-helix	His361(M)	Asp421(S)	83.67	85.71
		His361(S)	Asp421(S)	25.55	32.97

Table 8.4.4: Hydrogen bond analysis: $3CS9_{TYF}$ vs. $3CS9_{WT}$ (4 Å, 30°); S = sidechain, M = main chain.

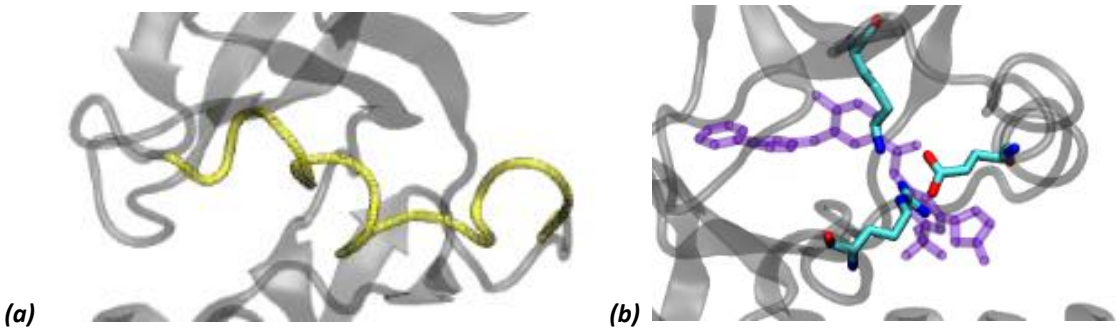


Figure 8.4.11: Denaturation of the αC-helix (a); Position of Glu286 forming a hydrogen bond with Arg386.

Stability and flexibility of nilotinib

Nilotinib demonstrated stable motions within the pocket, with a marginally higher RMSD compared to the WT structure.

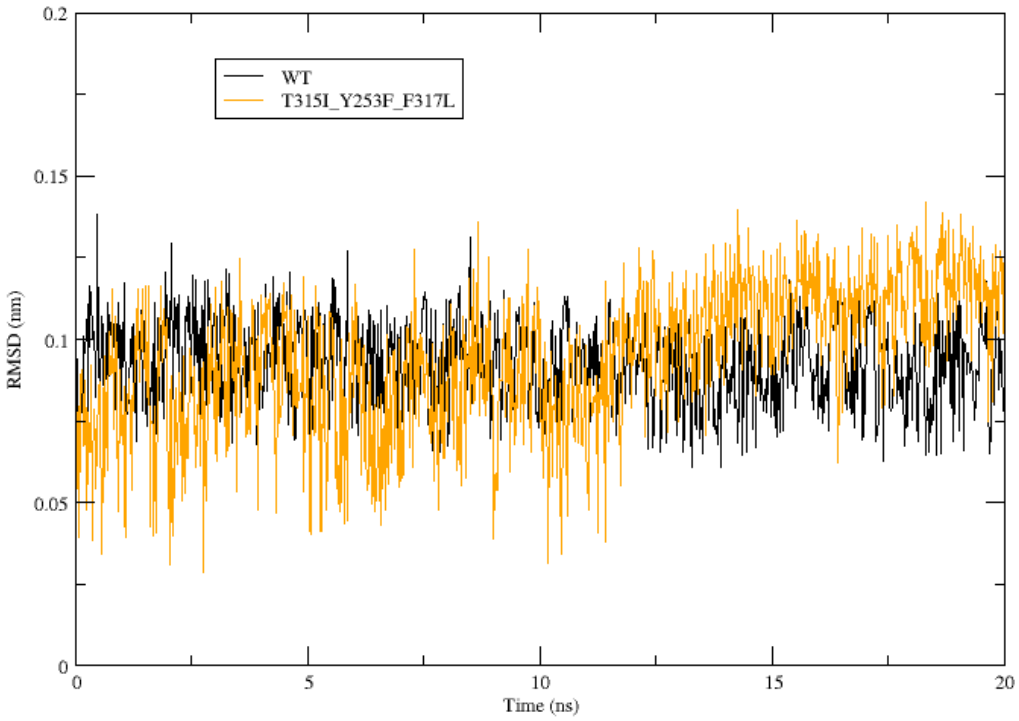


Figure 8.4.12: RMSD of nilotinib (3CS9$_{TYF}$)

RMSF analysis confirmed that fluctuations of the ligand were in line with that of the overall protein structure (Figure 8.4.14) and that the largest fluctuations were located at the trifluoromethyl and imidazole groups at the entrance to the binding site and the pyrimidine ring. Figure 8.4.15 illustrates the shift in the ligand position compared to the WT.

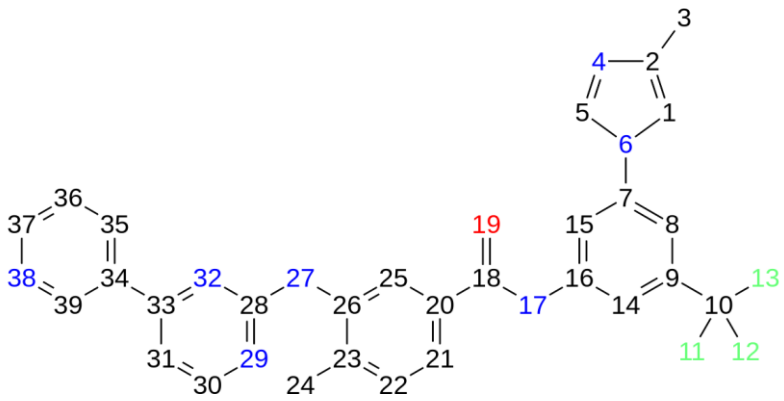


Figure 8.4.13: Index of atoms for nilotinib ($3CS9_{TYF}$)

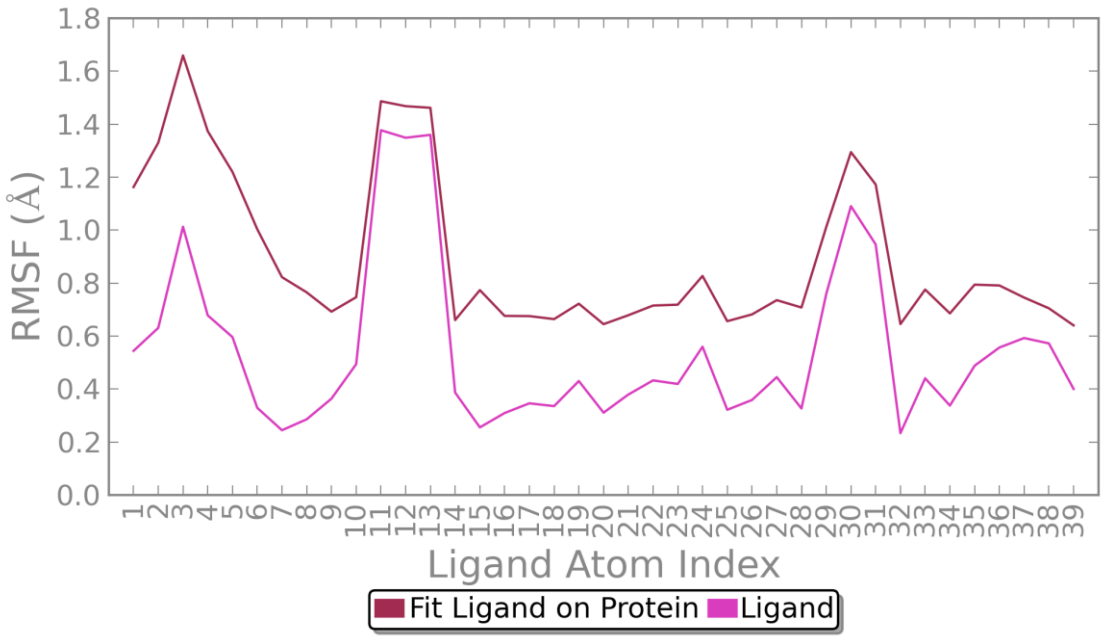


Figure 8.4.14: RMSF of nilotinib ($3CS9_{TYF}$)

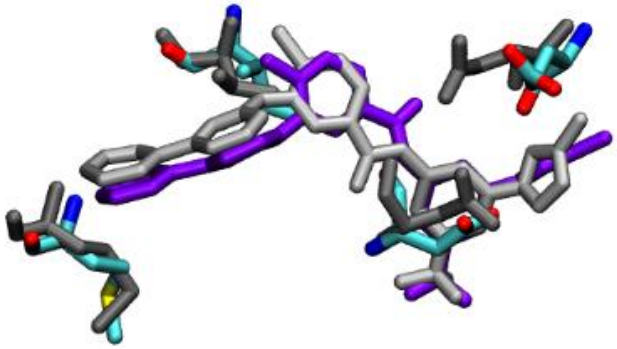


Figure 8.4.15: Comparison of nilotinib binding site in the WT and TYF mutant structures; grey/silver, WT; purple/cyan, TYF (15 ns)

Protein-ligand interactions

The triple mutation resulted in the loss of a hydrogen bond between nilotinib and Ile315 (Ile83 in Figure 8.4.16), and the loss of interaction with Leu317 (Leu85 in Figure 8.4.16), however some interaction were maintained with Phe253 (Phe21 in Figure 8.4.16).

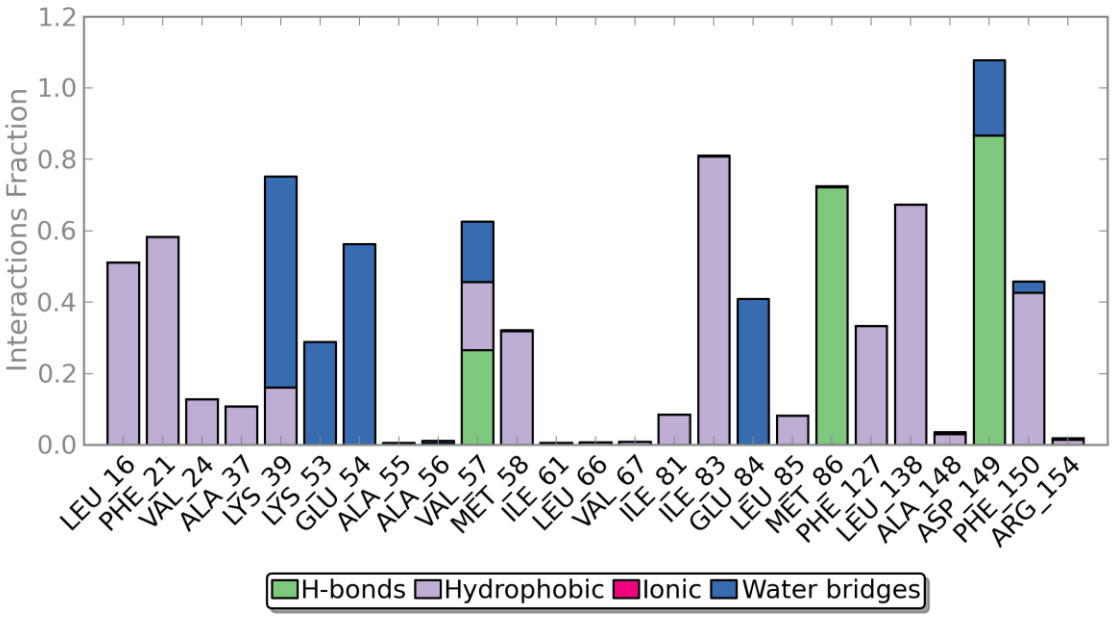


Figure 8.4.16: Protein-ligand interactions by residue (nilotinib, $3CS9_{TYF}$)

The new position of the ligand enabled a hydrogen bond between Val289 and nilotinib (Val57 in Figure 8.4.16)

8.5 Summary of molecular dynamics study

The analysis of the $1IEP_{YF}$ simulation suggests that the presence of both the Y253F and F317L mutations causes extensive disruption to the protein structure; increased flexibility was observed at the P-loop, DFG-loop and αC-helix. As in the single mutation study the initial loss of interaction disrupted the preferred induced fit of imatinib, causing the ligand to shift to a suboptimal position in the pocket. Imatinib appeared less stable in the pocket that in the WT simulation, but the new positioning enabled hydrogen bonding to be maintained with Asp381, Thr315, Glu286, Met290, Ile360 and His361.

Analysis of the $3CS9_{YF}$ simulation also suggests that the double mutation destabilises the kinase structure; due to the denaturation of the C-helix, the Glu286-Lys271 salt bridge is unformed at the start of the simulation and the Glu286 residue rotates away from the binding site. The loss of interaction also disrupted the conformation and flexibility of the P-loop but in contrast to the imatinib simulation the DFG-loop remained stable. The shift in the ligand conformation also resulted in the hydrogen bonds between nilotinib and the protein becoming diminished, particularly Glu286 and Met290; the ligand appeared to be becoming increasingly unstable despite the new position of nilotinib facilitating a stable water mediated hydrogen bond with Lys271.

The analysis of the $1IEP_{YF}$ simulation also revealed increased flexibility in the P-loop, however the DFG-loop and αC-helix were more stable than in the double mutation simulation. The ligand experienced a small in the pocket but appeared to be stable, however the new positioning resulted in diminished hydrogen bonding with Asp381.

The analysis of the $3CS9_{TYF}$ simulation suggested similar instability with the denaturation of the αC-helix resulting in the absence of the Glu286-Lys271 salt bridge and the rotation of Glu286 away from the binding site. As expected the triple mutation resulted in the loss of a hydrogen bond between nilotinib and residue 315, and the loss of interactions with residue 317, however some interactions were maintained with Phe253.

Given the instability demonstrated by the four simulations, the results suggest that the presence of the Y253F/F317L and T315I/Y253F/F317L compound mutations are not compatible with imatinib or nilotinib inhibition. The observed interactions with binding site residues are likely the result of increased fluctuation across the kinase structure, allowing for interactions to be

made that would be unlikely to form in the stable structure. Given the effects on the P-loop, DFG-loop and αC-helix the results suggest that the presence of the double and triple mutation could initiate the shift to the active conformation.

8.6 Ensemble docking

In Chapter 7, the results of the docking study with 3CS9 as a receptor correctly ranked the mutations according to experimental data, whereas the results from the docking study with 1IEP did not. Although experimental data for Y253F/F317L mutation is not currently available, based on the data that is available, it is estimated that the Y253F/F317L mutation is more resistant to imatinib than the single mutations F317L and Y253F, but not as resistant as the gatekeep mutation T315I.

8.6.1 Methodology

Imatinib and nilotinib were docked to the snapshot conformations of $1IEP_{Y253F/F317L}$, $1IEP_{T315I/Y253F/F317L}$, $3CS9_{Y253F/F317L}$ and $3CS9_{T315I/Y253F/F317L}$.

Docking simulations were completed as described in section 6.7.1.

8.6.2 Results

8.6.2.1 Y253F/F317L compound mutation

Imatinib

Table 8.6.1 and Table 8.6.2 compare the best GlideScores for imatinib docked to 1IEP and 3CS9 (WT, F317L, Y253F, T315I and YF) to what is known experimentally. Based on experimental data, the mutants were expected to be ranked in the following order: WT < F317L < Y253F < YF < T315I.

The results of the ensemble docking study using the 3CS9 snapshots as receptors correctly ranked the mutations according to experimental data (Table 8.6.2Table 7.7.2); however the docking scores from the 1IEP receptor study were not aligned with experimental data and the mutations were not correctly ranked in order of binding affinity as expected (Table 8.6.1).

	Experimental sensitivity	Experimental rank	GlideScore (kcal/mol)	GlideScore rank
WT		1	-13.40	4
F317L		2	-14.93	1
Y253F		3	-12.12	5
YF		4 (est)	-14.43	2
T315I		5	-13.96	3

Table 8.6.1: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental sensitivity	Experimental rank	GlideScore (kcal/mol)	GlideScore rank
WT		1	-14.78	1
F317L		2	-13.07	2
Y253F		3	-12.86	3
YF		4 (est)	-12.27	4
T315I		5	-11.71	5

Table 8.6.2: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Nilotinib

Table 8.6.3 and Table 8.6.4 compare the best GlideScores for nilotinib docked to the 1IEP and 3CS9 structures to what is known experimentally. The results should have correctly ranked the structures: WT < F317L < Y253F < YF < T315I. Furthermore the docking scores for nilotinib should have been better (more negative) than those for imatinib.

The results of the ensemble docking study using the 3CS9 snapshots as receptors correctly ranked the mutations according to experimental data and ranked nilotinib as a more potent inhibitor compared to imatinib (Table 8.6.4); however the docking scores from the 1IEP receptor study were not aligned with experimental data and the mutations were not correctly ranked in order of binding affinity as expected (Table 8.6.3).

	Experimental sensitivity	Experimental rank	GlideScore (1IEP) (kcal/mol)	GlideScore rank
WT		1	-13.78	4
F317L		2	-14.50	2
Y253F		3	-13.68	5
YF		4 (est)	-15.4	1
T315I		5	-14.17	3

Table 8.6.3: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental sensitivity	Experimental rank	GlideScore (3CS9) (kcal/mol)	GlideScore rank
WT		1	-15.41	1
F317L		2	-14.22	2
Y253F		3	-14.19	3
YF		4 (est)	-13.21	4
T315I		5	-12.97	5

Table 8.6.4: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

T315I/Y253F/F317L compound mutation

Imatinib

Table 8.6.5 and Table 8.6.6 compare the best GlideScores for imatinib docked to 1IEP and 3CS9 (WT, F317L, Y253F, T315I, YF and TYF) to what is known experimentally. The expected ranking of the mutants is: WT < F317L < Y253F < YF < T315I < TYF.

As with previous results, the docking scores using 1IEP were not in line with what was expected and the TYF mutation was not ranked as the most resistant mutation to imatinib (Table 8.6.5).

Interestingly, the docking score using 3CS9, which had previously produced results that were well aligned with experimental data, also did not rank TYF as the most resistant mutation to imatinib (Table 8.6.6).

	Experimental sensitivity	Experimental rank	GlideScore (kcal/mol)	GlideScore rank
WT		1	-13.40	5
F317L		2	-14.93	1
Y253F		3	-12.12	6
YF		4 (est)	-14.43	3
T315I		5	-13.96	4
TYF		6	-14.44	2

Table 8.6.5: GlideScores (best achieved) for imatinib docked to 1IEP (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental sensitivity	Experimental rank	GlideScore (kcal/mol)	GlideScore rank
WT		1	-14.78	1
F317L		2	-13.07	2
Y253F		3	-12.86	4
YF		4 (est)	-12.27	5
T315I		5	-11.71	6
TYF		6	-13.04	3

Table 8.6.6: GlideScores (best achieved) for imatinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM[331]). IC_{50} values from O'Hare et al, 2009 [333].

Nilotinib

Table 8.6.7 and Table 8.6.8 compare the best GlideScores for nilotinib docked to the 1IEP and 3CS9 structures to what is known experimentally. The results should have correctly ranked the structures: WT < F317L < Y253F < YF < T315I < TYF.

As with imatinib, the docking scores for nilotinib using the $1IEP_{TYF}$ and $3CS9_{TYF}$ structures were not aligned with the expected results (Table 8.6.7 and Table 8.6.8). In both cases the TYF mutation was ranked as less resistant than the single gatekeeper mutation.

	Experimental	Rank	GlideScore (1IEP)	Rank
WT		1	-13.78	5
F317L		2	-14.50	2
Y253F		3	-13.68	6
YF		4 (est)	-15.4	1
T315I		5	-14.17	3
TYF		6	-13.82	4

Table 8.6.7: GlideScores (best achieved) for nilotinib docked to 1IEP (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental	Rank	GlideScore (3CS9)	Rank
WT		1	-15.41	1
F317L		2	-14.22	3
Y253F		3	-14.19	4
YF		4 (est)	-13.21	5
T315I		5	-12.97	6
TYF		6	-14.65	2

Table 8.6.8: GlideScores (best achieved) for nilotinib docked to 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

8.6.3 Discussion

In the single mutation study (section 7.7) the docking scores from the 3CS9 dataset correctly ranked the mutants according to binding affinity. The results suggested that a docking score between -14.22 kcal/mol and -13.07 kcal/mol indicates moderate resistance to imatinib and nilotinib and a docking score of greater than -12.86kcal/mol, high resistance.

Experimental data is currently not available for the double mutation Y253F/F317L, but based on previous research it is estimated that this mutation will confer a higher level of resistance compared to the single mutations Y253F and F317L. The docking results for imatinib/nilotinib and the $3CS9_{Y253F/F317L}$ were in line with this estimate; GlideScores for imatinib/nilotinib were poorer when docked to $3CS9_{Y253F/F317L}$ than to $3CS9_{Y253F}$ and $3CS9_{F317L}$.

The docking scores for imatinib and nilotinib docked to $3CS9_{Y253F/F317L}$ were -12.27 kcal/mol and -13.21 kcal/mol respectively; using the single mutations as a guide we can estimate that the double compound mutation is highly resistant to imatinib and moderately resistant to nilotinib.

The triple mutation was not ranked as expected in the 3CS9 study (Figure 8.6.1). Previous research has demonstrated that the presence of the gatekeeper mutation in compound mutations results in extremely high resistance to both imatinib and nilotinib. Given that the results from the single and double compound mutations are considered to be sensible, this could mean that: a) the triple mutation is less resistant that expected, which goes against current

evidence b) there is an error in the simulation, or c) there is an artefact in this simulation that reduces the accuracy of the docking estimates.

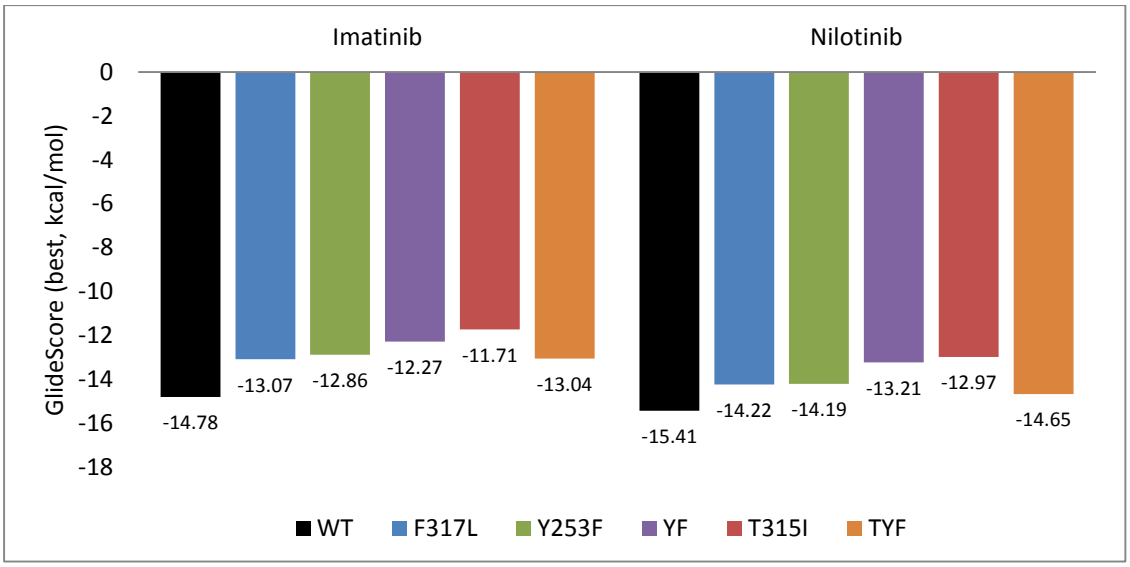


Figure 8.6.1: Summary of docking study using 3CS9 as the receptor

The results from the 1IEP single mutation study (section 7.7) were not well aligned with experimental data or the results of the 3CS9 study. As discussed in Chapters 6 and 7, this could be due to the omission of a crystallographic water molecule, W2, from the protein receptor structures; the exclusion of conserved waters is known to cause inaccuracies in the prediction of binding affinity values and docking scores. Consequently, the results for imatinib and nilotinib docked to the compound mutations $1IEP_{Y253F/F317L}$ and $1IEP_{T315I/Y253F/F317L}$ are not considered to be reliable.

8.7 MMGBSA calculations

8.7.1 Methodology

MM-GBSA calculations were performed on the 1IEP and 3CS9 trajectories were completed as described in section 6.8.1.

8.7.2 Results

8.7.2.1 Y253F/F317L compound mutation

The energy components of the MM-GBSA results for imatinib and 1IEP are displayed in Table 8.7.1. Table 8.7.2 compares the rankings from the MM-GBSA values to the experimental rankings based on IC_{50} values. Based on what is known from published literature, the mutants were expected to be ranked in the following order: WT < F317L < Y253F < YF < T315I.

As observed in the single mutation data set, the MMGBSA results for 1IEP were not well aligned with experimental data and the Y253F/F317L mutation was ranked as less resistant than the single mutations F317L and Y253F (Table 8.7.1 and Table 8.7.2).

Energy components (kcal/mol)	WT	F317L	Y253F	T315I	YF
Total ligand energy	-144.95	-145.99	-147.91	-147.35	-148.12
Total receptor energy	-6368.80	-6076.51	-6357.99	-6341.49	-6028.80
Total complex energy	-6586.41	-6317.04	-6580.38	-6563.31	-6265.74
Difference	-72.66	-94.53	-74.48	-74.48	-88.83

Table 8.7.1: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF) with imatinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-72.66	5
F317L		2	-94.53	1
Y253F		3	-74.48	=3
YF		4 (est)	-88.83	2
T315I		5	-74.48	=3

Table 8.7.2: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

The energy components of the MM-GBSA results for nilotinib and 3CS9 are displayed in Table 8.7.3. Table 8.7.4 compares the MMGBSA values to what is known experimentally. As with imatinib, the mutants were expected to be ranked in the following order: WT < F317L < Y253F < YF < T315I.

The MM-GBSA results for 3CS9 correctly ranked the Y253F/F317L mutation as more resistant than the single mutations F317L and Y253F (Table 8.7.4).

Energy components (kcal/mol)	WT	F317L	Y253F	T315I	YF
Total ligand energy	-223.99	-221.62	-223.40	-224.63	-223.04
Total receptor energy	-6413.74	-6422.64	-6380.38	-6372.00	-6378.58
Total complex energy	-6712.63	-6706.84	-6666.73	-6650.87	-6658.23
Difference	-74.90	-62.58	-62.95	-54.25	-56.61

Table 8.7.3: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF) with nilotinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-74.90	1
F317L		2	-62.58	3
Y253F		3	-62.95	2
YF		4 (est)	-56.61	4
T315I		5	-54.25	5

Table 8.7.4: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF) with nilotinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

8.7.2.2 F317L/Y253F/T315I compound mutation

Imatinib + 1IEP

The energy components of the MM-GBSA results for imatinib and 1IEP are displayed in Table 8.7.5. Table 8.7.6 compares the rankings from the MM-GBSA values to the experimental

rankings based on IC_{50} values. Based on what is known from published literature, the TYF mutant was estimated to be the most resistant mutation to imatinib.

As in the docking study for imatinib and 1IEP, the MM-GBSA results using 1IEP were not in line with what was expected and the T315I/Y253F/F317L mutation was not ranked as the most resistant mutation to imatinib (Table 8.7.6).

Energy components (kcal/mol)	WT	F317L	Y253F	T315I	YF	TYF
Total ligand energy	-144.95	-145.99	-147.91	-147.35	-148.12	-149.23
Total receptor energy	-6368.80	-6076.51	-6357.99	-6341.49	-6028.80	-5987.11
Total complex energy	-6586.41	-6317.04	-6580.38	-6563.31	-6265.74	-6220.72
Difference	-72.66	-94.53	-74.48	-74.48	-88.83	-84.38

Table 8.7.5: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF, TYF) with imatinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-72.66	5
F317L		2	-94.53	1
Y253F		3	-74.48	=4
YF		4 (est)	-88.83	2
T315I		5	-74.48	=4
TYF		6	-84.38	3

Table 8.7.6: MM-GBSA results for 1IEP (WT, F317L, Y253F, T315I, YF, TYF) with imatinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Nilotinib + 3CS9

The energy components of the MM-GBSA results for nilotinib and 3CS9 are displayed in Table 8.7.7. Table 8.7.8 compares the rankings from the MM-GBSA values to the experimental

rankings based on IC_{50} values. Based on what is known from published literature, the TYF mutant was estimated to be the most resistant mutation to imatinib.

As was observed in the docking study for nilotinib and 3CS9, the MM-GBSA results did not rank the TYF mutation as the most resistant mutation to nilotinib (Table 8.7.8).

Energy components (kcal/mol)	WT	F317L	Y253F	T315I	YF	TYF
Total ligand energy	-223.99	-221.62	-223.40	-224.63	-223.04	-222.19
Total receptor energy	-6413.74	-6422.64	-6380.38	-6372.00	-6378.58	-6365.57
Total complex energy	-6712.63	-6706.84	-6666.73	-6650.87	-6658.23	-6647.73
Difference	-74.90	-62.58	-62.95	-54.25	-56.61	-59.97

Table 8.7.7: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) with nilotinib

	Experimental sensitivity	Experimental rank	MM-GBSA (Difference, kcal/mol)	MM-GBSA rank
WT		1	-74.90	1
F317L		2	-62.58	3
Y253F		3	-62.95	2
YF		4 (est)	-56.61	5
T315I		5	-54.25	6
TYF		6	-59.97	4

Table 8.7.8: MM-GBSA results for 3CS9 (WT, F317L, Y253F, T315I, YF, TYF) with nilotinib vs. experimental rankings. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

8.7.3 Discussion

Both previous research and the results of the docking study suggested that the double mutation Y253F/F317L confers a higher level of resistance compared to the single mutations Y253F and F317L. The MM-GBSA results for nilotinib and $3CS9_{Y253F/F317L}$ support this estimate; MM-GBSA values for nilotinib and $3CS9_{Y253F/F317L\ were}$ poorer than for $3CS9_{Y253F}$ and $3CS9_{F317L}$.

The triple mutation was not ranked as expected in the 3CS9 study. As stated previously in section 8.6.3, given the results of from the single mutants and double mutant, this could mean that: a) the triple mutation is less resistant that expected, b) there is an error in the simulation, c) there is an artefact in this simulation that reduces the accuracy of the MM-GBSA estimates.

The MM-GBSA results from the 1IEP single mutation study were not well aligned with experimental data or the results of the 3CS9 study. As discussed in Chapters 6 and 7, this could be due to the omission of a crystallographic water molecule, W_2, from the protein receptor structures; the exclusion of conserved waters is known to cause inaccuracies in the prediction of binding affinity values and docking scores. Consequently, the MMGBSA results for imatinib and the compound mutations $1IEP_{Y253F/F317L}$ and $1IEP_{T315I/Y253F/F317L}$ are not considered to be reliable.

8.8 Chapter summary

The aim of this chapter was to investigate the effect of two compound mutations, Y253F/F317L and T315I/Y253F/F317L, on the conformation of Abl and the binding modes of imatinib and nilotinib. Having validated the methods by reproducing experimental rankings in Chapter 7, docking and MMGBSA calculations were utilised to predict the level of resistance conferred by the compound mutations. In the single mutation study it was shown that a single occurrence of these mutations has dramatic effects on the dynamics of the kinase as well as protein-ligand interactions. The Y253F and F317L mutations remain sensitive to nilotinib as single mutations, however recent studies have shown that the combination of mutations at these positions may confer resistance to nilotinib [331]. IC_{50} experimental data is not available for either of the compound mutations studied; however, based on previous research, the double mutation Y253F/F317L, was expected to confer a higher level of resistance compared to the Y253F and F317L mutations, and previous studies on double mutations that include T315I as a component have all demonstrated a high level of resistance to imatinib and nilotinib [330, 331]. Both compound mutations were expected to confer resistance to imatinib and nilotinib, and the triple mutation was predicted to be ranked the most resistant.

The results of the molecular dynamics study suggest that the presence of the Y253F/F317L and T315I/Y253F/F317L compound mutations are not compatible with imatinib or nilotinib inhibition. Both compound mutations, the double mutation in particular, caused significant disruption to the kinase structure and although hydrogen bonding was maintained, the observed interactions are likely the result of increased fluctuation across the kinase structure, allowing for interactions to be made that would be unlikely to form in the stable structure. Given the effects on the P-loop, DFG-loop and αC-helix the results suggest that the presence of the double and triple mutation could initiate the shift to the active conformation.

The results of the MM-GBSA and docking studies for are summarised in Table 8.8.1 (imatinib) and Table 8.8.2 (nilotinib). The results for $3CS9_{Y253F/F317L}$ support the estimate that the double mutation would be more resistant to imatinib and nilotinib than the Y253F and F317L single mutants. The best achieved GlideScores for imatinib and nilotinib docked to $3CS9_{Y253F/F317L}$ were -12.27 kcal/mol and -13.21 kcal/mol respectively; using the single mutations as a guide it is estimated that the double compound mutation is highly resistant to imatinib and moderately resistant to nilotinib.

For $3CS9_{T315I/Y253F/F317L}$ the MM-GBSA and docking method did not rank the triple compound mutation as the most resistant mutation as expected (Figure 8.6.1). Given that the results from the single and double compound mutations are considered to be sensible, this could mean that: a) the triple mutation is less resistant that expected, which goes against current evidence b) there is an error in the simulation, or c) there is an artefact in this simulation that reduces the accuracy of the docking estimates. In the absence of experimental data, repeating the study to include a larger and more varied dataset would be useful to inform further work on compound mutations.

Given that the results from the 1IEP single mutation study (section 7.7) were not well aligned with experimental data or the results of the 3CS9 study, the results for imatinib and nilotinib docked to the compound mutations $1IEP_{Y253F/F317L}$ and $1IEP_{T315I/Y253F/F317L}$ are not considered to be reliable.

	Experimental		Ensemble docking (GlideScore, kcal/mol)				MM-GBSA (kcal/mol)	
	Sensitivity	Rank	1IEP	Rank	3CS9	Rank	Difference	Rank
WT		1	-13.40	5	-14.78	1	-72.66	5
F317L		2	-14.93	1	-13.07	2	-94.53	1
Y253F		3	-12.12	6	-12.86	4	-74.48	=4
YF		4	-14.43	3	-12.27	5	-88.83	2
T315I		5	-13.96	4	-11.71	6	-74.48	=4
TYF		6	-14.44	2	-13.04	3	-84.38	3

Table 8.8.1: Results summary – Imatinib, single and compound mutation studies. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to imatinib (green: <1000 nM; orange: 1000-4000 nM; red: >4000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

	Experimental		Ensemble docking (GlideScore, kcal/mol)				MM-GBSA (kcal/mol)	
	Sensitivity	Rank	1IEP	Rank	3CS9	Rank	Difference	Rank
WT		1	-13.78	5	-15.41	1	-74.90	1
F317L		2	-14.50	2	-14.22	3	-62.58	3
Y253F		3	-13.68	6	-14.19	4	-62.95	2
YF		4	-15.4	1	-13.21	5	-56.61	4
T315I		5	-14.17	3	-12.97	6	-54.25	5
TYF		6	-13.82	4	-14.65	2	-74.90	1

Table 8.8.2: Results summary – Nilotinib, single and compound mutation studies. A colour gradient from green (sensitive) to orange (moderately resistant) to red (highly resistant) denotes the IC_{50} sensitivity to nilotinib (green: <200 nM; orange: 200-1000; nM; red: >1000 nM [331]). IC_{50} values from O'Hare et al, 2009 [333].

Chapter 9

9 Conclusions

Imatinib is a highly successful Abl kinase inhibitor widely used in the treatment of CML, however in a small percentage of CML patients point mutations develop in the Abl kinase resulting in varying levels of drug resistance. Nilotinib is a more a potent inhibitor of the Abl kinase than imatinib and is capable of inhibiting the majority of imatinib-resistant mutations, however mutations that are resistant to both imatinib and nilotinib persist in patients; these include double and triple compound mutations which can confer a much higher level of resistance than the individual single mutations. Further understanding regarding how mutations confer resistance and the ability to predict which compound mutations are likely to be highly resistant would enable drug design initiatives, both in CML and other cancers, to be improved in the future.

The aim of this research was to study the effect of 5 mutations (three single, one double and one triple) on the conformation of the Abl kinase domain and the binding modes of imatinib and nilotinib using computational methods. Previous research suggested that mutations affect the binding of these inhibitors via the loss of interactions in the binding site together with subtle conformational changes that destabilise the inactive conformation.

Molecular dynamics, using the NAMD molecular dynamics software, has been used to simulate the dynamics and flexibility of the wild-type and mutant protein structures. Trajectory (RMSD, RMSF and PCA) and binding site analysis methods (PocketAnalyzer, SID, ptraj and JAWS) have been applied to the subsequent MD trajectories to identify differences between the structures. To predict the relative resistance of mutations, protein-ligand docking studies have been completed for the wild-type and mutant structures via the docking program Glide, using an ensemble of protein conformations from the MD trajectories. In order to support the docking results, the binding free energy of the wild-type and mutant structures have been calculated from the MD trajectories using the MM-GBSA method.

In Chapter 5 the structure and dynamics of the active and inactive conformations were investigated using molecular dynamics simulations to confirm the stable characteristics of the active state. This was to act as a benchmark to the inactive and mutant studies and validate the molecular dynamics methods.

In Chapter 6, the structure and dynamics of the inactive conformation were investigated using molecular dynamics simulations to confirm the stable characteristics of the inactive state and compare the binding modes of imatinib and nilotinib. The differences in protein-ligand binding were then further investigated using ensemble docking and approximate free energy calculations (MD-GBSA). The results of the interaction analysis showed that nilotinib is the more stable of the two ligands and forms strong interactions with the kinase, supporting what is known experimentally regarding nilotinib being a superior inhibitor compared to imatinib; this was also supported by the docking and MM-GBSA results that correctly ranked nilotinib as having improved binding affinity compared to imatinib.

The methods described in Chapters 5 and 6 were then utilised to study the three single mutations T315I, Y253F and F317L on the imatinib and nilotinib in Chapter 7; molecular dynamics and subsequent analysis were utilised to understand the conformational changes induced by the mutation; binding site analysis enabled any loss of interaction with the ligand to be identified; and the docking and MMGBSA studies enabled the ranking of mutations, offering insight into how the binding affinities compare. The results from the docking and MM-GBSA studies of single mutants using mutated 3CS9 as the protein receptor correctly ranked the mutants in order of resistance according to experimental data and showed poorer docking scores for imatinib compared to nilotinib as expected, given nilotinib is a more potent inhibitor. The docking and MM-GBSA results using the 1IEP structures were poorly aligned with experimental data and the mutations were not correctly ranked.

The reasons behind this inconsistency could be either due to poor performance from the docking/MM-GBSA software or an error in the simulation. On review of the 1IEP MD simulation it was found that two water molecules interacted consistently with imatinib in the simulation. Further investigation and water molecule analysis suggested that including the W_1 and W_2 water molecules would have been appropriate for the 1IEP simulations, however there is no conclusive evidence to confirm that this is and not some other error source is the driver behind the unexpected docking and MM-GBSA results.

Having validated the methods in the single mutation study, the effects of the compound mutations Y253F/F317L and T315I/Y253F/F317L were investigated in Chapter 8. The analysis showed that the double mutation significantly destabilised the inactive conformation, considerably more than the individual mutations, in both imatinib and nilotinib complexes. The docking and MM-GBSA studies confirmed these results, ranking the Y253F/F317L mutation as more resistant than the single mutations Y253F and F317L. This has significant implications for the treatment of CML and other diseases where resistance can develop through mutations in

the kinase receptor; mutations that are considered sensitive to a given inhibitor, or those classified as moderately resistant, may become highly resistant when combined with another mutation. Based on the results for the single mutations, the docking results for the Y253F/F317L mutation suggested that this double mutation would confer moderate resistance to nilotinib and a high level of resistance to imatinib.

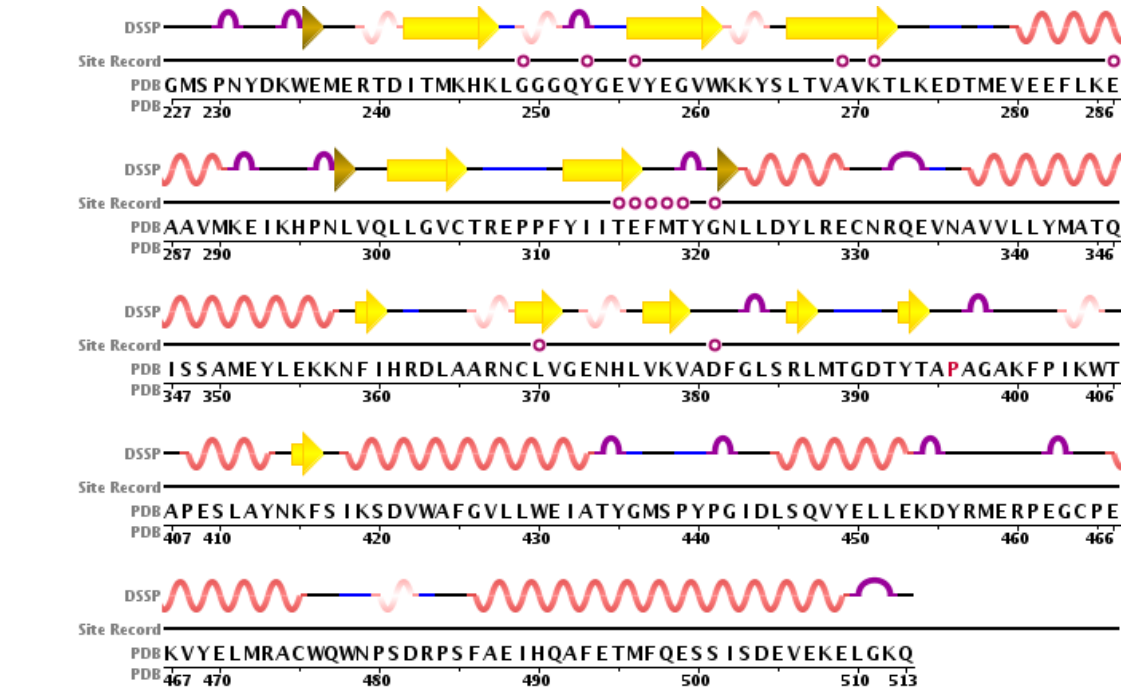
Although experimental results for the T315I/Y253F/F317L mutation are not available, previous research has reported that compound mutations that include the T315I mutation are highly resistant. Interestingly the docking results placed the triple mutation at a similar ranking to the single Y253F and F317L mutations and not as resistant as the single gatekeeper mutation. Given that the results from the single and double compound mutations are considered to be sensible, this suggests that either there is an error or artefact in the triple mutant simulation that means we are not seeing an accurate result, or that the triple mutation is in fact less resistant than expected. In the absence of experimental data, repeating the study to include a larger and more varied dataset would be useful to inform further work on compound mutations.

There are of course limitations with the methods discussed and as stated previously further work on additional single and compound mutations in Abl as well as other systems should be explored to validate this approach further; including a complex with ponatinib for example would be a valuable addition. One important limitation of the research discussed in this thesis is the omission of the conserved water molecule in the 1IEP simulations; there is a risk that aside from distorting the docking and free-energy estimations, not including this water molecule at the set up stage could significantly affect the interactions in the binding site. To confirm this, multiple simulations using the 1IEP structure, with and without water molecules W_1 and W_2, as well as alternative structures of the Abl kinase domain, such as 1OPJ, would need to be completed. This was unfortunately not possible to complete as part of this research due to time constraints.

In summary, this research has given insight into the conformational and binding effects of clinically relevant mutations in CML, using a mixed method approach and supporting what is known experimentally. The docking and MMGBSA methods used offer a fast method to rank mutations for various ligands, and this gives an indication of the level of resistance which can then be further investigated.

Appendices

Appendix A: The sequence and secondary structure of 2F4J.pdb



Site Record Legend

BINDING SITE FOR RESIDUE VX6 A 514 (SOFTWARE)

DSSP Legend

E: beta strand

T: turn

empty: no secondary structure assigned

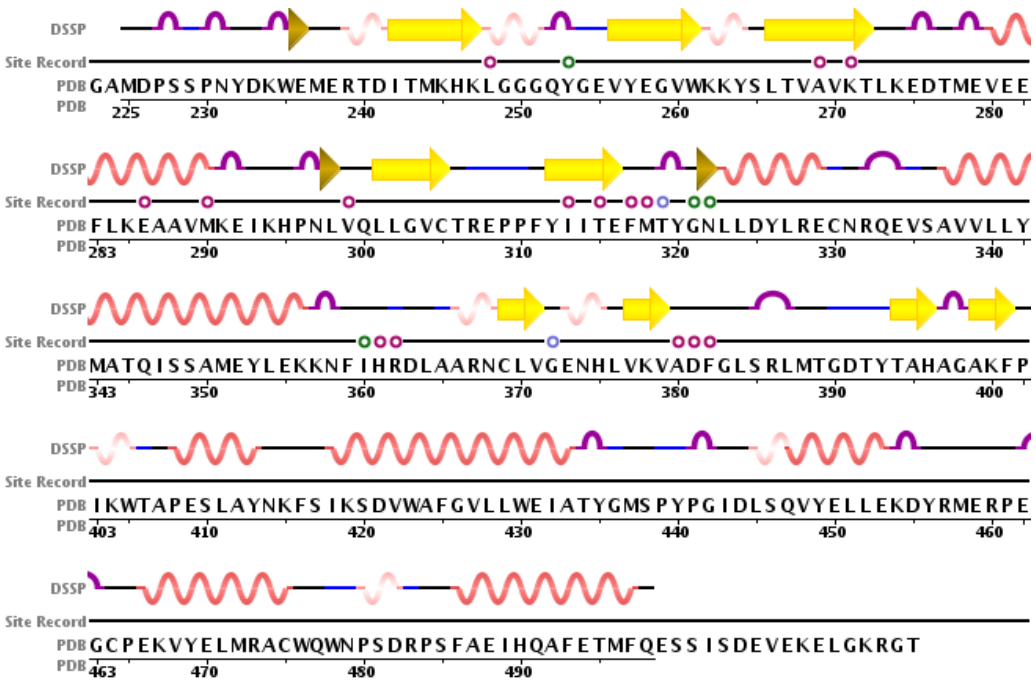
G: 3/10-helix

S: bend

B: beta bridge

H: alpha helix

Appendix B: The sequence and secondary structure of 1IEP.pdb



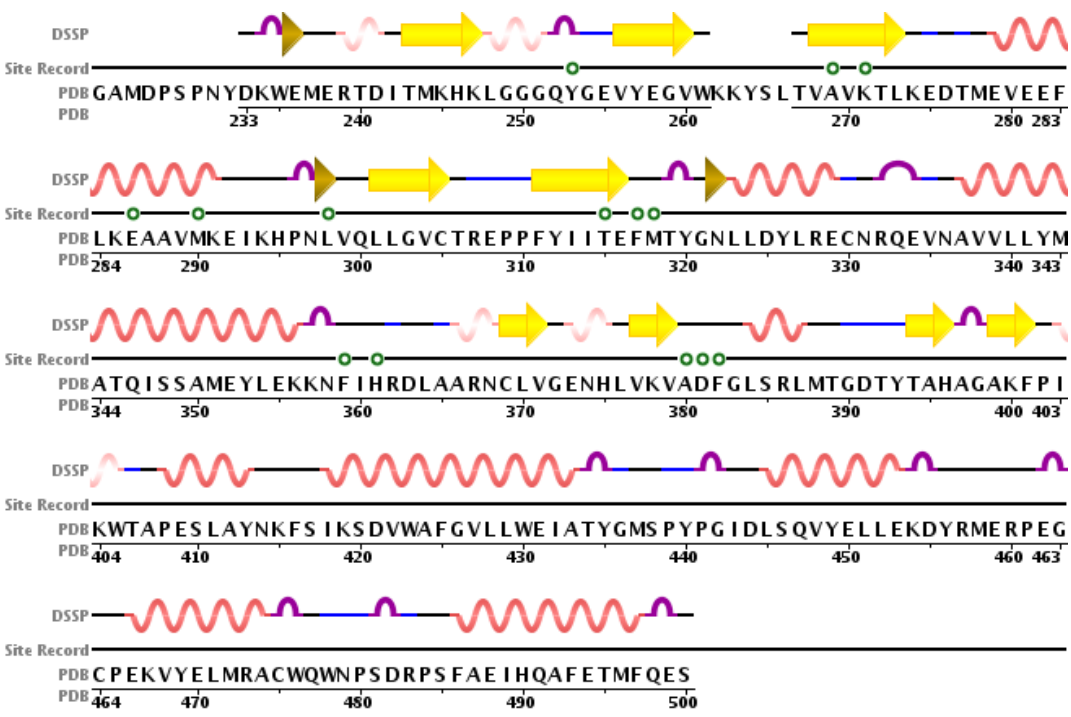
Site Record Legend

- BINDING SITE FOR RESIDUE CL A 4 (SOFTWARE)
- BINDING SITE FOR RESIDUE STI A 201 (SOFTWARE)
- BINDING SITE FOR RESIDUE CL A 2 (SOFTWARE)
- BINDING SITE FOR RESIDUE CL A 5 (SOFTWARE)

DSSP Legend

- E: beta strand
- T: turn
- empty: no secondary structure assigned
- G: 3/10-helix
- B: beta bridge
- S: bend
- H: alpha helix

Appendix C: The sequence and secondary structure of 3CS9.pdb



Site Record Legend

BINDING SITE FOR RESIDUE NIL D 600 (SOFTWARE)

DSSP Legend

E: beta strand

T: turn

empty: no secondary structure assigned

G: 3/10-helix

S: bend

B: beta bridge

H: alpha helix

Appendix D: 1ATP vs 2F4J (Needleman-Wunsch Sequence Alignment)

```
Query:      1ATP chain: E, Length: 350
Subject:    2F4J chain: A, Length: 287
Identities: 86/364, i.e., 24.57 % (query) and 29.97 % (subject)
Similars:   142/364, i.e., 40.57 % (query) and 49.48 % (subject)

    1ATP.E   1 GNAAAAKKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDRIKTLGTGSFGRVMLV  60
               | .                     |   |||    .     .      || | .| |
    2F4J.A   1 GMSP-----------------NYD---KWEMERTDITMKHK------LGGGQYGEVYEG  33

    1ATP.E  61 KHKESGNHYAMKILDKQKVVKLKQIEHTLNEKRILQAVNFPFLVKLEFSFKDNSNLYMVM 120
                 |.     |.| | .  .    ..|  | |  ... .  | ||.|          |..
    2F4J.A  34 VWKKYSLTVAVKTLKEDTM----EVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIIT  89

    1ATP.E 121 EYVAGGEMFSHLRRIGRFSEPHA---RFYAAQIVLTFEYLHSLDLIYRDLKPENLLIDQQ 177
               |..  | .  .||   |  | .|    . | ||    |||   . |.|||   | |. .
    2F4J.A  90 EFMTYGNLLDYLRECNR-QEVNAVVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGEN 148

    1ATP.E 178 GYIQVTDFGFAKRVKGRTWTL-CGT--P-EYLAPEIILSKGYNK---AVDWWALGVLIYE 230
                 ..| ||| .. . | |.|   |   | .. |||   |  |||     | || |||..|
    2F4J.A 149 HLVKVADFGLSRLMTGDTYTAPAGAKFPIKWTAPE---SLAYNKFSIKSDVWAFGVLLWE 205

    1ATP.E 231 MAA-GYPPFFADQPIQIYEKIVSG-KVRFPSHFSSDLKDLLRNLLQVDLTKR--FGNLKN 286
               .|  |  |.      |.|| .    ..  |      . .|.|   | . . |  |  .
    2F4J.A 206 IATYGMSPYPGIDLSQVYELLEKDYRMERPEGCPEKVYELMRACWQWNPSDRPSFAEI-- 263

    1ATP.E 287 GVNDIKNHKWFATTDWIAIYQRKVEAPFIPKFKGPGDTSNFDDYEEEEIRVSINEKCGKE 346
                     |. | |     ..|                 |.  |  |.|.        ||.
    2F4J.A 263 -------HQAFET-----MFQE----------------SSISDEVEKEL--------GKQ 287

    1ATP.E 347 FTEF 350

    2F4J.A 287 ---- 287
```

Legend: *Green - identical residues | Pink - similar residues | Blue - sequence mismatch | Brown - insertion/deletion |*

Appendix E: 2F4J vs 1IEP (Needleman-Wunsch Sequence Alignment)

```
Query:      2F4J chain: A, Length: 287
Subject:    1IEP chain: A, Length: 293
Identities: 283/293, i.e., 98.61 % (query) and 96.59 % (subject)
Similars:   285/293, i.e., 99.30 % (query) and 97.27 % (subject)

    2F4J.A   1 GM----SPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEE  56
               |     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
    1IEP.A   1 GAMDPSSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEE  60

    2F4J.A  57 FLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVNAVVLLY 116
               ||||||||||||||||||||||||||||||||||||||||||||||||||||.||||||
    1IEP.A  61 FLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVSAVVLLY 120

    2F4J.A 117 MATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAPAGAKFP 176
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
    1IEP.A 121 MATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFP 180

    2F4J.A 177 IKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPE 236
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    1IEP.A 181 IKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPE 240

    2F4J.A 237 GCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQESSISDEVEKELGKQ-- 287
               ||||||||||||||||||||||||||||||||||||||||||||||||||.
    1IEP.A 241 GCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQESSISDEVEKELGKRGT 293
```

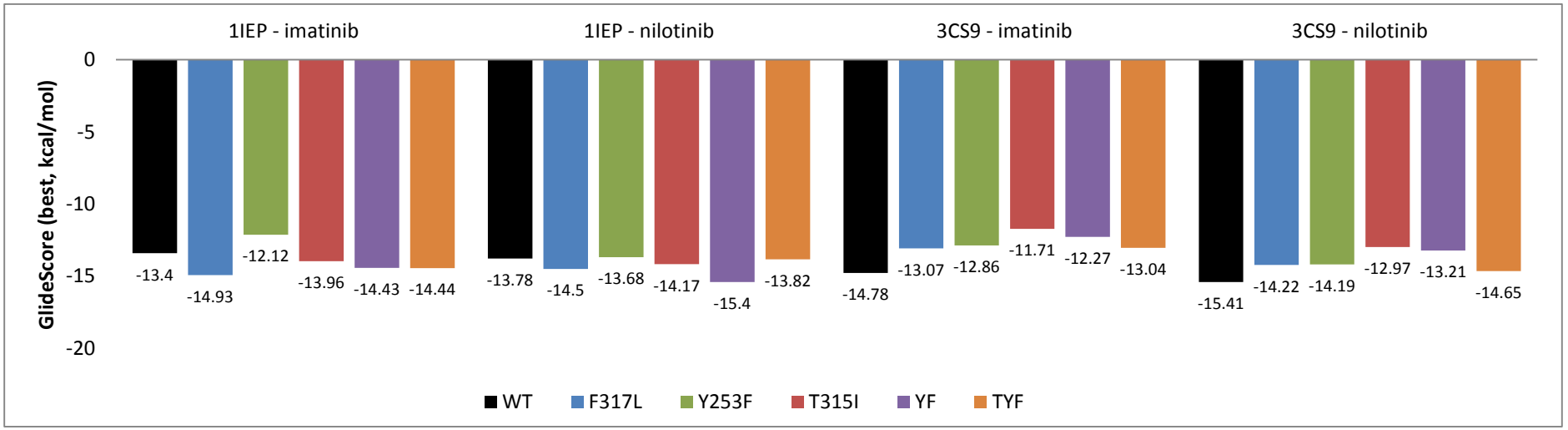
Legend: *Green - identical residues | Pink - similar residues | Blue - sequence mismatch | Brown - insertion/deletion |*

Appendix F: 1IEP vs 3CS9 (Needleman-Wunsch Sequence Alignment)

```
Query:      1IEP chain: A, Length: 293
Subject:    3CS9 chain: A, Length: 277
Identities: 276/293, i.e., 94.20 % (query) and 99.64 % (subject)
Similars:   277/293, i.e., 94.54 % (query) and 100.00 % (subject)

    1IEP.A   1 GAMDPSSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEE  60
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
    3CS9.A   1 GAMDPS-PNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEE  59

    1IEP.A  61 FLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVSAVVLLY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||.||||||
    3CS9.A  60 FLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVNAVVLLY 119

    1IEP.A 121 MATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFP 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    3CS9.A 120 MATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFP 179

    1IEP.A 181 IKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPE 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    3CS9.A 180 IKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPE 239

    1IEP.A 241 GCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQESSISDEVEKELGKRGT 293
               |||||||||||||||||||||||||||||||||||||
    3CS9.A 240 GCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQES--------------- 277
```

Legend: *Green - identical residues | Pink - similar residues | Blue - sequence mismatch | Brown - insertion/deletion |*

Appendix G: Summary of docking studies (best score achieved)



Bibliography

1. Carter, T.A., et al., *Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(31): p. 11011-11016.
2. Nardi, V., M. Azam, and G.Q. Daley, *Mechanisms and implications of imatinib resistance mutations in BCR-ABL.* Current Opinion in Hematology, 2004. **11**(1): p. 35-43.
3. Kobayashi, S., et al., *EGFR mutation and resistance of non-small-cell lung cancer to gefitinib.* N Engl J Med, 2005. **352**(8): p. 786-92.
4. Bikker, J.A., et al., *Kinase Domain Mutations in Cancer: Implications for Small Molecule Drug Design Strategies.* Journal of Medicinal Chemistry, 2009. **52**(6): p. 1493-1509.
5. Daub, H., K. Specht, and A. Ullrich, *Strategies to overcome resistance to targeted protein kinase inhibitors.* Nature Reviews Drug Discovery, 2004. **3**(12): p. 1001-1010.
6. O'Hare, T., et al., *In vitro activity of Bcr-Abl inhibitors AMN107 and BMS-354825 against clinically relevant imatinib-resistant Abl kinase domain mutants.* Cancer Res, 2005. **65**(11): p. 4500-5.
7. A, J.Y., et al., *Chronic Myeloid Leukemia Patients Sensitive and Resistant to Imatinib Treatment Show Different Metabolic Responses.* Plos One, 2010. **5**(10).
8. Heinrich, M.C., et al., *Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor.* J Clin Oncol, 2003. **21**(23): p. 4342-9.
9. Pricl, S., et al., *T315I-mutated Bcr-Abl in chronic myeloid leukemia and imatinib: insights from a computational study.* Mol Cancer Ther, 2005. **4**(8): p. 1167-74.
10. Liu, B., B. Bernard, and J.H. Wu, *Impact of EGFR point mutations on the sensitivity to gefitinib: insights from comparative structural analyses and molecular dynamics simulations.* Proteins-Structure Function and Bioinformatics, 2006. **65**(2): p. 331-46.
11. Hao, G.F., G.F. Yang, and C.G. Zhan, *Computational mutation scanning and drug resistance mechanisms of HIV-1 protease inhibitors.* Journal of Physical Chemistry B, 2010. **114**(29): p. 9663-76.
12. Levinson, N.M., et al., *A Src-like inactive conformation in the abl tyrosine kinase domain.* PLoS Biol, 2006. **4**(5): p. e144.
13. Aleksandrov, A. and T. Simonson, *Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases.* J Biol Chem, 2010. **285**(18): p. 13807-15.
14. Yang, L.J., et al., *Steered Molecular Dynamics Simulations Reveal the Likelier Dissociation Pathway of Imatinib from Its Targeting Kinases c-Kit and Abl.* Plos One, 2009. **4**(12): p. -.
15. Li, Y.Y., J. An, and S.J. Jones, *A large-scale computational approach to drug repositioning.* Genome Inform, 2006. **17**(2): p. 239-47.
16. Coleman, R.G., A.C. Salzberg, and A.C. Cheng, *Structure-based identification of small molecule binding sites using a free energy model.* Journal of Chemical Information and Modeling, 2006. **46**(6): p. 2631-7.
17. Jackson, R.M. and A.T.R. Laurie, *Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites.* Bioinformatics, 2005. **21**(9): p. 1908-1916.
18. Shan, Y., et al., *A conserved protonation-dependent switch controls drug binding in the Abl kinase.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 139-44.

19. Lin, J.H., et al., *Computational drug design accommodating receptor flexibility: the relaxed complex scheme.* Journal of the American Chemical Society, 2002. **124**(20): p. 5632-3.
20. Michel, J. and J.W. Essex, *Hit identification and binding mode predictions by rigorous free energy simulations.* Journal of Medicinal Chemistry, 2008. **51**(21): p. 6654-64.
21. Lee, T.S., et al., *Molecular basis explanation of imatinib resistance of Bcr-Abl due to T3151 and P-loop mutations from molecular dynamics simulations.* Blood, 2007. **110**(11): p. 858a-858a.
22. Lee, T.S., S.J. Potts, and M. Albitar, *Basis for Resistance to Imatinib in 16 BCR-ABL Mutants as Determined Using Molecular Dynamics.* Recent Patents on Anti-Cancer Drug Discovery, 2009. **4**(2): p. 164-173.
23. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.
24. Leach, A.R., *Molecular modelling : principles and applications*. 2nd ed. ed. 2001, Harlow: Prentice Hall. xxiv, 744 p.
25. Halperin, I., et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions.* Proteins-Structure Function and Bioinformatics, 2002. **47**(4): p. 409-43.
26. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein-ligand docking: current status and future challenges.* Proteins-Structure Function and Bioinformatics, 2006. **65**(1): p. 15-26.
27. Case, D.A., et al., *The Amber biomolecular simulation programs.* J Comput Chem, 2005. **26**(16): p. 1668-88.
28. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* Journal of Chemical Information and Modeling, 2011. **51**(1): p. 69-82.
29. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
30. Nagar, B., et al., *Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571).* Cancer Res, 2002. **62**(15): p. 4236-43.
31. Young, M.A., et al., *Structure of the kinase domain of an imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680.* Cancer Res, 2006. **66**(2): p. 1007-14.
32. Weisberg, E., et al., *Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl.* Cancer Cell, 2005. **7**(2): p. 129-41.
33. Daley, G.Q. and Y. Ben-Neriah, *Implicating the bcr/abl gene in the pathogenesis of Philadelphia chromosome-positive human leukemia.* Adv Cancer Res, 1991. **57**: p. 151-84.
34. Lo, H.W., S.C. Hsu, and M.C. Hung, *EGFR signaling pathway in breast cancers: from traditional signal transduction to direct nuclear translocalization.* Breast Cancer Res Treat, 2006. **95**(3): p. 211-8.
35. Fabbro, D. and F. McCormick, *Protein tyrosine kinases : from inhibitors to useful drugs*. 2006, Totowa, NJ: Humana Press. xiii, 290 p.
36. Matthews, D.J. and M.E. Gerritsen, *Targeting protein kinases for cancer therapy*. 2010, Oxford: Wiley.
37. Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nat Rev Cancer, 2009. **9**(1): p. 28-39.
38. Deininger, M., E. Buchdunger, and B.J. Druker, *The development of imatinib as a therapeutic agent for chronic myeloid leukemia.* Blood, 2005. **105**(7): p. 2640-2653.
39. Lodish, H.F., *Molecular cell biology*. 4th ed. ed. Vol. Section 8.1, Mutations: Types and Causes. 1999, New York ; Basingstoke: W.H. Freeman.

40. Chapman, P.B., et al., *Improved survival with vemurafenib in melanoma with BRAF V600E mutation.* N Engl J Med, 2011. **364**(26): p. 2507-16.
41. Kumar, R., et al., *BRAF mutations in metastatic melanoma: a possible association with clinical outcome.* Clinical Cancer Research, 2003. **9**(9): p. 3362-8.
42. Lopez-Rios, F., et al., *Comparison of testing methods for the detection of BRAF V600E mutations in malignant melanoma: pre-approval validation study of the companion diagnostic test for vemurafenib.* Plos One, 2013. **8**(1): p. e53733.
43. Quintas-Cardama, A., H.M. Kantarjian, and J.E. Cortes, *Mechanisms of primary and secondary resistance to imatinib in chronic myeloid leukemia.* Cancer Control, 2009. **16**(2): p. 122-31.
44. Azam, M., R.R. Latek, and G.Q. Daley, *Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL.* Cell, 2003. **112**(6): p. 831-843.
45. Heinrich, M.C., et al., *Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor.* Journal of Clinical Oncology, 2003. **21**(23): p. 4342-4349.
46. Fletcher, J.A. and B.P. Rubin, *KIT mutations in GIST.* Current Opinion in Genetics & Development, 2007. **17**(1): p. 3-7.
47. Paez, J.G., et al., *EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy.* Science, 2004. **304**(5676): p. 1497-1500.
48. Dowell, J.E. and J.D. Minna, *Chasing mutations in the epidermal growth factor in lung cancer.* New England Journal of Medicine, 2005. **352**(8): p. 830-832.
49. Liu, B., B. Bernard, and J.H. Wu, *Impact of EGFR point mutations on the sensitivity to gefitinib: Insights from comparative structural analyses and molecular dynamics simulations.* Proteins-Structure Function and Bioinformatics, 2006. **65**(2): p. 331-346.
50. Yun, C.H., et al., *The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2070-5.
51. Breitenlechner, C.B., D. Bossemeyer, and R.A. Engh, *Crystallography for protein kinase drug design: PKA and SRC case studies.* Biochim Biophys Acta, 2005. **1754**(1-2): p. 38-49.
52. Breitenlechner, C.B., et al., *Design and crystal structures of protein kinase B-selective inhibitors in complex with protein kinase A and mutants.* J Med Chem, 2005. **48**(1): p. 163-70.
53. Rhodes, G., *Crystallography made crystal clear : a guide for users of macromolecular models*. 3rd ed. ed. 2006, Amsterdam ; Oxford: Elsevier Academic Press.
54. Martz, E. *RCSB Protein Data Bank: Nature of 3D Structural Data*. [cited 2016 5/5/2016]; Available from: http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/nature_of_3d_structural_data.html.
55. Widmer, H. and W. Jahnke, *Protein NMR in biomedical research.* Cell Mol Life Sci, 2004. **61**(5): p. 580-99.
56. Vajpai, N., et al., *Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib.* Journal of Biological Chemistry, 2008. **283**(26): p. 18292-18302.
57. Vajpai, N., et al., *Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib.* Biomol NMR Assign, 2008. **2**(1): p. 41-2.
58. Fernandez, C. and W. Jahnke, *New approaches for NMR screening in drug discovery.* Drug Discov Today Technol, 2004. **1**(3): p. 277-83.
59. Gajiwala, K.S., et al., *KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients.* Proc Natl Acad Sci U S A, 2009. **106**(5): p. 1542-7.

60. Panjarian, S., et al., *Structure and dynamic regulation of Abl kinases.* J Biol Chem, 2013. **288**(8): p. 5443-50.
61. Konermann, L., J. Pan, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics.* Chemical Society Reviews, 2011. **40**(3): p. 1224-34.
62. Zhang, J.M., et al., *Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors.* Nature, 2010. **463**(7280): p. 501-U116.
63. Pao, W., et al., *Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain.* PLoS Med, 2005. **2**(3): p. e73.
64. Azam, M., et al., *A screen to identify drug resistant variants to target-directed anti-cancer agents.* Biological Procedures Online, 2003. **5**: p. 204-210.
65. Azam, M. and G.Q. Daley, *Anticipating clinical resistance to target-directed agents - The BCR-ABL paradigm.* Molecular Diagnosis & Therapy, 2006. **10**(2): p. 67-76.
66. Warmuth, M., et al., *Ba/F3 cells and their use in kinase drug discovery.* Current Opinion in Oncology, 2007. **19**(1): p. 55-60.
67. Lolli, G., et al., *Inhibitor affinity chromatography: profiling the specific reactivity of the proteome with immobilized molecules.* Proteomics, 2003. **3**(7): p. 1287-98.
68. Valsasina, B., H.M. Kalisz, and A. Isacchi, *Kinase selectivity profiling by inhibitor affinity chromatography.* Expert Rev Proteomics, 2004. **1**(3): p. 303-15.
69. Fedorov, O., et al., *A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases.* Proc Natl Acad Sci U S A, 2007. **104**(51): p. 20523-8.
70. Jacobson, M.P., et al., *On the role of the crystal environment in determining protein side-chain conformations.* J Mol Biol, 2002. **320**(3): p. 597-608.
71. Xiang, Z. and B. Honig, *Extending the accuracy limits of prediction for side-chain conformations.* J Mol Biol, 2001. **311**(2): p. 421-30.
72. Isralewitz, B., M. Gao, and K. Schulten, *Steered molecular dynamics and mechanical functions of proteins.* Curr Opin Struct Biol, 2001. **11**(2): p. 224-30.
73. Sliwoski, G., et al., *Computational methods in drug discovery.* Pharmacol Rev, 2014. **66**(1): p. 334-95.
74. Stahl, M., C. Taroni, and G. Schneider, *Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network.* Protein Eng, 2000. **13**(2): p. 83-8.
75. Laurie, A.T. and R.M. Jackson, *Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites.* Bioinformatics, 2005. **21**(9): p. 1908-16.
76. Massova, I. and P. Kollman, *Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding.* Perspectives in Drug Discovery and Design, 2000. **18**(1): p. 113-135.
77. Homeyer, N. and H. Gohlke, *Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method.* Molecular Informatics, 2012. **31**(2): p. 114-122.
78. Steinbrecher, T. and A. Labahn, *Towards accurate free energy calculations in ligand protein-binding studies.* Curr Med Chem, 2010. **17**(8): p. 767-85.
79. Carlson, H.A. and J.A. McCammon, *Accommodating protein flexibility in computational drug design.* Mol Pharmacol, 2000. **57**(2): p. 213-8.
80. Huang, S.Y. and X. Zou, *Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking.* Proteins-Structure Function and Bioinformatics, 2007. **66**(2): p. 399-421.
81. Dixit, A. and G.M. Verkhivker, *Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations.* PLoS Comput Biol, 2009. **5**(8): p. e1000487.

82. Mukherjee, S., et al., *A docking interaction study of the effect of critical mutations in ribonuclease a on protein-ligand binding.* Biochem Mol Biol Educ, 2005. **33**(5): p. 335-43.
83. Tutone, M., A. Lauria, and A.M. Almerico, *Study of the role of "gatekeeper" mutations V654A and T670I of c-kit kinase in the interaction with inhibitors by means mixed molecular dynamics/docking approach.* Bioinformation, 2011. **7**(6): p. 296-8.
84. Sousa, S.F., et al., *Protein-ligand docking in the new millennium--a retrospective of 10 years in the field.* Curr Med Chem, 2013. **20**(18): p. 2296-314.
85. Cornell WD, C.P., Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* J. Am. Chem. Soc., 1995(117): p. 5179–5197.
86. Kollman, P.A., *Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules.* Accounts of Chemical Research, 1996. **29**(10): p. 461-469.
87. MacKerell, A.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids.* Biopolymers, 2000. **56**(4): p. 257-265.
88. Patel, S. and C.L. Brooks, 3rd, *CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations.* J Comput Chem, 2004. **25**(1): p. 1-15.
89. Patel, S., A.D. Mackerell, Jr., and C.L. Brooks, 3rd, *CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model.* J Comput Chem, 2004. **25**(12): p. 1504-14.
90. Williams, S., *The Study of Conformational Motions using Enhanced Sampling Techniques*, in *School of Chemistry*. 2007, University of Southampton.
91. Hofmann, D.W.M., L.N. Kuleshova, and J. Apostolakis, *Data mining in crystallography*. 2010, Heidelberg: Springer.
92. Vahdati, N., *Application of simulation methods for the identification of allosteric binding site in human glucokinase*, in *School of Chemistry*. 2010, University of Southampton.
93. Atkins, P.W. and J. De Paula, *Physical chemistry for the life sciences*. 2006, Oxford: Oxford University Press.
94. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
95. Wang, J., Cieplak, P. and Kollman, P. A. , *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* J. Comput. Chem, 2000(21): p. 1049–1074.
96. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.* J Comput Chem, 2003. **24**(16): p. 1999-2012.
97. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-49.
98. Friesner, R.A., et al., *Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.* Journal of Medicinal Chemistry, 2006. **49**(21): p. 6177-96.
99. Halgren, T.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1750-9.
100. Kibble, T.W.B. and F.H. Berkshire, *Classical mechanics*. 5th ed. / Tom W.B. Kibble, Frank H. Berkshire. ed. 2004, London: Imperial College Press.

101. Verlet, L., *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.* Physical Review, 1967. **159**(1): p. 98.
102. Swope, W.C., Andersen, H.C., Berens, P. H., and Wilson, K.R. , *A computer simulation method for the calculation of equilibrium constants for the formation if physical clusters of molecules: Application to small water clusters.* Journal of Chemical Physics, 1982(76): p. 637-649.
103. Beeman, D., *Some multistep methods for use in molecular dynamics calculations.* Journal of Computational Physics, 1976. **20**: p. 130-139.
104. Hockney, R.W., *The potential calculation and some applications.* Methods Computational Physics, 1970: p. 135-211.
105. Ryckaert, J.P., Ciccotti, G., and Berendsen, H. J. C., *Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of N-Alkanes.* Journal of Computational Physics, 1977. **23**: p. 327-341.
106. Andersen, H.C., *Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations.* Journal of Computational Physics, 1983(52): p. 24-34.
107. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.
108. Laurendeau, N.M., *Statistical thermodynamics : fundamentals and applications*. 2005, Cambridge: Cambridge University Press.
109. Berendsen, H.J.C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R., *Molecular dynamics with coupling to an external bath.* Journal of Chemical Physics, 1984. **81**(8).
110. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
111. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature.* Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.
112. Grest, G.S., Kremer, K., *Molecular dynamics simulation for polymers in the presence of a heat bath.* Physical Review A, 1986. **33**(5): p. 3628 - 3631.
113. Nadler, W., et al., *Molecular and stochastic dynamics of proteins.* Proc Natl Acad Sci U S A, 1987. **84**(22): p. 7933-7.
114. Steinbach, P. *Langevin Dynamics (LD) Simulation*. 2010 13/06/14]; Available from: https://cmm.cit.nih.gov/intro_simulation/node24.html.
115. Hunenberger, B., *Thermostat Algorithms for Molecular Dynamics Simulations*, in *Advanced Computer Simulation*. 2005, Springer. p. 105-149.
116. Csaba, H. *Lecture 6 - Solvation models and homology modeling*. [cited 2013 30/3/13]; Available from: xray.bmc.uu.se/~csaba/lecture_notes_6.pdf.
117. Guillot, B., *A reappraisal of what we have learnt during three decades of computer simulations on water.* Journal of Molecular Liquids, 2002. **101**(1-3): p. 219-260.
118. Berendsen, H.J.C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J., *Intermolecular Forces - Interaction Models for Water in relation to Protein Hydration*. 1981.
119. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
120. Mahoney, M., Jorgensen, W, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* Journal of Chemical Physics, 2000. **112**(20).
121. Nezbeda, I., Lisal, M., Kolafa, J., *An examination of the five-site potential (TIP5P) for water.* Journal of Chemical Physics, 2002. **117**(19): p. 8892-8897.
122. Stillinger, F.H. and A. Rahman, *Improved simulation of liquid water by molecular dynamics.* The Journal of Chemical Physics, 1974. **60**(4): p. 1545-1557.
123. Koehl, P. *Implicit Solvent Models for Molecular Simulations*. 2002 12/04/2014]; Available from: http://csb.stanford.edu/~koehl/ProShape/born.php.

124. Eisenberg, D. and A.D. McLachlan, *Solvation energy in protein folding and binding.* Nature, 1986. **319**(6050): p. 199-203.
125. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**(5214): p. 1144-9.
126. Fogolari, F., A. Brigo, and H. Molinari, *The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology.* J Mol Recognit, 2002. **15**(6): p. 377-92.
127. Onufriev, A. *The generalized Born model; its foundation, applications, and limitations*. 2010.
128. Onufriev, A., D.A. Case, and D. Bashford, *Effective Born radii in the generalized Born approximation: the importance of being perfect.* J Comput Chem, 2002. **23**(14): p. 1297-304.
129. Still, W.C., Tempzyk, A., Hawley, R. C., Hendrickson, T., *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics* J. Am. Chem. Soc., 1990. **112**(16).
130. Koehl, P., *Electrostatics calculations: latest methodological advances.* Curr Opin Struct Biol, 2006. **16**(2): p. 142-51.
131. Genheden, S. and U. Ryde, *The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.* Expert Opin Drug Discov, 2015. **10**(5): p. 449-61.
132. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-18.
133. Keskin, O., R.L. Jernigan, and I. Bahar, *Proteins with similar architecture exhibit similar large-scale dynamic behavior.* Biophys J, 2000. **78**(4): p. 2093-106.
134. Fuglebakk, E., J. Echave, and N. Reuter, *Measuring and comparing structural fluctuation patterns in large protein datasets.* Bioinformatics, 2012. **28**(19): p. 2431-40.
135. Karplus, P.A.S.G.E., *Prediction of chain flexibility in proteins.* Naturwissenschaften, 1985. **72**(4): p. 212-213.
136. Vihinen, M., E. Torkkila, and P. Riikonen, *Accuracy of protein flexibility predictions.* Proteins, 1994. **19**(2): p. 141-9.
137. Radivojac, P., et al., *Protein flexibility and intrinsic disorder.* Protein Sci, 2004. **13**(1): p. 71-80.
138. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
139. Skjaerven, L., *Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit.* Proteins: Structure, Function and Bioinformatics, 2011. **79**: p. 232 - 243.
140. Kaiser, H.F., *Educational and psychological measurement*. Vol. 20. 1960: [Durham.
141. Cattell, R.B., *Handbook of multivariate experimental psychology*. 1966, Chicago: Rand McNally & Company.
142. Laurie, A.T. and R.M. Jackson, *Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening.* Curr Protein Pept Sci, 2006. **7**(5): p. 395-406.
143. Laskowski, R.A., *SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions.* J Mol Graph, 1995. **13**(5): p. 323-30, 307-8.
144. Hendlich, M., F. Rippmann, and G. Barnickel, *LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins.* J Mol Graph Model, 1997. **15**(6): p. 359-63, 389.
145. Brady, G.P., Jr. and P.F. Stouten, *Fast prediction and visualization of protein binding pockets with PASS.* J Comput Aided Mol Des, 2000. **14**(4): p. 383-401.
146. Binkowski, T.A., L. Adamian, and J. Liang, *Inferring functional relationships of proteins from local sequence and spatial surface patterns.* J Mol Biol, 2003. **332**(2): p. 505-26.
147. Binkowski, T.A., S. Naghibzadeh, and J. Liang, *CASTp: Computed Atlas of Surface Topography of proteins.* Nucleic Acids Res, 2003. **31**(13): p. 3352-5.

148. Weisel, M., E. Proschak, and G. Schneider, *PocketPicker: analysis of ligand binding-sites with shape descriptors.* Chem Cent J, 2007. **1**: p. 7.
149. Craig, I.R., et al., *Pocket-space maps to identify novel binding-site conformations in proteins.* Journal of Chemical Information and Modeling, 2011. **51**(10): p. 2666-79.
150. Bowers, K.J., et al. *Scalable algorithms for molecular dynamics simulations on commodity clusters*. in *SC 2006 Conference, Proceedings of the ACM/IEEE*. 2006. IEEE.
151. Research, D.E.S., *Maestro-Desmond Interoperability Tools*. 2013: New York.
152. Jorgensen, W.L. and L.L. Thomas, *Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria.* Journal of Chemical Theory and Computation, 2008. **4**(6): p. 869-876.
153. Zwanzig, R.W., *High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases.* The Journal of Chemical Physics, 1954. **22**(8): p. 1420-1426.
154. Mezei, M., *The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water.* The Journal of Chemical Physics, 1987. **86**(12): p. 7084-7088.
155. Woods, C.J., J.W. Essex, and M.A. King, *Enhanced Configurational Sampling in Binding Free-Energy Calculations.* The Journal of Physical Chemistry B, 2003. **107**(49): p. 13711-13718.
156. Woods, C.J., J.W. Essex, and M.A. King, *The Development of Replica-Exchange-Based Free-Energy Methods.* The Journal of Physical Chemistry B, 2003. **107**(49): p. 13703-13710.
157. Sugita, Y., A. Kitao, and Y. Okamoto, *Multidimensional replica-exchange method for free-energy calculations.* The Journal of Chemical Physics, 2000. **113**(15): p. 6042-6051.
158. Fukunishi, H., O. Watanabe, and S. Takada, *On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction.* The Journal of Chemical Physics, 2002. **116**(20): p. 9058-9067.
159. Guimarães, C.R.W. and A.M. Mathiowetz, *Addressing Limitations with the MM-GB/SA Scoring Procedure using the WaterMap Method and Free Energy Perturbation Calculations.* Journal of Chemical Information and Modeling, 2010. **50**(4): p. 547-559.
160. Graves, A.P., et al., *Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing.* Journal of Molecular Biology, 2008. **377**(3): p. 914-934.
161. Wolf, R., *pymdgbsa.py*. 2010, CADD Novartis: Basel. p. Python script for a complete MMGBSA evaluation of a single receptor-ligand trajectory, version 0.6, 2010.
162. Salomon-Ferrer, R., D.A. Case, and R.C. Walker, *An overview of the Amber biomolecular simulation package.* Wiley Interdisciplinary Reviews-Computational Molecular Science, 2013. **3**(2): p. 198-210.
163. Miller, B.R., et al., *MMPBSA.py: An Efficient Program for End-State Free Energy Calculations.* Journal of Chemical Theory and Computation, 2012. **8**(9): p. 3314-3321.
164. Huang, S.Y. and X. Zou, *Advances and challenges in protein-ligand docking.* Int J Mol Sci, 2010. **11**(8): p. 3016-34.
165. McGann, M.R., et al., *Gaussian docking functions.* Biopolymers, 2003. **68**(1): p. 76-90.
166. Bohm, H.J., *The computer program LUDI: a new method for the de novo design of enzyme inhibitors.* J Comput Aided Mol Des, 1992. **6**(1): p. 61-78.
167. Rarey, M., et al., *A fast flexible docking method using an incremental construction algorithm.* J Mol Biol, 1996. **261**(3): p. 470-89.
168. Ewing, T.J.A. and I.D. Kuntz, *Critical evaluation of search algorithms for automated molecular docking and database screening.* Journal of Computational Chemistry, 1997. **18**(9): p. 1175-1189.
169. Miller, M.D., et al., *FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure.* J Comput Aided Mol Des, 1994. **8**(2): p. 153-74.

170. Trosset, J.-Y. and H.A. Scheraga, *Prodock: Software package for protein modeling and docking.* Journal of Computational Chemistry, 1999. **20**(4): p. 412-427.
171. Abagyan, R., M. Totrov, and D. Kuznetsov, *ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation.* Journal of Computational Chemistry, 1994. **15**(5): p. 488-506.
172. Jones, G., P. Willett, and R.C. Glen, *Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation.* J Mol Biol, 1995. **245**(1): p. 43-53.
173. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking.* J Mol Biol, 1997. **267**(3): p. 727-48.
174. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
175. Goodsell, D.S., G.M. Morris, and A.J. Olson, *Automated docking of flexible ligands: applications of AutoDock.* J Mol Recognit, 1996. **9**(1): p. 1-5.
176. Morris, G.M., et al., *Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4.* J Comput Aided Mol Des, 1996. **10**(4): p. 293-304.
177. Baxter, C.A., et al., *Flexible docking using Tabu search and an empirical estimate of binding affinity.* Proteins-Structure Function and Bioinformatics, 1998. **33**(3): p. 367-82.
178. Murray, C.W., C.A. Baxter, and A.D. Frenkel, *The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase.* J Comput Aided Mol Des, 1999. **13**(6): p. 547-62.
179. Chen, H.M., et al., *SODOCK: swarm optimization for highly flexible protein-ligand docking.* J Comput Chem, 2007. **28**(2): p. 612-23.
180. Chen, K., T. Li, and T. Cao, *Tribe-PSO: A novel global optimization algorithm and its application in molecular docking.* Chemometrics and Intelligent Laboratory Systems, 2006. **82**(1–2): p. 248-259.
181. Namasivayam, V. and R. Gunther, *pso@autodock: a fast flexible molecular docking program based on Swarm intelligence.* Chemical Biology & Drug Design, 2007. **70**(6): p. 475-84.
182. Goodsell, D.S. and A.J. Olson, *Automated docking of substrates to proteins by simulated annealing.* Proteins-Structure Function and Bioinformatics, 1990. **8**(3): p. 195-202.
183. Meza, J.C. and M.L. Martinez, *Direct search methods for the molecular conformation problem.* J. Comput. Chem., 1994. **15**(6): p. 627-632.
184. Sabesan, M.N., *An energy mimimization technique using the least-squares method.* J Theor Biol, 1977. **68**(2): p. 241-6.
185. Erickson, J.A., et al., *Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy.* Journal of Medicinal Chemistry, 2004. **47**(1): p. 45-55.
186. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD.* Proteins-Structure Function and Bioinformatics, 2003. **52**(4): p. 609-23.
187. Ponder, J.W. and D.A. Case, *Force fields for protein simulations.* Adv Protein Chem, 2003. **66**: p. 27-85.
188. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.
189. Eldridge, M.D., et al., *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, 1997. **11**(5): p. 425-45.

190. Tao, P. and L. Lai, *Protein ligand docking based on empirical method for binding affinity estimation.* J Comput Aided Mol Des, 2001. **15**(5): p. 429-46.
191. Wang, R., et al., *SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex.* Molecular modeling annual, 1998. **4**(12): p. 379-394.
192. Rognan, D., et al., *Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins.* Journal of Medicinal Chemistry, 1999. **42**(22): p. 4650-8.
193. Wang, R., L. Lai, and S. Wang, *Further development and validation of empirical scoring functions for structure-based binding affinity prediction.* J Comput Aided Mol Des, 2002. **16**(1): p. 11-26.
194. Gohlke, H., M. Hendlich, and G. Klebe, *Knowledge-based scoring function to predict protein-ligand interactions.* J Mol Biol, 2000. **295**(2): p. 337-56.
195. Velec, H.F., H. Gohlke, and G. Klebe, *DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction.* Journal of Medicinal Chemistry, 2005. **48**(20): p. 6296-303.
196. Muegge, I., *PMF scoring revisited.* Journal of Medicinal Chemistry, 2006. **49**(20): p. 5895-902.
197. Muegge, I. and Y.C. Martin, *A general and fast scoring function for protein-ligand interactions: a simplified potential approach.* Journal of Medicinal Chemistry, 1999. **42**(5): p. 791-804.
198. DeWitte, R.S. and E.I. Shakhnovich, *SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence.* Journal of the American Chemical Society, 1996. **118**(47): p. 11733-11744.
199. Clark, R.D., et al., *Consensus scoring for ligand/protein interactions.* J Mol Graph Model, 2002. **20**(4): p. 281-95.
200. Baber, J.C., et al., *The use of consensus scoring in ligand-based virtual screening.* Journal of Chemical Information and Modeling, 2006. **46**(1): p. 277-88.
201. Feher, M., *Consensus scoring for protein-ligand interactions.* Drug Discov Today, 2006. **11**(9-10): p. 421-8.
202. Terp, G.E., et al., *A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities.* Journal of Medicinal Chemistry, 2001. **44**(14): p. 2333-43.
203. Bissantz, C., G. Folkers, and D. Rognan, *Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations.* Journal of Medicinal Chemistry, 2000. **43**(25): p. 4759-67.
204. Ferrari, A.M., et al., *Soft Docking and Multiple Receptor Conformations in Virtual Screening.* Journal of Medicinal Chemistry, 2004. **47**(21): p. 5076-5084.
205. Jiang, F. and S.H. Kim, *"Soft docking": matching of molecular surface cubes.* J Mol Biol, 1991. **219**(1): p. 79-102.
206. Leach, A.R., *Ligand docking to proteins with discrete side-chain flexibility.* J Mol Biol, 1994. **235**(1): p. 345-56.
207. Knegtel, R.M., I.D. Kuntz, and C.M. Oshiro, *Molecular docking to ensembles of protein structures.* J Mol Biol, 1997. **266**(2): p. 424-40.
208. Bottegoni, G., et al., *Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking.* Journal of Medicinal Chemistry, 2009. **52**(2): p. 397-406.
209. Broughton, H.B., *A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening.* J Mol Graph Model, 2000. **18**(3): p. 247-57, 302-4.
210. Nabuurs, S.B., M. Wagener, and J. de Vlieg, *A flexible approach to induced fit docking.* Journal of Medicinal Chemistry, 2007. **50**(26): p. 6507-18.

211. Wagener, M., J.D. Vlieg, and S.B. Nabuurs, *Flexible protein-ligand docking using the Fleksy protocol.* J Comput Chem, 2012.
212. Meiler, J. and D. Baker, *ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility.* Proteins-Structure Function and Bioinformatics, 2006. **65**(3): p. 538-48.
213. Babine, R.E. and S.L. Bender, *Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design.* Chem Rev, 1997. **97**(5): p. 1359-1472.
214. Roberts, B.C. and R.L. Mancera, *Ligand-protein docking with water molecules.* Journal of Chemical Information and Modeling, 2008. **48**(2): p. 397-408.
215. Huang, N. and B.K. Shoichet, *Exploiting ordered waters in molecular docking.* Journal of Medicinal Chemistry, 2008. **51**(16): p. 4862-5.
216. de Graaf, C., et al., *Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking.* Journal of Medicinal Chemistry, 2005. **48**(7): p. 2308-18.
217. Lam, P.Y., et al., *Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors.* Science, 1994. **263**(5145): p. 380-4.
218. Poornima, C.S. and P.M. Dean, *Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions.* J Comput Aided Mol Des, 1995. **9**(6): p. 500-12.
219. Verdonk, M.L., et al., *Modeling water molecules in protein-ligand docking using GOLD.* Journal of Medicinal Chemistry, 2005. **48**(20): p. 6504-15.
220. Barillari, C., et al., *Analysis of water patterns in protein kinase binding sites.* Proteins-Structure Function and Bioinformatics, 2011. **79**(7): p. 2109-21.
221. Barillari, C., et al., *Classification of water molecules in protein binding sites.* Journal of the American Chemical Society, 2007. **129**(9): p. 2577-87.
222. Abel, R., et al., *Role of the active-site solvent in the thermodynamics of factor Xa ligand binding.* Journal of the American Chemical Society, 2008. **130**(9): p. 2817-31.
223. Li, Z. and T. Lazaridis, *The effect of water displacement on binding thermodynamics: concanavalin A.* Journal of Physical Chemistry B, 2005. **109**(1): p. 662-70.
224. Li, Z. and T. Lazaridis, *Thermodynamics of buried water clusters at a protein-ligand binding interface.* Journal of Physical Chemistry B, 2006. **110**(3): p. 1464-75.
225. Young, T., et al., *Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 808-13.
226. Michel, J., J. Tirado-Rives, and W.L. Jorgensen, *Prediction of the water content in protein binding sites.* Journal of Physical Chemistry B, 2009. **113**(40): p. 13337-46.
227. Woo, H.J., A.R. Dinner, and B. Roux, *Grand canonical Monte Carlo simulations of water in protein environments.* Journal of Chemical Physics, 2004. **121**(13): p. 6392-400.
228. Speidel, J.A., J.R. Banfelder, and M. Mezei, *Automatic Control of Solvent Density in Grand Canonical Ensemble Monte Carlo Simulations.* Journal of Chemical Theory and Computation, 2006. **2**(5): p. 1429-1434.
229. Zhang, L. and J. Hermans, *Hydrophilicity of cavities in proteins.* Proteins-Structure Function and Bioinformatics, 1996. **24**(4): p. 433-8.
230. Gilson, M.K., et al., *The statistical-thermodynamic basis for computation of binding affinities: a critical review.* Biophysical Journal, 1997. **72**(3): p. 1047-69.
231. Hamelberg, D. and J.A. McCammon, *Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method.* Journal of the American Chemical Society, 2004. **126**(24): p. 7683-9.
232. Michel, J., J. Tirado-Rives, and W.L. Jorgensen, *Just add water: The prediction of the water content of binding sites.* Abstracts of Papers of the American Chemical Society, 2009. **238**.
233. Bodnarchuk, M., *Predicting the location and binding affinity of small molecules in protein binding sites* in *School of Chemistry*. 2012, University of Southampton.

234. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
235. Michel, J., J. Tirado-Rives, and W.L. Jorgensen, *Prediction of the Water Content in Protein Binding Sites.* Journal of Physical Chemistry B, 2009. **113**(40): p. 13337-13346.
236. Bodnarchuk, M.S., et al., *Strategies to calculate water binding free energies in protein-ligand complexes.* J Chem Inf Model, 2014. **54**(6): p. 1623-33.
237. Nowell, P.C. and D.A. Hungerford, *A Minute Chromosome in Human Chronic Granulocytic Leukemia.* Science, 1960.
238. Ren, R., *Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia.* Nat Rev Cancer, 2005. **5**(3): p. 172-83.
239. Nagar, B., et al., *Structural basis for the autoinhibition of c-Abl tyrosine kinase.* Cell, 2003. **112**(6): p. 859-71.
240. Hehlmann, R., A. Hochhaus, and M. Baccarani, *Chronic myeloid leukaemia.* Lancet, 2007. **370**(9584): p. 342-50.
241. Hubbard, S.R. and J.H. Till, *Protein tyrosine kinase structure and function.* Annu Rev Biochem, 2000. **69**: p. 373-98.
242. Hong, W.K., *Holland-Frei cancer medicine*. 8th ed. / by Waun Ki Hong ... [et al.]. ed. 2010, New York: McGraw-Hill Medical ; London : McGraw-Hill [distributor]. xxv, 2021 p.
243. Sirvent, A., C. Benistant, and S. Roche, *Cytoplasmic signalling by the c-Abl tyrosine kinase in normal and cancer cells.* Biol Cell, 2008. **100**(11): p. 617-31.
244. Shaul, Y., *c-Abl: activation and nuclear targets.* Cell Death Differ, 2000. **7**(1): p. 10-6.
245. Van Etten, R.A., *Cycling, stressed-out and nervous: cellular functions of c-Abl.* Trends Cell Biol, 1999. **9**(5): p. 179-86.
246. Woodring, P.J., T. Hunter, and J.Y. Wang, *Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases.* J Cell Sci, 2003. **116**(Pt 13): p. 2613-26.
247. Zipfel, P.A., et al., *Requirement for Abl kinases in T cell receptor signaling.* Curr Biol, 2004. **14**(14): p. 1222-31.
248. Ushio-Fukai, M., et al., *cAbl tyrosine kinase mediates reactive oxygen species- and caveolin-dependent AT1 receptor signaling in vascular smooth muscle: role in vascular hypertrophy.* Circulation Research, 2005. **97**(8): p. 829-36.
249. Hantschel, O. and G. Superti-Furga, *Regulation of the c-Abl and Bcr-Abl tyrosine kinases.* Nat Rev Mol Cell Biol, 2004. **5**(1): p. 33-44.
250. Smith, D.L., J. Burthem, and A.D. Whetton, *Molecular pathogenesis of chronic myeloid leukaemia.* Expert Rev Mol Med, 2003. **5**(27): p. 1-27.
251. Hantschel, O., et al., *A myristoyl/phosphotyrosine switch regulates c-Abl.* Cell, 2003. **112**(6): p. 845-857.
252. Gomperts, B.D., P.E.R. Tatham, and I.M. Kramer, *Signal transduction*. 2nd ed. ed. 2009, Burlington, M.A. ; London: Elsevier/Academic Press.
253. Nagar, B., et al., *Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase.* Mol Cell, 2006. **21**(6): p. 787-98.
254. McPherson, P.S., *Regulatory role of SH3 domain-mediated protein-protein interactions in synaptic vesicle endocytosis.* Cellular Signalling, 1999. **11**(4): p. 229-38.
255. Waksman, G. and J. Kuriyan, *Structure and specificity of the SH2 domain.* Cell, 2004. **116**(2 Suppl): p. S45-8, 3 p following S48.
256. Sawyer, T.K., *Src homology-2 domains: structure, mechanisms, and drug discovery.* Biopolymers, 1998. **47**(3): p. 243-61.
257. Goga, A., et al., *Oncogenic activation of c-ABL by mutation within its last exon.* Mol Cell Biol, 1993. **13**(8): p. 4967-75.
258. Hantschel, O. and G. Superti-Furga, *Regulation of the C-Abl and Bcr-Abl tyrosine kinases.* Nature Reviews Molecular Cell Biology, 2004. **5**(1): p. 33-44.

259. Pluk, H., K. Dorey, and G. Superti-Furga, *Autoinhibition of c-Abl.* Cell, 2002. **108**(2): p. 247-59.
260. Kornev, A.P., et al., *Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(47): p. 17783-17788.
261. Woodring, P.J., T. Hunter, and J.Y. Wang, *Mitotic phosphorylation rescues Abl from F-actin-mediated inhibition.* J Biol Chem, 2005. **280**(11): p. 10318-25.
262. Reddy, E.P. and A.K. Aggarwal, *The ins and outs of bcr-abl inhibition.* Genes Cancer, 2012. **3**(5-6): p. 447-54.
263. Craig, I.R., J.W. Essex, and K. Spiegel, *Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments.* Journal of Chemical Information and Modeling, 2010. **50**(4): p. 511-24.
264. Hantschel, O., *Mechanisms of activation of Abl family kinases*, in *Marie Curie Bioscience Database*. 2000, Landes Bioscience.
265. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-82.
266. Azam, M., et al., *Activation of tyrosine kinases by mutation of the gatekeeper threonine.* Nat Struct Mol Biol, 2008. **15**(10): p. 1109-18.
267. Kornev, A.P. and S.S. Taylor, *Defining the conserved internal architecture of a protein kinase.* Biochim Biophys Acta, 2010. **1804**(3): p. 440-4.
268. Schindler, T., et al., *Structural mechanism for STI-571 inhibition of abelson tyrosine kinase.* Science, 2000. **289**(5486): p. 1938-42.
269. Salomon, A.R., et al., *Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 443-8.
270. Zhao, X., et al., *Structure of the Bcr-Abl oncoprotein oligomerization domain.* Nat Struct Biol, 2002. **9**(2): p. 117-20.
271. Smith, K.M. and R.A. Van Etten, *Activation of c-Abl kinase activity and transformation by a chemical inducer of dimerization.* J Biol Chem, 2001. **276**(26): p. 24372-9.
272. Macmillan. *Macmillan Cancer Support*. Available from: http://www.macmillan.org.uk.
273. Institute., C.C.I.-N.C. *Chronic Myelogenous Leukemia Treatment (PDQ®) - National Cancer Institute.* 2006 6 September 2011]; Available from: http://www.cancer.gov/cancertopics/pdq/treatment/CML/Patient/page1.
274. Calabretta, B. and D. Perrotti, *The biology of CML blast crisis.* Blood, 2004. **103**(11): p. 4010-22.
275. Kolitz, J.E., *Current therapeutic strategies for acute myeloid leukaemia.* Br J Haematol, 2006. **134**(6): p. 555-72.
276. Stapnes, C., et al., *Targeted therapy in acute myeloid leukaemia: current status and future directions.* Expert Opin Investig Drugs, 2009. **18**(4): p. 433-55.
277. Villela, L. and J. Bolanos-Meade, *Acute myeloid leukaemia: optimal management and recent developments.* Drugs, 2011. **71**(12): p. 1537-50.
278. Cancer Research UK, C.U. *Staging for Chronic Myeloid Leukaemia (CML)*. 2007 7 September 2011]; Available from: http://cancerhelp.cancerresearchuk.org/type/cml/treatment/staging-for-chronic-myeloid-leukaemia.
279. *Treatment of chronic myeloid leukemia by phase*. Comprehensive Cancer Information - National Cancer Institute. Web. 06 24th June 2011 4th October 2011]; Available from: http://www.cancer.org/Cancer/Leukemia-ChronicMyeloidCML/OverviewGuide/leukemia-cml-overview-treating-treatment-by-phase.
280. Sokal, J.E., *Prognosis in chronic myeloid leukaemia: biology of the disease vs. treatment.* Baillieres Clin Haematol, 1987. **1**(4): p. 907-29.

281. Hasford, J., et al., *A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa. Writing Committee for the Collaborative CML Prognostic Factors Project Group.* J Natl Cancer Inst, 1998. **90**(11): p. 850-8.
282. Quintas-Cardama, A., et al., *Imatinib mesylate therapy may overcome the poor prognostic significance of deletions of derivative chromosome 9 in patients with chronic myelogenous leukemia.* Blood, 2005. **105**(6): p. 2281-6.
283. Quintas-Cardama, A., et al., *Prognostic impact of deletions of derivative chromosome 9 in patients with chronic myelogenous leukemia treated with nilotinib or dasatinib.* Cancer, 2011.
284. Huntly, B.J., A. Bench, and A.R. Green, *Double jeopardy from a single translocation: deletions of the derivative chromosome 9 in chronic myeloid leukemia.* Blood, 2003. **102**(4): p. 1160-8.
285. Huntly, B.J., et al., *Imatinib improves but may not fully reverse the poor prognosis of patients with CML with derivative chromosome 9 deletions.* Blood, 2003. **102**(6): p. 2205-12.
286. Pretscher, D. and M. Wilhelm, *Nilotinib and Dasatinib in the First-Line Therapy for chronic myeloid Leukemia.* Medizinische Klinik, 2010. **105**(8): p. 588-588.
287. Kantarjian, H., et al., *Dasatinib versus Imatinib in Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia.* New England Journal of Medicine, 2010. **362**(24): p. 2260-2270.
288. Saglio, G., et al., *Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia.* N Engl J Med, 2010. **362**(24): p. 2251-9.
289. Gleixner, K.V., et al., *Effects of Bosutinib (SKI-606) in CML: Kinase Target Profile, Effects on BCR/ABL Mutants, and Synergism with Dasatinib in T315I+Cells.* Blood, 2008. **112**(11): p. 1097-1097.
290. U.S. Food and Drug Administration, F. *Bosutinib tablets*. 2012 [cited 2013 5 March 2013]; FDA approval of Bosutinib]. Available from: www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm318203.htm.
291. Cantor, M.E. *ARIAD Announces Initiation of Ponatinib (AP24534) Pivotal Trial in Drug-Resistant or Intolerant Chronic Myeloid Leukemia*. 2010 [cited 2011 2/12/11]; Available from: http://www.businesswire.com/news/home/20100913005322/en/ARIAD-Announces-Initiation-Ponatinib-AP24534-Pivotal-Trial.
292. Cortes, J.E., et al., *Ponatinib in refractory Philadelphia chromosome-positive leukemias.* N Engl J Med, 2012. **367**(22): p. 2075-88.
293. O'Brien, S., et al., *NCCN clinical practice guidelines in oncology: chronic myelogenous leukemia.* J Natl Compr Canc Netw, 2009. **7**(9): p. 984-1023.
294. O'Brien, S.G., et al., *Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia.* N Engl J Med, 2003. **348**(11): p. 994-1004.
295. Druker, B.J., et al., *Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia.* N Engl J Med, 2006. **355**(23): p. 2408-17.
296. Hantschel, O., U. Rix, and G. Superti-Furga, *Target spectrum of the BCR-ABL inhibitors imatinib, nilotinib and dasatinib.* Leukemia & Lymphoma, 2008. **49**(4): p. 615-619.
297. Deininger, M.W., *Nilotinib.* Clinical Cancer Research, 2008. **14**(13): p. 4027-4031.
298. Aguilera, D.G. and A.M. Tsimberidou, *Dasatinib in chronic myeloid leukemia: a review.* Ther Clin Risk Manag, 2009. **5**(2): p. 281-9.
299. Cortes, J.E., et al., *Safety and efficacy of bosutinib (SKI-606) in chronic phase Philadelphia chromosome-positive chronic myeloid leukemia patients with resistance or intolerance to imatinib.* Blood, 2011. **118**(17): p. 4567-76.
300. Redaelli, S., et al., *Activity of Bosutinib, Dasatinib, and Nilotinib Against 18 Imatinib-Resistant BCR/ABL Mutants.* Journal of Clinical Oncology, 2009. **27**(3): p. 469-471.

301. Levinson, N.M. and S.G. Boxer, *Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain.* Plos One, 2012. **7**(4): p. e29828.
302. O'Hare, T., et al., *AP24534, a Pan-BCR-ABL Inhibitor for Chronic Myeloid Leukemia, Potently Inhibits the T315I Mutant and Overcomes Mutation-Based Resistance.* Cancer Cell, 2009. **16**(5): p. 401-412.
303. U.S. Food and Drug Administration, F. *Ponatinib tablets*. 2012 [cited 2016 26/03/16]; FDA label for ICLUSIG (ponatinib)]. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2012/203469lbl.pdf.
304. Baccarani, M., et al., *Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet.* Blood, 2006. **108**(6): p. 1809-20.
305. Burke, A.C., et al., *Current status of agents active against the T315I chronic myeloid leukemia phenotype.* Expert Opin Emerg Drugs, 2011. **16**(1): p. 85-103.
306. Hochhaus, A. and P. La Rosee, *Imatinib therapy in chronic myelogenous leukemia: strategies to avoid and overcome resistance.* Leukemia, 2004. **18**(8): p. 1321-1331.
307. Corbin, A.S., et al., *Several Bcr-Abl kinase domain mutants associated with imatinib mesylate resistance remain sensitive to imatinib.* Blood, 2003. **101**(11): p. 4611-4.
308. Lahaye, T., et al., *Response and resistance in 300 patients with BCR-ABL-positive leukemias treated with imatinib in a single center - A 4.5-year follow-up.* Cancer, 2005. **103**(8): p. 1659-1669.
309. Hughes, T., et al., *Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results.* Blood, 2006. **108**(1): p. 28-37.
310. Deininger, M.W., et al., *Detection of ABL kinase domain mutations with denaturing high-performance liquid chromatography.* Leukemia, 2004. **18**(4): p. 864-71.
311. Hochhaus, A., et al., *Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy.* Leukemia, 2002. **16**(11): p. 2190-6.
312. Gorre, M.E., et al., *Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification.* Science, 2001. **293**(5531): p. 876-80.
313. Bacher, U., et al., *Clonal aberrations in Philadelphia chromosome negative hematopoiesis in patients with chronic myeloid leukemia treated with imatinib or interferon alpha.* Leukemia, 2005. **19**(3): p. 460-463.
314. Burchert, A., et al., *Compensatory PI3-kinase/Akt/mTor activation regulates imatinib resistance development.* Leukemia, 2005. **19**(10): p. 1774-1782.
315. Wendel, H.G., et al., *Loss of p53 impedes the antileukemic response to BCR-ABL inhibition.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(19): p. 7444-7449.
316. Donato, N.J., et al., *Imatinib mesylate resistance through BCR-ABL independence in chronic myelogenous leukemia.* Cancer Research, 2004. **64**(2): p. 672-677.
317. Azam, M., et al., *Activation of tyrosine kinases by mutation of the gatekeeper threonine.* Nature Structural & Molecular Biology, 2008. **15**(10): p. 1109-1118.
318. von Bubnoff, N., et al., *BCR-ABL gene mutations in relation to clinical resistance of Philadelphia-chromosome-positive leukaemia to STI571: a prospective study.* Lancet, 2002. **359**(9305): p. 487-91.
319. Deininger, M., *Resistance to imatinib: mechanisms and management.* J Natl Compr Canc Netw, 2005. **3**(6): p. 757-68.
320. Kantarjian, H., et al., *High-dose imatinib mesylate therapy in newly diagnosed Philadelphia chromosome-positive chronic phase chronic myeloid leukemia.* Blood, 2004. **103**(8): p. 2873-2878.

321. Kantarjian, H.M., et al., *Dose escalation of imatinib mesylate can overcome resistance to standard-dose therapy in patients with chronic myelogenous leukemia.* Blood, 2003. **101**(2): p. 473-475.
322. Golemovic, M., et al., *AMN107, a novel aminopyrimidine inhibitor of Bcr-Abl, has in vitro activity against imatinib-resistant chronic myeloid leukemia.* Clinical Cancer Research, 2005. **11**(13): p. 4941-7.
323. Verstovsek, S., et al., *AMN107, a novel aminopyrimidine inhibitor of p190 Bcr-Abl activation and of in vitro proliferation of Philadelphia-positive acute lymphoblastic leukemia cells.* Cancer, 2005. **104**(6): p. 1230-6.
324. Estrov, Z., et al., *Dasatinib (BMS354825), a dual Src-Abl inhibitor, is active in Philadelphia chromosome-positive chronic myelogenous leukemia (Ph plus CML) following treatment failure with imatinib mesylate and AMN107.* Journal of Clinical Oncology, 2006. **24**(18): p. 344s-344s.
325. U.S Food and Drug Administration. *Nilotinib (Tasigna)*. 2010 13/07/12]; FDA approves nilotinib for first-line treatment of Ph+ CML]. Available from: www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm216218.htm.
326. U.S. Food and Drug Administration. *FDA approves addtional medical indication for Sprycel*. 2010 13/7/12]; FDA approves dasatinib (Sprycel) for first-line treatment of Ph+ CML]. Available from: www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2010/ucm231409.htm.
327. Soverini, S., et al., *BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet.* Blood, 2011. **118**(5): p. 1208-15.
328. Shah, N.P., et al., *Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency.* J Clin Invest, 2007. **117**(9): p. 2562-9.
329. Khorashad, J.S., et al., *BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant CML: frequency and clonal relationships.* Blood, 2013. **121**(3): p. 489-98.
330. Gibbons, D.L., et al., *Molecular dynamics reveal BCR-ABL1 polymutants as a unique mechanism of resistance to PAN-BCR-ABL1 kinase inhibitor therapy.* Proc Natl Acad Sci U S A, 2014. **111**(9): p. 3550-5.
331. Zabriskie, M.S., et al., *BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia.* Cancer Cell, 2014. **26**(3): p. 428-42.
332. Tanneeru, K. and L. Guruprasad, *Ponatinib is a pan-BCR-ABL kinase inhibitor: MD simulations and SIE study.* PLoS One, 2013. **8**(11): p. e78556.
333. O'Hare, T., C.A. Eide, and M.W. Deininger, *Bcr-Abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia.* Blood, 2007. **110**(7): p. 2242-9.
334. Roumiantsev, S., et al., *Clinical resistance to the kinase inhibitor STI-571 in chronic myeloid leukemia by mutation of Tyr-253 in the Abl kinase domain P-loop.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10700-5.
335. Kim, R. and J. Skolnick, *Assessment of programs for ligand binding affinity prediction.* J Comput Chem, 2008. **29**(8): p. 1316-31.
336. Greenidge, P.A., et al., *MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement.* Journal of Chemical Information and Modeling, 2013. **53**(1): p. 201-9.