11-2-2017

Computational study targeting anti-fungal Tavaborole analogs and anti-cancer BRACO19

Babitha Machireddy
Rowan University

COMPUTATIONAL STUDY TARGETING ANTI-FUNGAL TAVABOROLE ANALOGS AND ANTI-CANCER BRACO19

by

Babitha Machireddy

A Thesis

Submitted to the
Department of Chemistry and Biochemistry
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Pharmaceutical Sciences
at
Rowan University
October 18, 2017

Thesis Chair: Chun Wu, Ph.D.

Dedication

I would like to dedicate this manuscript to Professor Kandalam Ramanujachary, Ph.D. To whom, I am forever indebted. He is my role model and a source of inspiration. Nothing I can say could ever express my gratitude for knowing you.

Acknowledgements

I would like to take this opportunity to thank Dr. Chun Wu for guiding and motivating me through this course. His impossible standards for humbled me and encouraged me to become a better student. I would like to thank Dr. Subash C Jonnalagadda for giving me this incredible opportunity to work on this project. The level of perfection he anticipates from his students nurtured the perfectionist in me. I would like to thank Dr. Kandalam Ramanujachary for the constant encouragement, support and guidance. And I would like thank Gayathri Jampana for the ocean of kindness she showed to a total stranger. Finally, I would like to thank Rowan University, Department of Chemistry and Biochemistry for giving me this incredible opportunity.

This work was supported by Rowan Startup and SEED grant and the National Science Foundation under Grant MRI-1429467 and XSEDE MCB160004/160164/160173.

Abstract

Babitha Machireddy
COMPUTATIONAL STUDY TARGETING ANTI-FUNGAL TAVABOROLE ANALOGS AND ANTI-CANCER BRACO19
2017-2018
Chun Wu, Ph.D.
Master of Science in Pharmaceutical Sciences

This thesis comprises of three computer aided drug design studies utilizing molecular docking and molecular dynamic simulations: (i) a lead optimization study virtually screening an initial library of ~120000 lead compounds targeting fungal leucyl tRNA synthetase, (ii) an exploratory study to understand the binding pathway of BRACO19 to a parallel telomeric DNA G-quadruplex by MD simulations and compare with experimentally solved X-ray crystal structure (iii) a comparative study to understand the lack of selectivity of BRACO19 to various topologies of human telomeric DNA G-quadruplex over DNA duplex.

The first chapter provides the background information required to understand the molecular docking studies and molecular dynamics simulation (MD) studies conducted and discussed in this thesis. This introductory chapter is organized as follows: the first section is an introduction to molecular recognition in protein-ligand interactions, the second section introduces computer-aided drug design, the third section introduces homology modelling, the fourth section discusses molecular docking and virtual screening, the fifth section introduces methods for binding affinity prediction and the sixth section explains MD simulations.

The second chapter of this thesis proposes a library of compounds with enhanced activity compared to the parent molecule it had been modified from. Tavaborole, the

recently approved topological anti-fungal drug, inhibits leucyl tRNA synthetase by irreversible covalent bonding and hinders protein synthesis. The benzo-boroxole pharmacophore of tavaborole is responsible for its unique activity. This study theoretically proposes molecules with improved anti-fungal affinity.

The third chapter of this thesis explores the binding pathway of anti-cancer drug, BRACO19 and human telomeric DNA G-quadruplex. G-quadruplex specific ligands that stabilizes the G-quadruplex, have great potential to be developed as anticancer agents. A free human telomeric DNA G-quadruplex and an unbound BRACO19 are simulated and the resulting structure is then compared with an experimentally solved X-ray structure of human telomeric G-quadruplex with a bound BRACO19 intercalated within the G-quadruplex. Three binding modes have been identified: top end stacking, bottom intercalation and groove binding. Bottom intercalation mode (51% of the population) is identical to the binding pose in the X-ray solved crystal structure.

The fourth chapter of this thesis compares different topological folds of human telomeric DNA G-quadruplexes (parallel, antiparallel and hybrid) that have been experimentally solved using molecular dynamic simulation to understand the 62-fold preferential selectivity of BRACO19 towards human telomeric DNA G-quadruplex over DNA duplex. Groove binding mode was found to be the most stable binding mode for the duplex and top stacking mode for the G-quadruplexes. The non-existential binding selectivity of BRACO19 can be accounted to the similar groove binding to both the duplex and the G-quadruplex. For that reason, a modification should be induced such that this prospective ligand destabilizes binding to the duplex but stabilizes the G-quadruplex binding.

Table of Contents

Abstract v

List of Figures xiii

List of Tables xviii

Chapter 1 : Introduction to Computer-Aided Drug Design 1

1.1 Introduction 1

1.2 Molecular Recognition 2

1.2.1 Affinity and Specificity. 2

1.2.2 Challenges in Entropy Estimation. 4

1.3 Homology Modeling 4

1.4 Protein–Ligand Binding Models 5

1.4.1 The Lock-and-Key Model. 6

1.4.2 The Induced Fit Model. 6

1.4.3 The Conformational Selection Model 7

1.5 Computer-Aided Drug Design 7

1.5.1 Motivation. 7

1.5.2 Protein–Ligand Docking. 8

Table of Contents (Continued)

1.5.3 Binding Affinity Predictions. 12

1.6 Molecular Dynamics Simulations 16

Chapter 2 : CADD in Boron Therapeutics 19

2.1 Abstract 19

2.2 Introduction 20

2.3 Tavaborole 20

2.3.1 Onychomycosis. 20

2.3.2 Mechanism of Action. 21

2.4 Hypothesis 21

2.5 Role of LeuRS in Protein Synthesis 22

2.5.1 Amino-Acylation Reaction. 22

2.6 Tavaborole and LeuRS 25

2.6.1 Structure-Activity Relationships (SAR) of Benzo-Boroxoles. 26

2.7 Computational Approach 27

2.7.1 Challenges and Assumptions. 27

2.8 Comparison with Human LeuRS – Selectivity 29

Table of Contents (Continued)

2.9 Methods.....32

2.9.1 Inherent Inaccuracies in Experimental Data.....32

2.9.2 Validation.....32

2.9.3 Screening the Given Library.....34

2.9.4 Binding Site Analysis and Enumeration.....42

2.9.5 Results.....43

Chapter 3 : Probing the Binding Pathway of BRACO19 to a Parallel-Stranded Human Telomeric G-Quadruplex Using Molecular Dynamics Binding Simulation with AMBER DNA OL15 and Ligand GAFF2 Force Fields.....152

3.1 Abstract.....152

3.2 Introduction.....153

3.3 Methods.....162

3.3.1 Simulation Systems.....162

3.3.2 Simulation Protocols.....164

3.3.3 Convergence of Simulations.....165

3.3.4 Binding Mode Identification.....166

3.3.5 Parameters for Characterizing DNA-Drug Binding Pathway.....169

Table of Contents (Continued)

3.4 Results 171

3.4.1 The Crystal Bottom Intercalation Mode of BRACO19 was Stable. 173

3.4.2 The Crystal Top Mode of BRACO19 Changed Significantly, Caused by Lack of Crystal Packing Constraints. 175

3.4.3 Three Drug Binding Modes were Observed in Free Binding DNA-Ligand Simulations. 177

3.4.4 The Stable G-Quadruplex Backbone Scaffold is Maintained as Indicated by the Flat and Small RMSD of ~2.5 Å in All Trajectories (Figure 60). 178

3.4.5 Bottom Intercalation of BRACO19 Follows the Flip-Insertion Mechanism. 181

3.4.6 Top Stacking of BRACO19 Leads to the Formation of ATAT-Tetrad in a Trajectory. 185

3.4.7 Groove Binding of BRACO19 Might be an Intermediate State for the End Binding Pose. 189

3.4.8 MM-PBSA Binding Energy Data Ranks the Bottom Intercalation Mode as the most stable mode among the three binding modes. 192

3.5 Discussion 193

3.6 Conclusions 197

Chapter 4 : Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent 198

4.1 Abstract 198

Table of Contents (Continued)

4.2 Introduction 199

4.3 Methods 206

4.3.1 Simulation Systems. 206

4.3.2 Simulation Protocols. 207

4.3.3 Convergence of Simulations. 208

4.3.4 Binding Mode Identification. 209

4.3.5 Order Parameters. 209

4.4 Results 211

4.4.1 Multiple Drug Binding Modes were Observed in Binding Simulations. ...212

4.4.2 VDW Interaction Contributes Most to the Total Binding Energy, Ranking the Binding Poses for Each DNA-Ligand System. 219

4.4.3 BRACO19 Binds to the Duplex DNA, Without Inducing Structural Changes. 221

4.4.4 Groove Binding Might be an Intermediate State for the Top Stacking Mode 223

4.4.5 BRACO19 Binds to the Anti-Parallel G-Quadruplex, Without Inducing Structural Changes. 227

4.4.6 BRACO19 Binds to the Hybrid Telomeric G-Quadruplex DNA, Without Inducing Structural Fluctuation. 230

Table of Contents (Continued)

4.5 Discussion.....233

4.6 Conclusion.....237

Publications Note.....238

References.....239

Appendix A: RMSD and Contact Plots from Chapter 3.....264

Appendix B: RMSD and Contact Plots from Chapter 4.....266

Appendix C: 2D Interactions and H-bond Network from Chapter 3.....274

List of Figures

Figure Page

Figure 1. Three protein models; (A) lock-and-key model, (B) induced fit model and (C) conformation selection model. 6

Figure 2. Endpoint methods. 13

Figure 3. Step 1 of amino-acylation. 23

Figure 4. Step 2 of amino-acylation. 24

Figure 5. Inhibition of leucylation due to the formation of boronic ester 25

Figure 6. Structure of Tavaborole. 26

Figure 7. 3D representation of homology modeled fungal LeuRS complexed with tRNA non-covalently interacting with tavaborole. 28

Figure 8. tRNA - Tavaborole complex adjusted from covalently bonded complex to non-covalently interacting moieties. 29

Figure 9. 3D representation of superimposed structures of homology modeled fungal and human LeuRS complexed with tRNA non-covalently interacting with tavaborole. ..30

Figure 10. Sequence alignment of fungal LeuRS and human LeuRS; residues interacting with tavaborole at the binding site of fungal LeuRS are depicted in yellow and human LeuRS in blue 30

Figure 11. Graph plotting Docking score of the molecules listed in table 4 vs binding experimental binding affinity of the same. 33

Figure 12. Graph plotting MM-GBSA values of the molecules listed in table 4 vs binding experimental binding affinity of the same. 34

List of Figures (Continued)

Figure Page

Figure 13. 2D structure of LL1_5 illustrating all the prospective modification sites. 43

Figure 14. 2D structure of LL1_6 illustrating all the prospective modification sites. 50

Figure 15. 2D structure of LL1_7 illustrating all the prospective modification sites. 58

Figure 16. 2D structure of LL1_9 illustrating all the prospective modification sites. 65

Figure 17. 2D structure of LL1_12 illustrating all the prospective modification sites. 76

Figure 18. 2D structure of LL1_13 illustrating all the prospective modification sites. 82

Figure 19. 2D structure of LL1_14 illustrating all the prospective modification sites. 91

Figure 20. 2D structure of LL2_1 illustrating all the prospective modification sites. 93

Figure 21. 2D structure of LL2_2 illustrating all the prospective modification sites. 99

Figure 22. 2D structure of LL2_3 illustrating all the prospective modification sites. 107

Figure 23. 2D structure of LL2_4 illustrating all the prospective modification sites. 112

Figure 24. 2D structure of LL2_6 illustrating all the prospective modification sites. 119

Figure 25. 2D structure of LL3_4 illustrating all the prospective modification sites. 125

Figure 26. 2D structure of LL3_5 illustrating all the prospective modification sites. 131

Figure 27. 2D structure of LL3_6 illustrating all the prospective modification sites. 138

List of Figures (Continued)

Figure Page

Figure 28. 2D structure of LL3_9 illustrating all the prospective modification sites. 145

Figure 29. Chemical structure of BRACO19 (3^{+}). 155

Figure 30. Comparison between apo and holo crystal structure of a parallel telomeric DNA G-quadruplex 157

Figure 31. Quadruplex-ligand-quadruplex interface in crystal structure. 158

Figure 32. The initial configuration of the simulation system (DNA Quadruplex + Unbound BRACO19) 162

Figure 33. Last snapshots of the ten quadruplex-BRACO19 simulations 165

Figure 34. Backbone Torsion Angles of DNA. 171

Figure 35. Stability simulation of the bottom binding mode in the crystal structure (PDB: 3CE5). 172

Figure 36. Stability simulation of the top binding mode in the crystal structure (PDB: 3CE5). Top: Representative structures with time annotation. 174

Figure 37. The MD relaxed crystal structure of (A) bottom intercalation mode and (B) top stacking mode (PDB ID: 3CE5) and the major binding modes, (C) bottom intercalation mode (D) top binding mode and (E) groove binding mode from the binding simulations. 176

Figure 38. A representative trajectory of the bottom intercalation mode. 180

Figure 39. Comparison for the backbone torsion angles of residue T7 between the free ligand binding simulation (red) and the stability simulation of the crystal pose (black) for the bottom intercalation mode. 182

List of Figures (Continued)

Figure Page

Figure 40. A representative trajectory of the top stacking mode. 184

Figure 41. Comparison for the backbone torsion angles of residue T7 between the free ligand binding simulation (red) and the stability simulation of the crystal pose (black) for the top stacking mode. 186

Figure 42. A representative trajectory of the groove binding mode 188

Figure 43. Comparison for the backbone torsion angles of residue T18 between the free ligand binding simulation (red) of the groove binding mode and the stability simulation of the crystal pose (black) of the bottom binding mode. 190

Figure 44. H-bond network indicating maximum number of H-bonds possible in G-tetrad (A), ATAT-tetrad (B) and T-dyad (C) respectively. 194

Figure 45. Structure of human telomeric DNA duplex (A), human telomeric parallel DNA quadruplex (PDB ID: 1KF1) (B), human telomeric antiparallel DNA quadruplex (PDB ID: 143D) (C), and human telomeric hybrid DNA quadruplex (PDB ID: 2HY9) (D). 204

Figure 46. Initial configuration of the simulation systems. 206

Figure 47. Simulated structures of human telomeric DNA duplex (A), human telomeric parallel DNA quadruplex (PDB ID: 1KF1) (B), human telomeric antiparallel DNA quadruplex (PDB ID: 143D) (C), and human telomeric hybrid DNA quadruplex (PDB ID: 2HY9) (D) in complex with BRACO19.....211

Figure 48. Last snapshots of 10 telomeric DNA duplex and BRACO19 simulations. ...213

Figure 49. Last snapshots of 10 parallel telomeric DNA G-quadruplex and BRACO19 simulations. 214

List of Figures (Continued)

Figure Page

Figure 50. Last snapshots of 10 anti-parallel telomeric DNA G-quadruplex and BRACO19 simulations. 215

Figure 51. Last snapshots of 10 hybrid telomeric DNA G-quadruplex and BRACO19 simulations.. 216

Figure 52. A representative groove binding trajectory of the DNA duplex. 221

Figure 53. A representative top stacking trajectory of the parallel G-quadruplex.....223

Figure 54. Comparison for the backbone torsion angles of residue T6 between the free ligand binding simulation (red) of the top stacking mode of the parallel G-quadruplex and the stability simulation of the crystal structure (black) of the parallel G-quadruplex 225

Figure 55. A representative top stacking trajectory of the anti-parallel G-quadruplex...227

Figure 56. Comparison for the backbone torsion angles of residue T05 between the free ligand binding simulation (red) of the top stacking trajectory of the anti-parallel G-quadruplex and the stability simulation of the crystal structure (black) of the anti-parallel G-quadruplex. 228

Figure 57. A representative top binding trajectory of the hybrid G-quadruplex. 231

Figure 58. Comparison for the backbone torsion angles of residue T8 between the free ligand binding simulation (red) of the top binding trajectory of the hybrid G-quadruplex and the stability simulation of the crystal structure (black) of the hybrid G-quadruplex.....232

Figure 59. (A) The experimental binding mode (PDB ID: 3CE5) of double stranded parallel telomeric DNA G-quadruplex (B) Major binding pose of simulated single stranded parallel telomeric DNA G-quadruplex. 234

List of Tables

Table Page

Table 1. Biological activity of benzo-boroxoles in anticancer cell-lines. 22

Table 2. Homology, identity and similarity statistics of human LeuRS compared to fungal LeuRS. 30

Table 3. List of the interacting residues (highlighted in figure 10) at the binding site of fungal LeuRS compared with residues interacting at the binding site of human LeuRS; highlighted cells indicate residues that might be responsible for high hydrophobicity of tavaborole binding site in fungal LeuRS. 31

Table 4. List of IC_{50} values of compounds structurally similar to tavaborole and their corresponding docking scores and MM-GBSA values against fungal LeuRS. The highlighted cells depicts activity of tavaborole. 33

Table 5. List of the docked molecules of ligand library 1 (LL1) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose. 35

Table 6. List of the docked molecules of ligand library 2 (LL2) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose. 38

Table 7. List of the docked molecules of ligand library 3 (LL3) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose. 39

Table 8. The surfaces generated by binding site analysis at the LeuRS-LL1_5 binding site and arrows point to modification site of R_0, R_1, R_2, R_3 and R_4. 43

Table 9. Hits from single site enumeration at R_0 of LL1_5. 44

Table 10. Hits from single site enumeration at R_1 of LL1_5. 45

List of Tables (Continued)

Table Page

Table 11. Hits from single site enumeration at R_2 of LL1_5. 46

Table 12. Hits from single site enumeration at R_3 of LL1_5. 46

Table 13. Hits from single site enumeration at R_4 of LL1_5. 47

Table 14. Hits from enumeration at all modification sites of LL1_5 and comparison with human LeuRS. 48

Table 15. The surfaces generated by binding site analysis at the LeuRS-LL1_6 binding site and arrows point to modification site of R_0, R_1, R_2, and R_3. 50

Table 16. Hits from single site enumeration at R_0 of LL1_6. 50

Table 17. Hits from single site enumeration at R_1 of LL1_6. 51

Table 18. Hits from single site enumeration at R_3 of LL1_6. 53

Table 19. Hits from enumeration at all modification sites of LL1_6 and comparison with human LeuRS 56

Table 20. The surfaces generated by binding site analysis at the LeuRS-LL1_7 binding site and arrows point to modification site of R_0, R_1 and R_2. 58

Table 21. Hits from single site enumeration at R_0 of LL1_7. 58

Table 22. Hits from single site enumeration at R_1 of LL1_7. 60

Table 23. Hits from single site enumeration at R_2 of LL1_7. 61

Table 24. Hits from enumeration at all modification sites of LL1_7. 63

List of Tables (Continued)

Table Page

Table 25. The surfaces generated by binding site analysis at the LeuRS-LL1_9 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3. 65

Table 26. Hits from single site enumeration at R_0 of LL1_9. 65

Table 27. Hits from single site enumeration at R_1 of LL1_9. 67

Table 28. Hits from single site enumeration at R_2 of LL1_9. 68

Table 29. Hits from single site enumeration at R_3 of LL1_9. 69

Table 30. Hits from enumeration at all modification sites of LL1_9 and comparison with human LeuRS. 72

Table 31. The surfaces generated by binding site analysis at the LeuRS-LL1_10 binding site and arrows point to modification site of R_1 and 2D structure of LL1_10 illustrating its prospective modification sites. 74

Table 32. Hits from single site enumeration at R1 of LL1_10 and comparison with human LeuRS. 74

Table 33. The surfaces generated by binding site analysis at the LeuRS-LL1_12 binding site and arrows point to modification site of R0, R1 and R2. 76

Table 34. Hits from single site enumeration at R_0 of LL1_12. 76

Table 35. Hits from single site enumeration at R_1 of LL1_12. 77

Table 36. Hits from single site enumeration at R_2 of LL1_12. 78

Table 37. Hits from enumeration at all modification sites of LL1_12. 80

List of Tables (Continued)

Table Page

Table 38. The surfaces generated by binding site analysis at the LeuRS-LL1_13 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3. 82

Table 39. Hits from single site enumeration at R_1 of LL1_13. 82

Table 40. Hits from single site enumeration at R_2 of LL1_13. 86

Table 41. Hits from enumeration at all modification sites of LL1_13 and comparison with human LeuRS. 89

Table 42. The surfaces generated by binding site analysis at the LeuRS-LL1_14 binding site and arrows point to modification site of R_0 and R_1. 91

Table 43. Hits from single site enumeration at R_1 of LL1_14. 91

Table 44. The surfaces generated by binding site analysis at the LeuRS-LL2_1 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4 and R_5. 93

Table 45. Hits from single site enumeration at R_0 of LL2_1. 93

Table 46. Hits from single site enumeration at R_1 of LL2_1. 94

Table 47. Hits from single site enumeration at R_2 of LL2_1. 95

Table 48. Hits from single site enumeration at R_3 of LL2_1. 95

Table 49. Hits from single site enumeration at R_4 of LL2_1. 96

Table 50. Hits from single site enumeration at R_5 of LL2_1. 96

List of Tables (Continued)

Table Page

Table 51. Hits from enumeration at all modification sites of LL2_1 and comparison with human LeuRS. 97

Table 52. The surfaces generated by binding site analysis at the LeuRS-LL2_2 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3. 99

Table 53. Hits from single site enumeration at R0 of LL2_2. 99

Table 54. Hits from single site enumeration at R_1 of LL2_2. 101

Table 55. Hits from single site enumeration at R_2 of LL2_2. 102

Table 56. Hits from single site enumeration at R_3 of LL2_2. 103

Table 57. Hits from enumeration at all modification sites of LL2_2 and comparison with human LeuRS. 105

Table 58. The surfaces generated by binding site analysis at the LeuRS-LL2_3 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4 and R_5. 107

Table 59. Hits from single site enumeration at R_0 of LL2_3. 107

Table 60. Hits from single site enumeration at R_1 of LL2_3. 108

Table 61. Hits from single site enumeration at R_2 of LL2_3. 108

Table 62. Hits from single site enumeration at R_3 of LL2_3. 109

Table 63. Hits from single site enumeration at R_4 of LL2_3. 110

List of Tables (Continued)

Table Page

Table 64. Hits from enumeration at all modification sites of LL2_3 and comparison with human LeuRS. 111

Table 65. The surfaces generated by binding site analysis at the LeuRS-LL2_4 binding site and arrows point to modification site of R_0, R_1, R_2, R_3 and R_4....112

Table 66. Hits from single site enumeration at R_0 of LL2_4. 112

Table 67. Hits from single site enumeration at R_1 of LL2_4. 113

Table 68. Hits from single site enumeration at R_2 of LL2_4. 114

Table 69. Hits from single site enumeration at R_3 of LL2_4. 115

Table 70. Hits from single site enumeration at R_4 of LL2_4. 116

Table 71. Hits from enumeration at all modification sites of LL2_4 and comparison with human LeuRS. 117

Table 72. The surfaces generated by binding site analysis at the LeuRS-LL2_6 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4, R_5, R_6 and R_7. 119

Table 73. Hits from single site enumeration at R1 of LL2_6. 119

Table 74. Hits from single site enumeration at R_2 of LL2_6. 120

Table 75. Hits from single site enumeration at R_3 of LL2_6. 122

Table 76. Hits from single site enumeration at R_4 of LL2_6. 122

Table 77. Hits from single site enumeration at R_6 of LL2_6. 123

List of Tables (Continued)

Table Page

Table 78. Hits from enumeration at all modification sites of LL2_6 and comparison with human LeuRS. 124

Table 79. The surfaces generated by binding site analysis at the LeuRS-LL3_4 binding site and arrows point to modification site of R_1, R_2, R_3, R_4 and R_5. 125

Table 80. Hits from single site enumeration at R_1 of LL3_4. 125

Table 81. Hits from single site enumeration at R_2 of LL3_4. 126

Table 82. Hits from single site enumeration at R_3 of LL3_4. 127

Table 83. Hits from single site enumeration at R_4 of LL3_4. 127

Table 84. Hits from single site enumeration at R_5 of LL3_4. 128

Table 85. Hits from enumeration at all modification sites of LL3_4. 128

Table 86. The surfaces generated by binding site analysis at the LeuRS-LL3_5 binding site and arrows point to modification site of R_1, R_2, R_3, R_4, R_5 and R_6. 131

Table 87. Hits from single site enumeration at R_1 of LL3_5. 131

Table 88. Hits from single site enumeration at R_2 of LL3_5. 132

Table 89. Hits from single site enumeration at R_3 of LL3_5. 132

Table 90. Hits from single site enumeration at R_4 of LL3_5. 133

Table 91. Hits from single site enumeration at R_5 of LL3_5. 134

List of Tables (Continued)

Table Page

Table 92. Hits from single site enumeration at R_6 of LL3_5. 135

Table 93. Hits from enumeration at all modification sites of LL3_5. 135

Table 94. The surfaces generated by binding site analysis at the LeuRS-LL3_6 binding site and arrows point to modification site of R_1, R_2, R_3, R_4, R_5, R_6, and R_7. 138

Table 95. Hits from single site enumeration at R_1 of LL3_6. 138

Table 96. Hits from single site enumeration at R_2 of LL3_6. 139

Table 97. Hits from single site enumeration at R_3 of LL3_6. 140

Table 98. Hits from single site enumeration at R_4 of LL3_6. 140

Table 99. Hits from single site enumeration at R_5 of LL3_6. 141

Table 100. Hits from single site enumeration at R_6 of LL3_6. 141

Table 101. Hits from single site enumeration at R_7 of LL3_6. 142

Table 102. Hits from enumeration at all modification sites of LL3_6. 142

Table 103. The surfaces generated by binding site analysis at the LeuRS-LL3_9 binding site and arrows point to modification site of R_1, R_2, R_3, R_4 and R_5.....145

Table 104. Hits from single site enumeration at R_1 of LL3_9. 145

Table 105. Hits from single site enumeration at R_2 of LL3_9. 146

List of Tables (Continued)

Table Page

Table 106. Hits from single site enumeration at R_3 of LL3_9. 147

Table 107. Hits from single site enumeration at R_4 of LL3_9. 147

Table 108. Hits from single site enumeration at R_5 of LL3_9. 148

Table 109. Hits from enumeration at all modification sites of LL3_9. 149

Table 110. In vivo activity of BRACO19 against various cancer cell lines 156

Table 111. Molecular dynamics simulations 162

Table 112. Representative structures of the most populated complex structure families (population ≥ 1 %) from the clustering analysis of the combined binding trajectories. 5' and 3' of the telomeric G-quadruplex DNA are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red. 167

Table 113. MM-PBSA energy of BRACO19 to the G-Quadruplex DNA in different binding modes. 191

Table 114. MM-GBSA binding energy (kcal/mol) of BRACO19 to human telomeric DNA duplex and Quadruplexes. 219

Chapter 1

Introduction to Computer-Aided Drug Design

1.1 Introduction

Based on principles of molecular recognition, computer-aided drug design (CADD) utilizes the increasing computational power to develop and employ various theoretical models for drug discovery and design. Over the years, computer-aided drug design (CADD) demonstrated to be effective and instrumental in influencing drug discovery and molecular recognition. Even though the fast computational tools are not absolutely accurate due to the resources, time and manpower required to perform experimental methods to gain the same insights renders CADD very valuable for drug discovery and design.(Tang, 2010)

The subject of designing drugs with high affinity for specific biological receptors is of continuing intellectual and practical interest. Molecular simulations and molecular modelling studies provide insights about the interactions contributing to the association of biological molecules. Before designing a model that can simulate association or dissociation of biological molecules, a systematic and exhaustive understanding of molecular recognition is essential. (Lamb & Jorgensen, 1997) The molecular establishment of many ubiquitous and crucial biological functions is formed by protein-ligand interactions. A rational guide to therapeutic drug design is attained by understanding the qualitative and quantitative components of the physical forces governing the protein-ligand interactions. Therefore, it is of immense scientific and practical importance to understand the role of molecular recognition in protein-ligand interactions.

1.2 Molecular Recognition

Molecular recognition is the non-covalent interactions between two or more molecules through van der Waals forces, hydrogen bonding, metal coordination and π-π, hydrophobic, or electrostatic interactions. Molecular recognition mediates interactions between receptors and ligands, proteins and proteins, nucleic acids and proteins, antigens and antibodies, enzymes and substrates etc. (Cleaves, 2011) Molecular recognition is defined by two characteristics: (i) affinity; governed by the strength of non-covalent interactions and (ii) specificity; relative strength of those non-covalent interactions with respect to another ligand/receptor. (Demchenko, 2001) Understanding the mechanisms of protein function is to understand the protein–ligand interactions. Therefore, it is crucial to understand the characterization and quantification of the energetics that govern the formation of a protein–ligand complex (Perozzo, Folkers, & Scapozza, 2004).

1.2.1 Affinity and specificity. The change in binding free energy of the complex compared with other potential targets determine the affinity and specificity of protein-ligand interactions. (Tang, 2010) One of the most significant thermodynamic quantities used to characterize the driving force is Gibbs free (binding) energy (G°). It indicates the capacity of a thermodynamic system to do maximum or reversible work at isothermal and isobaric conditions. A protein–ligand–solvent system is considered as a simple solute-solvent thermodynamic system, where the protein and ligand molecules are solutes and liquid water and buffer ions make a solvent system. Accordingly, thermodynamic laws dictate the interactions between them, resulting heat transfer and how these energy changes between the solutes and solvent correlate to association of protein and ligand (Gilson & Zhou, 2007).

Change in binding (Gibbs) free energy ($\Delta G°$) depends on two independent thermodynamics entities: change in enthalpy ($\Delta H°$) and change in entropy ($\Delta S°$) written as,

$$\Delta G° = \Delta H° - T\Delta S° \text{ (Pirzadeh, Beaudoin, \& Kusalik, 2012)} \quad (1.1)$$

It should be noted that the free energy ($\Delta G°$) is defined merely by the initial and final thermodynamic states, regardless of the pathway connecting these two states. (Du et al., 2016)

1.2.1.1 Enthalpy. Enthalpy ($\Delta H°$) is the total energy of the system, i.e., the sum of the internal energies of the solute and solvent and the energy required to solvate the system. (Li, Xie, Liu, & Liu, 2014) In general, the binding enthalpy of a thermodynamic system is the change in energy as a consequence of noncovalent interactions formed (van der Waals forces, hydrogen bonding, metal coordination and π-π, hydrophobic, or electrostatic interactions) at the binding site. (Perozzo et al., 2004)

1.2.1.2 Entropy. Entropy ($\Delta S°$) is the disorder or randomness of atoms and molecules in the system. The binding entropy ($\Delta S°$) (the total entropy change associated with binding) can be broken down into solvent entropy change ($\Delta S_S°$), conformational entropy change ($\Delta S_C°$) and translational and rotational degrees of freedom lost due to the formation of protein-ligand complex ($\Delta S_T°$):

$$\Delta S° = \Delta S_S° + \Delta S_C° + \Delta S_T° \text{ (Du et al., 2016)} \quad (1.2)$$

In protein-ligand complex formation, the binding entropy ($\Delta S°$) is mostly derived from solvation, de-solvation and the degrees of freedom of both ligand and protein during complex formation. When the ligand is transferred from the hydrophilic solvent to the predominantly hydrophobic binding site, entropy change of the ligand (ΔS_L) can be split

into conformational entropy ($\Delta S_C°$) (accessible rotamers), configurational entropy ($\Delta S_T°$) (translational and rotational), and vibrational entropy ($\Delta S_V°$) as follows,

$$\Delta S_L° = \Delta S_C° + \Delta S_T° + \Delta S_V° \quad (1.3)$$

Entropy change of the protein (ΔS_P) can also be split into the same components as the ligand. However, under the assumption that the entropy change in the protein is negligible, essential simplification of the model does not consider entropy change in the protein (ΔS_P) when bound to different ligands.

1.2.2 Challenges in entropy estimation. As mentioned in 1.2.1.2 solvation and de-solvation of both protein and ligand contribute to the entropic changes. Additionally, multiple binding states, entropy-entropy compensation, and configurational entropy also contribute to the entropic changes. These entropic changes are both challenging and energetically critical to accurately model the heuristic methods. Various theoretical methods were employed to estimate the solvation energies involved in protein-ligand interactions. Theoretical studies including MD simulations and intrinsic solvent representation (MM-GBSA, MM-PBSA, etc.) have been developed and applied.

While docking the ligands into the binding site, the flexibility of the same is ignored. Accounting to this oversimplification to reduce the computational expense, docking ignores enthalpy-entropy compensation. This compensation, an effect of the receptor's assumed rigidity, contributes to inaccurate entropy estimation.

1.3 Homology Modeling

3D structure of a protein can be obtained using X-ray crystallography or NMR spectroscopy studies are available in the Protein Data Bank (PDB) (Westbrook, Feng, Chen, Yang, & Berman, 2003), http://www.rcsb.org/pdb. However, when the 3D structure

of a protein is unknown, a theoretical protein model can be generated based on a homologous protein with known 3D structure. A technique known as homology modeling predicts the 3D structure of the protein utilizing its amino-acid sequence and the topology of the homologous protein (template). (Withana-Gamage, Hegedus, Qiu, & Wanasundara, 2011) This technique is based on the structural similarity of evolutionarily related proteins. (Chandonia & Brenner, 2005; Vitkup, Melamud, Moult, & Sander, 2001)

Generating a homology model of an amino-acid sequence is a multi-step process: (i) template identification, (ii) sequence alignments and (iii) model building (Joo, Lee, & Lee, 2012) (Vyas, Ukawala, Ghate, & Chintha, 2012).

1.4 Protein–Ligand Binding Models

Binding mechanisms of proteins and ligands have been explained by three known models; the lock-and-key (E. Fischer, 1894), induced fit (Koshland, 1958) and conformational selection (Csermely, Palotai, & Nussinov, 2010; Du et al., 2016; Ma, Kumar, Tsai, & Nussinov, 1999; Tobi & Bahar, 2005; C.-J. Tsai, S. Kumar, B. Ma, & R. Nussinov, 1999).

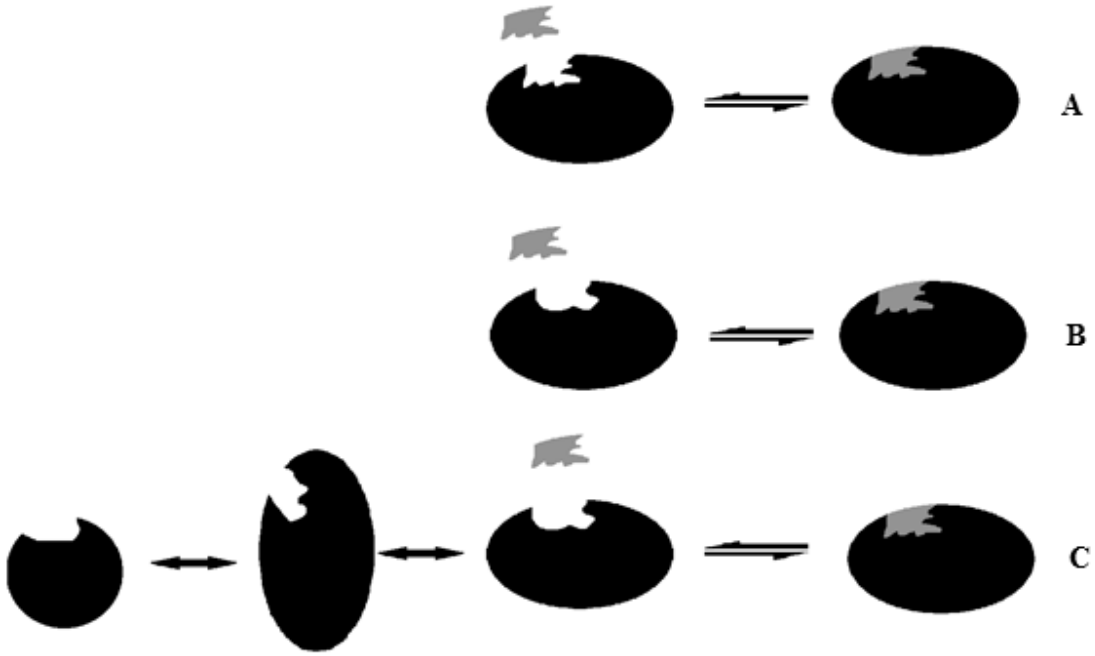


Figure 1. Three protein models; (A) lock-and-key model, (B) induced fit model and (C) conformation selection model.

1.4.1 The lock-and-key model. The lock-and-key model (Figure1A) employs a rigid ligand being fit into a rigid binding site of a rigid protein, where the ligand fits into the binding pocket like a key in a lock. However, this mechanism conflicted the experimental evidence demonstrated by the protein-ligand complexes whose initial protein and ligand structures did not resemble the final structure.

1.4.2 The induced fit model. Contrary to the lock-and-key model, the induced fit model (Figure 1B) employs a flexible binding site in the protein and a conformational change is induced at the binding pocket by the approaching ligand. This model illustrates the binding mechanism of the protein-ligand complex demonstrating minor conformational changes after the ligand binding at the binding site.

1.4.3 The conformational selection model. One main characteristic assumption in both induced fit and lock-and-key models is that the protein adopts a singular, stable conformation throughout the binding process. But, most proteins are inherently dynamic. And the conformational selection model considers this inherent flexibility.

The conformational selection model (Figure 1C) defines protein with its implicit flexibility based on the free energy landscape (FEL) theory of protein structure and dynamics. (Bryngelson, Onuchic, Socci, & Wolynes, 1995; Frauenfelder, Sligar, & Wolynes, 1991; Henzler-Wildman & Kern, 2007; Miller & Dill, 1997) This model hypothesizes the protein to be an assembly of conformations existing in equilibrium with different population distributions. Therefore, an approaching ligand has the opportunity to choose the most appropriate conformation and shift the equilibrium accordingly.

1.5 Computer-Aided Drug Design

Computer-aided drug design (CADD) is utilized to illustrate the application of molecular modeling methods and computational chemistry to drug design. Escalation of computational influence enabled CADD to study more complex biomolecular systems and to define, develop and apply more physically accurate models.

1.5.1 Motivation. Employing experimental methods such as X-ray crystallography or NMR solvation techniques to atomically resolve ligand bound complexes at an atomic level is extremely time-consuming and laborious and therefore limits their applicability in drug design. With that said and done, virtual docking and in-silico screening provides rapid and relatively accurate resolution for rationalization and visualization.

1.5.2 Protein–ligand docking. The most economical and relatively fast computational tool used for in-silico prediction of binding modes and affinities is molecular docking (Sousa et al., 2013). In the contemporary drug discovery process, protein–ligand docking (a sub-category of molecular docking) signifies a particularly essential methodology. (S.-Y. Huang & Zou, 2010; Manly, Chandrasekhar, Ochterski, Hammer, & Warfield, 2008; Sousa et al., 2013) Protein-ligand docking is utilized to virtually-screen large libraries of prospective ligands and identify the lead compounds (Sergio, Nuno, Pedro, & Maria Joao, 2010). Consequently, over the past 20 years, protein–ligand docking played an active role in pharmaceutical research. There is a great assortment of docking software packages available for academic and commercial use. Well-known among them are AutoDock (G. Jones, Willett, Glen, Leach, & Taylor, 1997; Morris et al., 2009), DOCK (Ewing, Makino, Skillman, & Kuntz, 2001; S. Mukherjee, Balius, & Rizzo, 2010), FlexX (Rarey, Kramer, Lengauer, & Klebe, 1996), Glide (Friesner et al., 2006) and GOLD (G. Jones et al., 1997).

There are two essential components in protein–ligand docking: (i) the search algorithm; searches for good binding poses of the ligand with respect to the receptor in the binding pocket (ii) the scoring function; estimates the binding affinity of the generated binding poses, ranks them, and identifies the most favorable binding pose(s) of the ligand with respect to the receptor in the binding pocket. (Du et al., 2016)

The search algorithms have evolved with the protein-ligand binding mechanisms; from the completely rigid-body methods to the flexible ligand-rigid protein, and then to the flexible ligand–flexible protein methods. (Sousa et al., 2013)

1.5.2.1 Scoring functions. To assess the binding affinity of the ligand towards the protein, fast and approximate mathematical methods, scoring functions, are used. (Ewing et al., 2001) Scoring functions render a compromise between speed and accuracy based on various approximations. Scoring functions are thus particularly suitable for high-throughput tasks, such as molecular docking, virtual screening, library design, and so on. (Liu & Wang, 2015)

Many scoring functions are available for protein–ligand docking studies. They are classified into three general classes based on how they were devised: the force-field-based, the empirical-based, and the knowledge-based scoring functions. (Du et al., 2016)

1.5.2.1.1 Force-field-based scoring functions. The force-field defines the potential energy of the system. In the force-field-based scoring functions, the binding affinities are estimated based on force-field parameters (physics-based functions and parameters) derived from quantum mechanical calculation of non-covalent interactions (N. Huang, Kalyanaraman, Irwin, & Jacobson, 2006). Including the molecular interactions induced by binding, changes induced in the solvent and particularly, the entropic effects would give a more accurately estimated binding affinity. Explicit treatment of water molecules or utilizing implicit solvent models can justify the solvent effect. Implicit solvent models such as Poisson–Boltzmann surface area (PB/SA) model (Rocchia et al., 2002; J. Wang, Morin, Wang, & Kollman, 2001) and the generalized-Born surface area (GB/SA) model. (G. D. Hawkins, C. J. Cramer, & D. G. Truhlar, 1995; Still, Tempczyk, Hawley, & Hendrickson, 1990)

$$\Delta G_{Binding} = \Delta E_{vdW} + \Delta E_{Electrostatic} + \Delta E_{H\text{-}bond} + \Delta G_{De\text{-}solvation} \quad (1.4)$$

1.5.2.2 Empirical scoring functions. Empirical scoring functions utilize either machine learning methods or regression to parameterize the interactions as favorable or unfavorable (penalty) energy terms. (Eldridge, Murray, Auton, Paolini, & Mee, 1997; Grinter & Zou, 2014) These energy terms include contributions from hydrophilic contacts, hydrophobic contacts, electrostatic and van der Waals energies, number of hydrogen bonds, number of rotatable bonds that are immobilized upon complex formation, or change in solvent accessible surface area (SASA) upon complex formation (Tang, 2010). PLP, (G. Verkhivker, Appelt, Freer, & Villafranca, 1995) ChemScore, (Eldridge et al., 1997; Murray, Auton, & Eldridge, 1998) X-Score, (R. Wang, Lai, & Wang, 2002) and GlideScore (Friesner et al., 2004; Friesner et al., 2006) are popular examples of empirical scoring functions.

1.5.2.2.1 Xtra precision glide score. XP Glide docking function employed in this thesis falls under semi-empirical scoring functions which employs H_2O de-solvation energy terms. It was reported that in XP glide docking, the scoring function reproduced experimental binding affinities of 198 ligands on various complexes with known experimental binding affinities. Out of the 198 ligands 132 ligands docked agreeably with root mean square deviations and average absolute deviations of 1.73 kcal/mol and 1.34 kcal/mol respectively.(Friesner et al., 2006)

Glidescore employed by Glide software is formulated as follows

$$\text{XP Glidescore} = E_{Coul} + E_{vdW} + E_{Bind} + E_{Penalty} \quad (1.5)$$

$$E_{Penalty} = E_{De\text{-}solvation} + E_{Ligand\text{-}strain} \text{ (Friesner et al., 2006)} \quad (1.6)$$

1.5.2.3 Knowledge-based scoring functions. The knowledge-based scoring functions assume that the close inter-atomic interactions between the protein and the ligand occurring more frequently can be energetically favorable than those anticipated by a random distribution and for that reason, contribute favorably to the binding affinity (Muegge, 2006). In other words, the statistical potentials are derived from the close contacts statistically studied in a training set containing suitable samples.

$$A = \sum_{i}^{lig} \sum_{j}^{prot} \omega_{ij}\,(r) \tag{1.7}$$

Each scoring function has its own pros and cons and none of them are neither accurate nor generally applicable. Therefore, the idea of using a combination of scores from multiple scoring functions, consensus scoring strategy, has been introduced to improve the accuracy and applicability. (Charifson, Corkery, Murcko, & Walters, 1999; S.-Y. Huang, Grinter, & Zou, 2010; G. M. Verkhivker et al., 2000)

1.5.2.4 Limitations and practical considerations. Most docking methods employ various limiting assumptions and oversimplifications, such as rigid binding site, inaccurate solvent representation, random probability distribution functions used in conformational searches etc. Many virtual screening studies identified a high percentage of false positives because of these often-necessary simplifications and inherently inaccurate implementations. The utility of docking studies conducted to identify novel potent ligands are limited by the rigid depiction of the ligand binding site (neither side-chain nor backbone flexibility) as the ligand may bind to the protein by an induced fit mechanism. This unsophisticated oversimplification is ignoring certain important energetic modifications induced by potential structural changes usually observed in the ligand binding process.

Additionally, a rigid-binding site does not account for enthalpy-entropy compensation (discussed in Section 1.2.2).

In docking, implicit solvent representations are used. Docking methods ignore the critical role of bound water molecules, accounting to the lack of explicit solvent representation. This misrepresentation might lead to inaccurate prediction of the binding pose. In these cases, including explicit waters at the binding site might achieve more accurate docking predictions. Accurate binding affinity estimations can be achieved by accurate physical representation of solvation and de-solvation effects.

1.5.3 Binding affinity predictions. One of the most critical and challenging components to structure-based CADD is predicting binding affinity. (Ajay & Murcko, 1995; Gohlke & Klebe, 2002) Predicting accurate binding affinity is essential to various applications including identification of native binding mode using molecular docking, identification of lead compounds by virtual screening of ligand libraries, and increasing target specificity and enhancing binding affinity for lead optimization. (Kitchen, Decornez, Furr, & Bajorath, 2004; Lyne, 2002; Shoichet, 2004) Even though first-principle methods such as free energy perturbation (FEP), (Kollman, 1993) linear interaction energy (LIE), (Hansson, Marelius, & Åqvist, 1998) and MM-PBSA/GBSA (P. A. Kollman et al., 2000) have gone through significant developments to predict accurate binding affinity (Beveridge & DiCapua, 1989; Hansson et al., 1998; P. A. Kollman et al., 2000), fast and relatively accurate empirical scoring functions are still widely used in drug discovery. (Böhm & Stahl, 2003)

1.5.3.1 Free energy calculations. Principles of statistical thermodynamics are utilized in protein-ligand binding free energy calculations. These are extensive computational simulations (Molecular Dynamics or Monte Carlo) based calculations and require computational efforts of higher magnitude by several orders than the traditional scoring functions. As a reward for the highly intensive computation, the results of free energy calculations ought to be reliable and almost quantitative.

The free energy calculations carry an advantage over the faster scoring functions by including both the energetic (solvation energy and potential energy) and entropic (solvent effects and flexibility/dynamics of both protein and ligand) contributions. And the free energy calculations do not require case-by-case parameter fitting. (de Ruiter & Oostenbrink, 2011; S. Thomas & Andreas, 2010)

The three main types of free energy calculations: the alchemical calculation, the path sampling, and the endpoint methods. Many factors like the length of the simulation, whether the absolute or relative binding free energy was calculated and whether an implicit or explicit solvent was used would affect the efficiency and accuracy of the calculations.

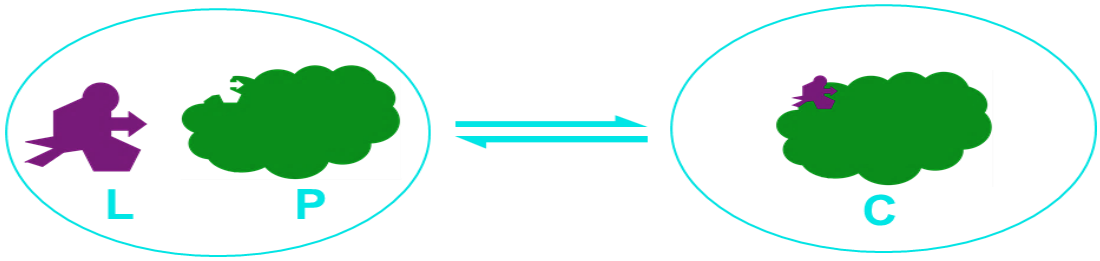


Figure 2. Endpoint methods.

1.5.3.1.1 Endpoint methods. The endpoint method calculates the binding free energies of the unbound state and bound state only. The intermediates stages are not considered. It is to be noted that endpoint method could be the efficient method of all three. The most endpoint methods applied to the binding free energy calculations, are molecular mechanics generalized Born surface area (MM-GBSA) and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA). (P. A. Kollman et al., 2000; Srinivasan, Cheatham, Cieplak, Kollman, & Case, 1998) In the GB/SA and PB/SA method, the binding free energy is calculated as:

$$\Delta G_{Bind} = G_C - (G_P + G_L) \quad (1.4)$$

Where G_C is the free energy of the protein-ligand complex's molecular system, G_P is the free energy of the protein molecular system and G_L is the free energy of the ligand molecular system. (Joseph M. Hayes, 2012) The free energy of each system is defined as:

$$G = E_{MM} + G_{Solv} - TS \quad (1.5)$$

Where E_{MM} is the total molecular mechanics energy of molecular system in the gas phase, G_{Solv} is a solvation free energy of the molecular system in solvent and T is temperature and S is the entropy of the molecular system (Joseph M. Hayes, 2012) (as explained in section 1.2.1.2). The total molecular mechanics energy of molecular system is defined as the sum total of energies contributed by covalent interactions (E_{Bonded}), electrostatic interactions (E_{Elec}) and van der Waals interactions (E_{VdW}). (Joseph M. Hayes, 2012)

$$E_{MM} = E_{Bonded} + E_{Elec} + E_{VdW} \quad (1.6)$$

The molecular mechanics energy of the system is computed by the molecular mechanics energy function known as force field. The solvation free energy constitutes

polar ($G_{GB/PB}$) and non-polar (G_{SASA}) contributions from the solvent. (Joseph M. Hayes, 2012)

$$G_{Solv} = G_{GB/PB} + G_{SASA} \quad (1.7)$$

The polar component is interpreted by the generalized Born (GB)/Poisson or Poisson-Boltzmann (PB) model. The non-polar component is considered to be proportional to solvent-accessible surface area (SASA). (Joseph M. Hayes, 2012)

GB/SA and PB/SA methods are more relevant to ranking of ligand binding affinities rather than to predict absolute binding free energies owing to its intrinsic approximations. Although, incorporating solute entropy (Foloppe & Hubbard, 2006) and solvent effects (Singh & Warshel, 2010) in binding affinity calculations is challenging, many studies applied PB/SA and GB/SA methods successfully and have generated some promising results. (Joseph M. Hayes, 2012) As MM-GBSA calculation ranges between their intermediate position between the empirical scoring and rigorous alchemical calculation methods in terms of both accuracy and computational intensity, PB/SA and GB/SA methods could be useful for post-processing of the docked structures or be used to rationalize the observed differences. (Genheden & Ryde, 2015)

1.5.3.1.2 Limitations and practical considerations. Although empirical scoring functions came a long way, there is still room for significant improvement in both applicability and accuracy. Accurate ranking of binding poses based on the relative affinities is still a challenge. The inability of the scoring functions used in virtual-screening studies to characterize the accurate binding nature constitutes as another limitation, especially the unrepresented systems of the training sets. Considering a best-case scenario where the docking method successfully predicted an accurate binding pose, the rigidity of

the protein might hinder the accuracy of the scoring function and generate false positives and false negatives. Under the assumption of correct binding pose prediction, this limitation could be circumvented by estimating binding affinities by utilizing first-principle methods.

1.6 Molecular Dynamics Simulations

As mentioned in section 1.5.2.4 the rigidity of protein is a limitation to the CADD and needs to be dealt with. However, to deal with this and develop a computational technique that can simulate protein dynamics, highly complicated and computationally demanding quantum-mechanics (QM) based calculations pertaining large molecular systems are required.

Molecular dynamics (MD) simulations, first developed in the late 1970's (McCammon, Gelin, & Karplus, 1977), simulate atomic motions utilizing Newtonian physics based on unassuming approximations thereby reducing the computational intricacy.

Initially, NMR spectroscopic, X-ray crystallographic, or homology-modeling data, in that preferential order, is utilized to formulate a molecular model of the molecular system. A potential energy estimation is made by formulating the forces acting on every atom of the system (Cornell et al., 1995). In short, covalent and non-covalent interactions of the system. Simple virtual springs, were utilized to model chemical bonds; sinusoidal function that approximates the energy differences between eclipsed and staggered conformations was utilized to model dihedral angles and atomic angles. The Lennard-Jones 6- 12 potential (J. E. Jones, 1924) was utilized to model van der Waals interactions and Coulomb's law for electrostatic interactions. These energetic terms need to be

parameterized to fit QM calculations and experimental data to be able to simulate the natural dynamics of the molecules. All these parameters are collectively known as a 'force-field'. Because these parameters define the forces that control and effect the dynamics simulation. Commonly known MD simulation force-fields are AMBER (Cornell et al., 1995; J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, & D. A. Case, 2004), CHARMM (Brooks et al., 1983), GROMOS (Christen et al., 2005) , NAMD (Kale et al., 1999; Phillips et al., 2005) etc. They differ only in the approach of parameterization.

Once the force-field has been defined, all atoms of the system are stimulated according to Newton's laws of motion. The molecular system will be simulated often by only 1-2 quadrillionths of a second, and this process is repeated, typically for a million times. As it is apparent that MD simulations require so many calculations, computer clusters or super-computers utilizing multiple processors in parallel are used to conduct them. One of the applications of MD simulations is validation of a force-field by comparing simulated data with experimental data (van Gunsteren, Dolenc, & Mark, 2008).

Many properties such as the time evolved root-mean-squared coordinate deviation (RMSD) to the initial structure, distance time series, angle time series, energy time series, H-bond time series etc., are used to characterize these MD simulation trajectories. However, extracting and evaluating some properties such as entropies and enthalpies can be time-consuming. Moreover, interpretation of these properties and the intrinsic relationships between the molecular configurations could be lost in the complexity of the data. To simplify this complexity a data mining tool called clustering analysis is used. (Karpen, Tobias, & Brooks, 1993; Shao, Tanner, Thompson, & Cheatham, 2007; Shenkin & McDonald, 1994) Clustering analysis groups MD simulated conformations based on

homogenous structures. (Barnard & Downs, 1992) Every conformation is divided into various groups known as clusters based on relative similarity. Each cluster has MD simulated conformations that resemble each other or are distinct from every other cluster. (Hartigan & Wong, 1979)

Regardless of the flaws in conformational sampling and current force fields, the insights offered by MD simulations into protein dynamics play essential roles in CADD. In the drug-binding process, the small molecule (drug) does not find a distinct rigid structure, rather a large dynamic molecular system in constant motion. MD simulations succeed where NMR, X-ray crystallography, and homology modeling generated rigid-receptor models fail to reproduce the dynamic molecular recognition and drug binding processes. Whether it's a lock-key model, where the drug might bind to a rigid binding pocket and the receptor dynamics are limited (E. Fischer, 1894) or a more common induced-fit model, where the drug binds only to a handful of conformations sampled by its dynamic receptor (Kumar, Ma, Tsai, Wolfson, & Nussinov, 1999; Ma et al., 1999; Ma, Shatsky, Wolfson, & Nussinov, 2002; C. J. Tsai, S. Kumar, B. Ma, & R. Nussinov, 1999) or it's inducing more conformational changes that cannot be sampled in its absence (Koshland, 1958), dynamics of the receptor play a crucial role in drug-binding process.

Chapter 2

CADD in Boron Therapeutics

2.1 Abstract

Tavaborole, the recently approved topological anti-fungal drug, inhibits leucyl tRNA synthetase by bonding covalently and hinders protein synthesis. The benzo-boroxole pharmacophore of tavaborole is responsible for its unique activity. With proper understanding of the tavaborole binding site, designing a molecule to enhance the binding affinity of tavaborole analogs should be quite achievable. The 3D crystal structure of fungal leucyl tRNA synthetase has not been solved yet. So, a theoretical 3D model of fungal leucyl tRNA synthetase has been generated and a combinatorial library has been generated by optimizing libraries of already synthesized drugs designed based on biological activity of amino-benzo-boroxoles on anti-cancer cell lines using the previously generated 3D model. The synthesized drug library mimics the activity of bortezomib. However, the pharmacophore benzo-boroxole would be more relevant to tavaborole rather than to bortezomib with a boronic acid pharmacophore. This study theoretically proposes molecules with prospective improved affinity towards fungal leucyl tRNA synthetase. To improve the selectivity of these molecules a theoretical 3D model of human leucyl tRNA synthetase has been generated and the hits from fungal leucyl tRNA synthetase are analyzed at the human leucyl tRNA binding site.

2.2 Introduction

The element boron is not very commonly found in living bodies however, it has been gaining a lot of attention recently, accounting to its potential for new therapeutic biological activity and drug design. The attention may be new but utilization of boron containing compounds started long ago with boric acid and borax.(Tibi, 2006) Although boron-chemistry started with inorganic boric acid compounds and borax, it has now progressed to boron based organic chemistry (Baker et al., 2009; Baker, Tomsho, & Benkovic, 2011; Das et al., 2013; Hernandez et al., 2013) and has been approved as an anticancer (bortezomib (Adams et al., 1998)) and antifungal (tavaborole (Leśnikowski, 2016; Rock et al., 2007)) agents in 2003 and 2014 respectively.

2.3 Tavaborole

Tavaborole (5-fluoro-1,3-dihydro-1-2,1-benzoxaborole/AN2690) is one of the most effective treatment against onychomycosis. Tavaborole is a topically acting, broad-spectrum antifungal agent. (Baker et al., 2006; A. K. Gupta & Simpson, 2012)

2.3.1 Onychomycosis. A fungal infection of the nail plate or the nail bed is known as onychomycosis. (Seebacher et al., 2007; J. Thomas et al., 2010) 80-90% of the documented onychomycosis cases are assessed to be caused by the dermatophytes Trichophyton rubrum and Trichophyton mentagrophytes. (J. Thomas et al., 2010) When left untreated, the nail plate gradually deteriorates and separates from the nail bed. It has been reported that the incidence rate of onychomycosis is ~10% of the worldwide population, ~20% for the >60 years old population and ~50% for people aged >70 years. (Elewski, 2000; J. Thomas et al., 2010; Westerberg & Voyack, 2013) While onychomycosis is not life threatening, it can lead to the cause of lesions in other regions

and can have grave consequences in combination with various other conditions like diabetes and contribute to poor quality of life. (Drake et al., 1999; A. P. Gupta, Verma, & Ikram, 2000) There is always a chance of infecting others and becoming a public health hazard.

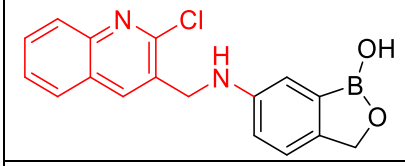
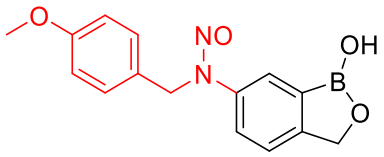
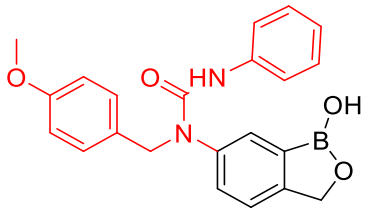
2.3.2 Mechanism of action. The mechanism of action of the fungicidal tavaborole has been explained by crystallographic, biochemical, and chemical studies. Tavaborole has been demonstrated to be effective against Trichophyton rubrum and Trichophyton mentagrophytes by targeting their leucyl tRNA synthetase (LeuRS). (Adamczyk-Woźniak, Komarovska-Porokhnyavets, Misterkiewicz, Novikov, & Sporzyński, 2012; Baker et al., 2006; Rock et al., 2007) LeuRS belongs to the class of aminoacyl-tRNA synthetases, a class of enzymes crucial for protein synthesis.

2.4 Hypothesis

Pathi et al., reported cell viability assays for synthesized amino-benzo-boroxoles (tavaborole analogs) and in-vitro IC_{50} for compounds with promising anti-cancer activity (Suman, Patel, Kasibotla, Solano, & Jonnalagadda, 2015) mimicking boronic acid containing bortezomib and ixazomib. However, the pharmacophore, benzo-boroxole is more similar to tavaborole. So, the goal of this study is to optimize anti-fungal activity of tavaborole analogs reported in table 1. This study attempts to optimize the antifungal activity of these molecules by modifying them and identify lead compounds by virtual screening.

Table 1

Biological activity of benzo-boroxoles in anticancer cell-lines.

Structure	MIA PaCa-2			MDA-MB-231		
	% Cell Viability		IC_{50}	% Cell Viability		IC_{50}
	50 µM	12.5 µM		50 µM	12.5 µM	
	28.6	28.2	8.3	44.8	44.3	11.5
	61.0	83.0		123.2	118.7	
	17.5	22.4	2.7	53.9	63.0	11.9

2.5 Role of LeuRS in Protein Synthesis

Major steps of protein synthesis are initiation, elongation, termination and folding. Amino-acylation reaction triggers the protein synthesis, followed by elongation of protein chain by formation of several peptide bonds and elongating the protein. The elongation is then terminated by the termination codon of mRNA and the newly synthesized protein is released which is consequently folded into its tertiary structure. (Banik & Nandi, 2013)

2.5.1 Amino-acylation reaction. The amino-acylation reaction binds an amino-acid with the transfer RNA. These amino-acids are attached to the transfer RNA (tRNA) by a class of enzymes called aminoacyl tRNA synthetases. The tRNA then transfers those amino acids onto the protein. (Ibba & Söll, 2000) There are two classes of aminoacyl tRNA synthetases. The difference between the classes being the transfer of the amino acid onto

2’-hydroxyl group in class I and onto 3’-hydroxyl group in class II. Most cells have at least 20 different aminoacyl tRNA synthetases, one for each essential amino acid. (Khan et al., 2011) However, many cells have additional aminoacyl tRNA synthetases. LeuRS belongs class II of the aminoacyl tRNA synthetases.

The aminoacylation of tRNA is a 2-step process: (i) the formation of aminoacyl adenylate from the amino acid in question triggers the 2nd step. The α-carboxylate group of the amino acid and the α-phosphate group of ATP forms aminoacyl adenylate, a mixed anhydride, in the presence of divalent magnesium (Mg^{2+}) ions and releases pyrophosphate. This pyrophosphate is further hydrolyzed and the equilibrium shifts forward. (Figure 3)

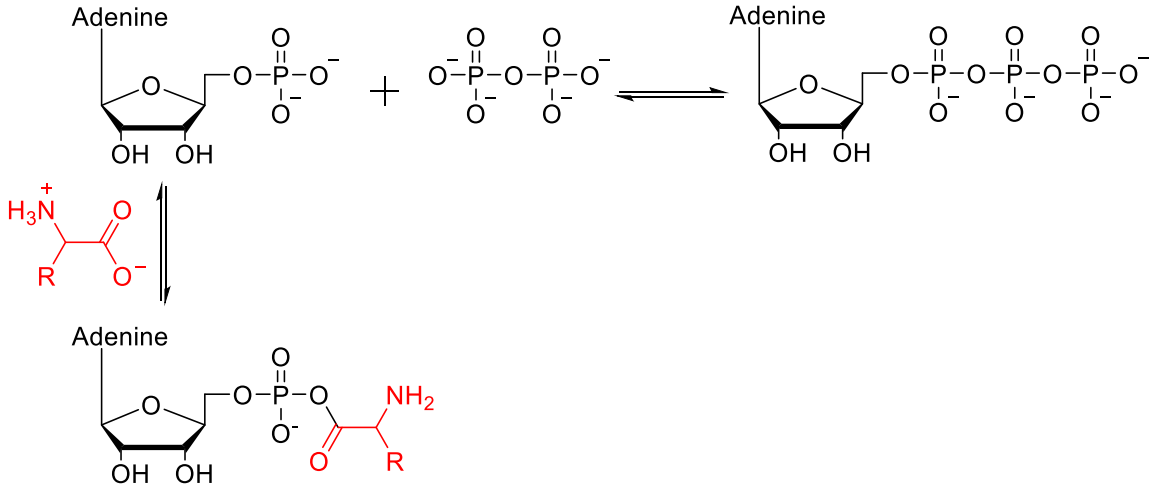


Figure 3. Step 1 of amino-acylation.

(ii) The amino acid in aminoacyl adenylate is transferred onto the 2′ or 3′ sugar hydroxyl group of the 3′-terminal adenosine nucleotide of the tRNA. (Figure 4) The accuracy of this process is very essential in ensuring the fidelity of the genetic code which

would otherwise lead to the synthesis of nonsensical proteins (Hong et al., 1996). To ensure this accuracy, most of the aminoacyl tRNA synthetases have an editing site to rectify an incorrectly aminoacylated tRNA (Schimmel & Schmidt, 1995). LeuRS has same proofreading mechanism. The synthetic and editing domains of LeuRS are separated by more than 30 Å (Cusack, Yaremchuk, & Tukalo, 2000; Fukunaga & Yokoyama, 2005).

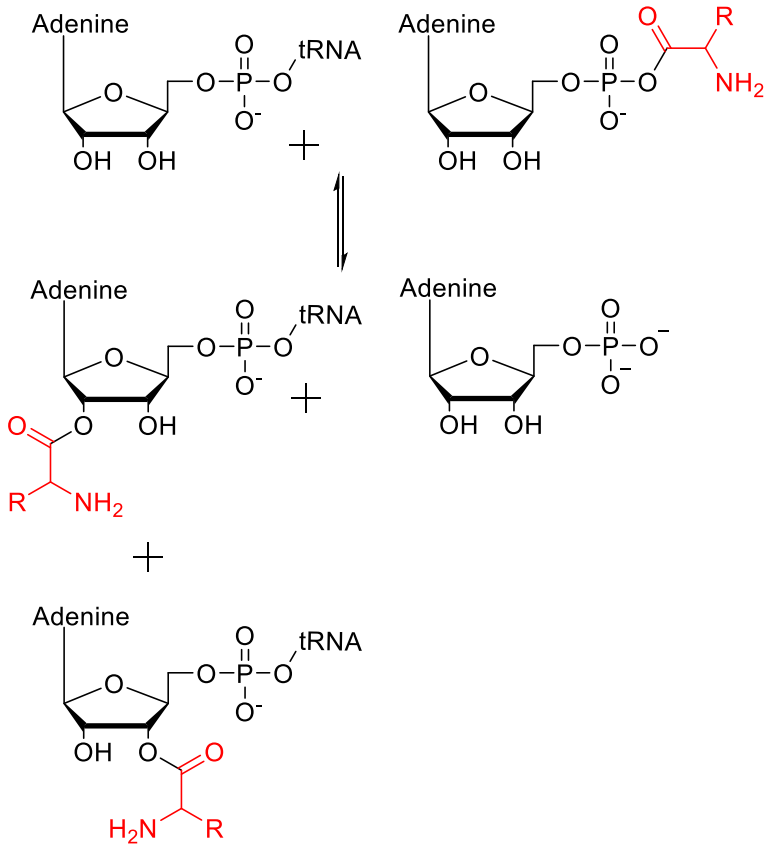


Figure 4. Step 2 of amino-acylation.

Many factors affect the selectivity for amino-acids; including amino-acid size, shape along with presence of an editing domain. (Guo & Schimmel, 2013) Concluding that aminoacyl tRNA synthetases are crucial for protein synthesis and cellular viability.

2.6 Tavaborole and LeuRS

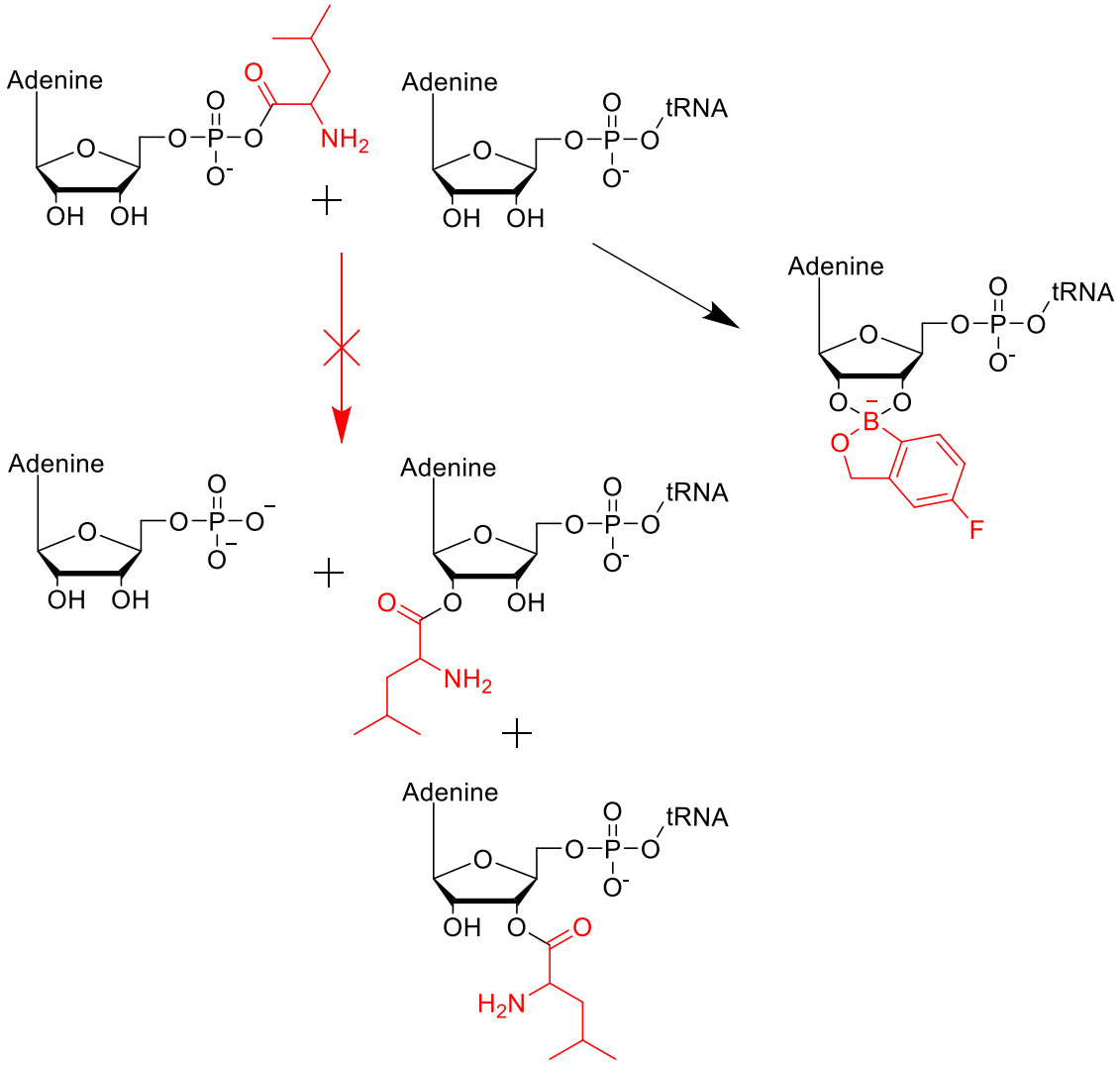


Figure 5. Inhibition of leucylation due to the formation of boronic ester

Tavaborole selectively binds to the editing domain of LeuRS. Tavaborole slowly and strongly binds to the binding site of leucine and renders the whole protein useless. (Figure 5) This subsequently stops protein synthesis or leads to synthesis of proteins with incorrect amino acid sequence. Eventually leading to apoptosis one way or the other.

(Lincecum et al., 2003) Rock et al reported that based on the X-ray crystallographic studies on LeuRS of Thermus thermophiles complexed with tavaborole, a tRNA-tavaborole spiro-ester adduct was formed by sp^2 hybridized boron from the boroxole ring and the two 2', 3'-hydroxyl groups on the terminal adenosine where boron is sp^3 hybridized with tetrahedral structure. The two hydroxyl groups which are essential to the amino-acylation reaction. This adduct formation is further stabilized by two H-bonds with threonine peptide and H_2O molecule. (Rock et al., 2007) This process is commonly referred to as the oxaborole tRNA trapping (OBORT) mechanism. (Baker et al., 2011) This stable complex has a half-life of ~7 hours at the active site. (Rock et al., 2007)

2.6.1 Structure-activity relationships (SAR) of benzo-boroxoles. Based on the SAR studies the 5-membered boroxole ring in which the boronic acid is embedded is critical for the therapeutic activity of the benzo-boroxoles. Comparative biochemical assays indicated substantial loss of antifungal activity with 6-membered ring and acyclic boronic acids analogs. (Rock et al., 2007)

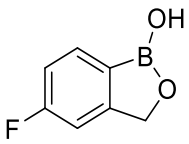


Figure 6. Structure of Tavaborole.

This unusual activity of the boroxole ring is fascinating, since the reaction coefficient of the very well-known boric, boronic and borinic ester formation with alcohols in aqueous solution by the corresponding acid is pKa dependent not the structure. (Martínez-Aguirre, Villamil-Ramos, Guerrero-Alvarez, & Yatsimirsky, 2013) Benzo-

boroxoles certainly have more pKa than acyclic boronic acids. (Yamaguchi et al., 2012) This however, cannot be correlated with disproportional increase of antifungal activity when compared with their corresponding acyclic boronic acids. Although in the absence of LeuRS, both benzo-boroxoles and acyclic boronic acids promptly forms esters interchangeably in neutral aqueous medium with mono-alcoholic and di-alcoholic compounds due to their obviously low association constants. (Martínez-Aguirre et al., 2013; Tomsho & Benkovic, 2012) Therefore, it is presumed that the hydrolysis of the boronic ester that usually occurs in the aqueous solution is prevented by the hydrophobic binding site of the editing domain in LeuRS. Thus the benzo-boroxole-LeuRS complex is selectively stabilized. (Baker et al., 2006)

2.7 Computational Approach

The goal of the study is to suggest modifications to the given library so as to optimize the fungicidal activity. The approach would be to

(i) validate the protein and ligand model,

(ii) virtually dock the ligands at the binding site and analyze the corresponding interactions,

(iii) identify the modification site,

(iv) enumerate the fragment library at individual modification sites and rank the fragments with respect to the fragment site,

(v) enumerate fragments at all sites and score the final modifications.

2.7.1 Challenges and assumptions. As mentioned in section 2.1.1.1 80-90% of onychomycosis is caused by Trichophyton mentagrophytes and Trichophyton rubrum. However, the 3D structure of LeuRS of neither is available in the protein data bank.

2.7.1.1 Homology modeling. A theoretical 3D structure of the Trichophyton rubrum has been generated by Prime (Jacobson, Friesner, Xiang, & Honig, 2002; Jacobson et al., 2004) (Schrodinger) software utilizing its amino-acid sequence from UniProt Consortium© (Magrane & Consortium, 2011) and the template, PDB ID: 2V0G from PDB (Berman et al., 2000). 2V0G is the 3D structure of LeuRS from Thermus Thermophilus, a Gram negative eubacterium, complexed with tRNA and characterizing formation of spiro-ester adduct of tavaborole with the ribose of adenosine- 76 at the editing site of the LeuRS.

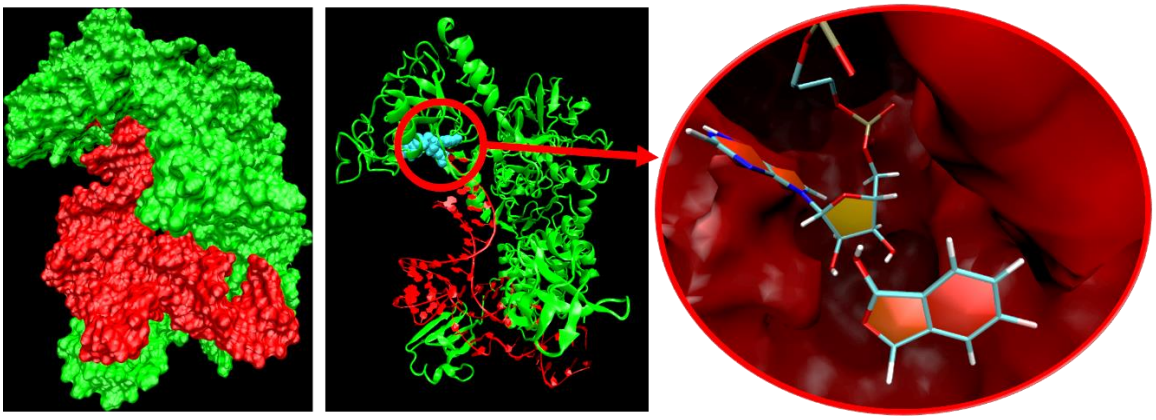


Figure 7. 3D representation of homology modeled fungal LeuRS complexed with tRNA non-covalently interacting with tavaborole.

The Glide (Friesner et al., 2004; Friesner et al., 2006; Halgren et al., 2004) software utilized to do molecular docking can form one covalent bond (covalent docking (Zhu et al., 2014)) with the amino-acid residues but the software has not been developed to conduct two subsequent covalent bonds formation with nucleic acids as observed with the spiro-ester formation.

To overcome this challenge only the approachability of the ligand is studied. As the

covalent bond formation is a fast and exothermic reaction which facilitates subsequent covalent bonds. The only challenge is approaching the binding site.

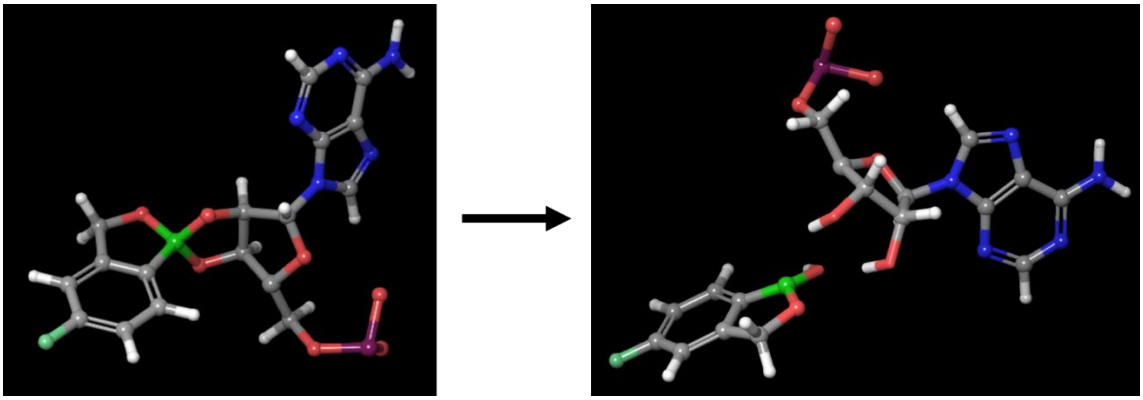


Figure 8. tRNA - Tavaborole complex adjusted from covalently bonded complex to non-covalently interacting moieties.

2.8 Comparison with Human LeuRS – Selectivity

Tavaborole is a topologically acting drug. Systemic activity can be induced when the ligands have preferential activity towards fungal LeuRS rather than human LeuRS. With this goal, the binding sites of both fungal and human LeuRS with tavaborole are modelled and compared.

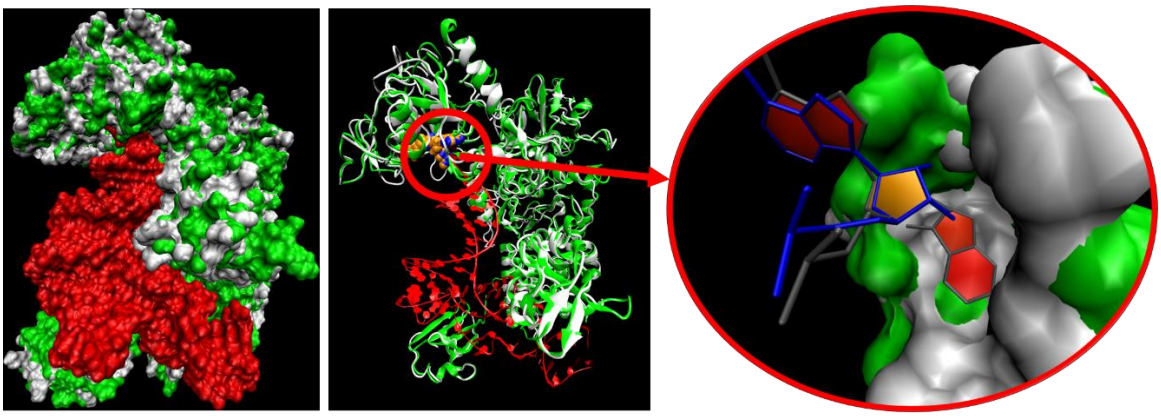


Figure 9. 3D representation of superimposed structures of homology modeled fungal and human LeuRS complexed with tRNA non-covalently interacting with tavaborole.

As mentioned in section 2.1.1.3 the SAR studies suggest the binding site of the fungal LeuRS and tavaborole to be hydrophobic. This is explained by comparison of the amino-acids interacting at the binding sites of both fungal and human LeuRS.

Table 2

Homology, identity and similarity statistics of human LeuRS compared to fungal LeuRS.

	Human LeuRS
Homology	45 %
Identity	34 %
Similarity	46 %

Figure 10. Sequence alignment of fungal LeuRS and human LeuRS; residues interacting with tavaborole at the binding site of fungal LeuRS are depicted in yellow and human LeuRS in blue. Residues depicted in green are common in both fungal and human LeuRS-tavaborole binding site.

The highlighted cells in the table 3 indicate amino-acids that might be responsible for high hydrophobicity at the binding site of tavaborole with fungal LeuRS compared to the human LeuRS.

Table 3

List of the interacting residues (highlighted in figure 10) at the binding site of fungal LeuRS compared with residues interacting at the binding site of human LeuRS; highlighted cells indicate residues that might be responsible for high hydrophobicity of tavaborole binding site in fungal LeuRS.

Fungal	**Human**
VAL	VAL
VAL	***LYS***
GLU	GLU
ALA	LEU
PRO	MET
SER	GLU
ILE	LEU
GLY	GLY
VAL	VAL
GLN	
LEU	
ALA	
GLN	PRO
LYS	LYS
ASP	GLU
LEU	***LYS***
ALA	ILE
TYR	TYR

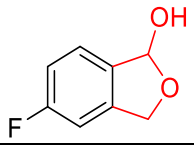
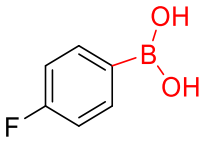
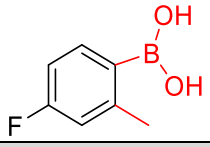
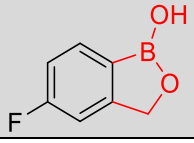
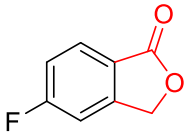
2.9 Methods

2.9.1 Inherent inaccuracies in experimental data. As mentioned in section 2.6.1.1 a theoretical 3D structure of LeuRS of Trichophyton rubrum has been generated by homology modeling using the template, PDB ID: 2V0G from PDB (Berman et al., 2000). The Protein Data Bank (PDB) has a comprehensive collection of X-ray and NMR solved biomolecular structures. (Westbrook et al., 2003) However, the experimental conditions like pH, temperature, salt concentrations etc., induced to crystallize the protein might differ from the actual physiological conditions. The minimal resolution of the model might result in inaccuracies on an atomic level. (Tang, 2010) For these reasons, this structure needs to be prepared to increase the accuracy and structural correctness of the 3D protein model. Schrodinger Maestro's protein preparation wizard was used to prepare the structure, optimize the H-bond network and minimize the potential energy of the protein model. ("Maestro© 2014 Schrödinger, LLC. Manuals,")

2.9.2 Validation. The protein-ligand model, 3D structure generated from homology modelling in complex with tavaborole, was validated by conducting molecular docking studies on compounds with reported inhibitory activity against cytoplasmic LeuRS of Saccharomyces cerevisiae S288c (Rock et al., 2007).

Table 4

List of IC_{50} values of compounds structurally similar to tavaborole and their corresponding docking scores and MM-GBSA values against fungal LeuRS. The highlighted cells depicts activity of tavaborole.

Structure	IC_{50}	Target Name	Docking score	MM-GBSA
	>100000 nM	Leucyl-tRNA synthetase, cytoplasmic	-4.348	-48.626
	>100000 nM	Leucyl-tRNA synthetase, cytoplasmic	-4.728	-43.9
	>100000 nM	Leucyl-tRNA synthetase, cytoplasmic	-4.515	-46.239
	=2100 nM	Leucyl-tRNA synthetase, cytoplasmic	-3.19	-56.181
	>100000 nM	Leucyl-tRNA synthetase, cytoplasmic	-3.86	-52.873

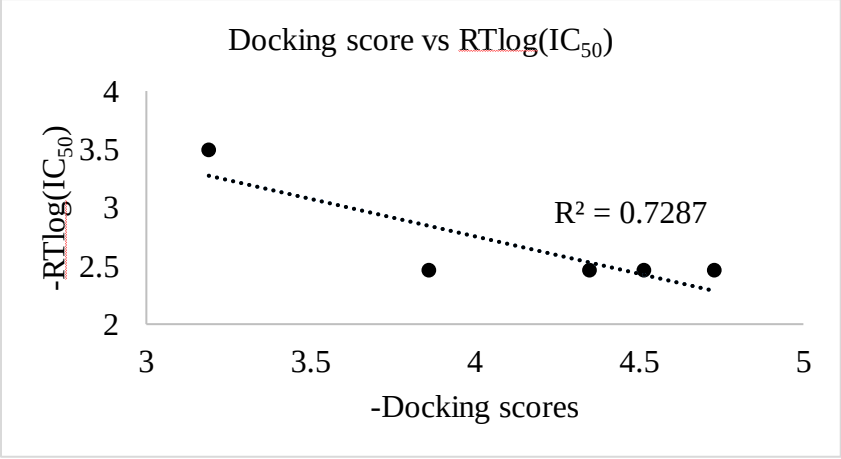


Figure 11. Graph plotting Docking score of the molecules listed in table 4 vs binding experimental binding affinity of the same.

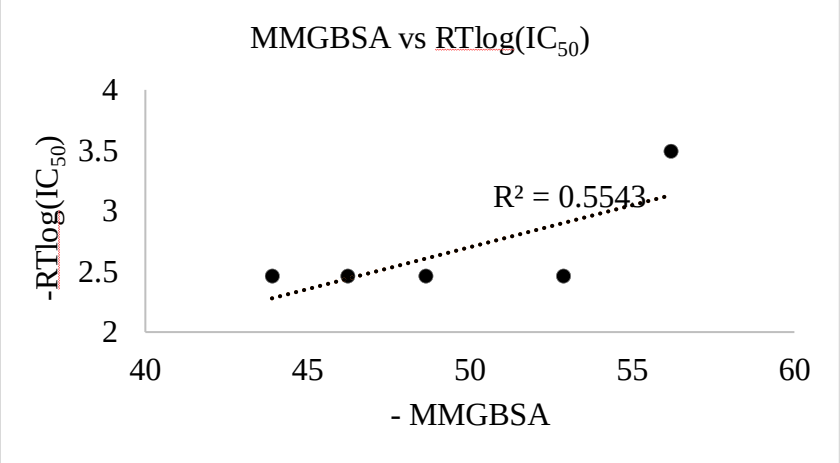


Figure 12. Graph plotting MM-GBSA values of the molecules listed in table 4 vs binding experimental binding affinity of the same.

The XP docking score generated by Glide and MM-GBSA free binding energy generated by Prime are used to validate the binding affinities. The docking score of the active compound is lowest when compared with other structurally similar compounds (indicated in Figure 6) however, the MM-GBSA binding energies show that the active compound has highest binding affinity when compared with the same (indicated in Figure 7). So, MM-GBSA free binding energy is used to validate the relative affinity of the ligands.

2.9.3 Screening the given library. The given libraries are screened based on the binding pose; extra precision glide docking is used to generate a binding pose for each ligand. The generated binding pose is then compared by super imposing with the binding pose of tavaborole. The ligands with binding poses that does not facilitate the formation of the spiro-adduct (boronic ester) are screened out. Since covalent bond formation between the boroxole of the ligands and the di-hydroxyl groups of the ribose from tRNA is essential for the spiro-adduct formation, the ligand binding pose that does not super impose its

boroxole with the boroxole of tavaborole is eliminated.

2.9.3.1 Ligand library 1 (LL1). In the tables 5, 6 and 7 listed below the highlighted cells have structures that successfully reproduced the binding pose that facilitates the covalent bond formation between boron and the ribose of adenosine of the tRNA.

Table 5

List of the docked molecules of ligand library 1 (LL1) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose.

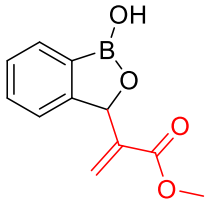
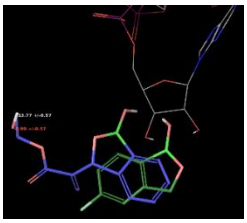
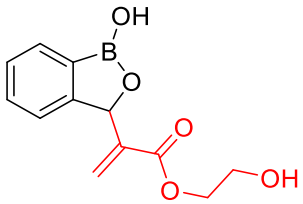
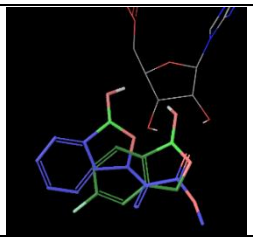
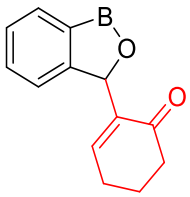
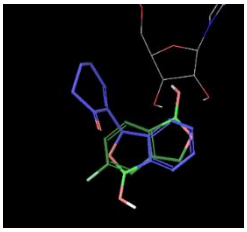
LL1_1	Binding Pose
	
LL1_2	Binding Pose
	
LL1_3	Binding Pose
	

Table 5 (continued)

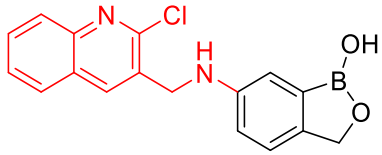
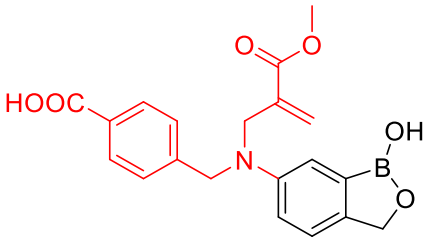
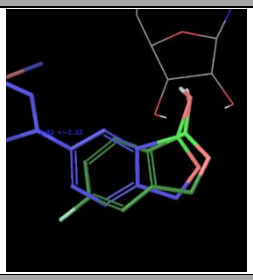
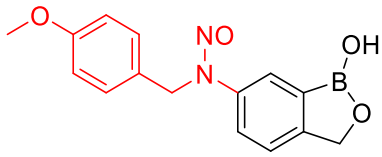
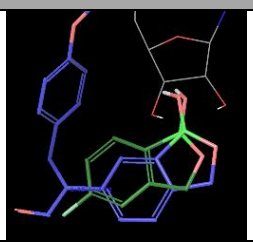
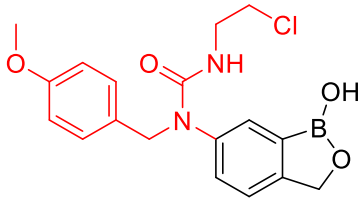
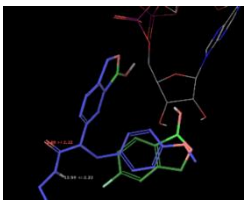
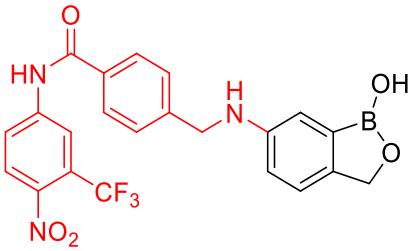
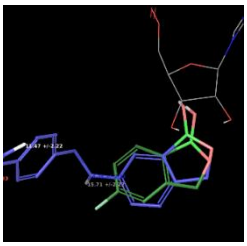
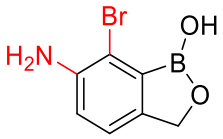
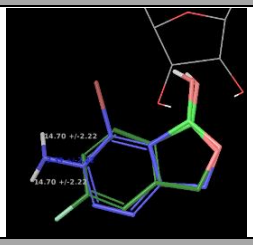
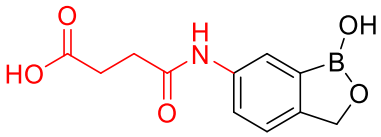
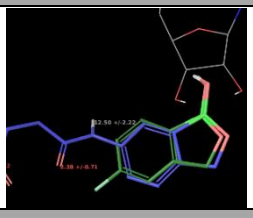
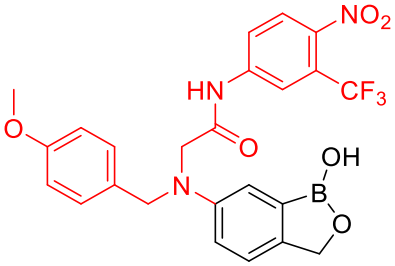
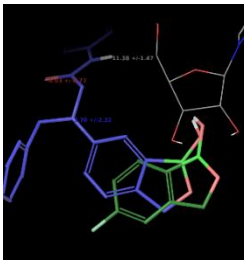
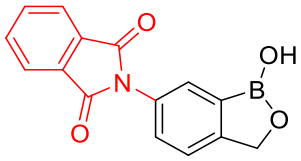
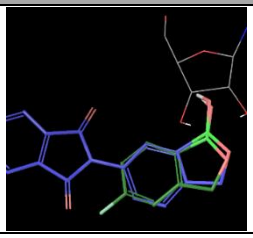
LL1_5	Binding Pose
	
LL1_6	Binding Pose
	
LL1_7	Binding Pose
	
LL1_8	Binding Pose
	
LL1_9	Binding Pose
	

Table 5 (continued)

LL1_10	Binding Pose
	
LL1_12	**Binding Pose**
	
LL1_13	**Binding Pose**
	
LL1_14	**Binding Pose**
	

2.9.3.2 Ligand library 2 (LL2).

Table 6

List of the docked molecules of ligand library 2 (LL2) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose.

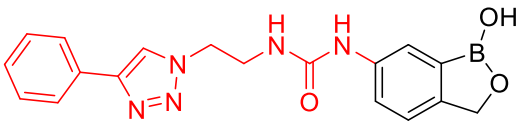
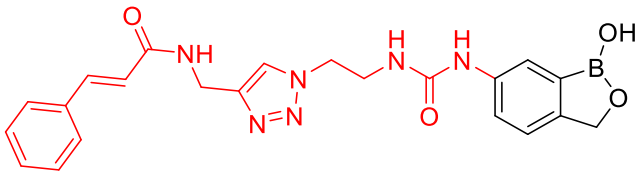
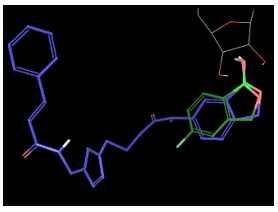
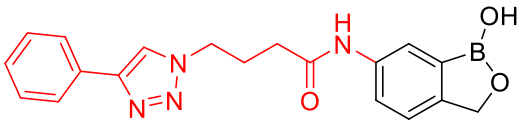
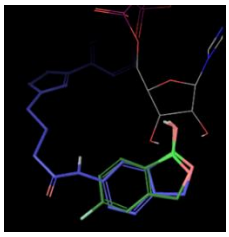
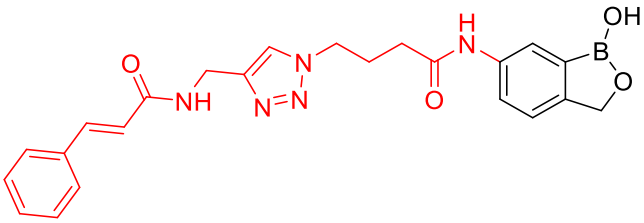
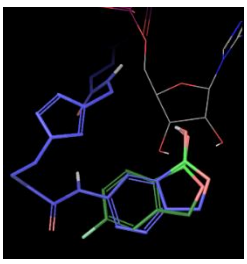
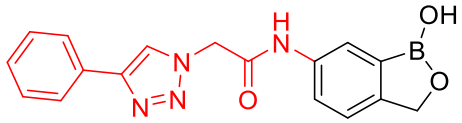
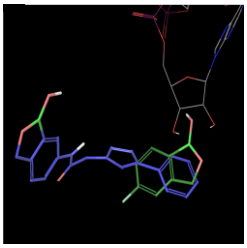
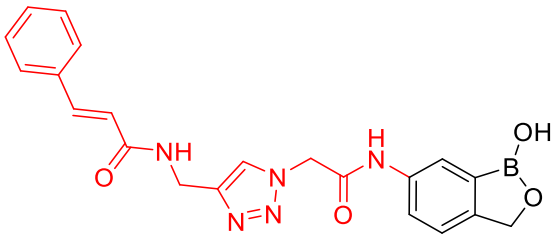
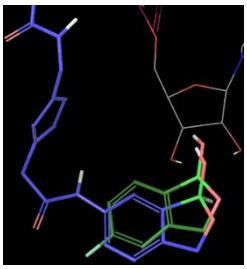
LL2_1	Binding Pose
	
LL2_2	Binding Pose
	
LL2_3	Binding Pose
	
LL2_4	Binding Pose
	

Table 6 (continued)

LL2_5	Binding Pose
	
LL2_6	**Binding Pose**
	

2.9.3.3 Ligand library 3 (LL3).

Table 7

List of the docked molecules of ligand library 3 (LL3) and their binding pose in complex with fungal LeuRS. The highlighted cells indicate molecules with appropriate binding pose.

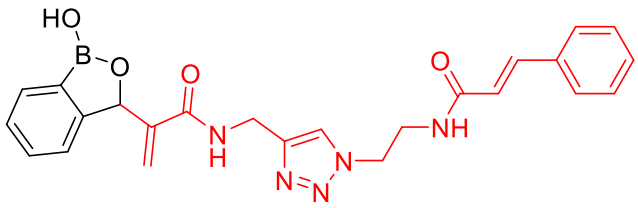
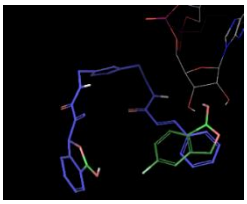
LL3_1	Binding Pose
	

Table 7 (continued)

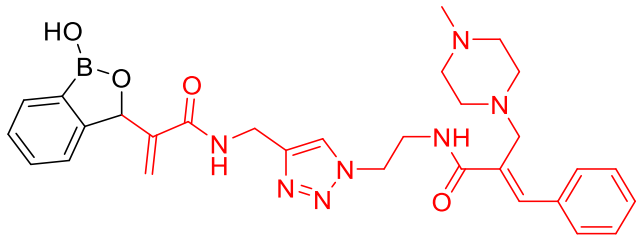
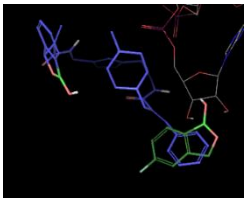
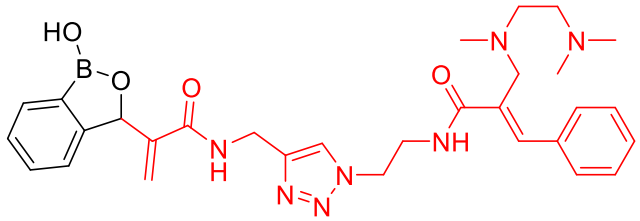
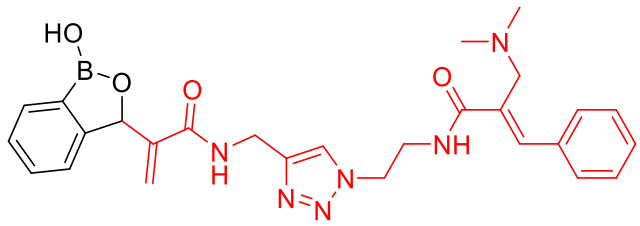
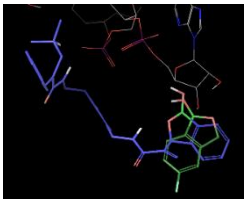
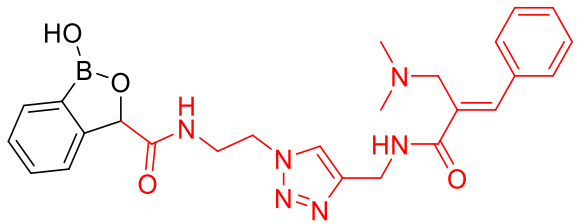
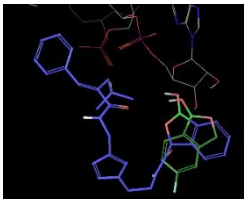
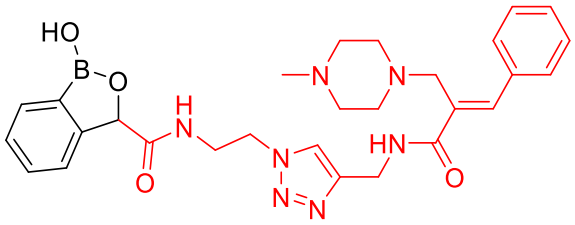
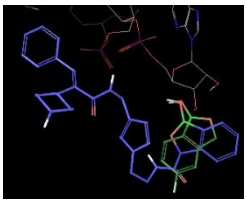
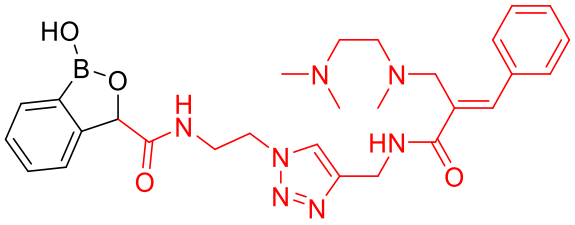
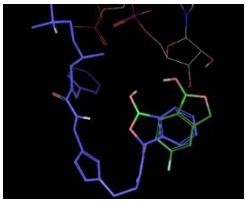
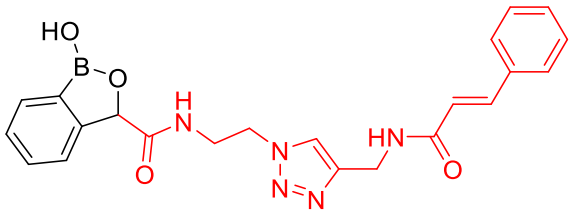
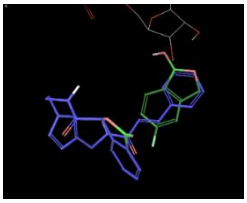
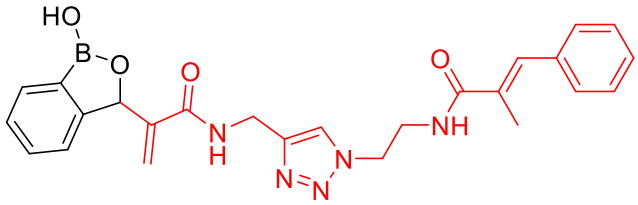
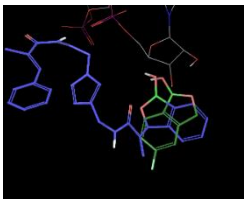
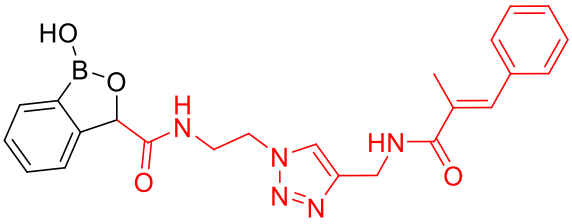
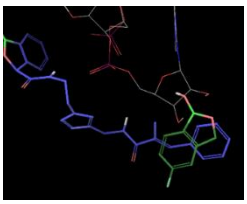
LL3_2	Binding Pose
	
LL3_3	Binding Pose
	
LL3_4	Binding Pose
	
LL3_5	Binding Pose
	

Table 7 (continued)

LL3_6	Binding Pose
	
LL3_7	Binding Pose
	
LL3_8	Binding Pose
	
LL3_9	Binding Pose
	
LL3_10	Binding Pose
	

2.9.4 Binding site analysis and enumeration. The binding site analysis provided by maestro is used to analyze and define prospective R-group positions to optimize the non-covalent interactions without disrupting the covalent bond formation. This binding site analysis indicates voids in the protein-ligand complex at the binding site. (i) Defining an R-group at each of these voids would optimize the affinity of the ligand without disrupting the covalent bond formation. Once these sites have been identified, (ii) the fragment library provided by maestro is used to enumerate R-groups at those defined positions. (iii) The resulting ligand library is docked into the binding site. (iv) The resulting docked binding poses are screened based on their comparability with binding pose of tavaborole. (v) Free energy binding affinities of these screened molecules (MM-GBSA ΔG_{Bind}) are calculated. 5 best fragments with high affinity binding poses are selected. After filtering out top affinity generating fragments at all predefined R-group sites, (vi) these fragments are then enumerated at their respective enumeration site. (vii) Once the generated library is docked and the incomparable binding poses are filtered out, a final active ligand library is generated. (viii) MM-GBSA ΔG_{Bind} is then calculated for this final ligand library.

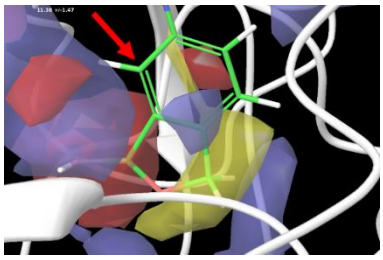
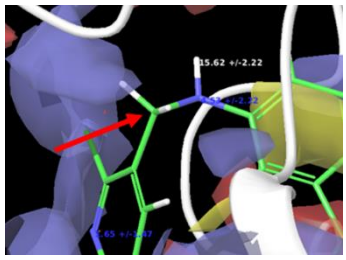
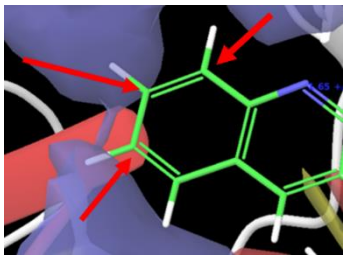
To achieve selectivity over human LeuRS, (ix) this final ligand library is docked into the binding site of the human LeuRS-tavaborole complex, which has been homology modelled using the same template used for generating fungal LeuRS-tavaborole complex. (x) Then a maximum of six best ligands displaying major affinity difference between fungal and human LeuRS and obviously favorable towards fungal LeuRS are screened. (Listed in results (2.9.5) section for every active ligand listed in tables 5, 6 and 7. For those ligands which gave no positive hits against human LeuRS, 6 best ligands with high MM-GBSA ΔG_{bind} are listed.)

2.9.5 Results.

2.9.5.1 Modifications on LL1_5. As mentioned in section 2.3.2 (i) Table 8 depicts the surfaces generated by binding site analysis and the voids represented by these surfaces. And these voids can accommodate a new R-group that optimize the affinity without disrupting the binding pose. All the prospective sites for modification are depicted in figure 13.

Table 8

The surfaces generated by binding site analysis at the LeuRS-LL1_5 binding site and arrows point to modification site of R_0, R_1, R_2, R_3 and R_4.

R_0	R_1	R_2, R_3 and R_4
		

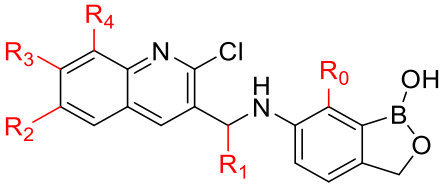


Figure 13. 2D structure of LL1_5 illustrating all the prospective modification sites.

Once the modification sites have been identified, these sites have been enumerated with fragment library provided by Schrodinger followed by steps (iii), (iv) and (v) (listed

in 2.9.4 section). The best fragments for ligand LL1_5 are listed in tables 9, 10, 11, 12 and 13.

Table 9

Hits from single site enumeration at R_0 of LL1_5.

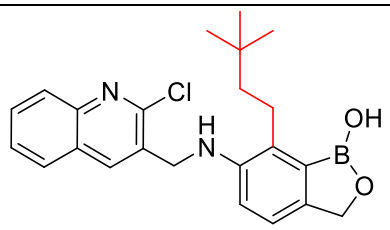
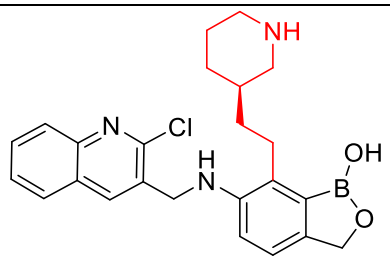
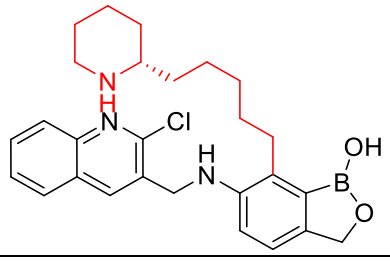
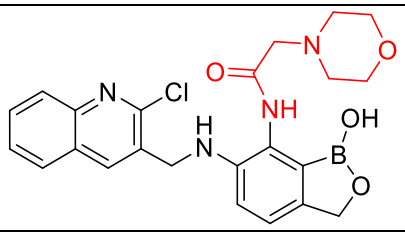
Structure	MM-GBSA ΔG_{Bind}
	-86.14
	-87.13
	-81.23
	-81.85

Table 10

Hits from single site enumeration at R_1 of LL1_5.

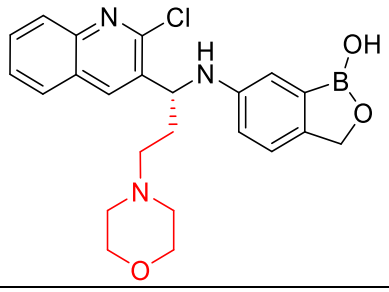
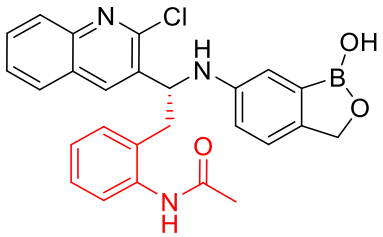
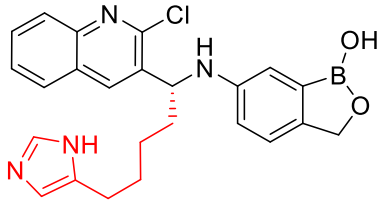
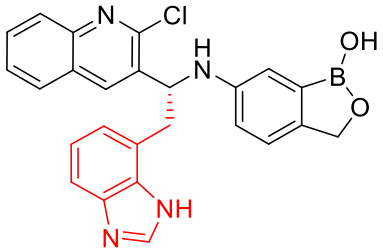
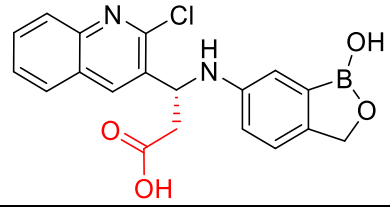
Structure	MM-GBSA ΔG_{Bind}
	-72.67
	-78.85
	-75.09
	-91.14
	-93.29

Table 11

Hits from single site enumeration at R_2 of LL1_5.

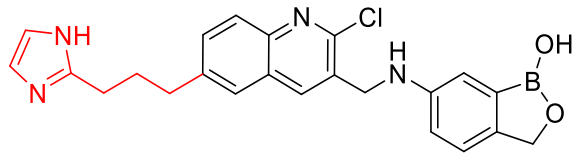
Structure	MM-GBSA ΔG_{Bind}
	-78.40

Table 12

Hits from single site enumeration at R_3 of LL1_5.

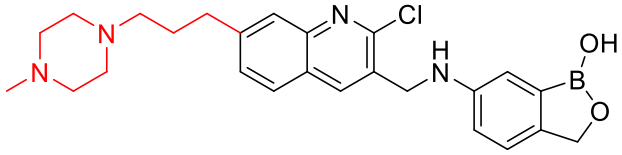
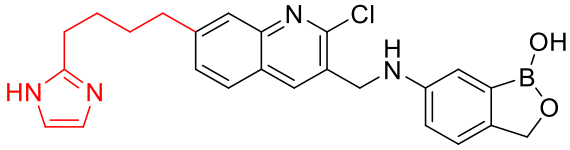
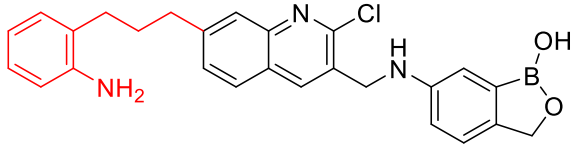
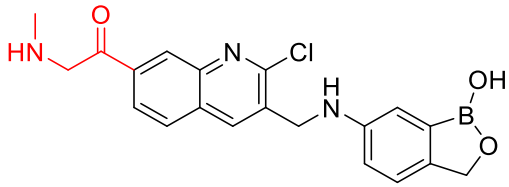
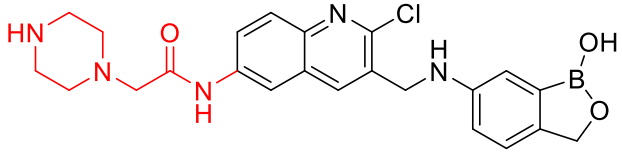
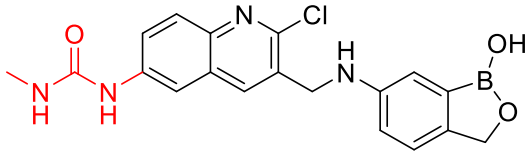
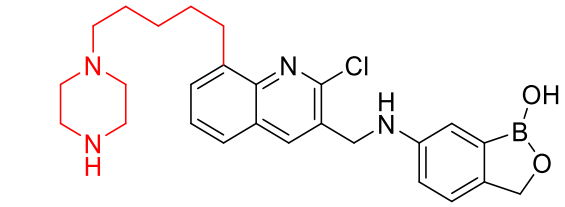
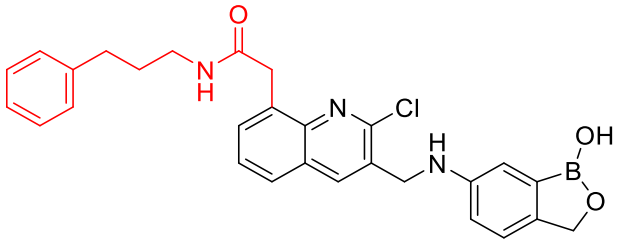
Structure	MM-GBSA ΔG_{Bind}
	-79.17
	-78.49
	-99.07
	-72.56
	-74.74
	-75.16

Table 13

Hits from single site enumeration at R_4 of LL1_5.

Structure	MM-GBSA ΔG_{Bind}
	-83.84
	-80.39

Once the top fragments have been identified, the steps (vii) (viii) (ix) and (x) (as listed in section 2.9.4) are carried out and the final hits are listed below

Table 14

Hits from enumeration at all modification sites of LL1_5 and comparison with human LeuRS.

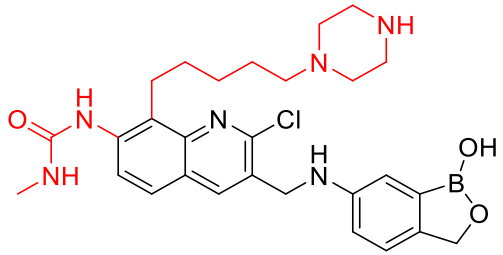
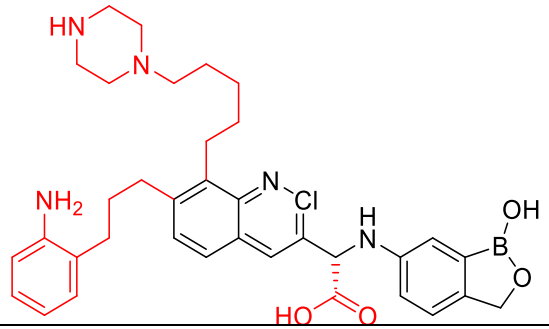
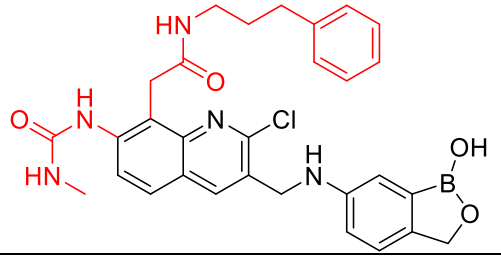
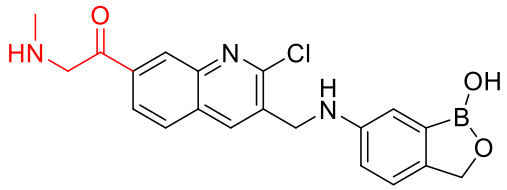
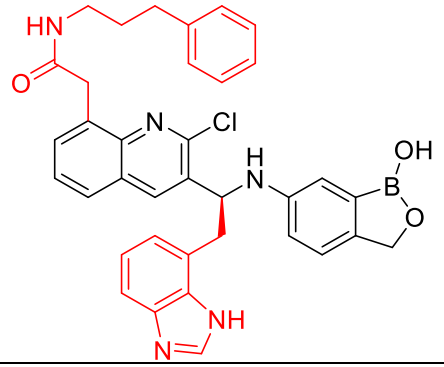
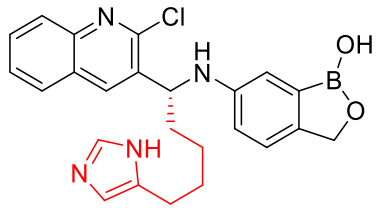
Structure	MM-GBSA ΔG_{Bind}			Docking Score		
	Fungal	Human	Difference	Fungal	Human	Difference
	-89.48	-42.9	-46.59	-8.718	-5.51	-3.208
	-85.83	-26.77	-59.06	-8.386	-8.105	-0.281
	-67.85	-49.19	-18.66	-4.707	-5.497	0.79

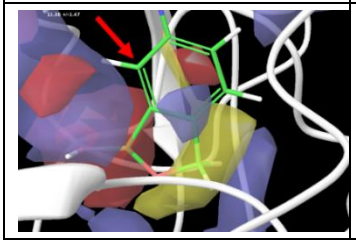
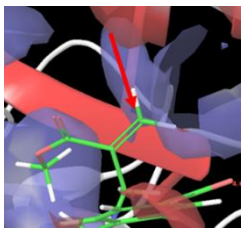
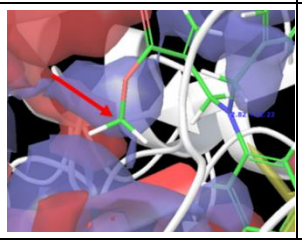
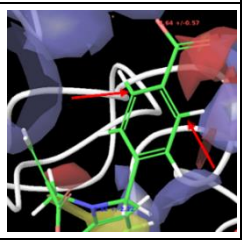
Table 14 (continued)

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-67.48	-57.09	-10.4	-6.485	-7.741	1.256
	-64.65	-57.11	-7.54	-4.979	-5.325	0.346
	-64.02	-54.67	-9.34	-4.103	-6.108	2.005

2.9.5.2 *LL1_6.*

Table 15

The surfaces generated by binding site analysis at the LeuRS-LL1_6 binding site and arrows point to modification site of R_0, R_1, R_2, and R_3.

R_0	R_1	R_2	R_3
			

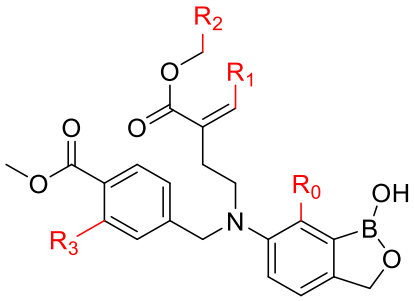


Figure 14. 2D structure of LL1_6 illustrating all the prospective modification sites.

Table 16

*Hits from single site enumeration at R_0 of LL1*_6.

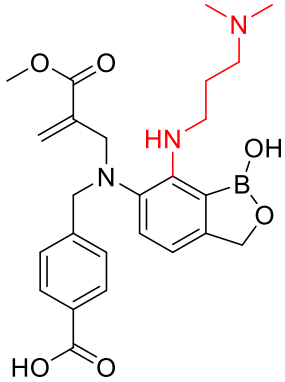
Structure	MM-GBSA ΔG_{Bind}
	-77.07

Table 17

Hits from single site enumeration at R_1 of LL1_6.

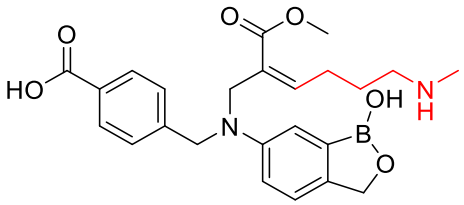
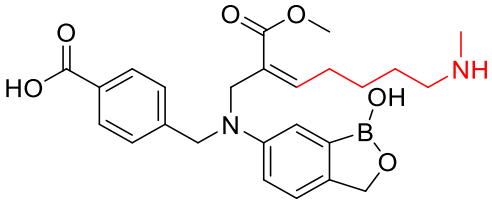
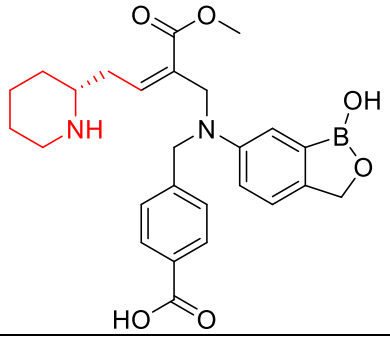
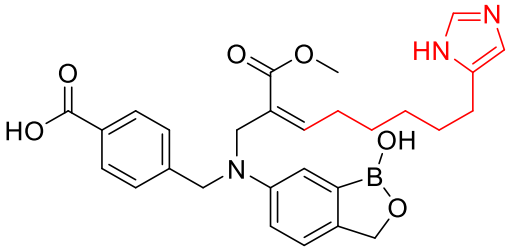
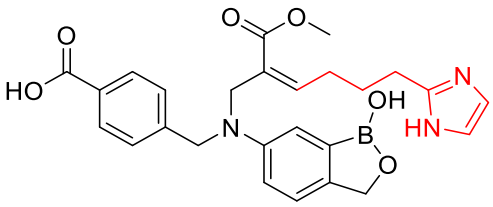
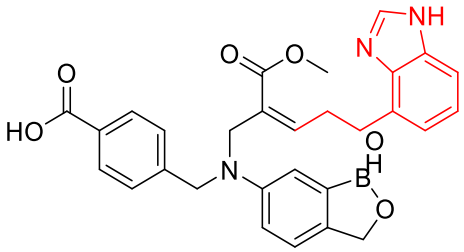
Structure	MM-GBSA ΔG_{Bind}
	-64.97
	-69.83
	-72.02
	-71.15
	-64.96
	-77.18

Table 17 (continued)

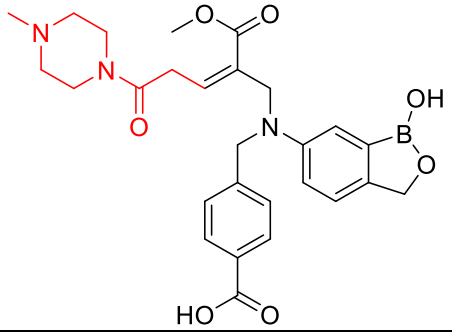
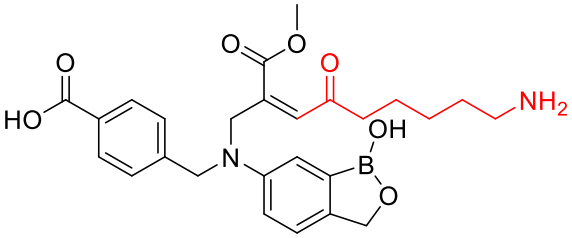
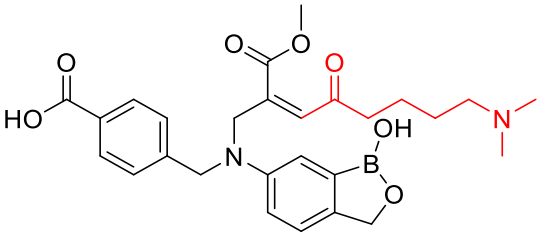
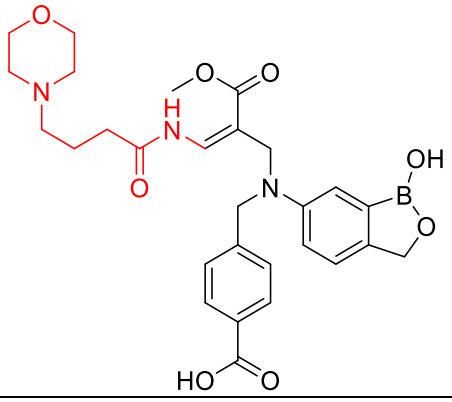
Structure	MM-GBSA ΔG_{Bind}
	-73.77
	-63.06
	-74.05
	-80.13

Table 18

Hits from single site enumeration at R_2 of LL1_6.

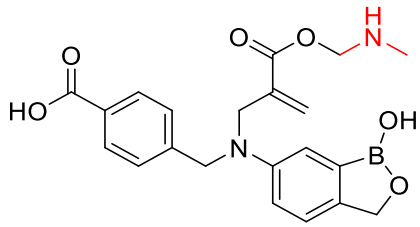
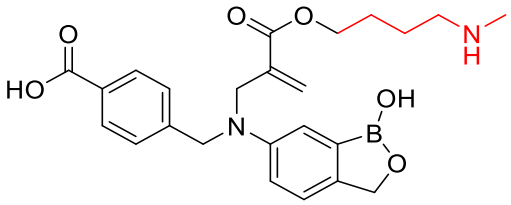
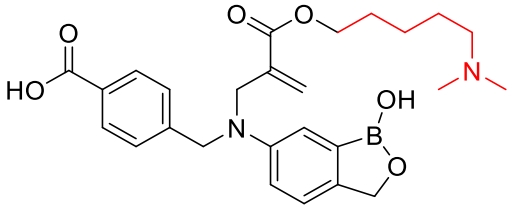
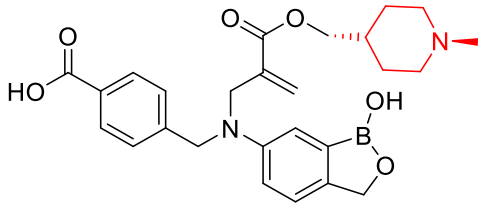
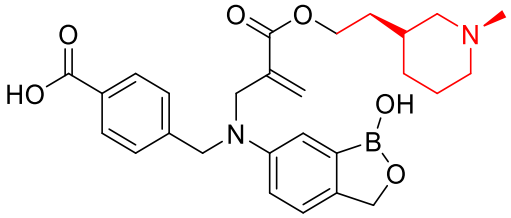
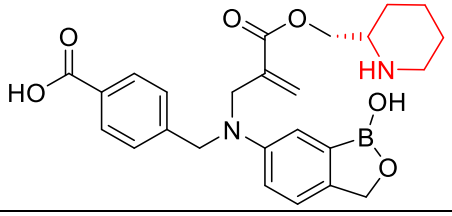
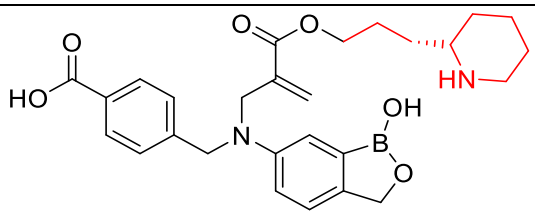
Structure	MM-GBSA ΔG_{Bind}
	-63.67
	-86.56
	-79.08
	-71.17
	-81.34
	-83.17
	-68.28

Table 18 (continued)

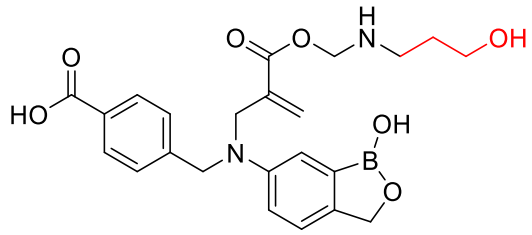
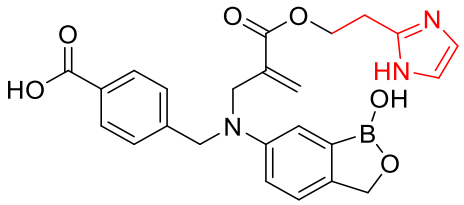
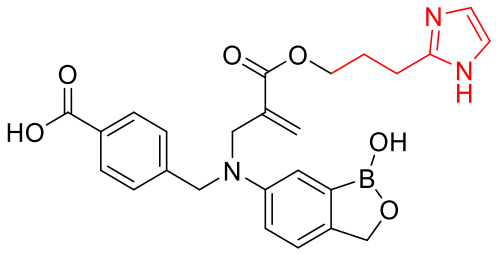
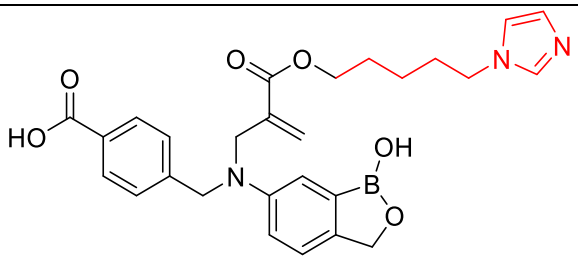
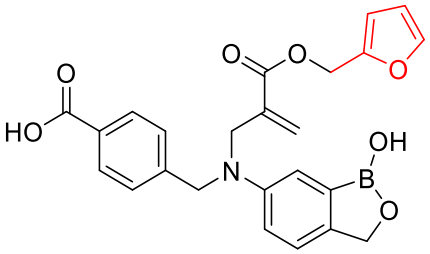
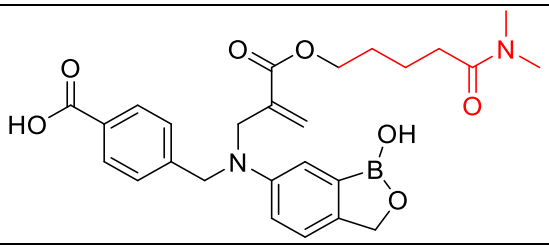
Structure	MM-GBSA ΔG_{Bind}
	-68.01
	-78.39
	-63.59
	-69.64
	-67.78
	-66.16

Table 18 (continued)

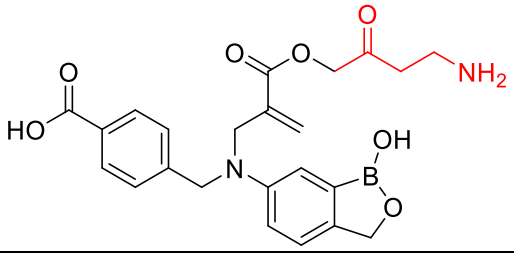
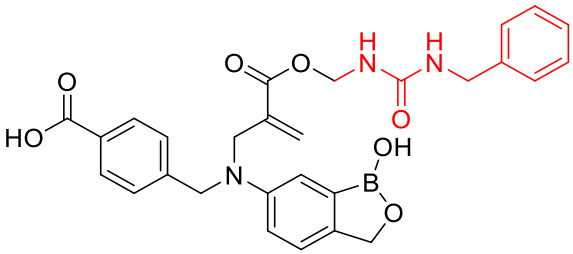
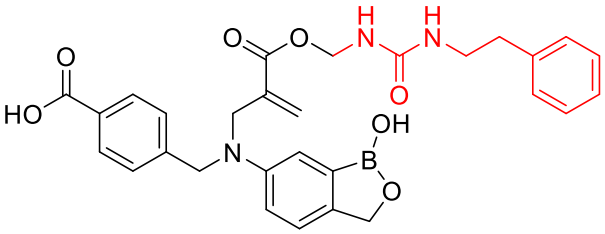
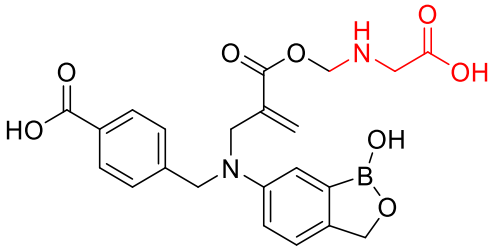
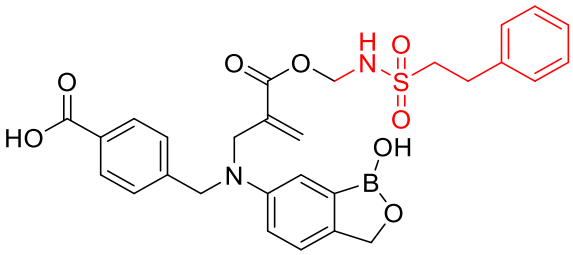
Structure	MM-GBSA ΔG_{Bind}
	-74.72
	-63.98
	-63.94
	-65.77
	-76.74

Table 19

Hits from enumeration at all modification sites of LL1_6 and comparison with human LeuRS.

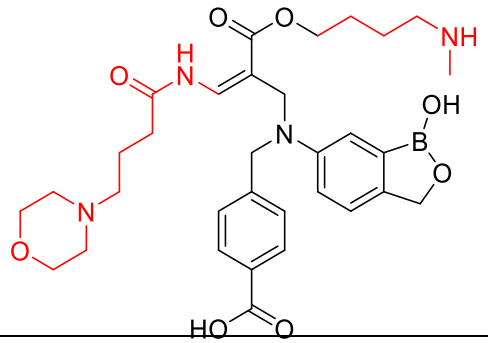
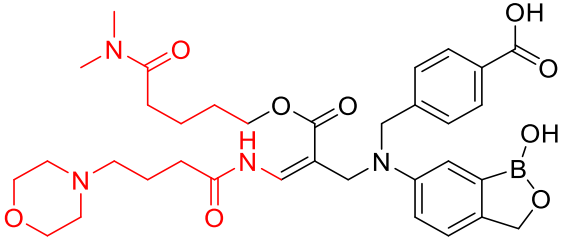
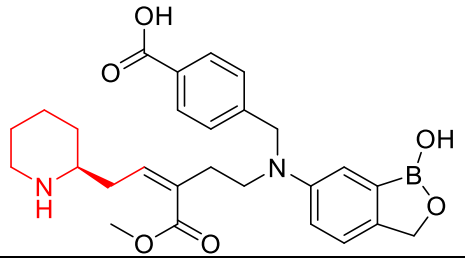
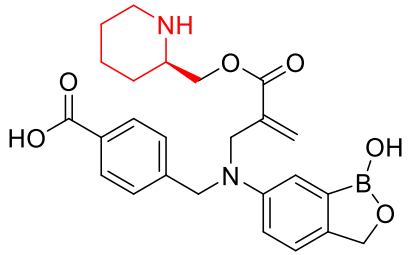
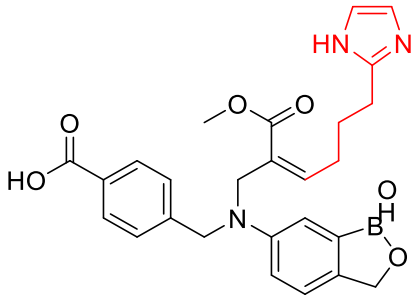
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-95.53	-56.41	-39.12	-9.908	-7.785	-2.123
	-76.58	-27.17	-49.4	-5.179	-3.221	-1.958
	-74.55	-65.76	-8.79	-6.009	-7.339	1.33

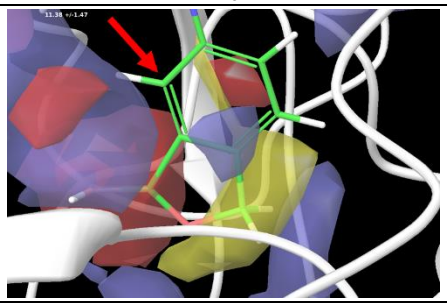
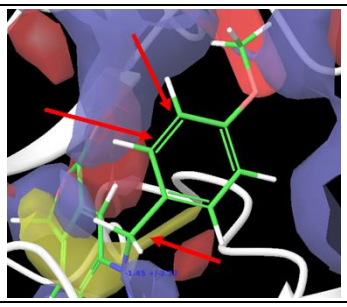
Table 19 (continued)

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-67.96	-10.86	-57.1	-4.336	-5.057	0.721
	-66.68	-42.45	-24.23	-4.538	-5.517	0.979

2.9.5.3 LL1_7.

Table 20

The surfaces generated by binding site analysis at the LeuRS-LL1_7 binding site and arrows point to modification site of R_0, R_1 and R_2.

R_0	R_1 and R_2
	

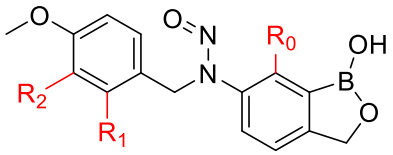


Figure 15. 2D structure of LL1_7 illustrating all the prospective modification sites.

Table 21

Hits from single site enumeration at R_0 of LL1_7.

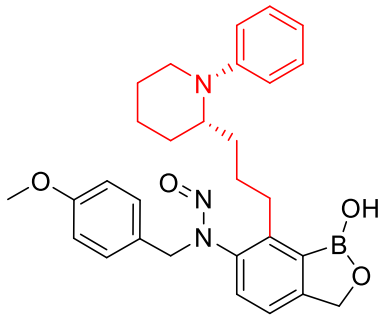
Structure	MM-GBSA ΔG_{Bind}
	-62.33

Table 21 (continued)

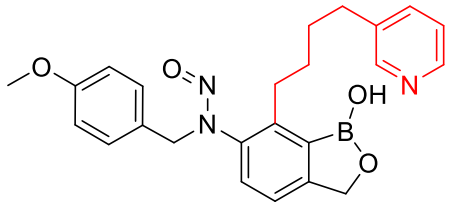
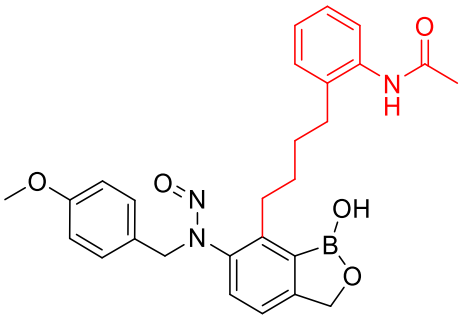
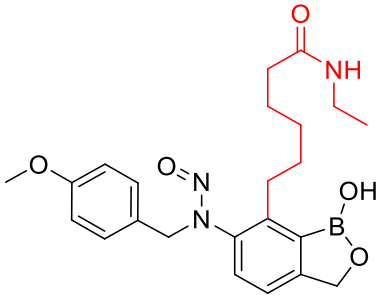
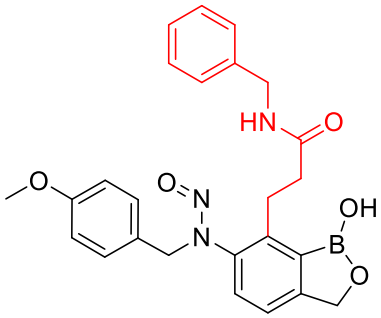
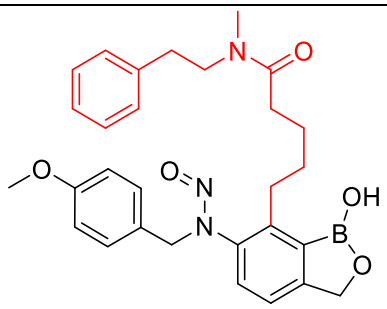
Structure	MM-GBSA ΔG_{Bind}
	-65.10
	-73.06
	-63.99
	-87.76
	-72.13

Table 21 (continued)

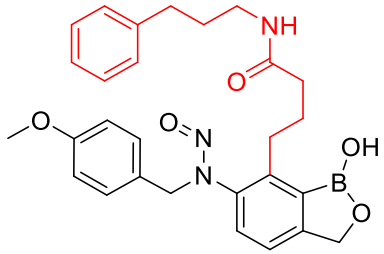
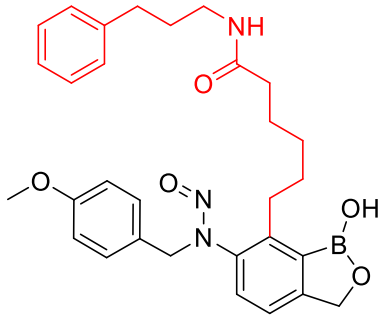
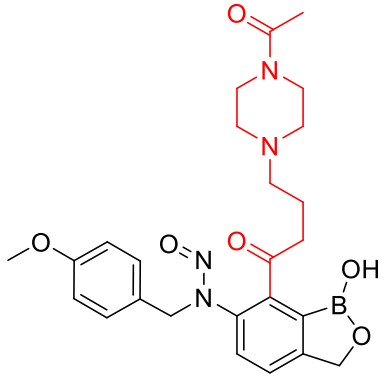
Structure	MM-GBSA ΔG_{Bind}
	-74.76
	-61.71
	-28.98

Table 22

Hits from single site enumeration at R_1 of LL1_7.

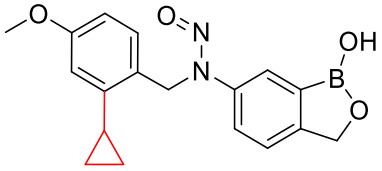
Structure	MM-GBSA ΔG_{Bind}
	-62.69

Table 22 (continued)

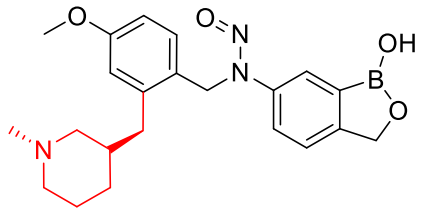
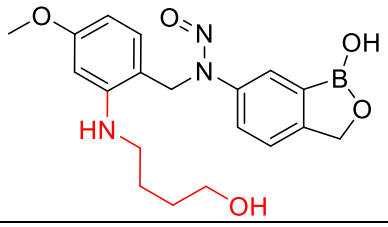
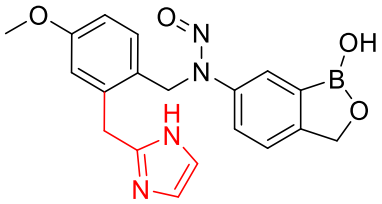
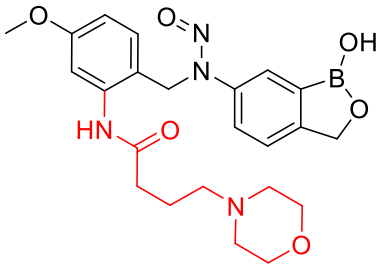
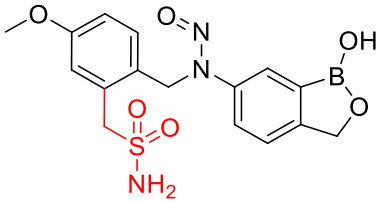
Structure	MM-GBSA ΔG_{Bind}
	-80.19
	-62.75
	-62.41
	-61.52
	-61.23

Table 23

Hits from single site enumeration at R_2 of LL1_7.

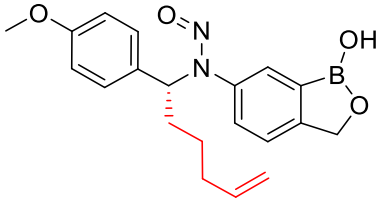
Structure	MM-GBSA ΔG_{Bind}
	-62.77

Table 23 (continued)

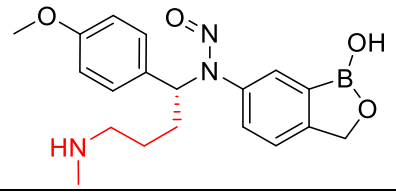
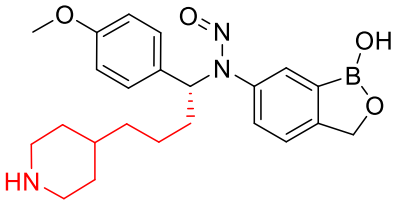
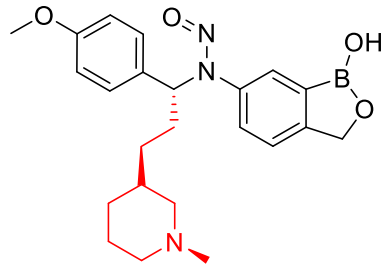
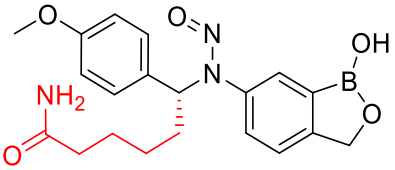
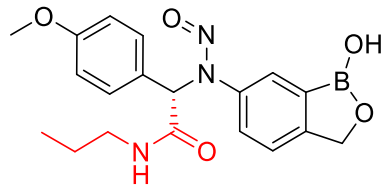
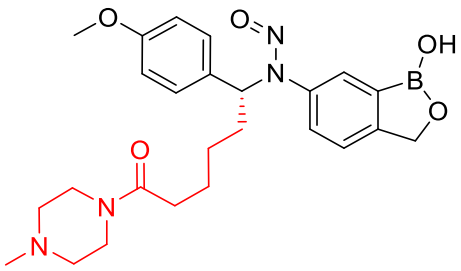
Structure	MM-GBSA ΔG_{Bind}
	-68.73
	-63.95
	-84.85
	-64.40
	-60.05
	-61.39

Table 24

Hits from enumeration at all modification sites of LL1_7.

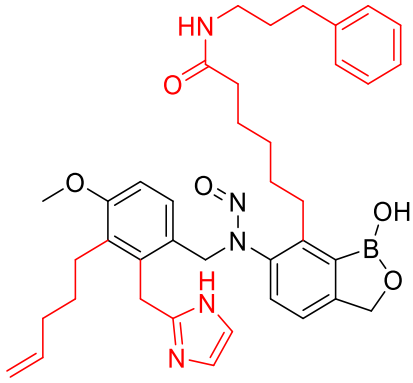
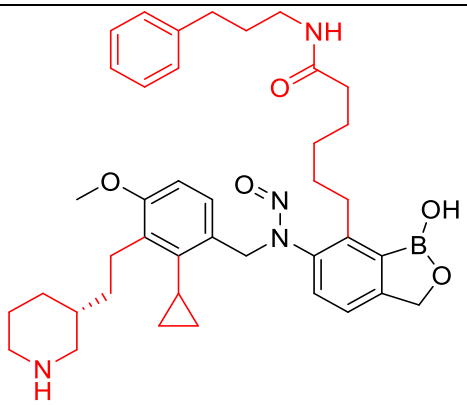
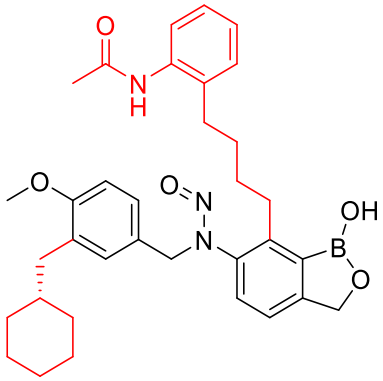
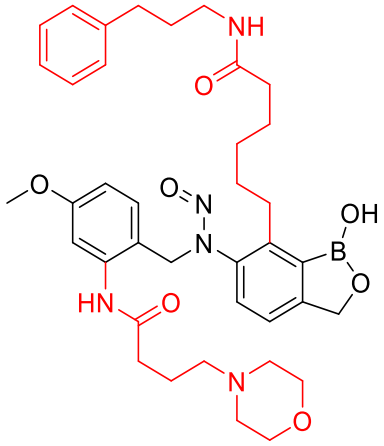
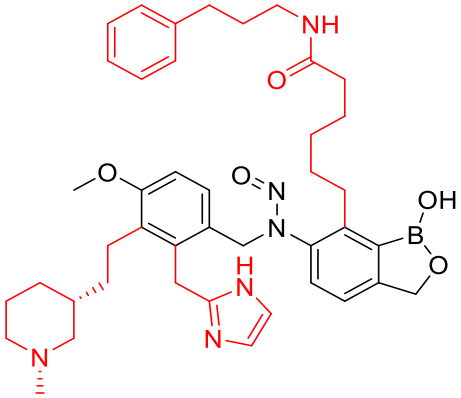
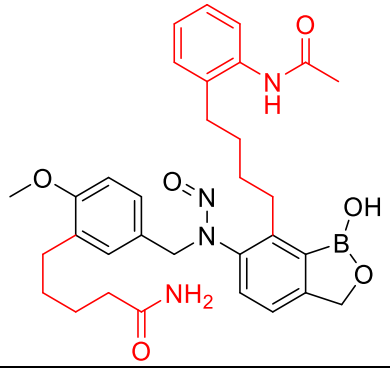
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-117.821	-4.312
	-116.719	-3.972
	-115.106	-6.325

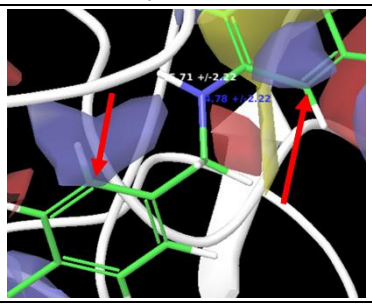
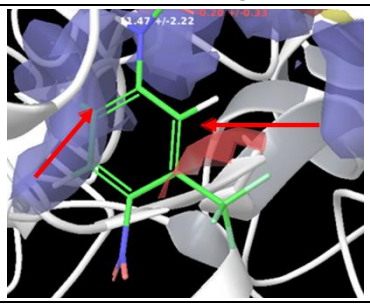
Table 24 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-113.406	-4.537
	-109.529	-6.998
	-101.104	-5.017

2.9.5.4 LL1_9.

Table 25

The surfaces generated by binding site analysis at the LeuRS-LL1_9 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3.

R_0 and R_1	R_2 and R_3
	

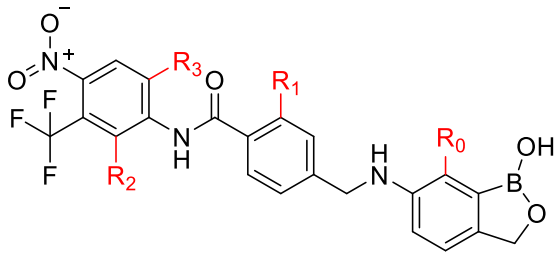


Figure 16. 2D structure of LL1_9 illustrating all the prospective modification sites.

Table 26

Hits from single site enumeration at R_0 of LL1_9.

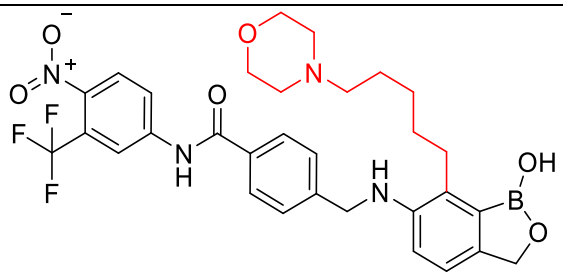
Structure	MM-GBSA ΔG_{Bind}
	-113.32

Table 26 (continued)

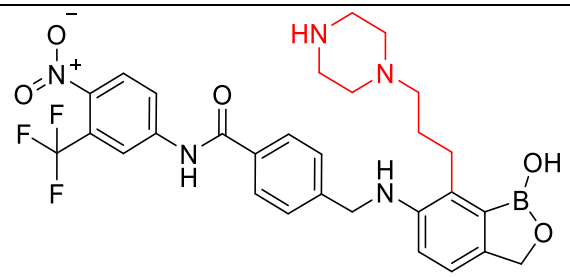
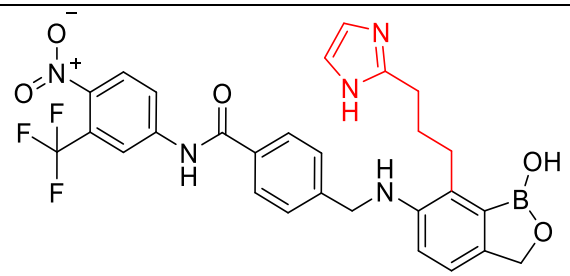
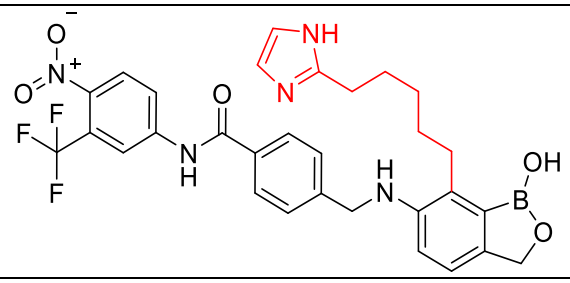
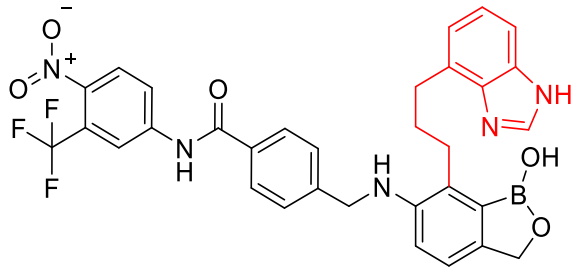
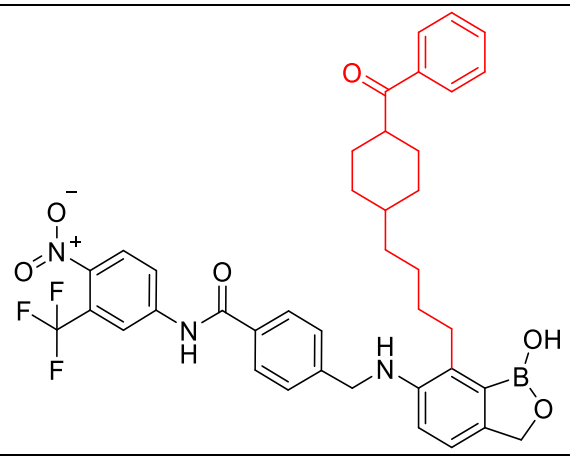
	-96.72
	-90.34
	-98.52
	-96.99
	-101.01

Table 27

Hits from single site enumeration at R_1 of LL1_9.

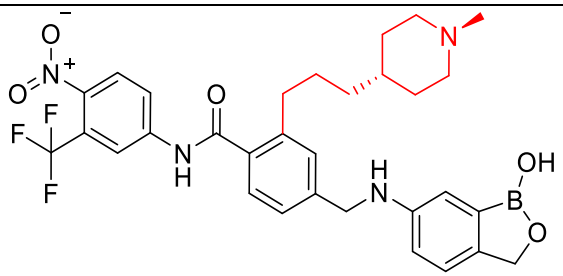
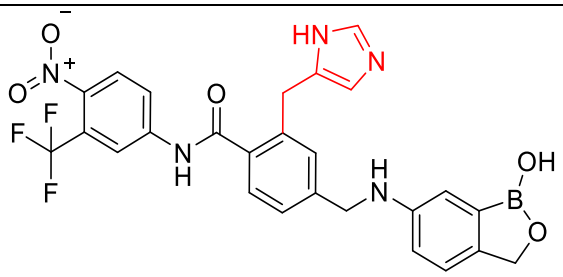
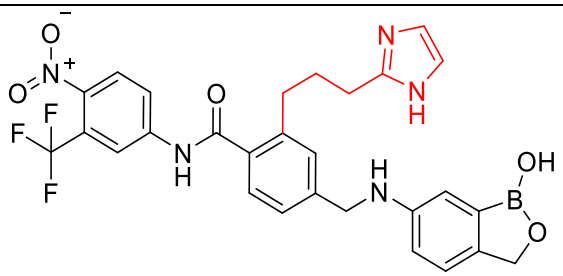
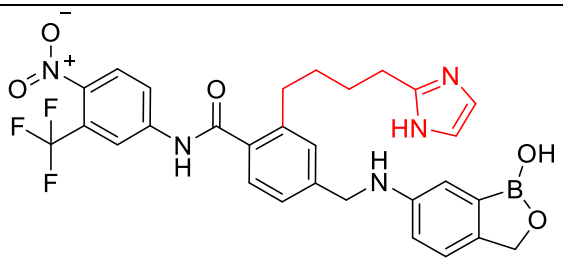
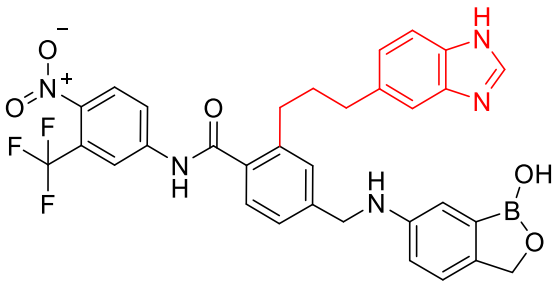
Structure	MM-GBSA ΔG_{Bind}
	-94.43
	-86.27
	-86.71
	-99.59
	-105.43

Table 27 (continued)

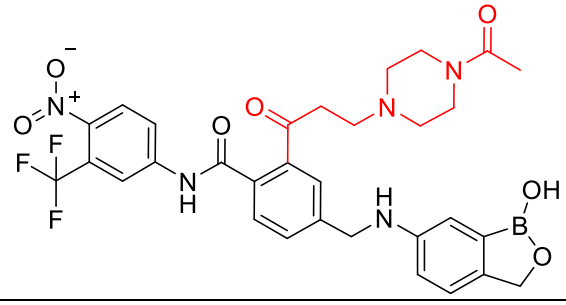
Structure	MM-GBSA ΔG_{Bind}
	-90.19

Table 28

Hits from single site enumeration at R_2 of LL1_9.

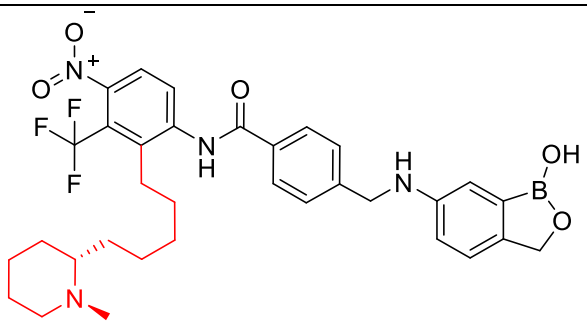
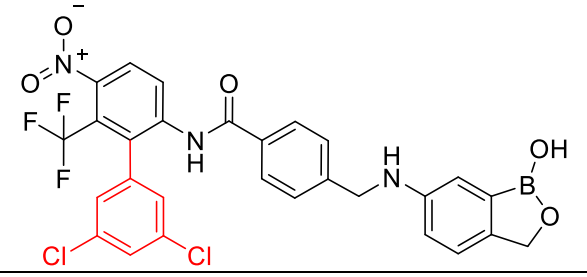
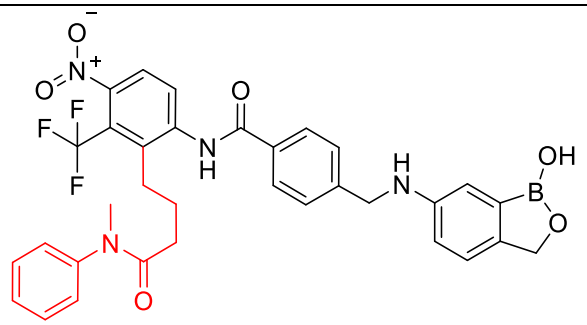
Structure	MM-GBSA ΔG_{Bind}
	-91.07
	-88.80
	-87.26

Table 28 (continued)

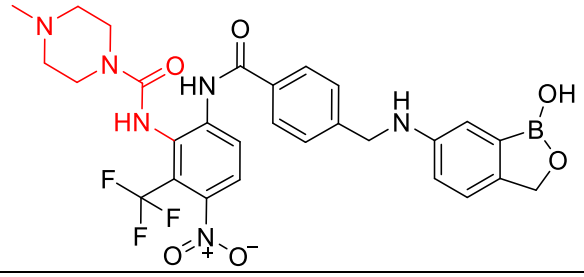
Structure	MM-GBSA ΔG_{Bind}
	-86.28

Table 29

Hits from single site enumeration at R_3 of LL1_9.

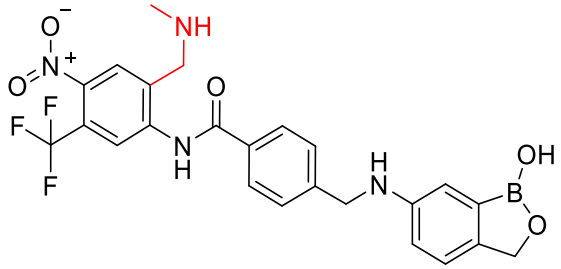
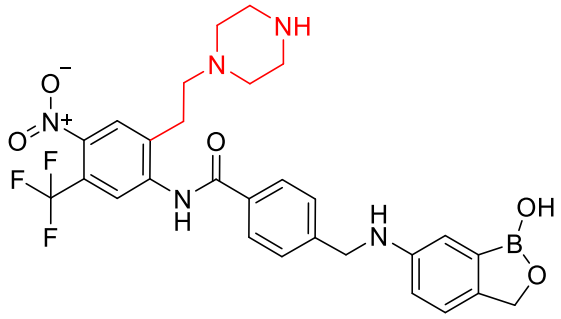
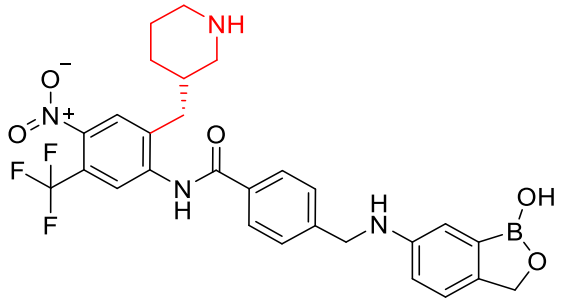
Structure	MM-GBSA ΔG_{Bind}
	-84.51
	-90.28
	-99.97

Table 29 (continued)

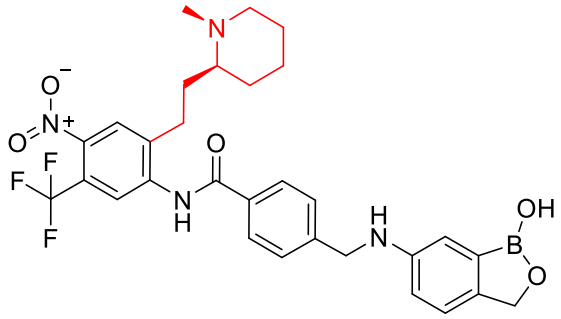
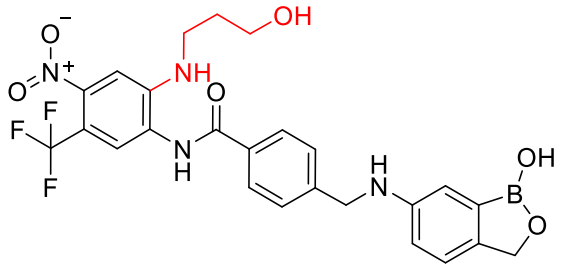
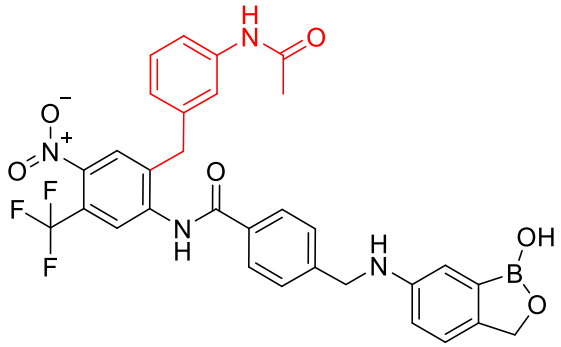
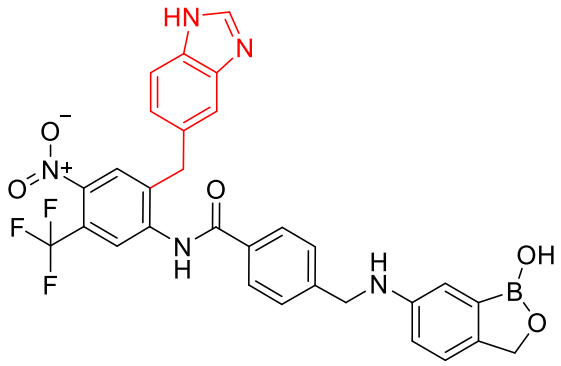
Structure	MM-GBSA ΔG_{Bind}
	-91.22
	-91.56
	-86.53
	-93.29

Table 29 (continued)

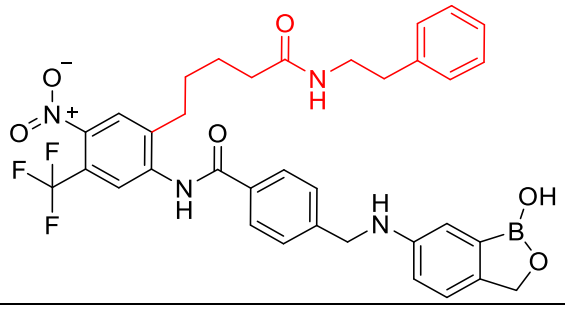
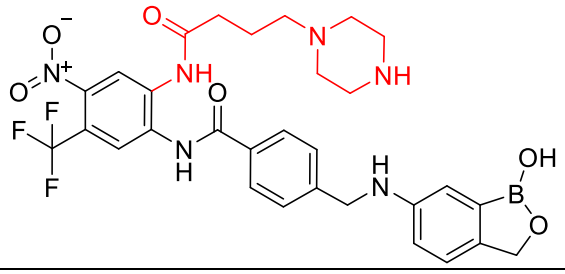
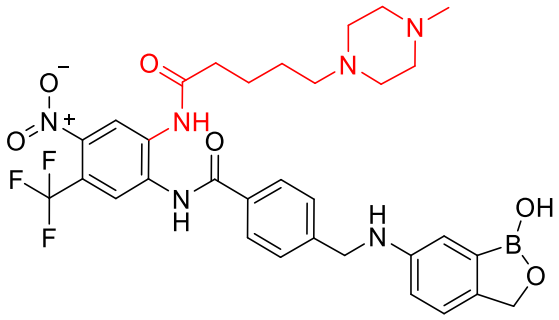
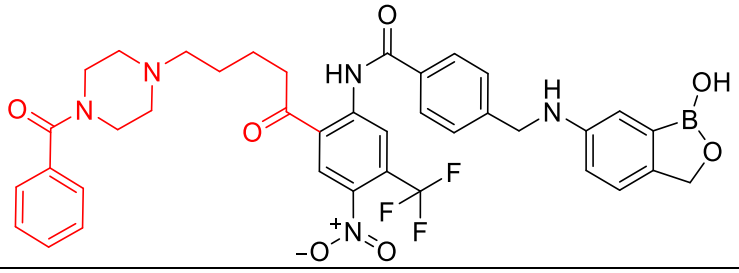
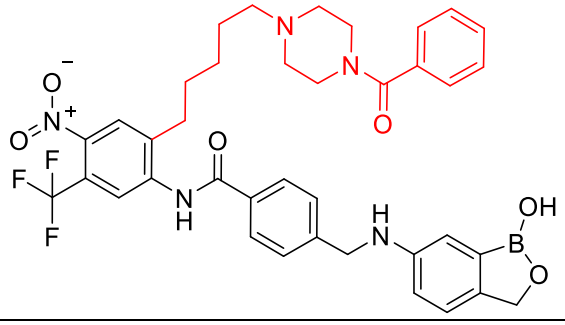
Structure	MM-GBSA ΔG_{Bind}
	-93.51
	-98.35
	-93.34
	-113.36
	-106.35

Table 30

Hits from enumeration at all modification sites of LL1_9 and comparison with human LeuRS.

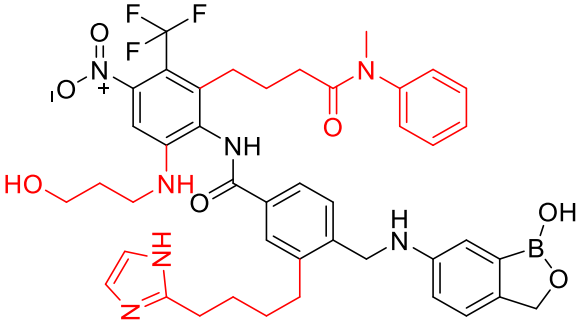
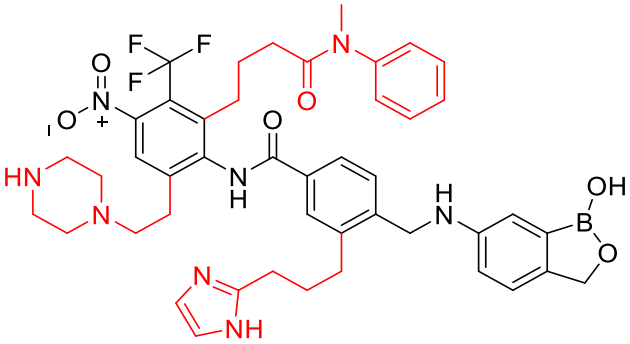
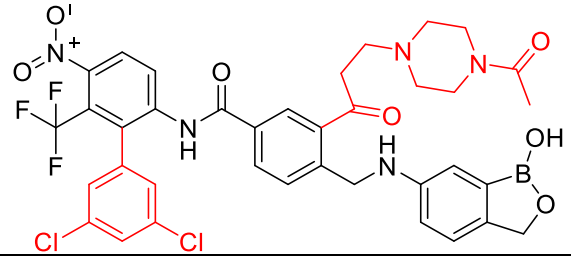
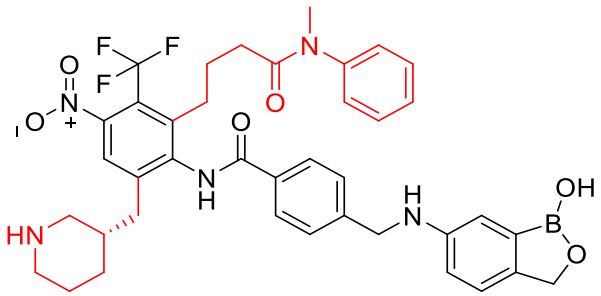
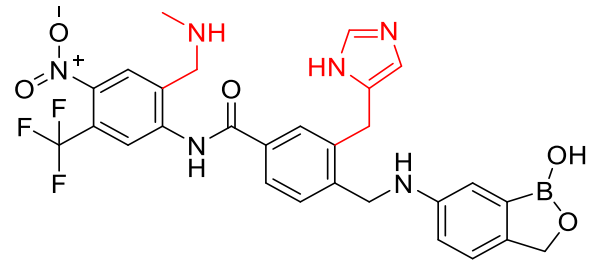
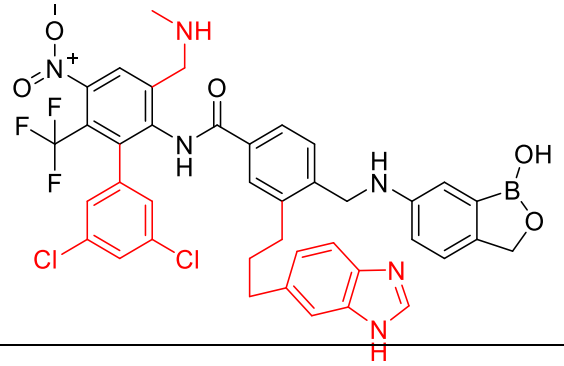
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-100.05	-75.93	-24.12	-7.401	-9.476	2.075
	-94.1	-66.53	-27.57	-7.567	-9.586	2.019
	-85.86	-50.92	-34.94	-5.228	-4.823	-0.405

Table 30 (continued)

Structure	MM-GBSA ΔG_{Bind}			Docking Score		
	Fungal	Human	Difference	Fungal	Human	Difference
	-84.46	-71.54	-12.91	-5.789	-7.402	1.613
	-84	-69.2	-14.8	-9.217	-7.919	-1.298
	-78.39	-60.36	-18.03	-8.122	-8.472	0.35

2.9.5.5 LL1_10.

Table 31

The surfaces generated by binding site analysis at the LeuRS-LL1_10 binding site and arrows point to modification site of R_1 and 2D structure of LL1_10 illustrating its prospective modification sites.

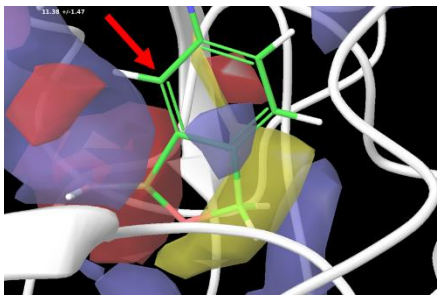
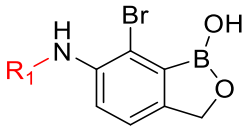
R_1	
	

Table 32

Hits from single site enumeration at R_1 of LL1_10 and comparison with human LeuRS.

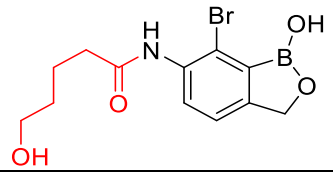
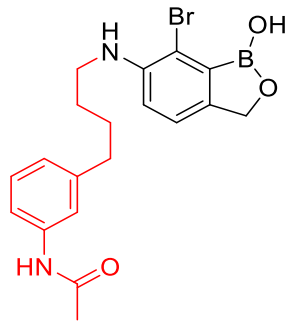
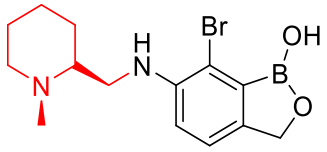
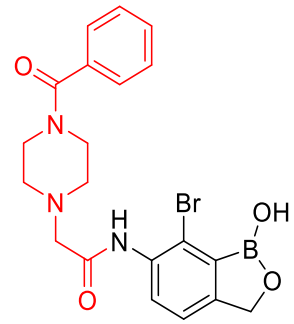
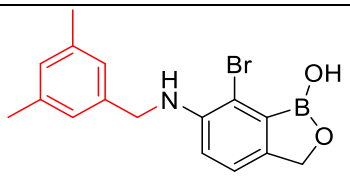
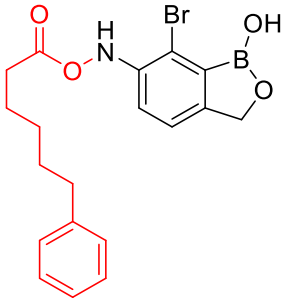
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Diff
	-84.665	-38.779	-45.886	-5.589	-4.29	-1.299
	-83.961	-34.274	-49.687	-5.264	-4.651	-0.613
	-77.476	-30.159	-47.317	-5.114	-5.536	0.422

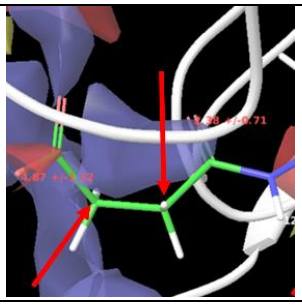
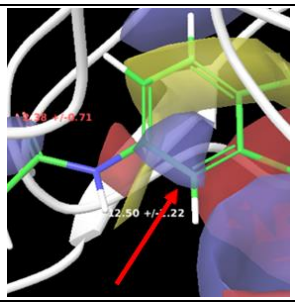
Table 32 (continued)

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Diff	Fungal	Human	Diff
	-77.14	-24.27	-52.87	-4.594	-5.939	1.345
	-68.825	-25.042	-43.783	-4.469	-4.427	-0.042
	-67.991	-25.829	-42.162	-4.226	-3.67	-0.556

2.9.5.6 LL1_12.

Table 33

The surfaces generated by binding site analysis at the LeuRS-LL1_12 binding site and arrows point to modification site of R_0, R_1 and R_2.

R_0 and R_1	R_2
	

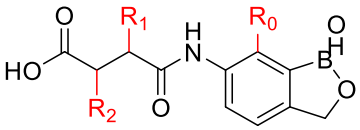


Figure 17. 2D structure of LL1_12 illustrating all the prospective modification sites.

Table 34

Hits from single site enumeration at R_0 of LL1_12.

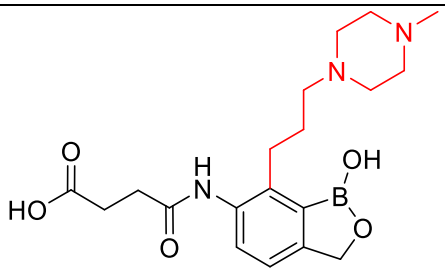
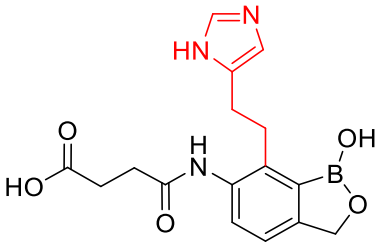
Structure	MM-GBSA ΔG_{Bind}
	-64.78
	-57.30

Table 35

Hits from single site enumeration at R_1 of LL1_12.

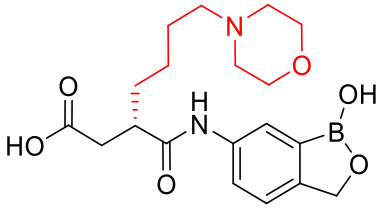
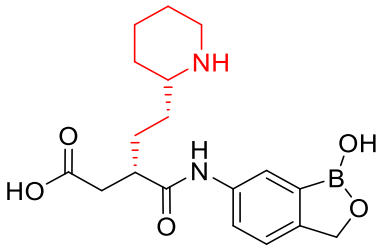
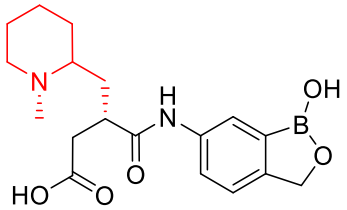
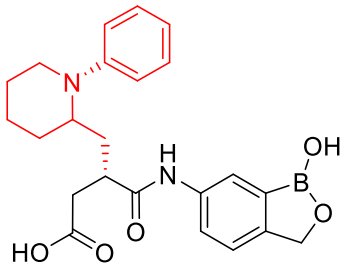
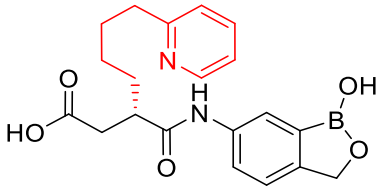
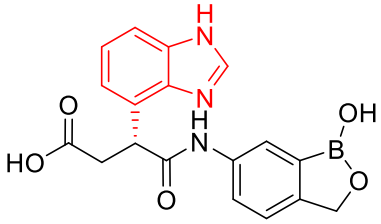
Structure	MM-GBSA ΔG_{Bind}
	-79.07
	-76.12
	-82.66
	-87.53
	-76.32
	-76.82

Table 35 (continued)

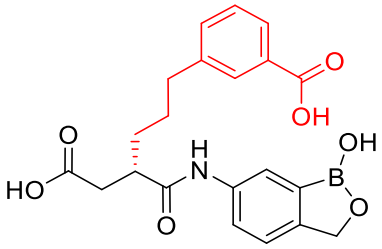
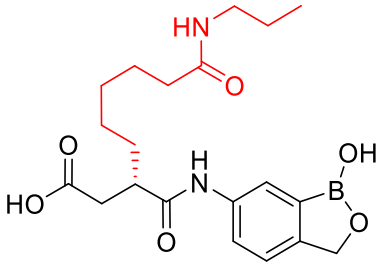
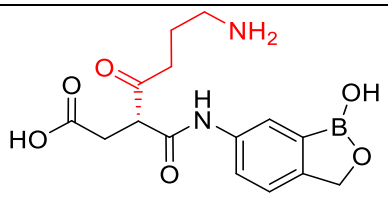
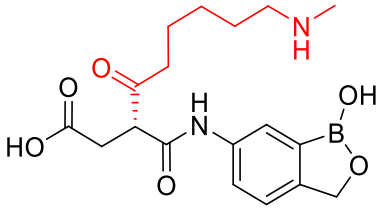
Structure	MM-GBSA ΔG_{Bind}
	-80.02
	-78.16
	-76.98
	-72.87

Table 36

Hits from single site enumeration at R_2 of LL1_12.

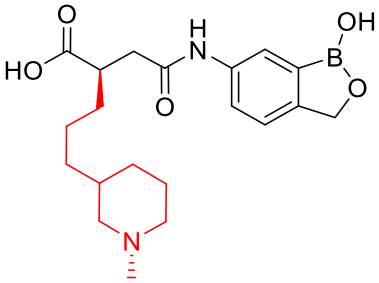
Structure	MM-GBSA ΔG_{Bind}
	-83.15

Table 36 (continued)

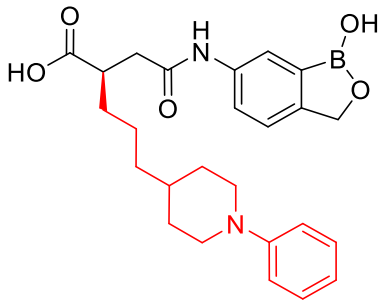
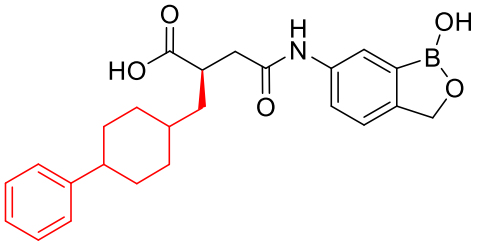
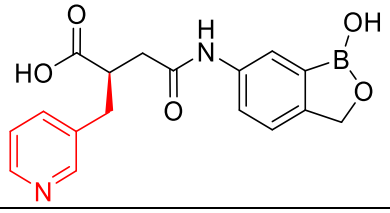
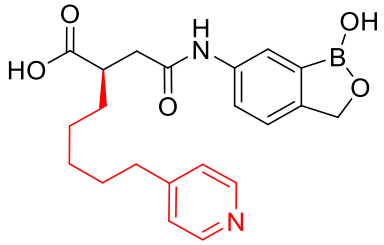
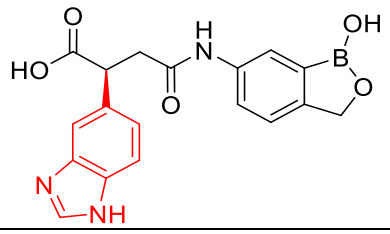
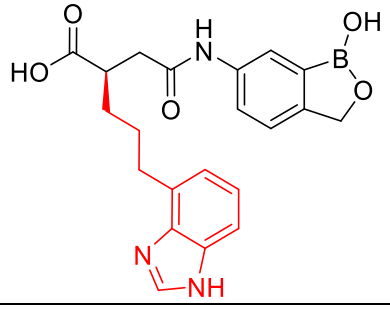
Structure	MM-GBSA ΔG_{Bind}
	-84.46
	-77.90
	-84.41
	-94.13
	-91.39
	-83.35

Table 36 (continued)

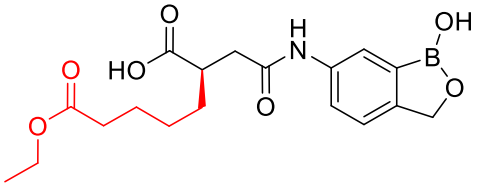
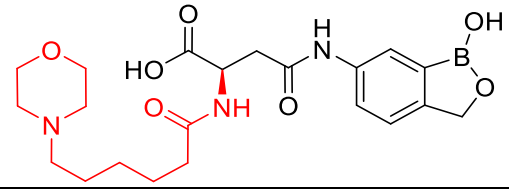
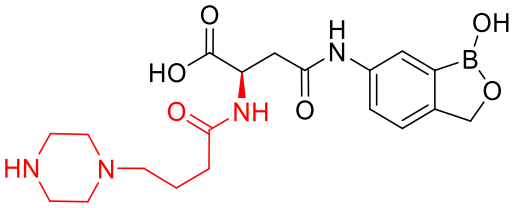
Structure	MM-GBSA ΔG_{Bind}
	-77.10
	-75.77
	-76.07

Table 37

Hits from enumeration at all modification sites of LL1_12.

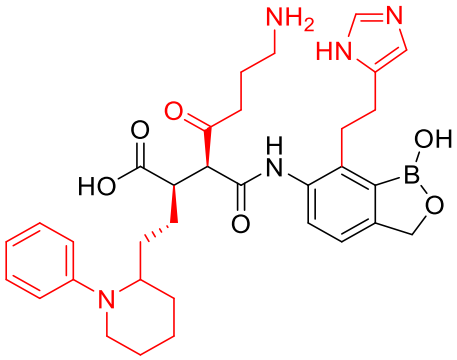
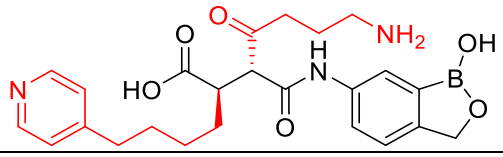
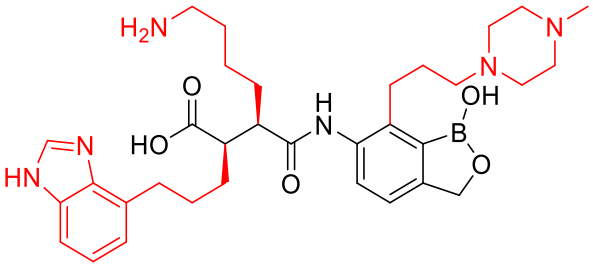
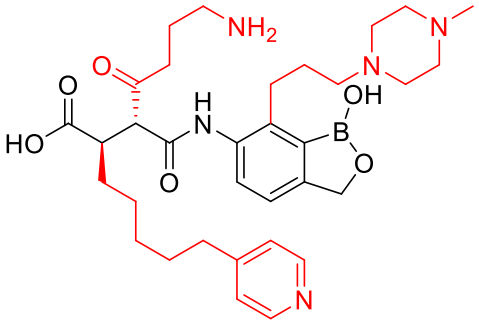
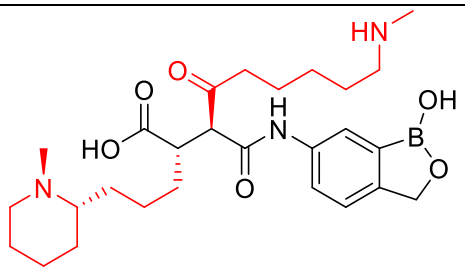
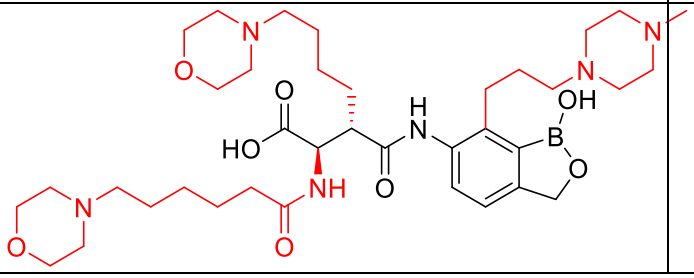
Structure	MM-GBSA ΔG_{Bind}	Docking Score
	-66.025	-8.314
	-65.823	-6.706

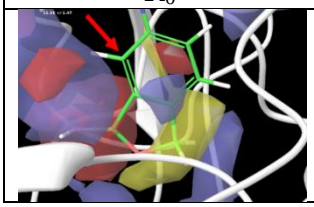
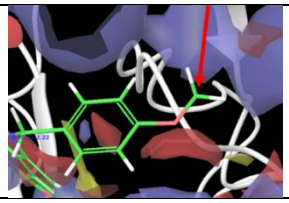
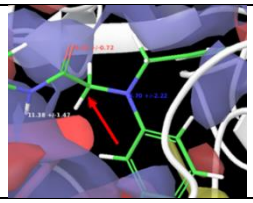
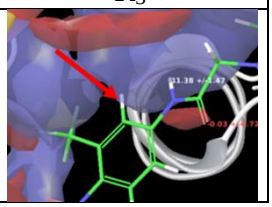
Table 37 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking Score
	-65.038	-7.699
	-64.912	-8.272
	-63.531	-7.143
	-62.649	-4.704

2.9.5.7 LL1_13.

Table 38

The surfaces generated by binding site analysis at the LeuRS-LL1_13 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3.

R_0	R_1	R_2	R_3
			

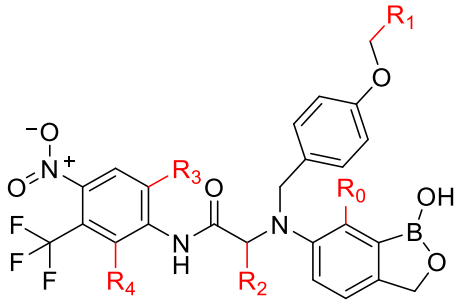


Figure 18. 2D structure of LL1_13 illustrating all the prospective modification sites.

Table 39

Hits from single site enumeration at R_1 of LL1_13.

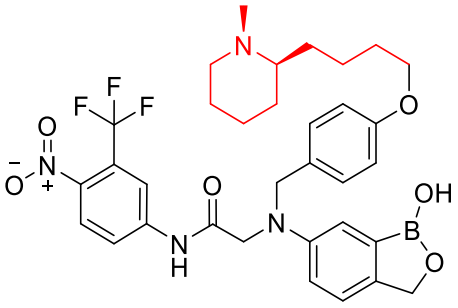
Structure	MM-GBSA ΔG_{Bind}
	-106.41

Table 39 (continued)

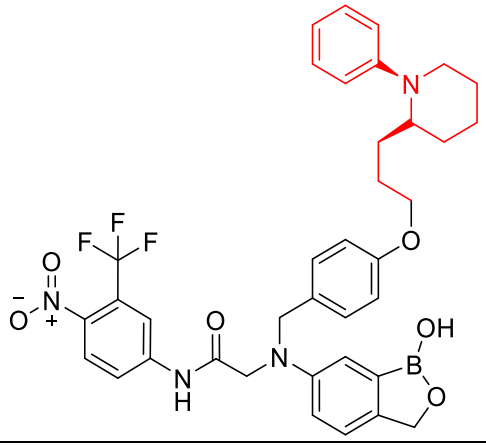
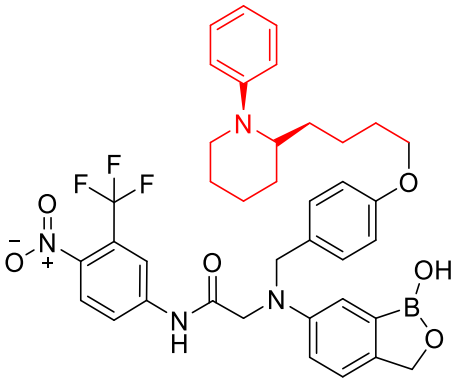
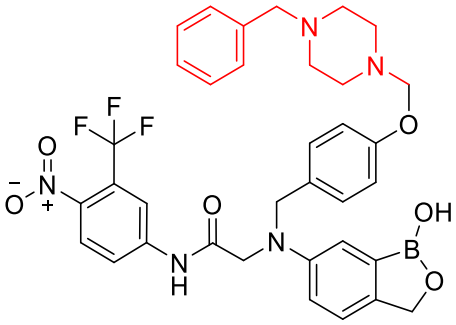
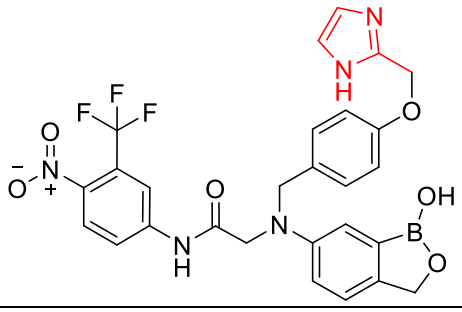
Structure	MM-GBSA ΔG_{Bind}
	-85.77
	-81.84
	-88.51
	-85.98

Table 39 (continued)

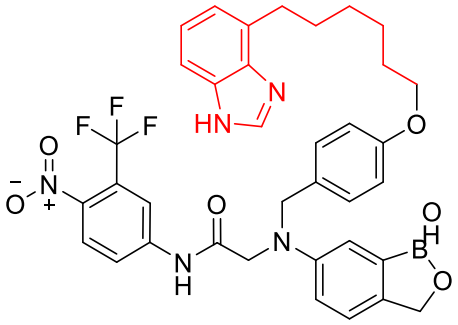
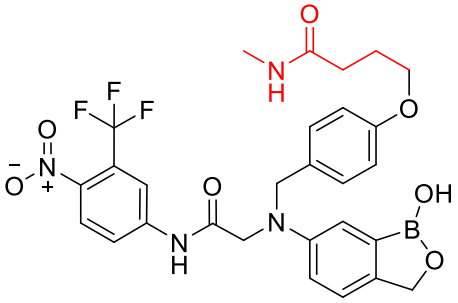
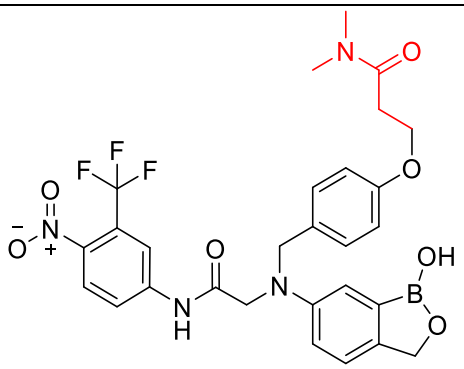
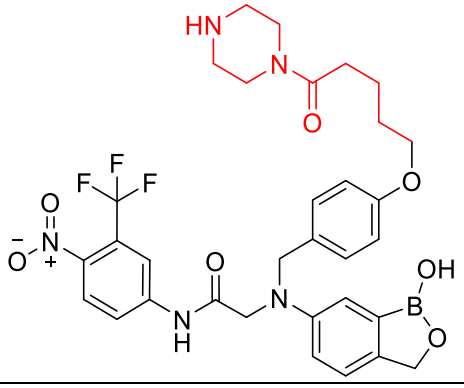
Structure	MM-GBSA ΔG_{Bind}
	-83.05
	-88.55
	-93.76
	-86.65

Table 39 (continued)

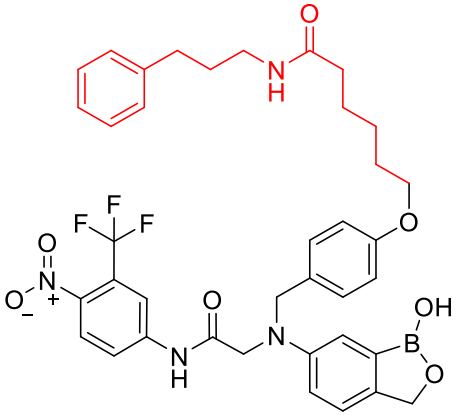
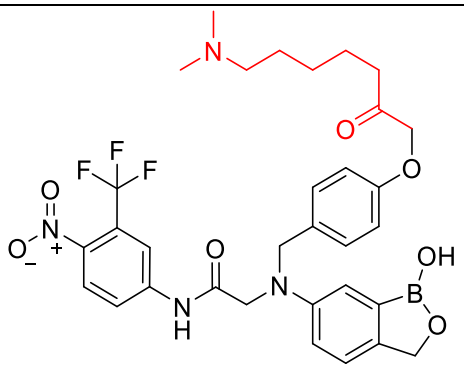
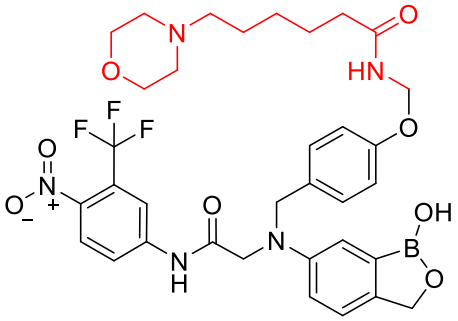
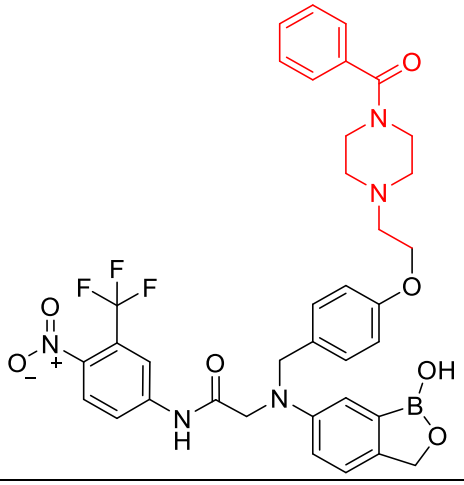
Structure	MM-GBSA ΔG_{Bind}
	-87.43
	-83.09
	-101.24
	-80.49

Table 40

Hits from single site enumeration at R_2 of LL1_13.

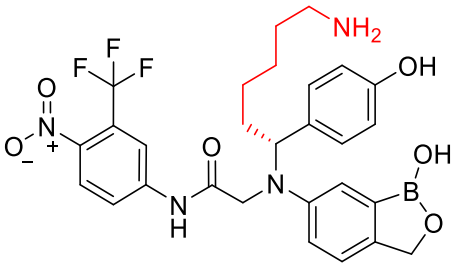
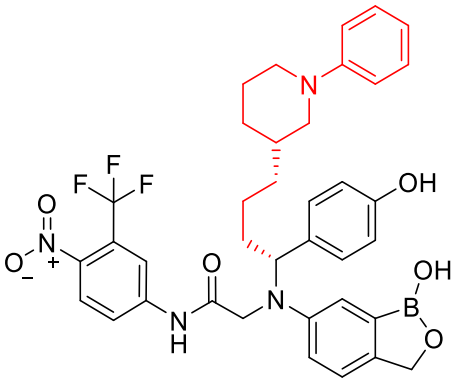
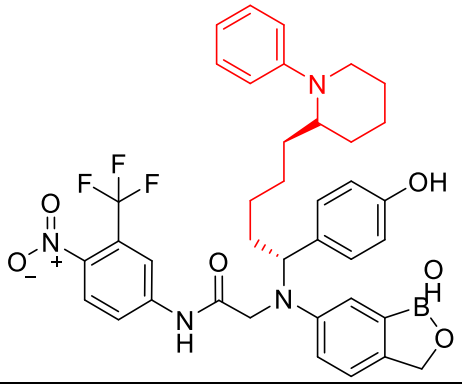
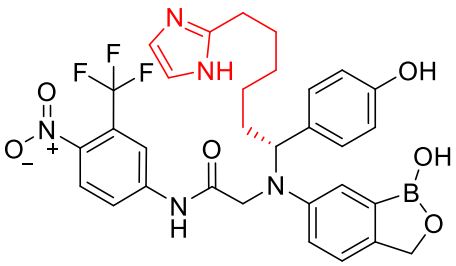
Structure	MM-GBSA ΔG_{Bind}
	-89.58
	-76.07
	-91.40
	-88.51

Table 40 (continued)

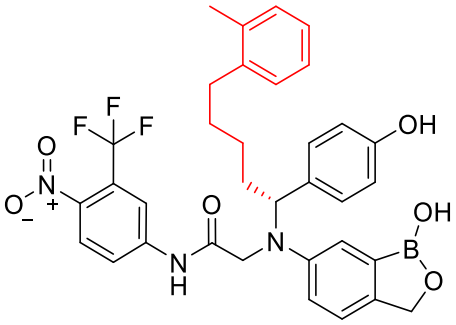
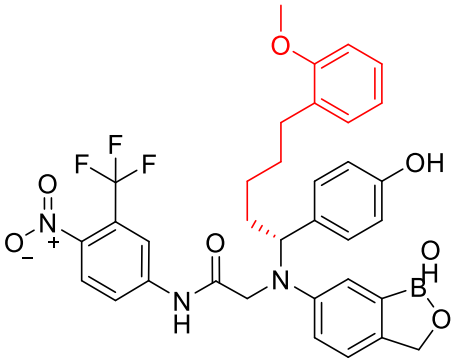
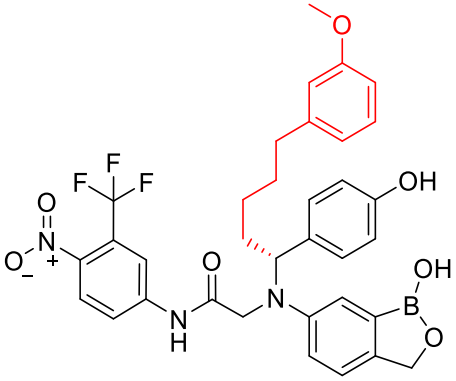
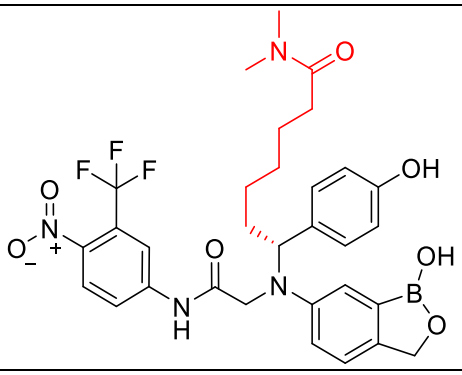
Structure	MM-GBSA ΔG_{Bind}
	-75.18
	-76.62
	-84.52
	-88.24

Table 40 (continued)

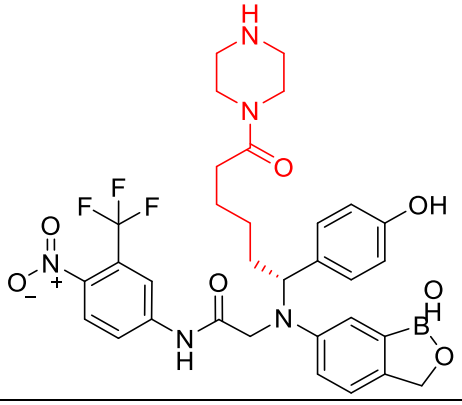
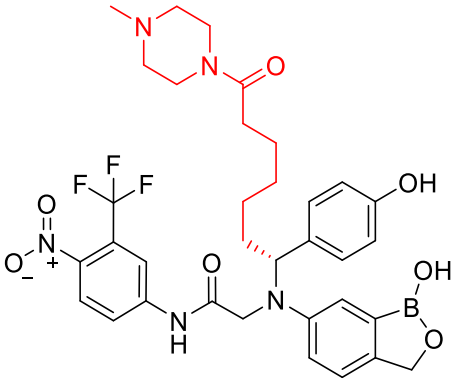
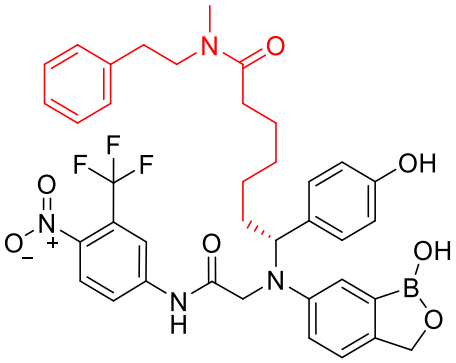
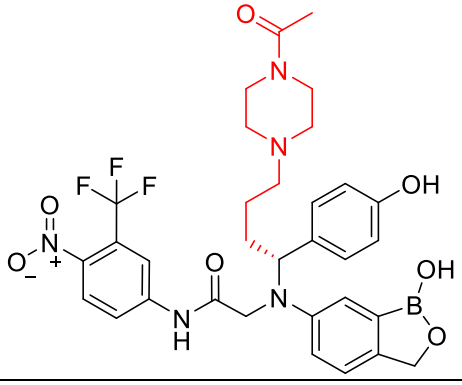
Structure	MM-GBSA ΔG_{Bind}
	-79.80
	-76.37
	-92.66
	-83.70

Table 41

Hits from enumeration at all modification sites of LL1_13 and comparison with human LeuRS.

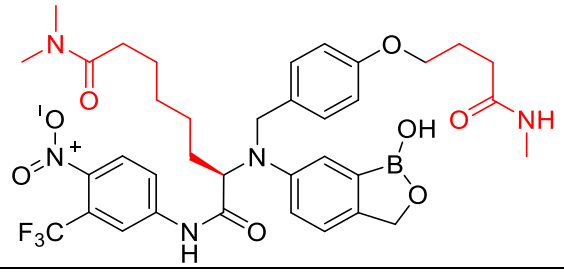
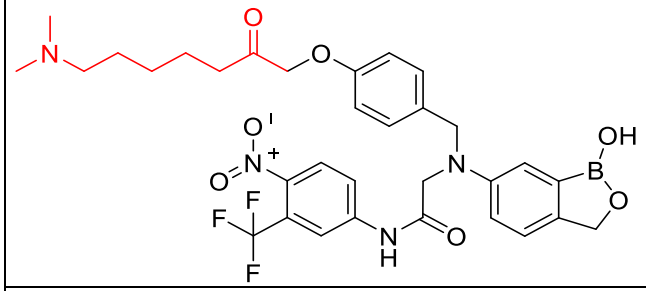
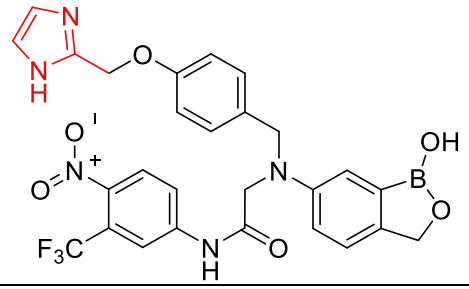
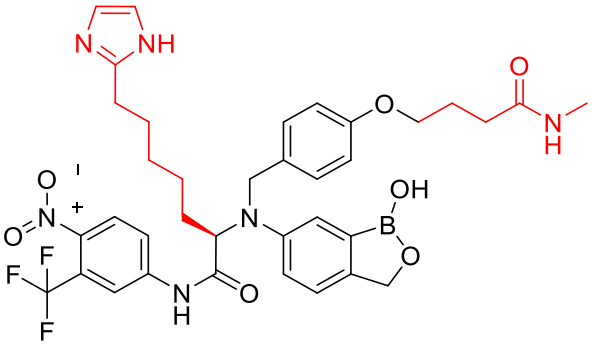
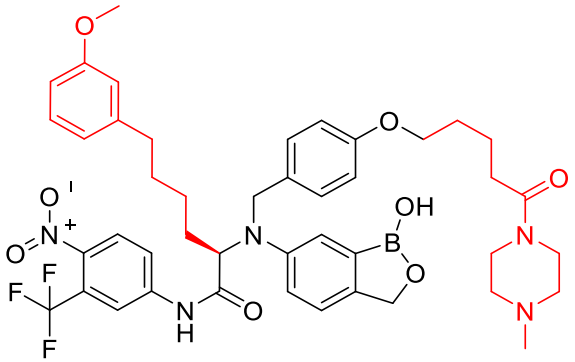
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-82	-80.35	-1.65	-5.041	-7.28	2.239
	-81.07	-69.61	-11.46	-7.962	-9.069	1.107
	-79.52	-63.05	-16.48	-4.693	-6.972	2.279

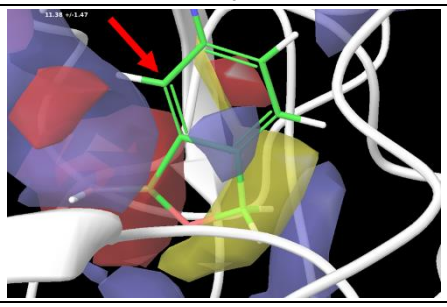
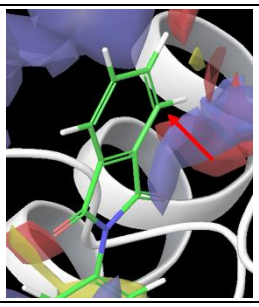
Table 41 (continued)

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-79.43	-70.14	-9.29	-4.889	-6.416	1.527
	-76.18	-64.74	-11.44	-4.751	-5.794	1.043

2.9.5.8 LL1_14.

Table 42

The surfaces generated by binding site analysis at the LeuRS-LL1_14 binding site and arrows point to modification site of R_0 and R_1.

R_0	R_1
	

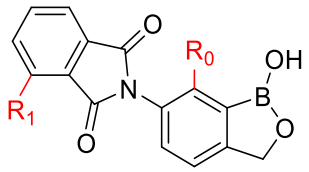


Figure 19. 2D structure of LL1_14 illustrating all the prospective modification sites.

Table 43

Hits from single site enumeration at R_1 of LL1_14.

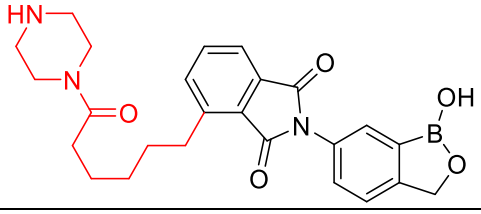
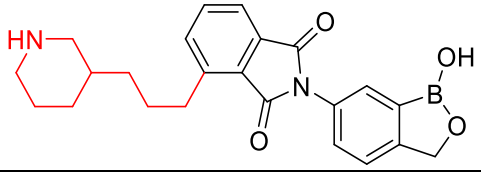
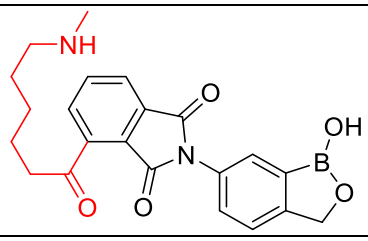
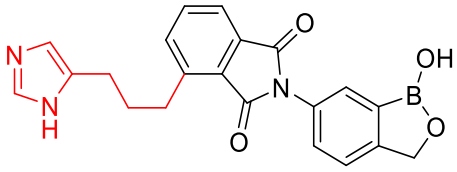
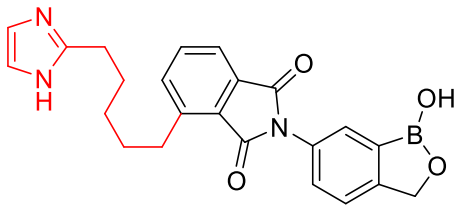
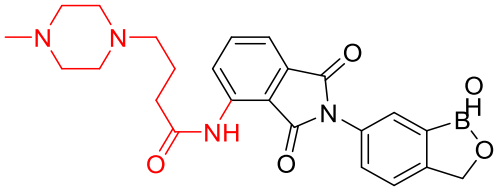
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-99.76	-7.363
	-92.273	-7.616

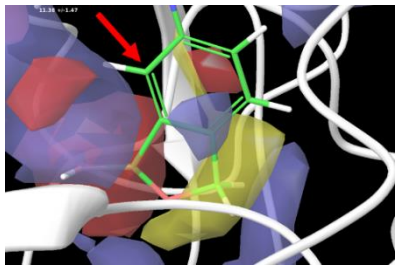
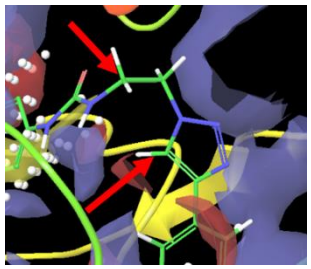
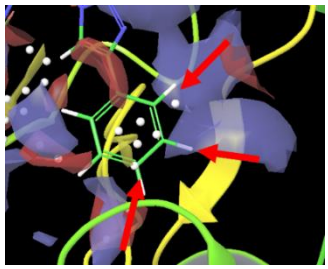
Table 43 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-90.439	-8.352
	-89.443	-7.57
	-89.001	-7.891
	-88.878	-5.312

2.9.5.9 LL2_ 1.

Table 44

The surfaces generated by binding site analysis at the LeuRS-LL2_1 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4 and R_5.

R_0	R_1, R_2	R_3, R_4 and R_5
		

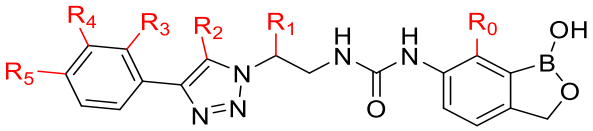


Figure 20. 2D structure of LL2_1 illustrating all the prospective modification sites.

Table 45

Hits from single site enumeration at R_0 of LL2_1.

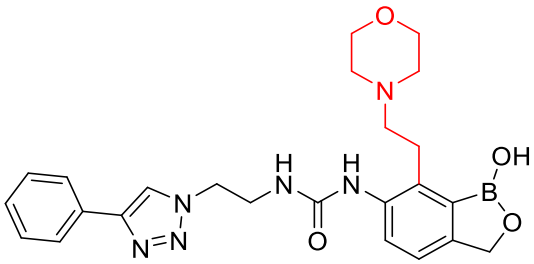
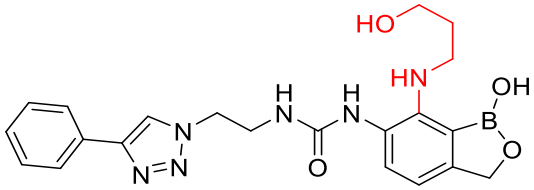
Structure	MM-GBSA ΔG_{Bind}
	-104.90
	-103.07

Table 45 (continued)

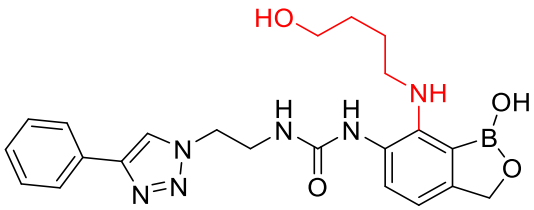
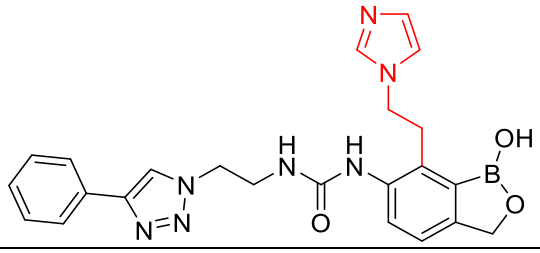
Structure	MM-GBSA ΔG_{Bind}
	-104.48
	-100.39

Table 46

Hits from single site enumeration at R_1 of LL2_1.

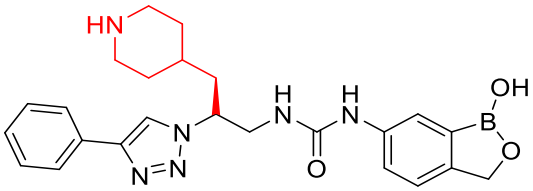
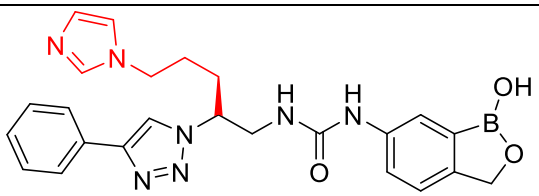
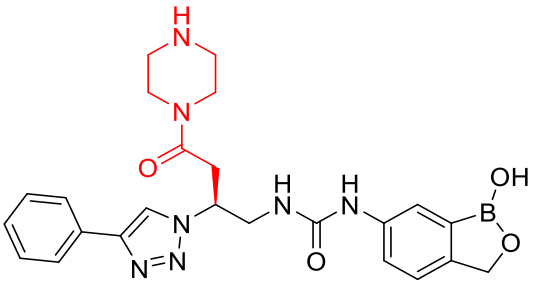
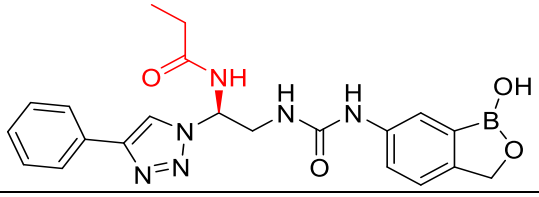
Structure	MM-GBSA ΔG_{Bind}
	-95.04
	-99.37
	-84.04
	-86.52

Table 47

Hits from single site enumeration at R_2 of LL2_1.

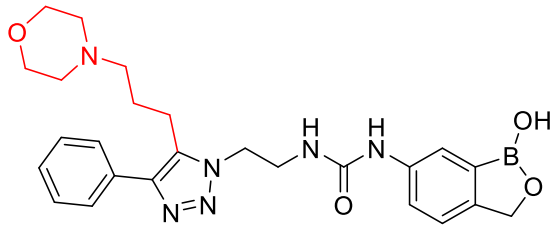
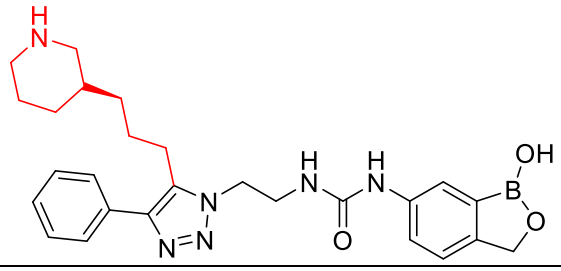
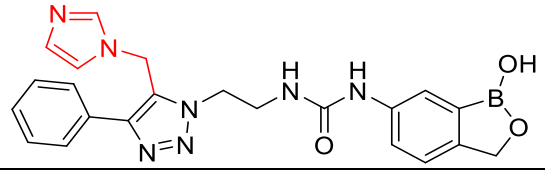
Structure	MM-GBSA ΔG_{Bind}
	-92.71
	-88.10
	-96.72

Table 48

Hits from single site enumeration at R_3 of LL2_1.

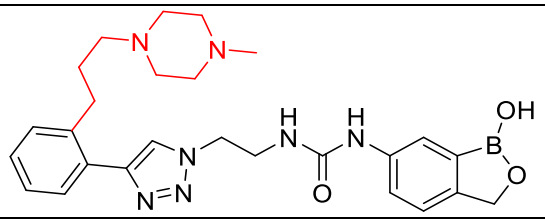
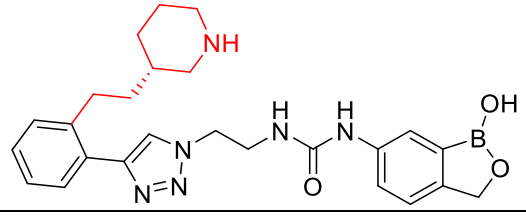
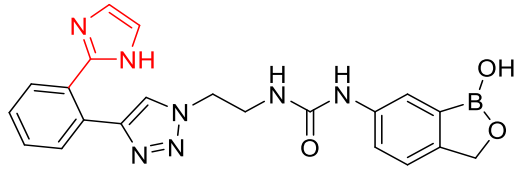
Structure	MM-GBSA ΔG_{Bind}
	-103.72
	-97.12
	-107.19

Table 49

Hits from single site enumeration at R_4 of LL2_1.

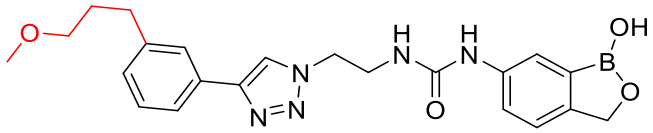
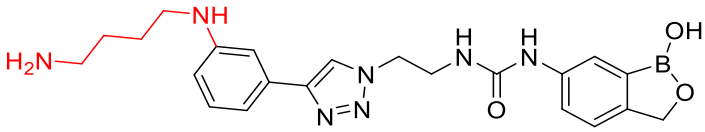
Structure	MM-GBSA ΔG_{Bind}
	-98.14
	-98.92

Table 50

Hits from single site enumeration at R_5 of LL2_1.

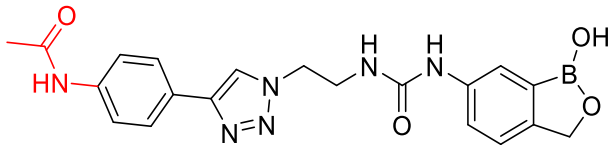
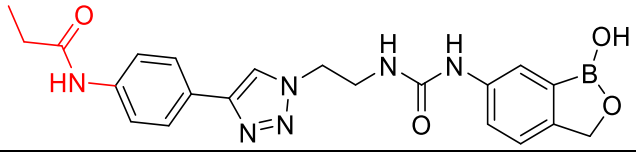
Structure	MM-GBSA ΔG_{Bind}
	-84.05
	-81.70

Table 51

Hits from enumeration at all modification sites of LL2_1 and comparison with human LeuRS.

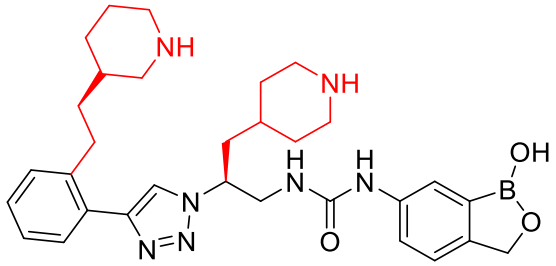
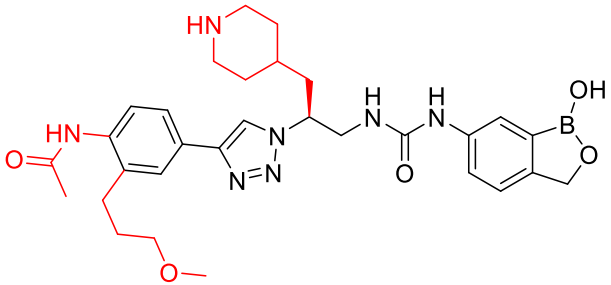
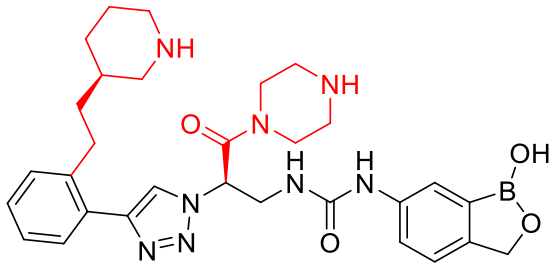
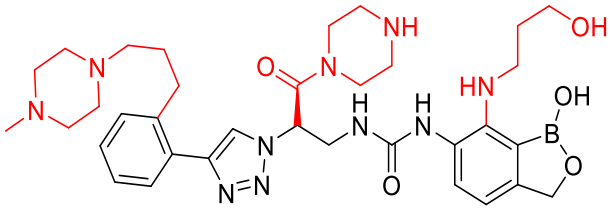
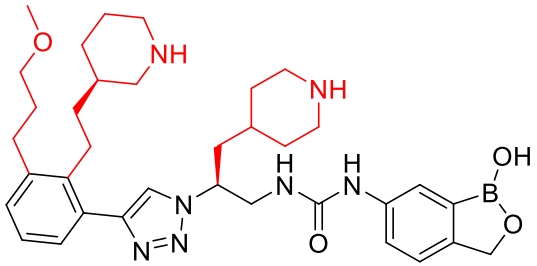
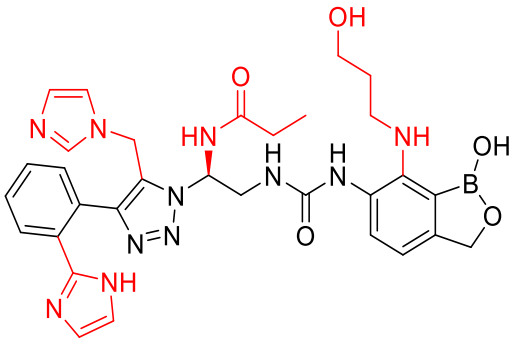
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-122.15	-74.41	-47.74	-8.916	-7.238	-1.678
	-108.73	-47.39	-61.34	-9.422	-7.543	-1.879
	-106.04	-41.12	-64.92	-10.023	-6.56	-3.464

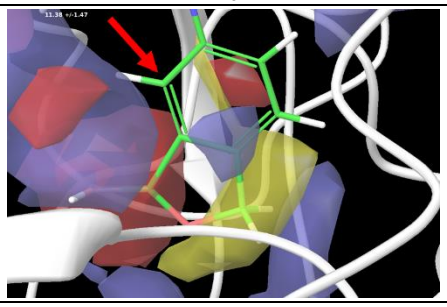
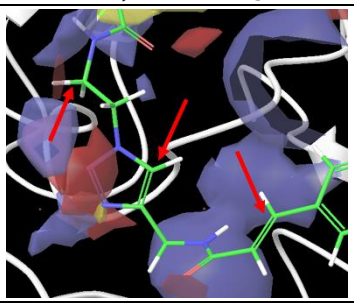
Table 51 (continued)

Structure	MM-GBSA ΔG_{Bind}			Docking Score		
	Fungal	Human	Difference	Fungal	Human	Difference
	-100.73	-15.03	-85.7	-5.908	-4.461	-1.447
	-91.83	-46.5	-45.33	-7.242	-7.213	-0.029
	-76.56	-15.67	-60.88	-5.346	-4.496	-0.85

2.9.5.10 LL2_2.

Table 52

The surfaces generated by binding site analysis at the LeuRS-LL2_2 binding site and arrows point to modification site of R_0, R_1, R_2 and R_3.

R_0	R_1, R_2 and R_3
	

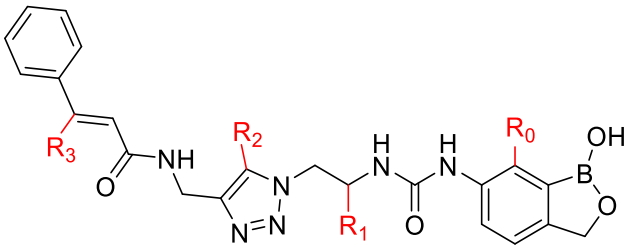


Figure 21. 2D structure of LL2_2 illustrating all the prospective modification sites.

Table 53

Hits from single site enumeration at R_0 of LL2_2.

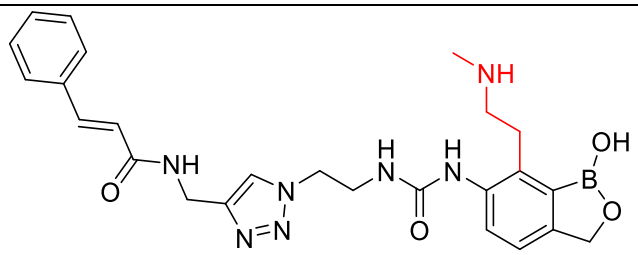
Structure	MM-GBSA ΔG_{Bind}
	-101.86

Table 53 (continued)

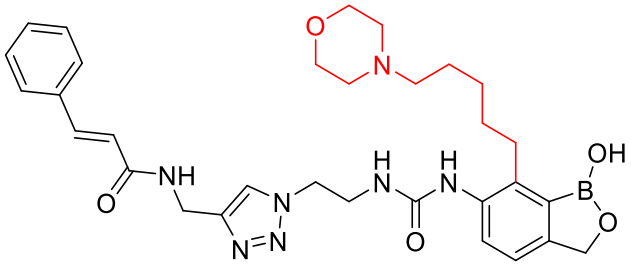
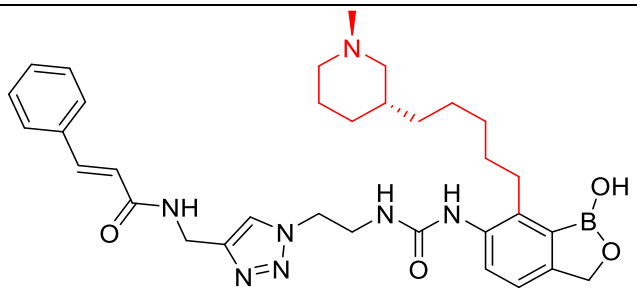
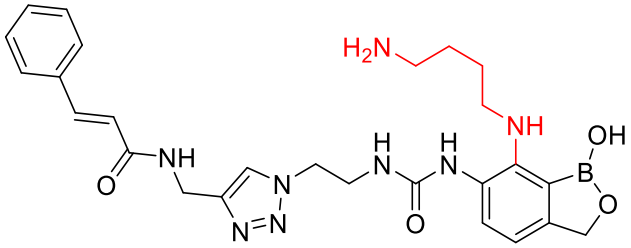
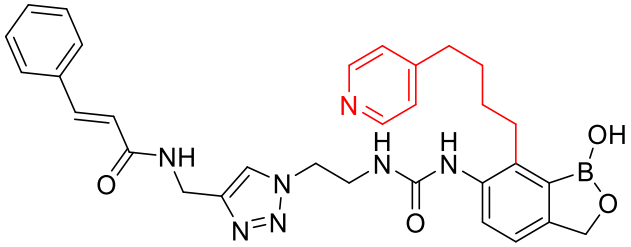
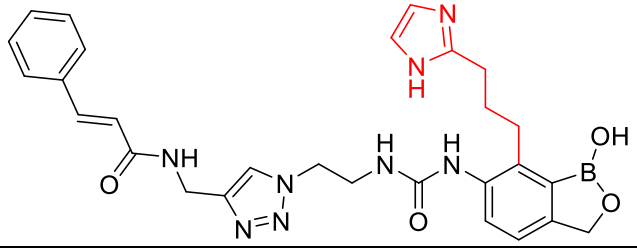
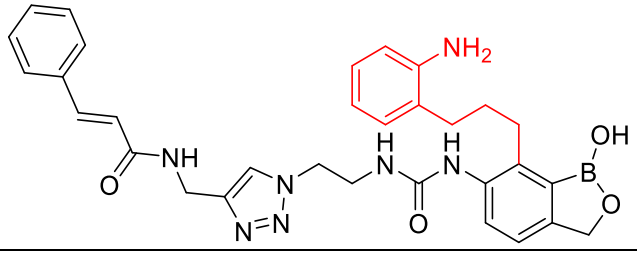
Structure	MM-GBSA ΔG_{Bind}
	-102.86
	-104.17
	-118.17
	-106.42
	-107.17
	-112.48

Table 53 (continued)

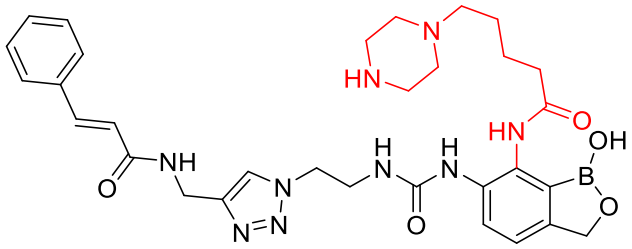
Structure	MM-GBSA ΔG_{Bind}
	-104.67

Table 54

Hits from single site enumeration at R_1 of LL2_2.

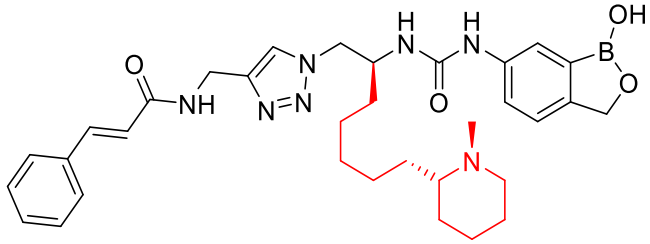
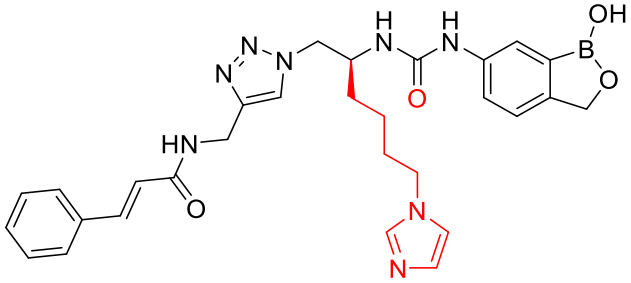
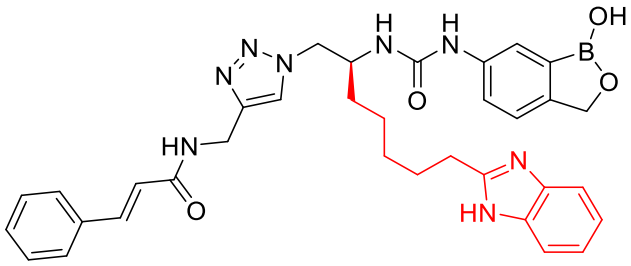
Structure	MM-GBSA ΔG_{Bind}
	-104.59
	-106.08
	-102.93

Table 54 (continued)

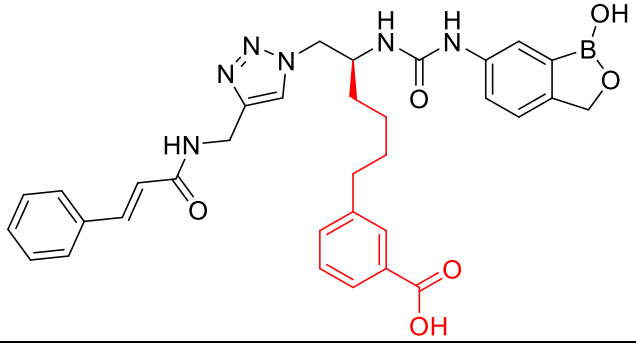
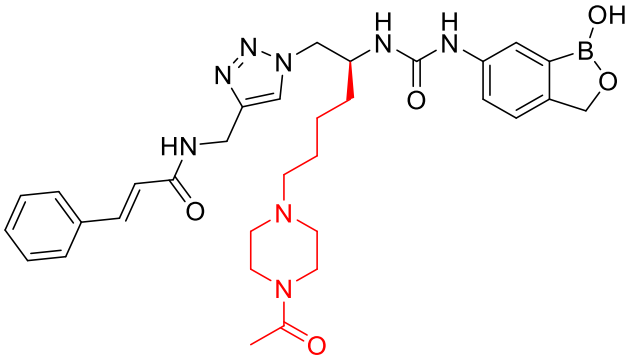
Structure	MM-GBSA ΔG_{Bind}
	-104.49
	-97.09

Table 55

Hits from single site enumeration at R_2 of LL2_2.

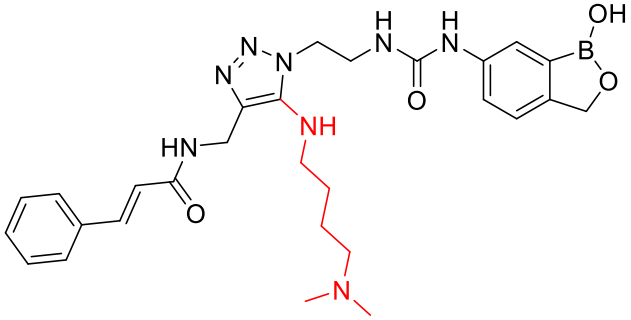
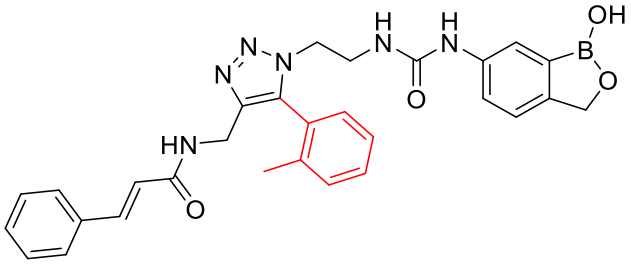
Structure	MM-GBSA ΔG_{Bind}
	-86.42
	-82.09

Table 55 (continued)

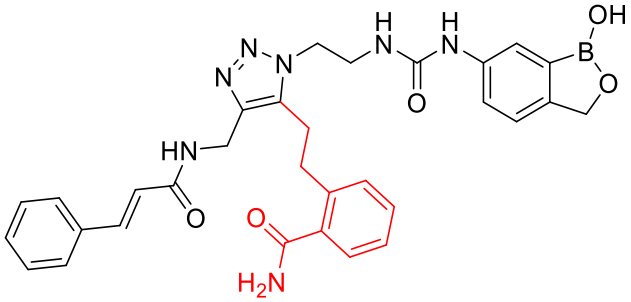
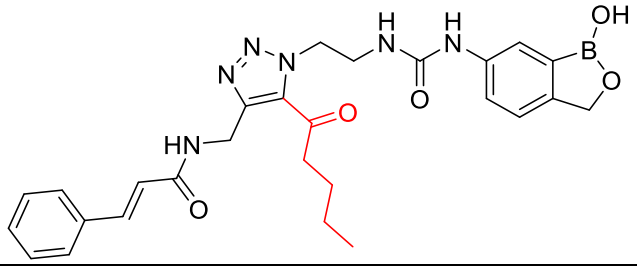
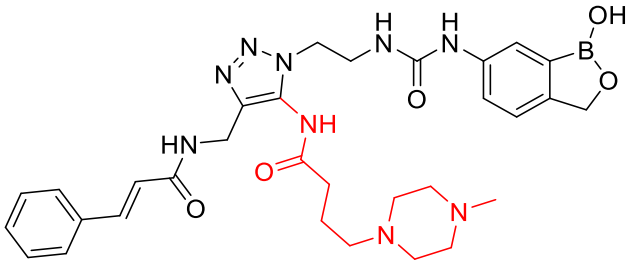
Structure	MM-GBSA ΔG_{Bind}
	-91.10
	-72.06
	-84.49

Table 56

Hits from single site enumeration at R_3 of LL2_2.

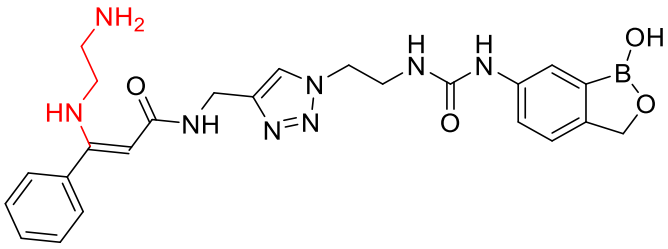
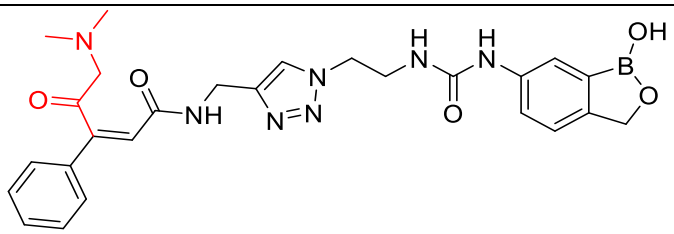
Structure	MM-GBSA ΔG_{Bind}
	-91.36
	-90.75

Table 56 (continued)

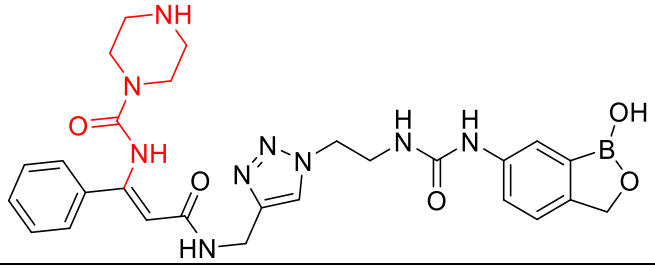
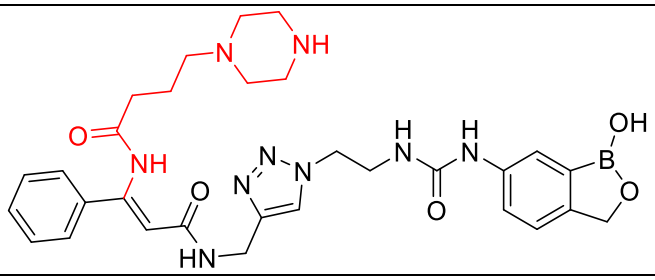
Structure	MM-GBSA ΔG_{Bind}
	-94.73
	-105.95

Table 57

Hits from enumeration at all modification sites of LL2_2 and comparison with human LeuRS.

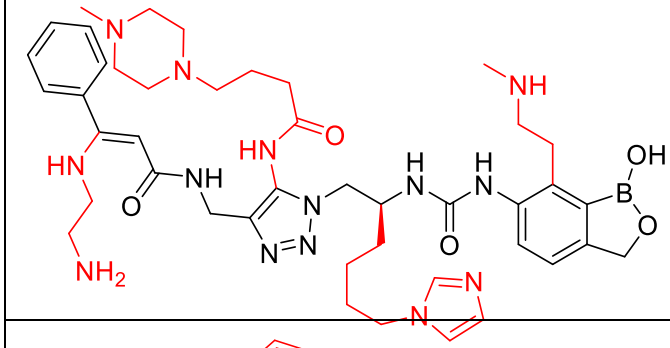
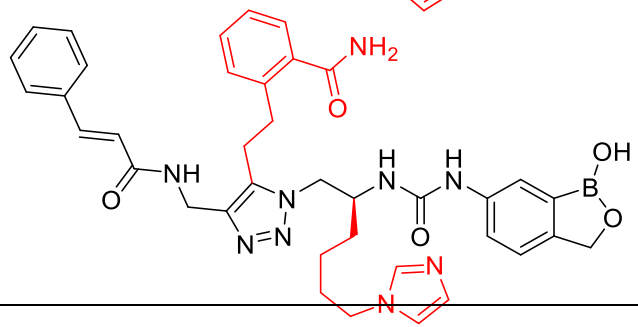
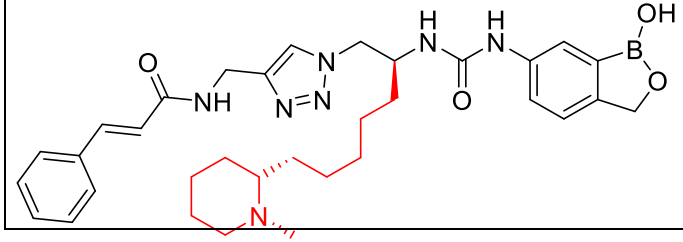
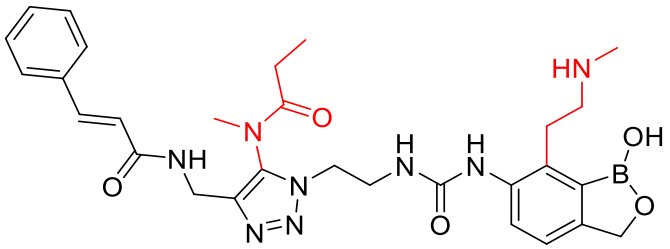
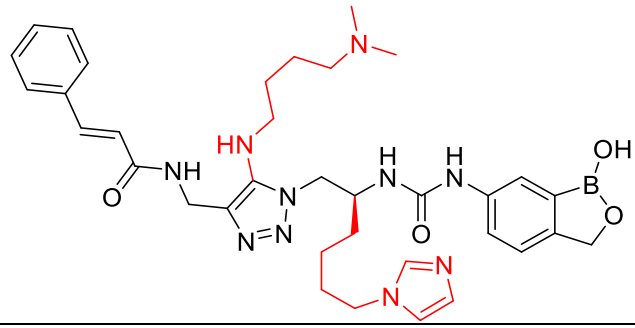
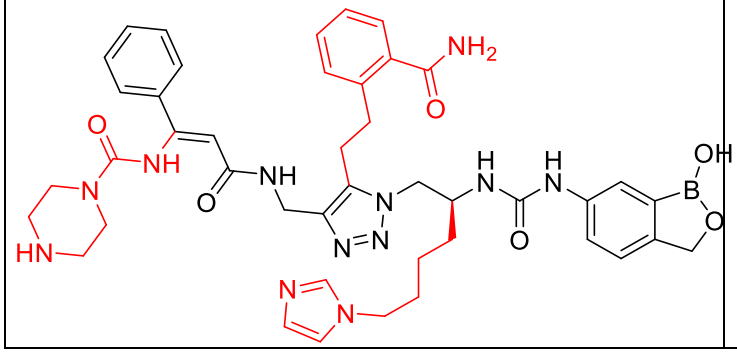
	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-133.89	-112.74	-21.16	-8.154	-5.973	-2.181
	-100.76	-71.36	-29.4	-7.603	-8.437	0.834
	-94.19	-70.02	-24.16	-6.012	-7.301	1.289

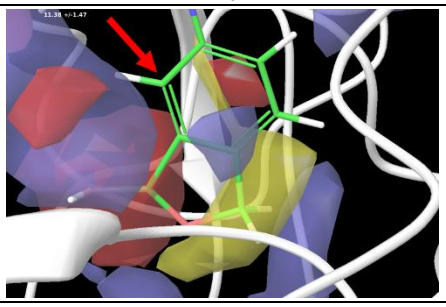
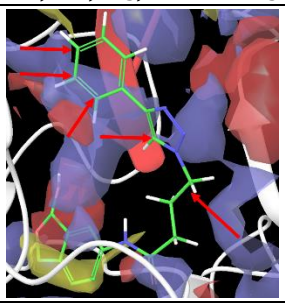
Table 57 (continued)

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-93.13	-19.83	-73.3	-5.6	-6.124	0.524
	-87.09	-52.08	-35.01	-10.069	-6.002	-4.067
	-85.51	-70.99	-14.53	-6.193	-6.774	0.581

2.9.5.11 LL2_3.

Table 58

The surfaces generated by binding site analysis at the LeuRS-LL2_3 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4 and R_5.

R_0	R_1, R_2, R_3, R_4 and R_5
	

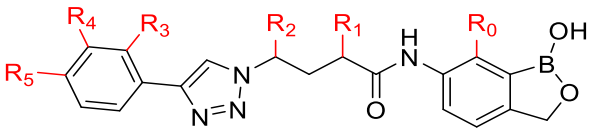


Figure 22. 2D structure of LL2_3 illustrating all the prospective modification sites.

Table 59

Hits from single site enumeration at R_0 of LL2_3.

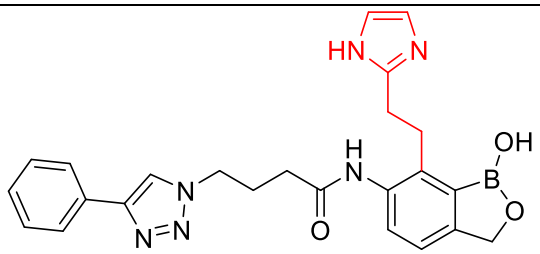
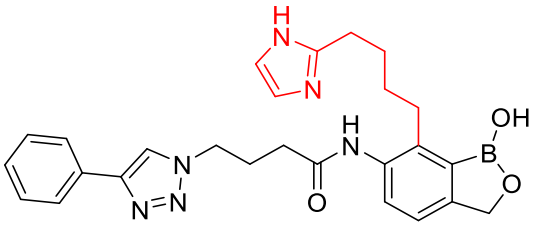
Structure	MM-GBSAΔG_{Bind}
	-88.83
	-94.16

Table 59 (continued)

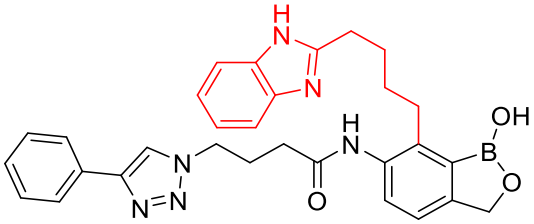
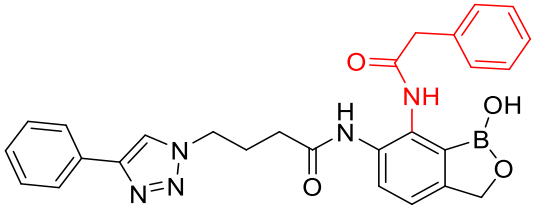
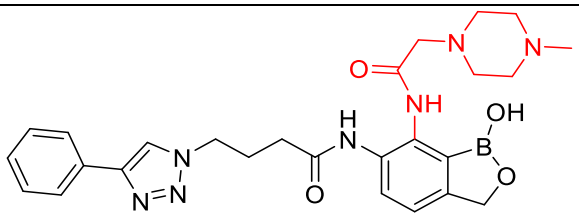
Structure	MM-GBSAΔG_{Bind}
	-93.02
	-93.97
	-92.53

Table 60

Hits from single site enumeration at R_1 of LL2_3.

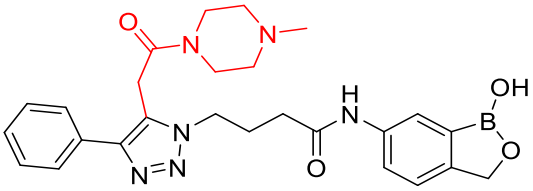
Structure	MM-GBSA ΔG_{Bind}
	-76.29

Table 61

Hits from single site enumeration at R_2 of LL2_3.

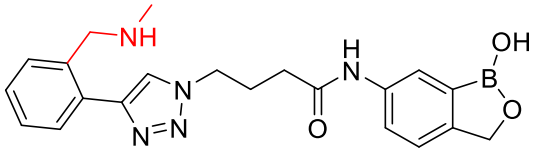
Structure	MM-GBSA ΔG_{Bind}
	-71.98

Table 61 (continued)

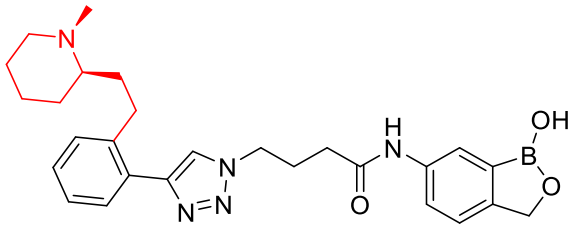
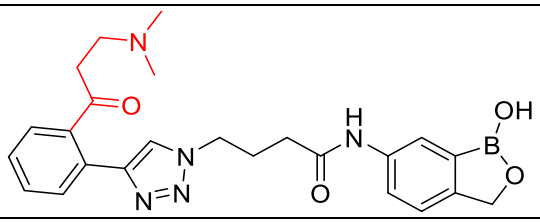
Structure	MM-GBSAΔG_{Bind}
	-71.17
	-70.06

Table 62

Hits from single site enumeration at R_3 of LL2_3.

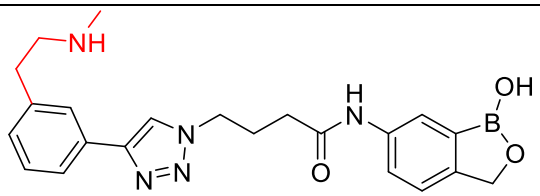
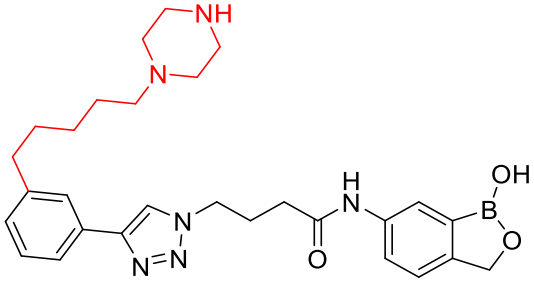
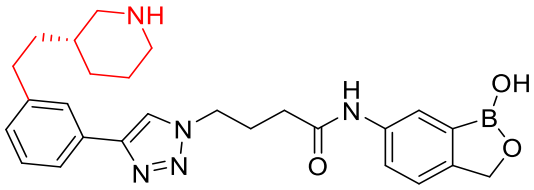
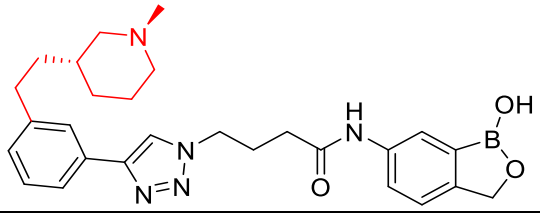
Structure	MM-GBSA ΔG_{Bind}
	-88.66
	-88.94
	-91.37
	-92.47

Table 62 (continued)

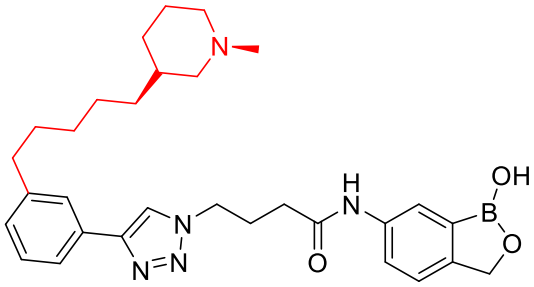
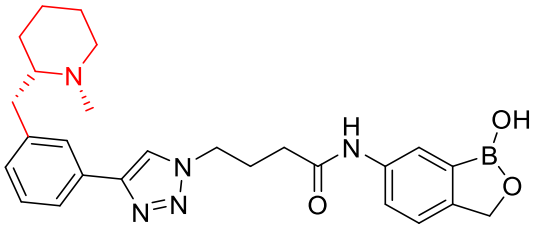
Structure	MM-GBSA ΔG_{Bind}
	-102.30
	-93.91

Table 63

Hits from single site enumeration at R_4 of LL2_3.

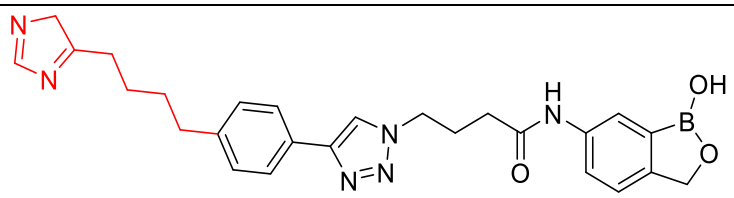
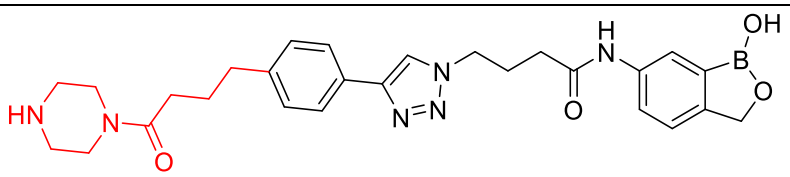
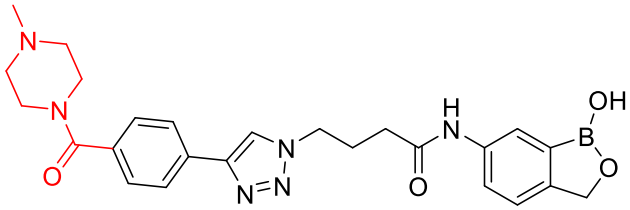
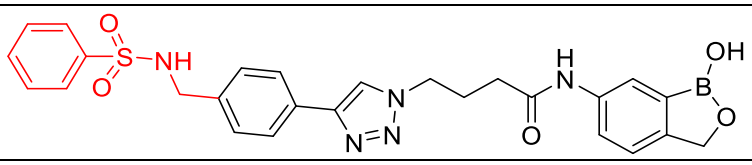
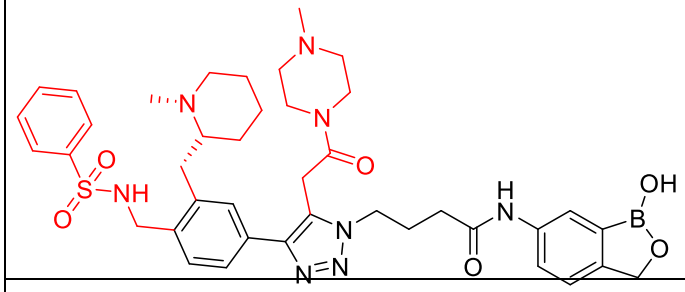
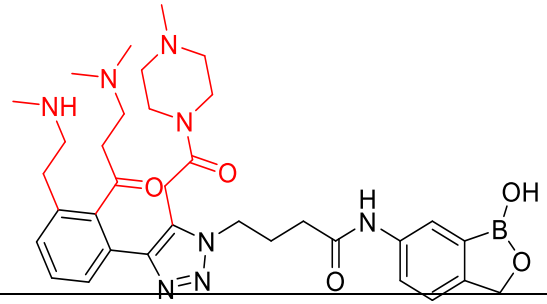
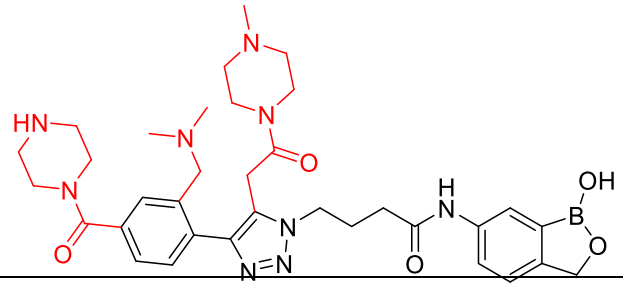
Structure	MM-GBSA ΔG_{Bind}
	-89.52
	-94.02
	-87.92
	-90.91

Table 64

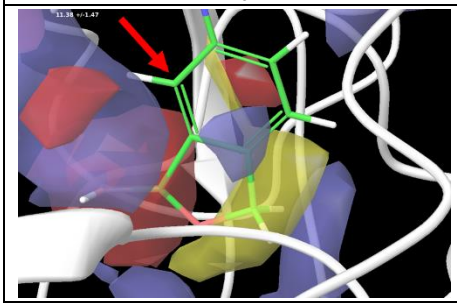
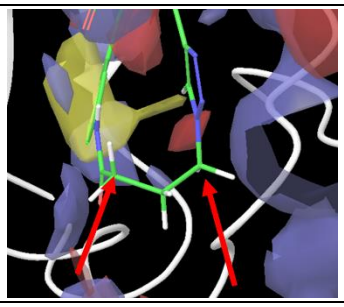
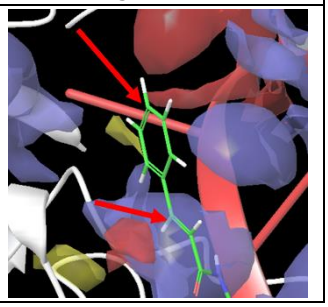
Hits from enumeration at all modification sites of LL2_3 and comparison with human LeuRS.

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-94.25	-92.79	-1.46	-4.946	-8.785	3.839
	-79.88	-40.94	-38.94	-9.587	-6.25	-3.337
	-68.05	-52.47	-15.58	-7.605	-8.164	0.559

2.9.5.12 LL2_4.

Table 65

The surfaces generated by binding site analysis at the LeuRS-LL2_4 binding site and arrows point to modification site of R_0, R_1, R_2, R_3 and R_4.

R_0	R_1 and R_2	R_3 and R_4
		

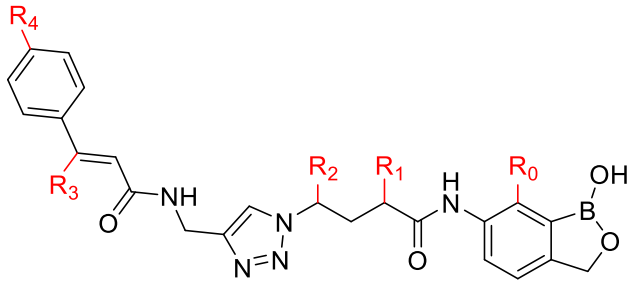


Figure 23. 2D structure of LL2_4 illustrating all the prospective modification sites.

Table 66

Hits from single site enumeration at R_0 of LL2_4.

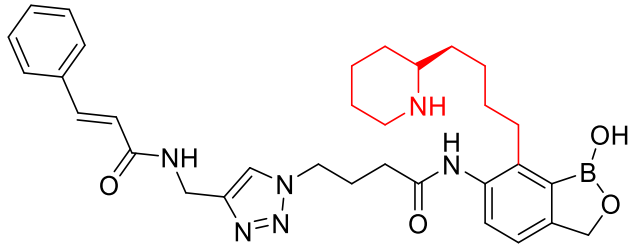
Structure	MM-GBSA ΔG_{Bind}
	-102.98

Table 66 (continued)

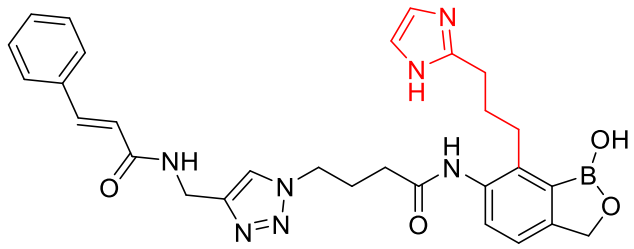
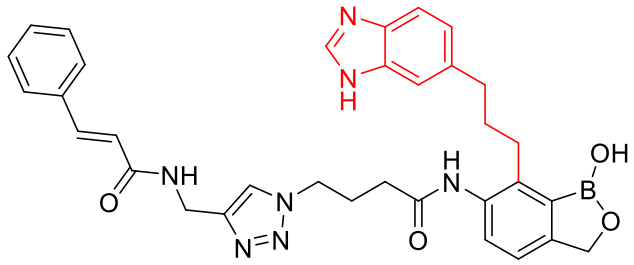
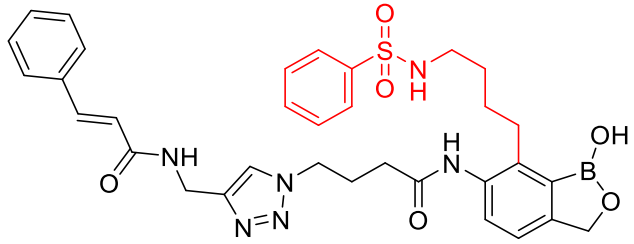
Structure	MM-GBSA ΔG_{Bind}
	-103.72
	-94.68
	-98.42

Table 67

Hits from single site enumeration at R_1 of LL2_4.

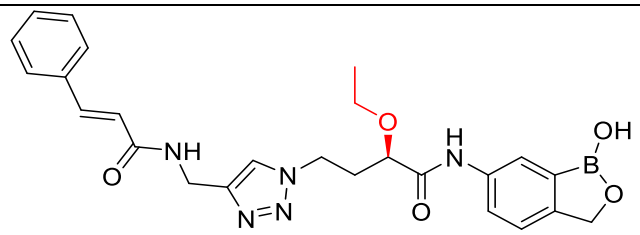
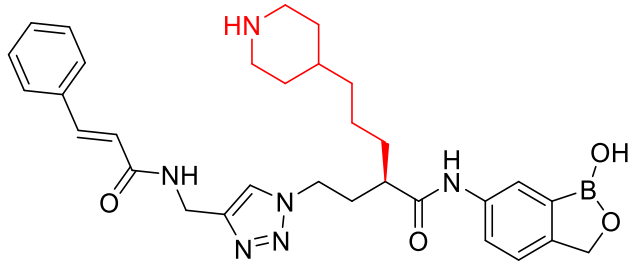
Structure	MM-GBSA ΔG_{Bind}
	-91.12
	-89.21

Table 67 (continued)

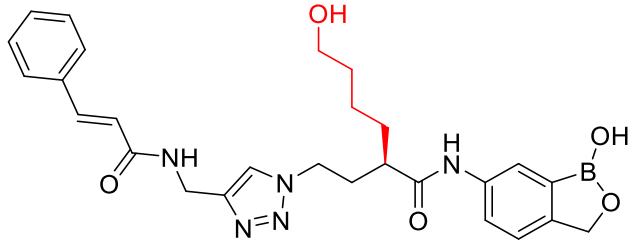
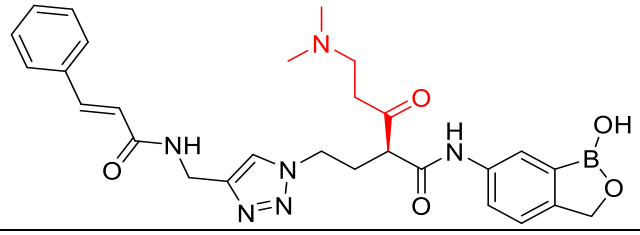
Structure	MM-GBSA ΔG_{Bind}
	-87.33
	-86.63

Table 68

Hits from single site enumeration at R_2 of LL2_4.

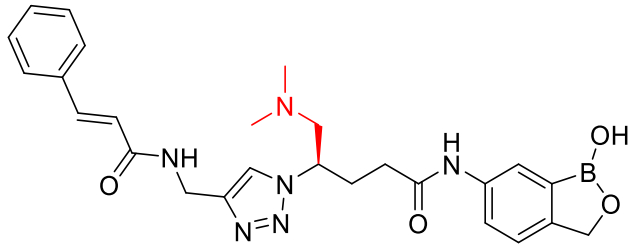
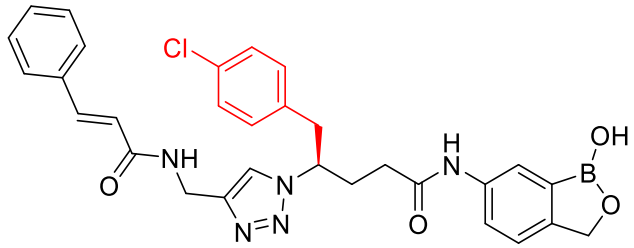
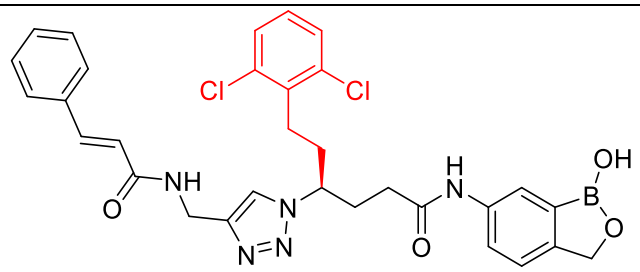
Structure	MM-GBSA ΔG_{Bind}
	-93.28
	-95.30
	-103.54

Table 68 (continued)

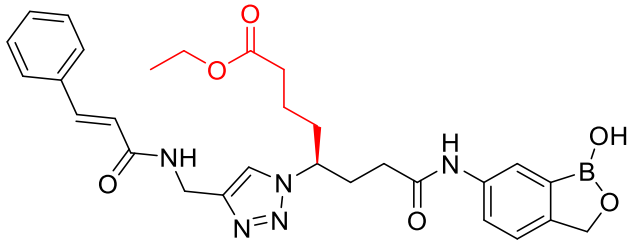
Structure	MM-GBSA ΔG_{Bind}
	-95.73

Table 69

Hits from single site enumeration at R_3 of LL2_4.

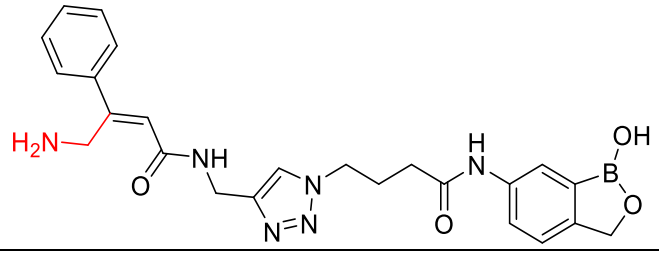
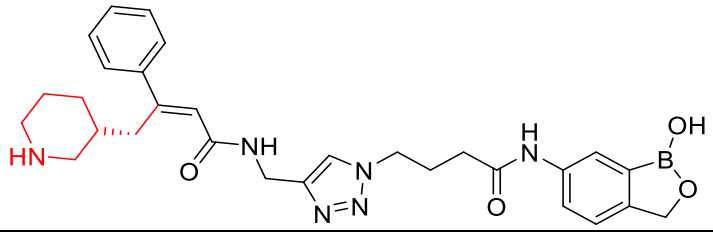
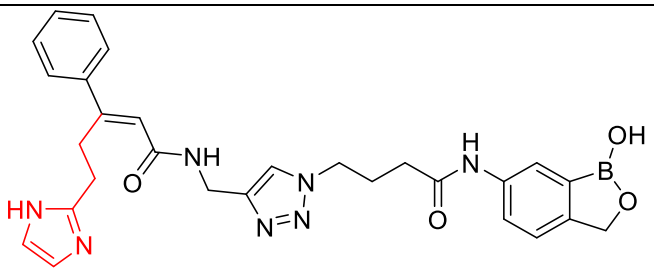
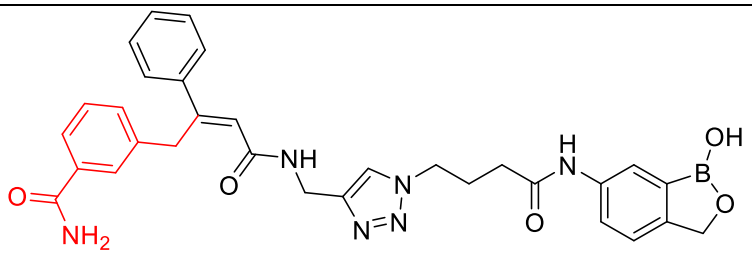
Structure	MM-GBSA ΔG_{Bind}
	-94.93
	-97.01
	-96.14
	-108.11

Table 70

Hits from single site enumeration at R_4 of LL2_4.

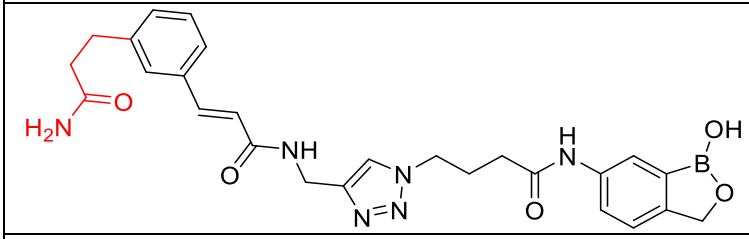
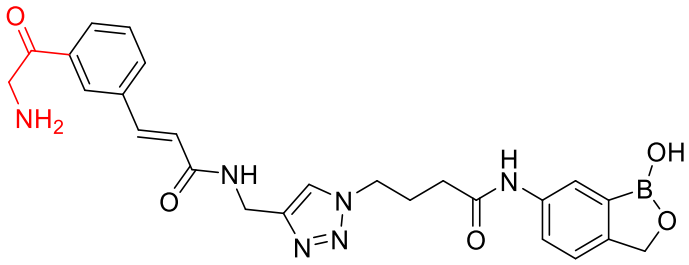
Structure	MM-GBSA ΔG_{Bind}
	-76.79
	-73.48

Table 71

Hits from enumeration at all modification sites of LL2_4 and comparison with human LeuRS.

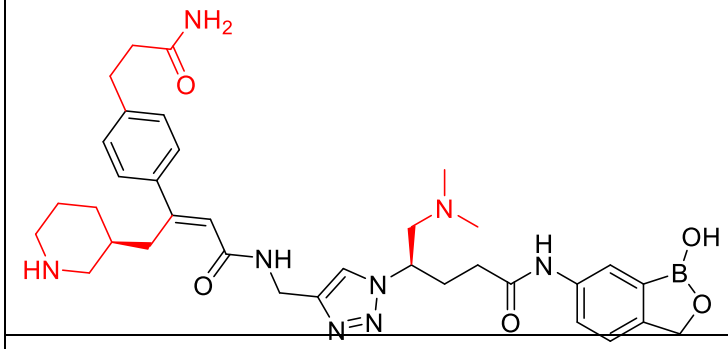
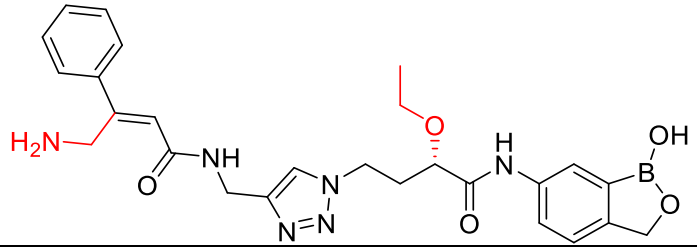
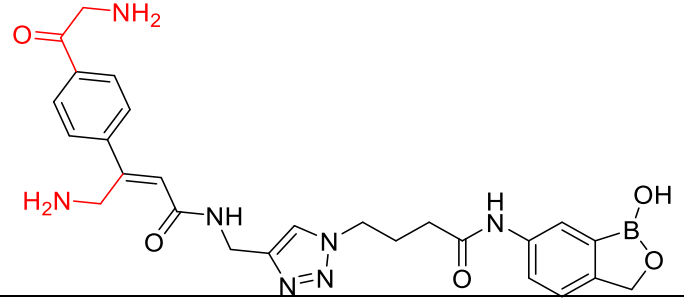
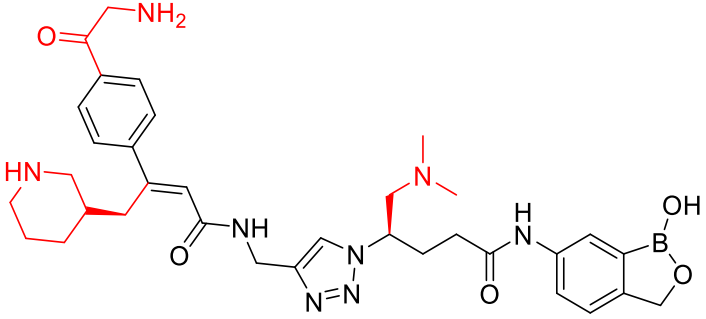
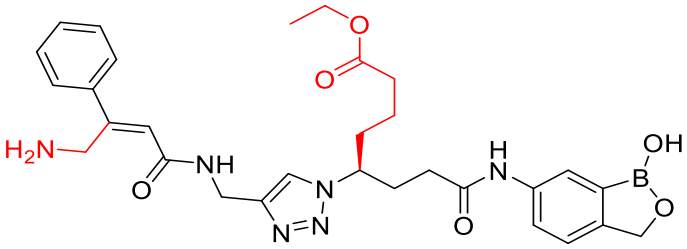
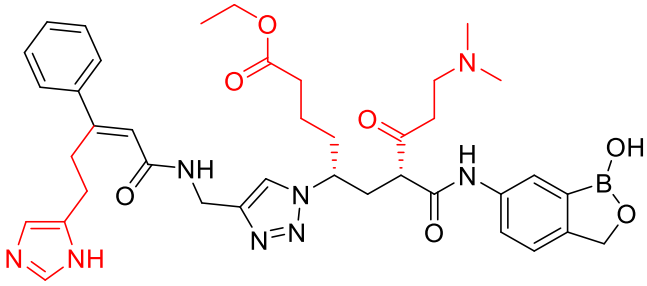
Structure	MM-GBSA ΔG_{Bind}			Docking Score		
	Fungal	Human	Difference	Fungal	Human	Difference
	-117.77	-86.57	-31.2	-9.666	-7.197	-2.469
	-115.46	-39.34	-76.12	-9.943	-7.298	-2.645
	-115.1	-53.6	-61.5	-6.555	-6.421	-0.134

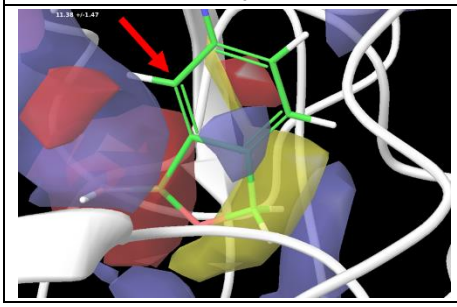
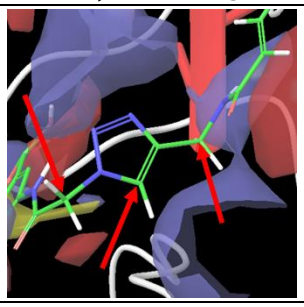
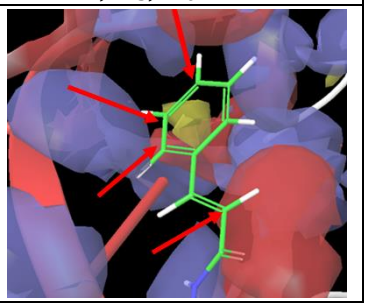
Table 71 (continued)

Structure	MM-GBSA ΔG_{Bind}			Docking Score		
	Fungal	Human	Difference	Fungal	Human	Difference
	-103.24	-53.34	-49.9	-9.602	-6.017	-3.585
	-100.51	-52.67	-47.85	-7.786	-5.209	-2.577
	-99.78	-50.59	-49.19	-10.286	-8.138	-2.148

2.9.5.13 LL2_6.

Table 72

The surfaces generated by binding site analysis at the LeuRS-LL2_6 binding site and arrows point to modification site of R_0, R_1, R_2, R_3, R_4, R_5, R_6 and R_7.

R_0	R_1, R_2 and R_3	R_4, R_5, R_6 and R_7
		

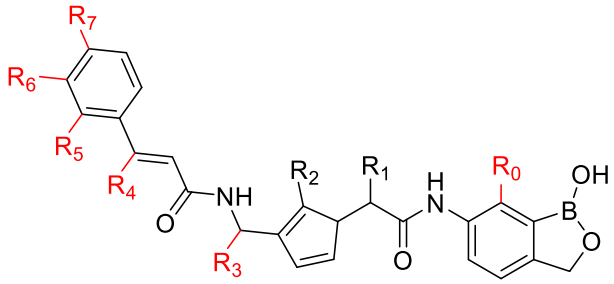


Figure 24. 2D structure of LL2_6 illustrating all the prospective modification sites.

Table 73

Hits from single site enumeration at R_0 of LL2_6.

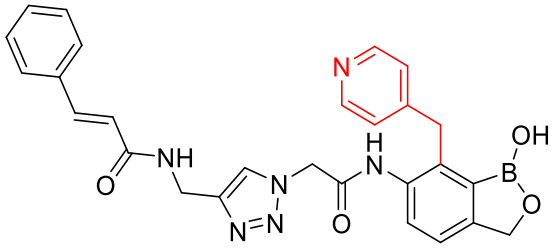
Structure	MM-GBSA ΔG_{Bind}
	-71.54

Table 73 (continued)

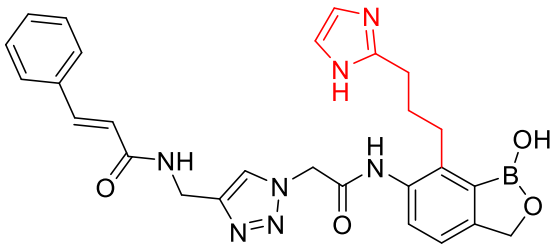
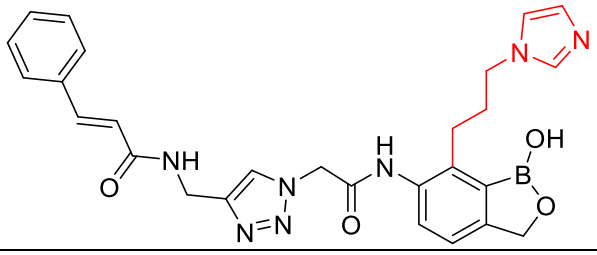
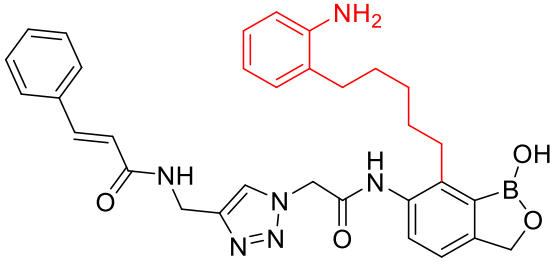
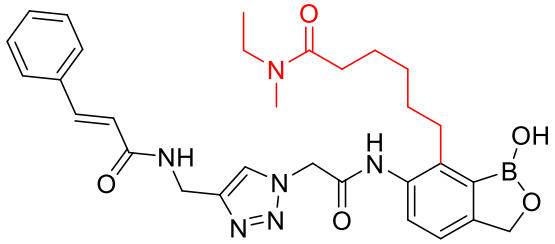
Structure	MM-GBSA ΔG_{Bind}
	-77.73
	-76.81
	-83.20
	-72.03

Table 74

Hits from single site enumeration at R_2 of LL2_6.

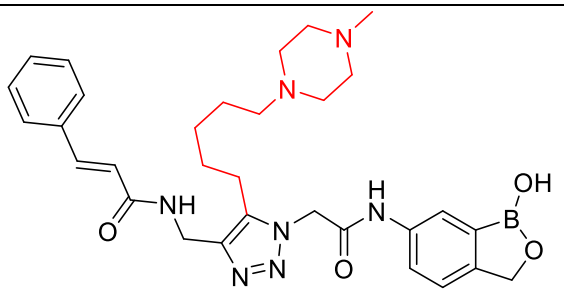
Structure	MM-GBSA ΔG_{Bind}
	-72.63

Table 74 (continued)

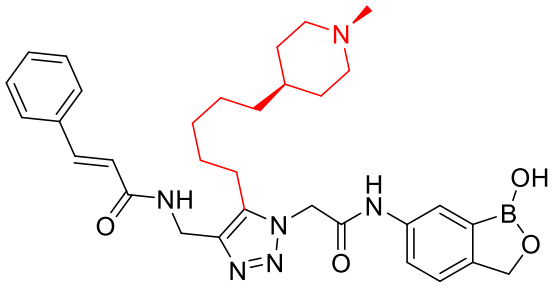
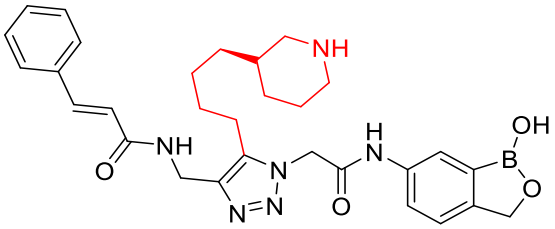
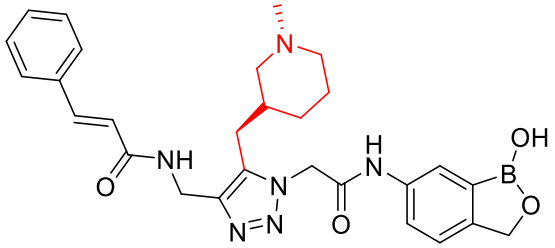
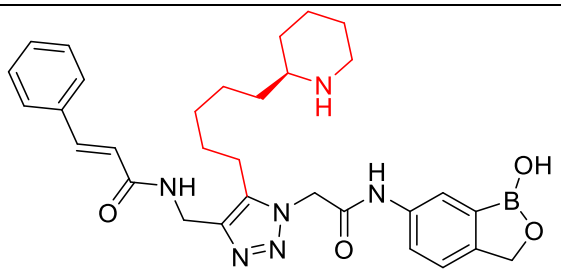
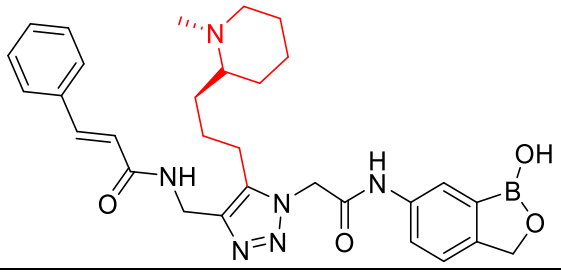
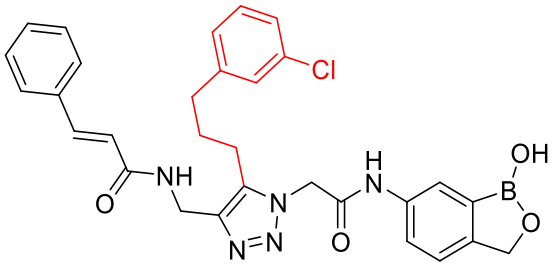
Structure	MM-GBSA ΔG_{Bind}
	-90.37
	-83.51
	-75.85
	-99.39
	-74.21
	-75.28

Table 75

Hits from single site enumeration at R_3 of LL2_6.

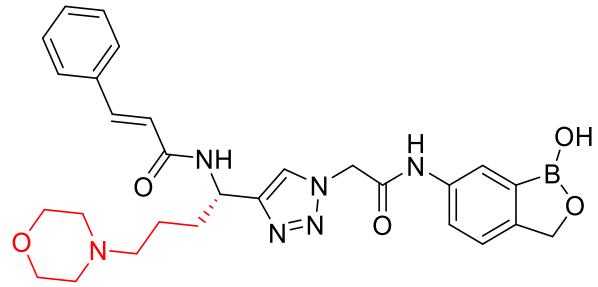
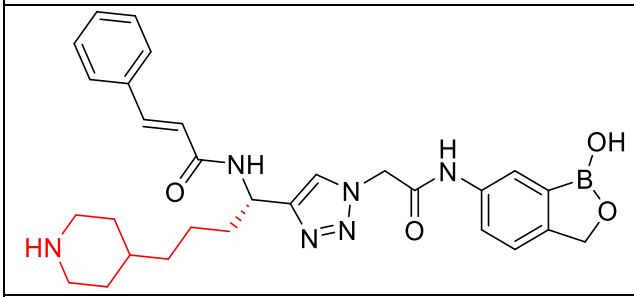
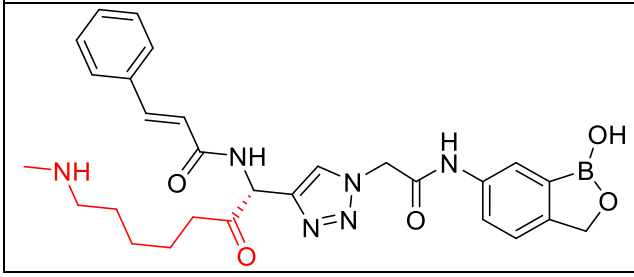
Structure	MM-GBSA ΔG_{Bind}
	-85.83
	-72.36
	-79.76

Table 76

Hits from single site enumeration at R_4 of LL2_6.

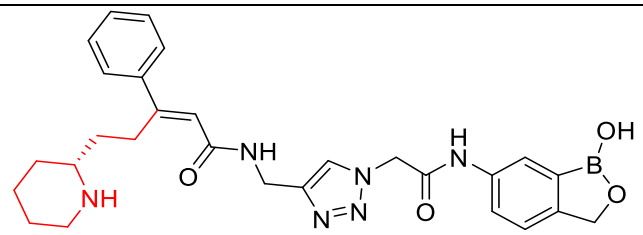
Structure	MM-GBSA ΔG_{Bind}
	-76.97

Table 76 (continued)

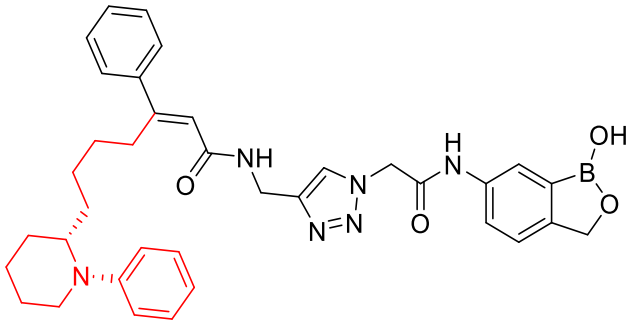
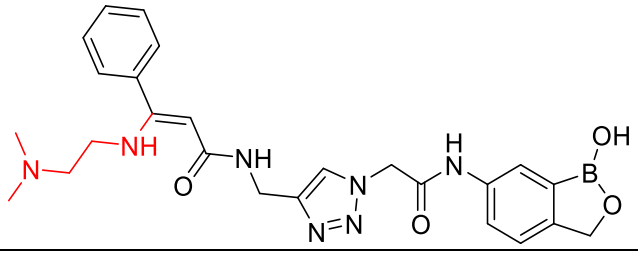
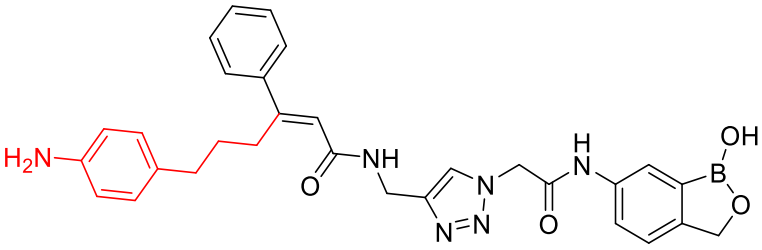
Structure	MM-GBSA ΔG_{Bind}
	-83.40
	-72.18
	-83.23

Table 77

Hits from single site enumeration at R_6 of LL2_6.

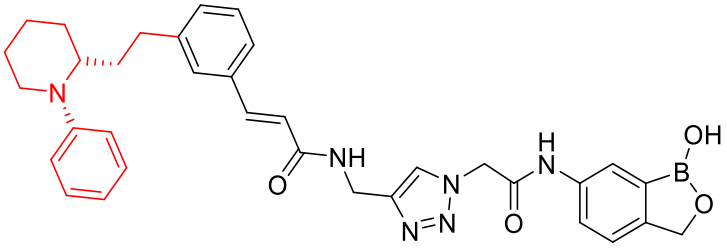
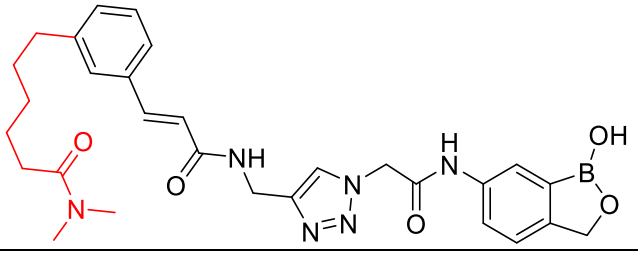
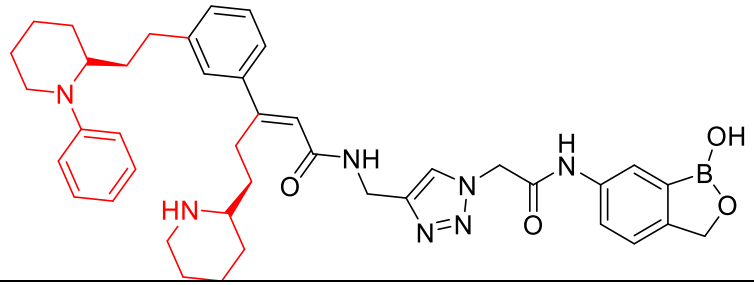
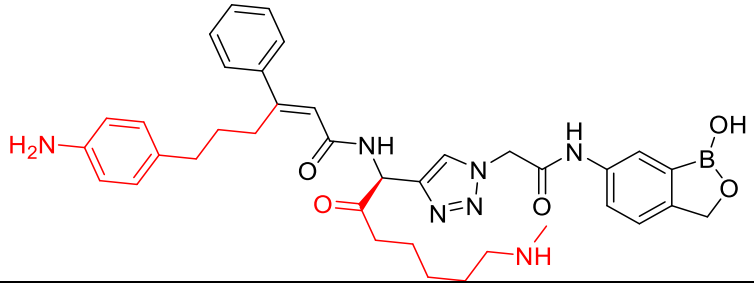
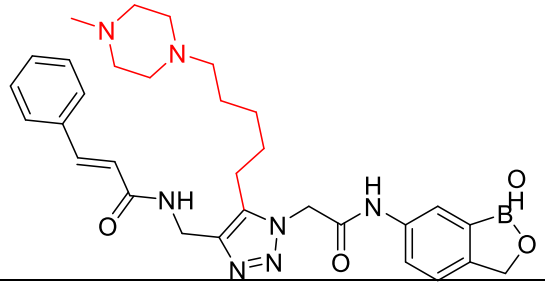
Structure	MM-GBSA ΔG_{Bind}
	-74.13
	-67.79

Table 78

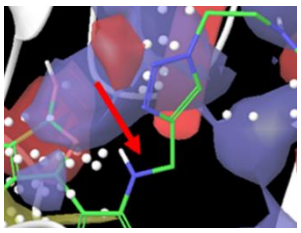
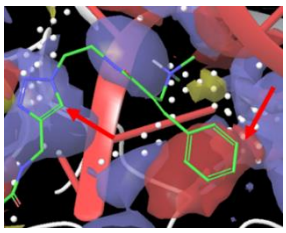
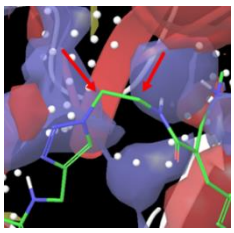
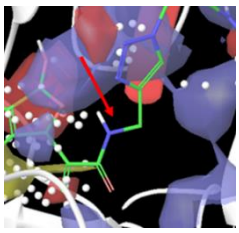
Hits from enumeration at all modification sites of LL2_6 and comparison with human LeuRS.

	MM-GBSA ΔG_{Bind}			Docking Score		
Structure	Fungal	Human	Difference	Fungal	Human	Difference
	-88.15	-50.07	-38.08	-6.012	-7.596	1.584
	-83.94	-72.23	-11.71	-9.159	-8.112	-1.047
	-72.85	-70.38	-2.47	-5.292	-7.55	2.258

2.9.5.14 LL3_4.

Table 79

The surfaces generated by binding site analysis at the LeuRS-LL3_4 binding site and arrows point to modification site of R_1, R_2, R_3, R_4 and R_5.

R_1	R_2	R_3 and R_4	R_5
			

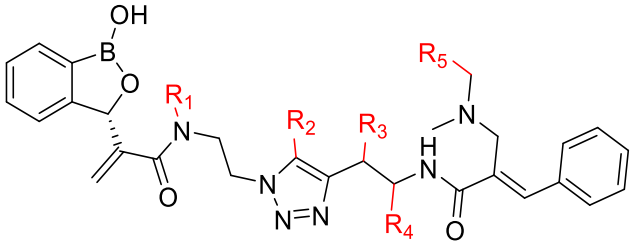


Figure 25. 2D structure of LL3_4 illustrating all the prospective modification sites.

Table 80

Hits from single site enumeration at R_1 of LL3_4.

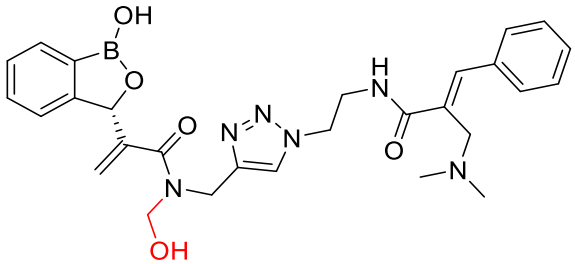
Structure	MM-GBSA ΔG_{Bind}
	-54.01

Table 80 (continued)

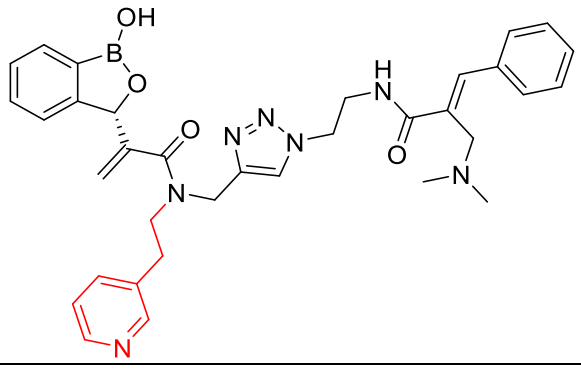
Structure	MM-GBSA ΔG_{Bind}
	-72.46

Table 81

Hits from single site enumeration at R_2 of LL3_4.

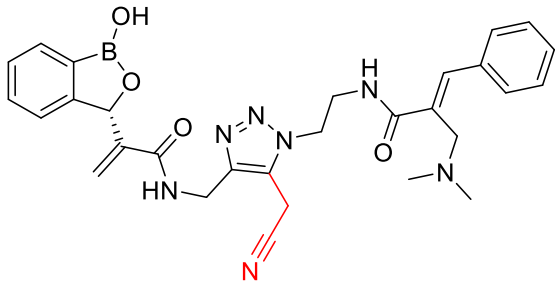
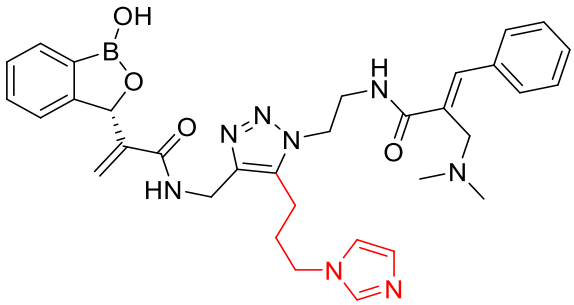
Structure	MM-GBSA ΔG_{Bind}
	-69.64
	-106.00

Table 82

Hits from single site enumeration at R_3 of LL3_4.

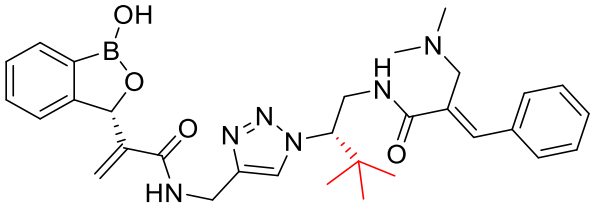
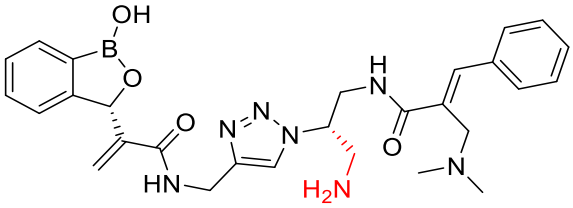
Structure	MM-GBSA ΔG_{Bind}
	-69.86
	-76.06

Table 83

Hits from single site enumeration at R_4 of LL3_4.

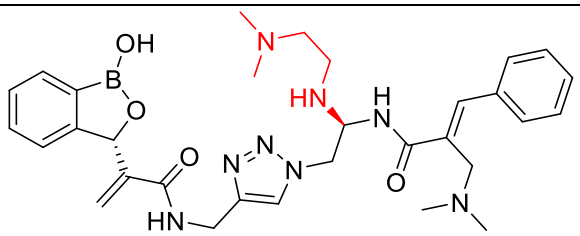
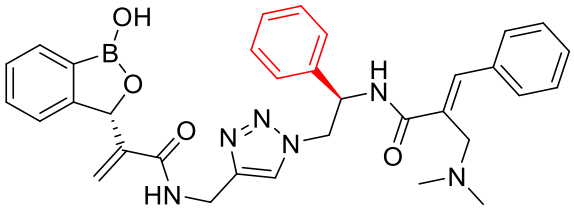
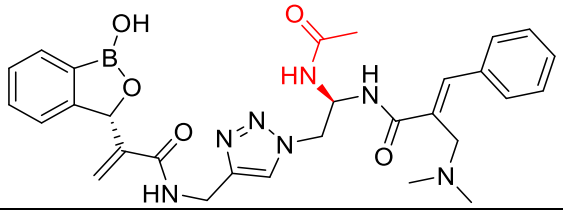
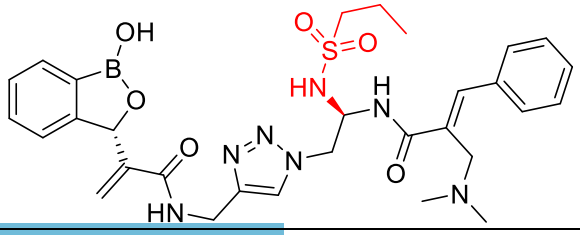
Structure	MM-GBSA ΔG_{Bind}
	-60.50
	-91.22
	-79.08
	-74.02

Table 83 (continued)

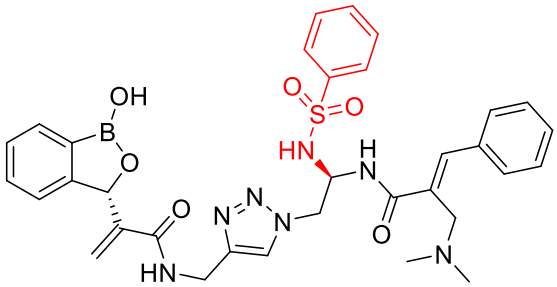
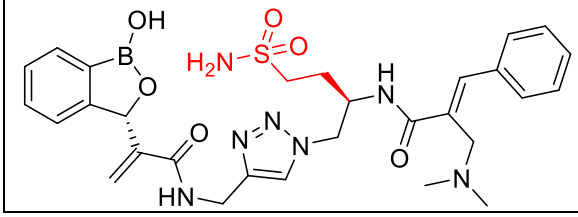
Structure	MM-GBSA ΔG_{Bind}
	-112.13
	-70.44

Table 84

Hits from single site enumeration at R_5 of LL3_4.

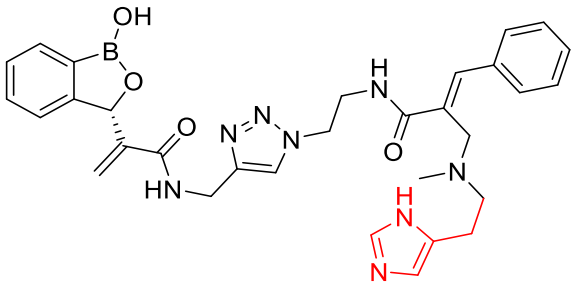
Structure	MM-GBSA ΔG_{Bind}
	-68.06

Table 85

Hits from enumeration at all modification sites of LL3_4.

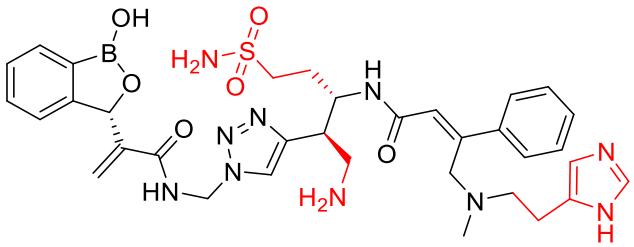
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-104.593	-10.997

Table 85 (continued)

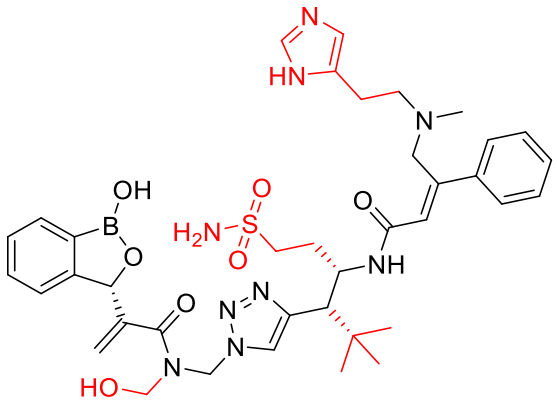
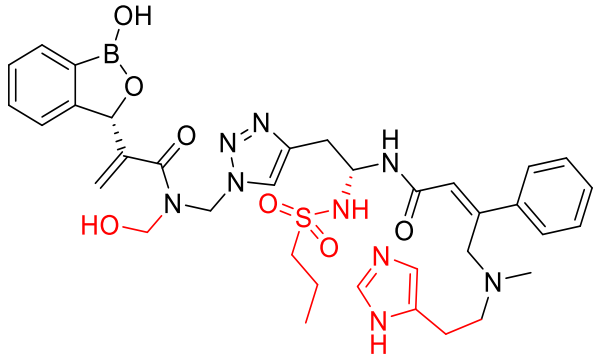
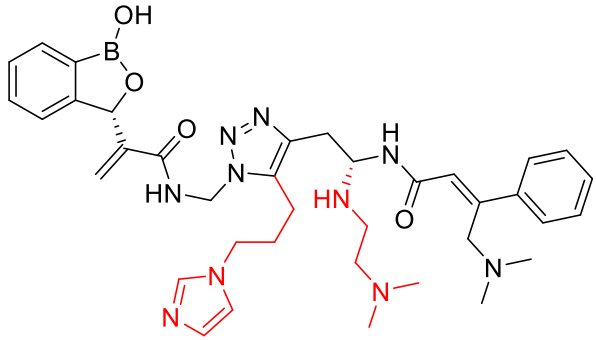
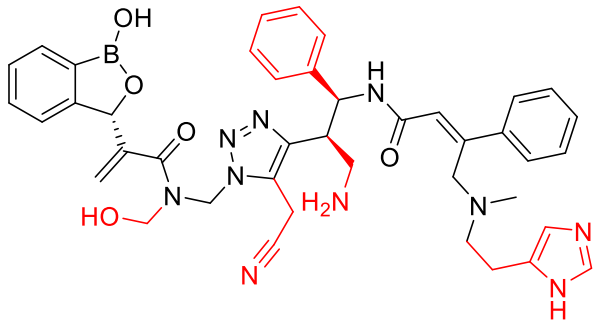
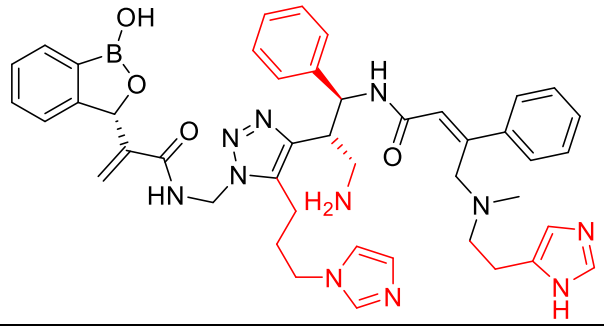
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-102.667	-8.022
	-94.382	-6.692
	-87.388	-8.358
	-86.375	-10.729

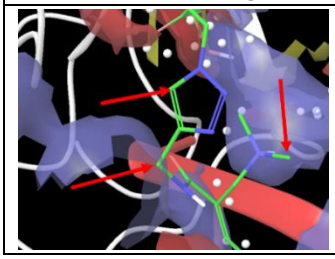
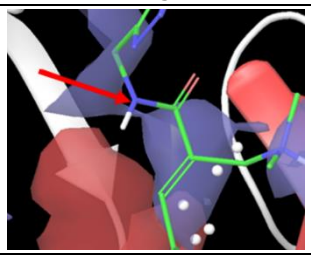
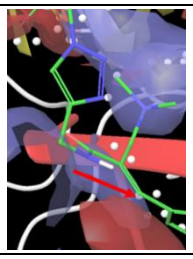
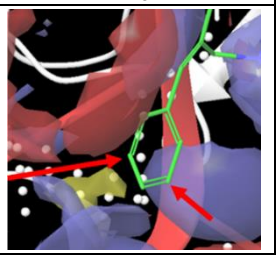
Table 85 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-81.409	-8.276

2.9.5.15 LL3_5.

Table 86

The surfaces generated by binding site analysis at the LeuRS-LL3_5 binding site and arrows point to modification site of R_1, R_2, R_3, R_4, R_5 and R_6.

R_1 R_2 and R_5	R_3	R_4	R_6
			

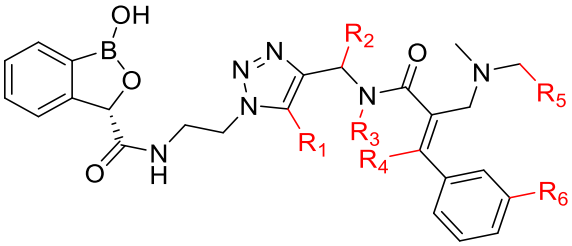


Figure 26. 2D structure of LL3_5 illustrating all the prospective modification sites.

Table 87

Hits from single site enumeration at R_1 of LL3_5.

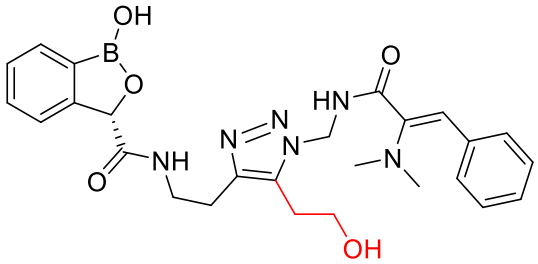
Structure	MM-GBSA ΔG_{Bind}
	-75.77

Table 87 (continued)

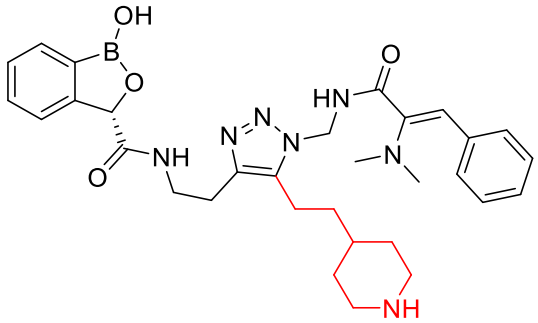
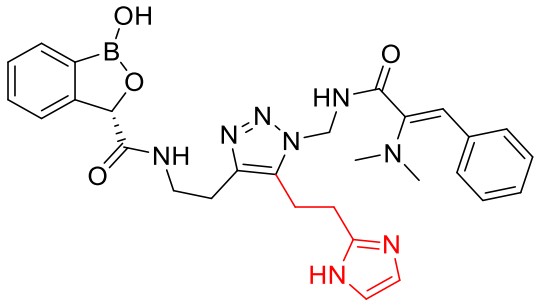
Structure	MM-GBSA ΔG_{Bind}
	-84.14
	-88.56

Table 88

Hits from single site enumeration at R_2 of LL3_5.

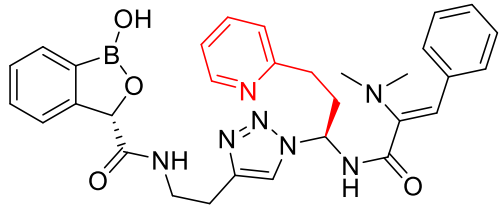
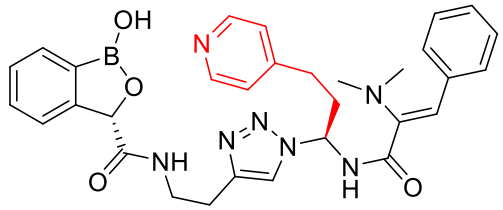
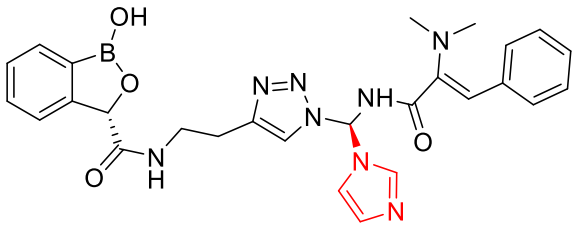
Structure	MM-GBSA ΔG_{Bind}
	-73.57
	-72.40
	-70.88

Table 89

Hits from single site enumeration at R_3 of LL3_5.

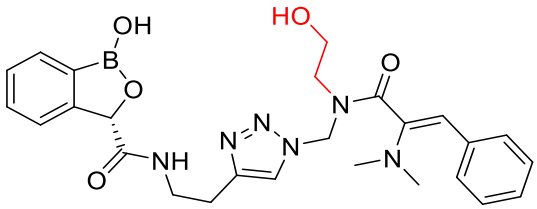
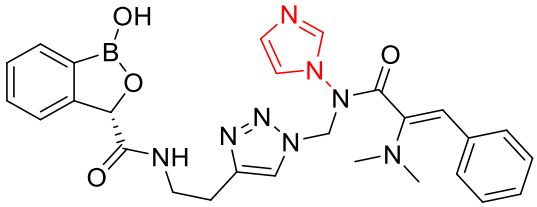
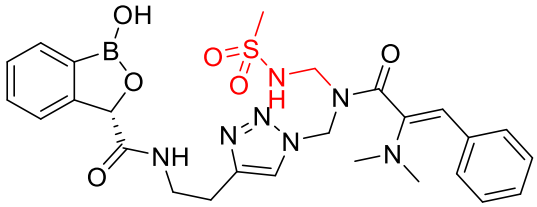
Structure	MM-GBSA ΔG_{Bind}
	-67.53
	-65.46
	-60.40

Table 90

Hits from single site enumeration at R_4 of LL3_5.

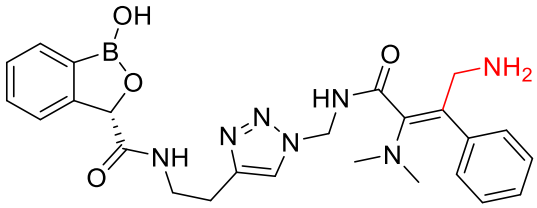
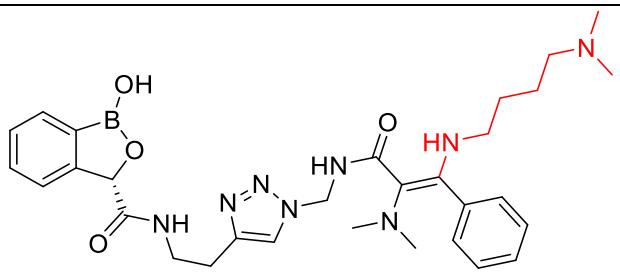
Structure	MM-GBSA ΔG_{Bind}
	-90.68
	-77.73

Table 91

Hits from single site enumeration at R_5 of LL3_5.

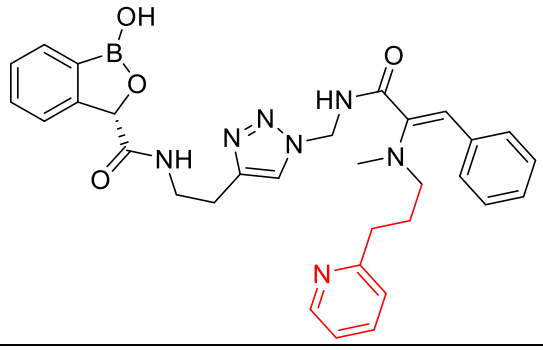
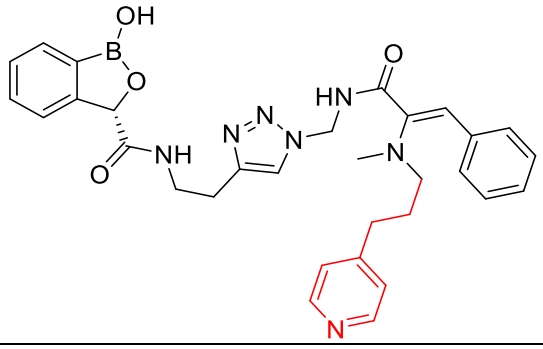
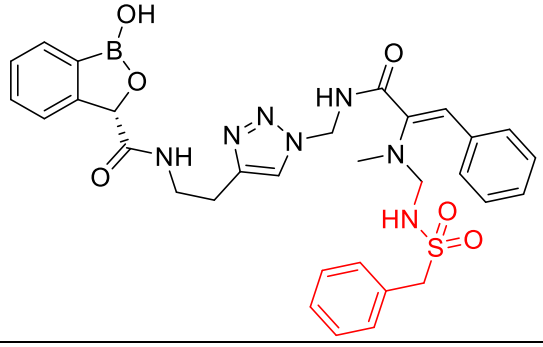
Structure	MM-GBSA ΔG_{Bind}
	-73.12
	-84.29
	-72.97

Table 92

Hits from single site enumeration at R_6 of LL3_5.

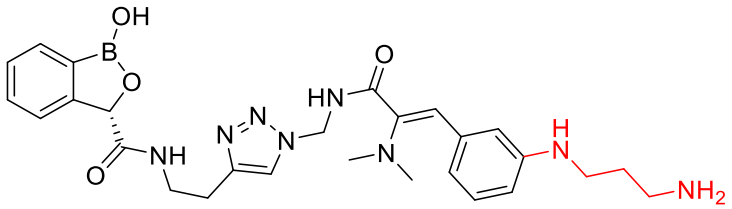
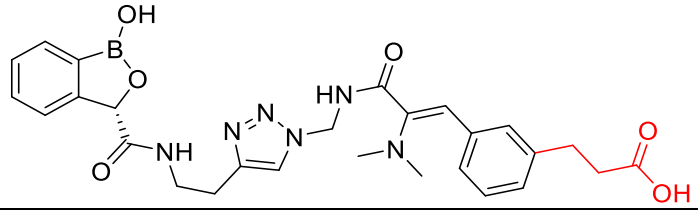
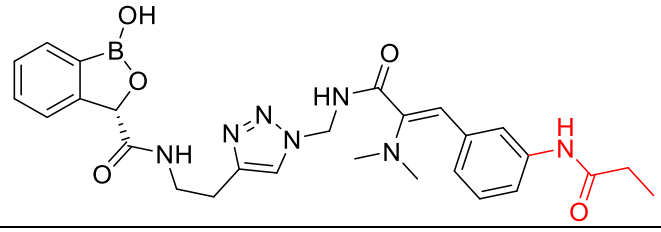
Structure	MM-GBSA ΔG_{Bind}
	-94.28
	-80.17
	-90.77

Table 93

Hits from enumeration at all modification sites of LL3_5.

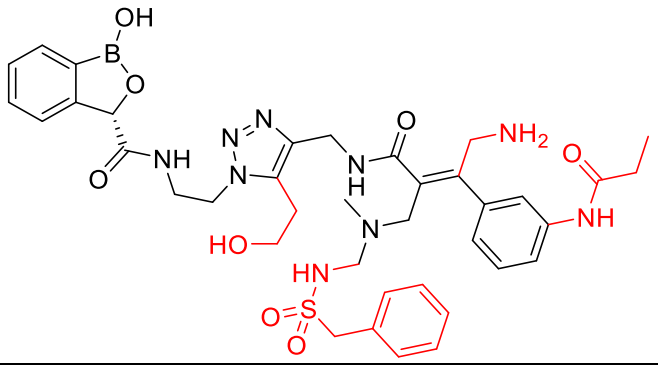
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-99.802	-11.367

Table 93 (continued)

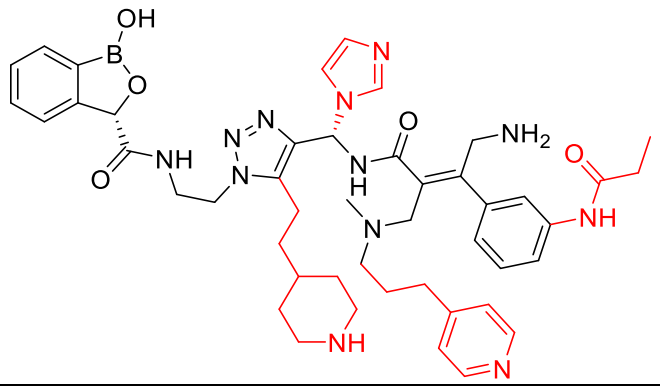
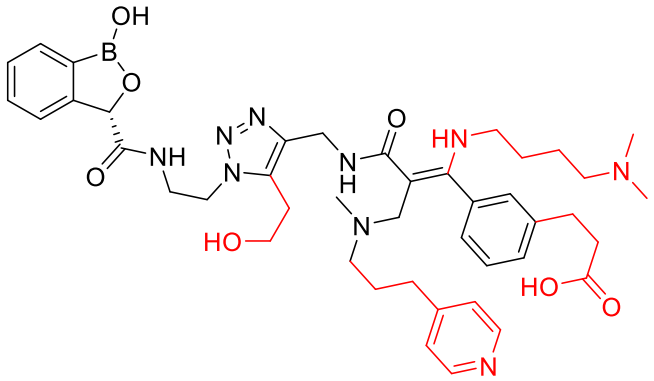
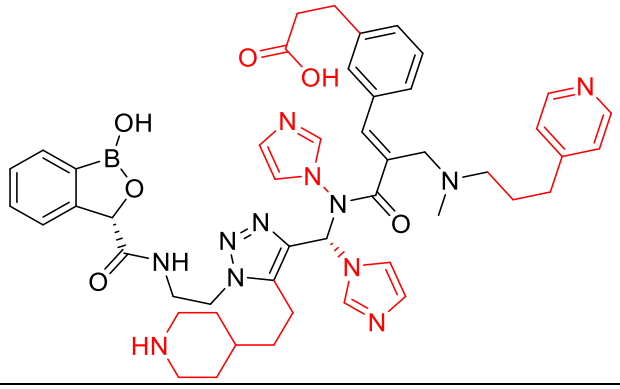
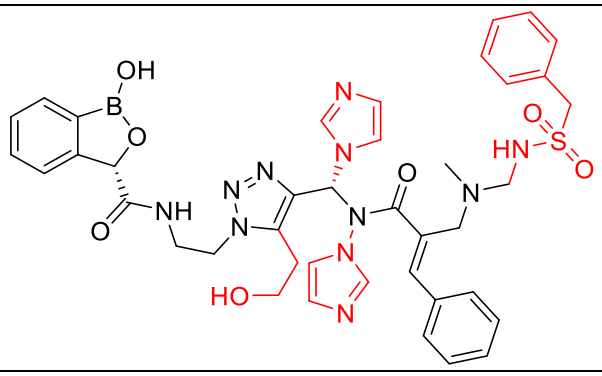
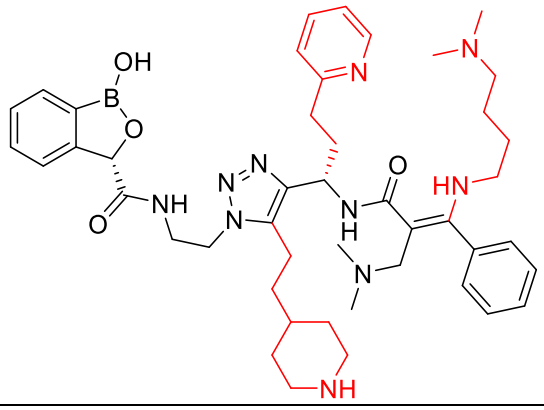
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-92.449	-6.889
	-90.937	-6.307
	-90.255	-7.82
	-88.603	-8.928

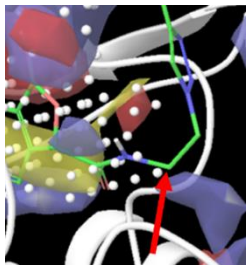
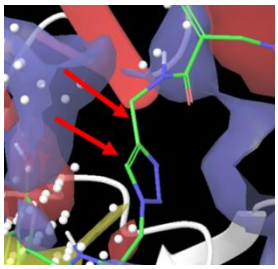
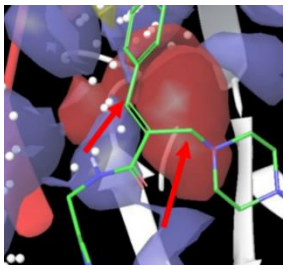
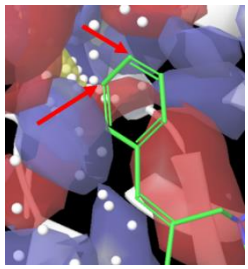
Table 93 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-88.418	-9.65

2.9.5.16 LL3_6.

Table 94

The surfaces generated by binding site analysis at the LeuRS-LL3_6 binding site and arrows point to modification site of R_1, R_2, R_3, R_4, R_5, R_6, and R_7.

R_1	R_2 and R_3	R_4 and R_5	R_6 and R_7
			

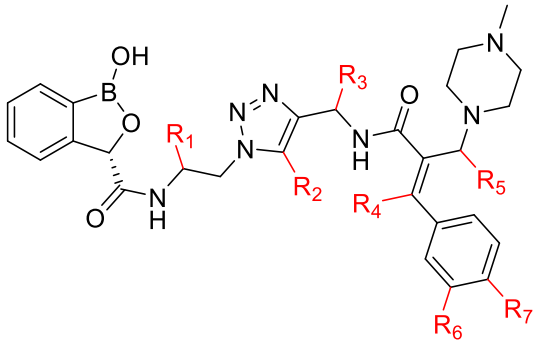


Figure 27. 2D structure of LL3_6 illustrating all the prospective modification sites.

Table 95

Hits from single site enumeration at R_1 of LL3_6.

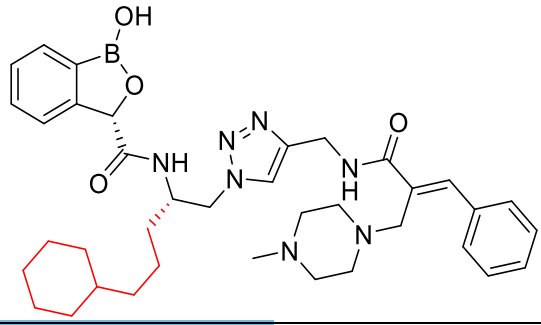
Structure	MM-GBSA ΔG_{Bind}
	-73.37

Table 95 (continued)

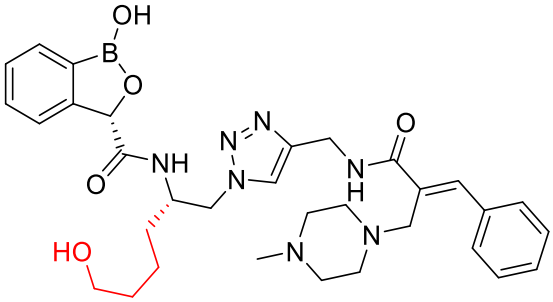
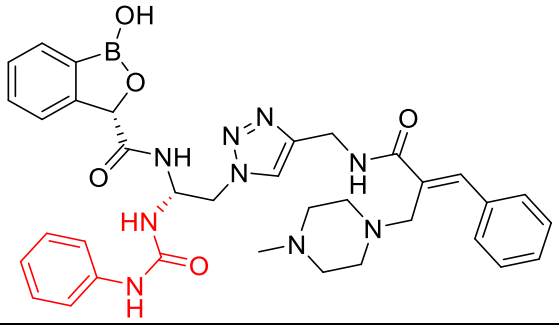
Structure	MM-GBSA ΔG_{Bind}
	-72.51
	-75.01

Table 96

Hits from single site enumeration at R_2 of LL3_6.

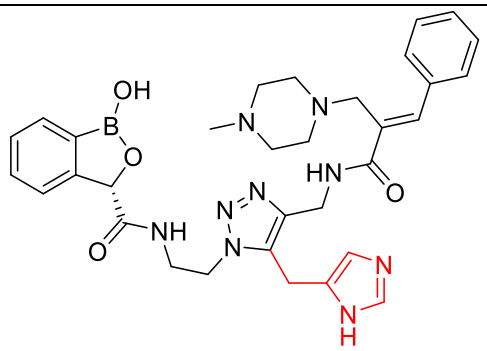
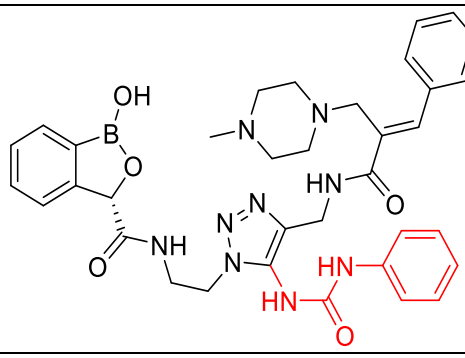
Structure	MM-GBSA ΔG_{Bind}
	-69.48
	-77.17

Table 97

Hits from single site enumeration at R_3 of LL3_6.

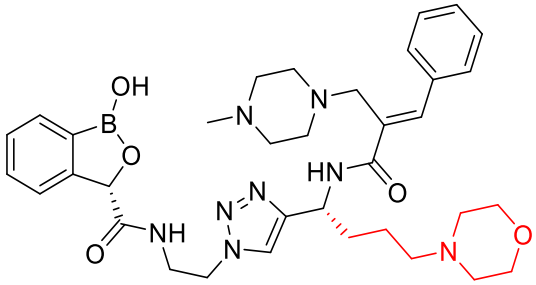
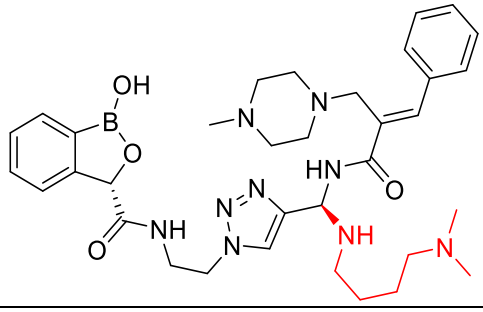
Structure	MM-GBSA ΔG_{Bind}
	-72.40
	-73.00

Table 98

Hits from single site enumeration at R_4 of LL3_6.

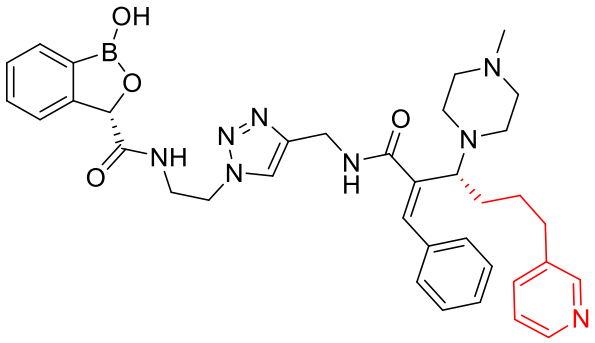
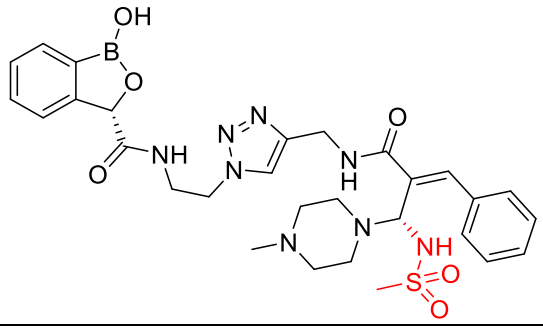
Structure	MM-GBSA ΔG_{Bind}
	-69.97
	-63.59

Table 99

Hits from single site enumeration at R_5 of LL3_6.

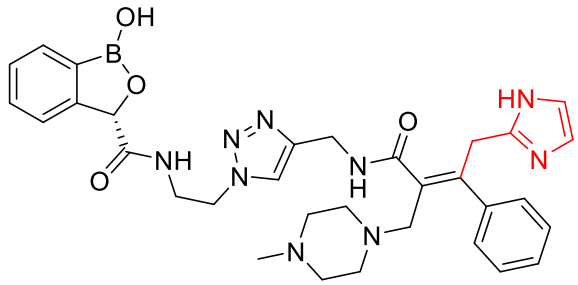
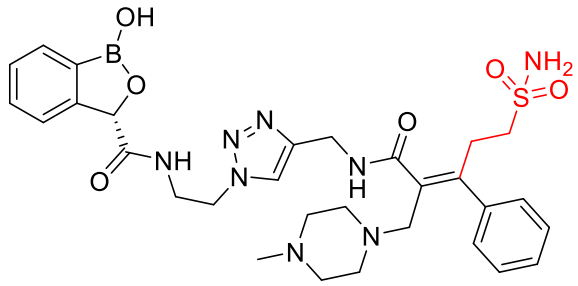
Structure	MM-GBSA ΔG_{Bind}
	-78.42
	-82.01

Table 100

Hits from single site enumeration at R_6 of LL3_6.

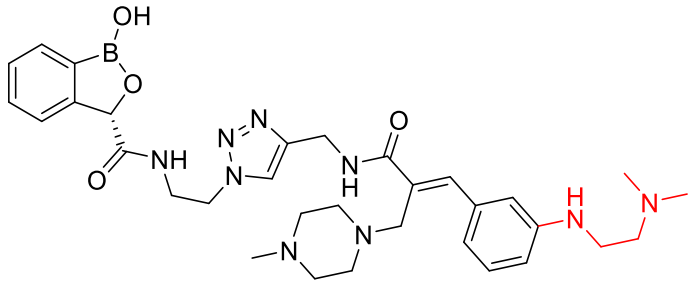
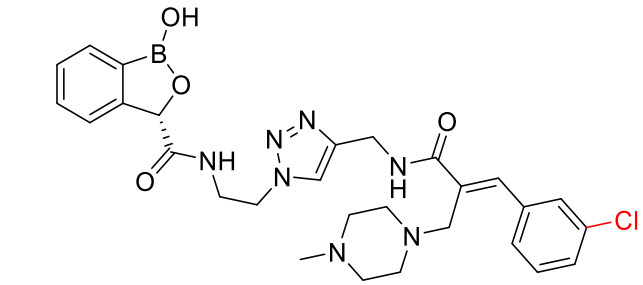
Structure	MM-GBSA ΔG_{Bind}
	-78.45
	-67.57

Table 101

Hits from single site enumeration at R_7 of LL3_6.

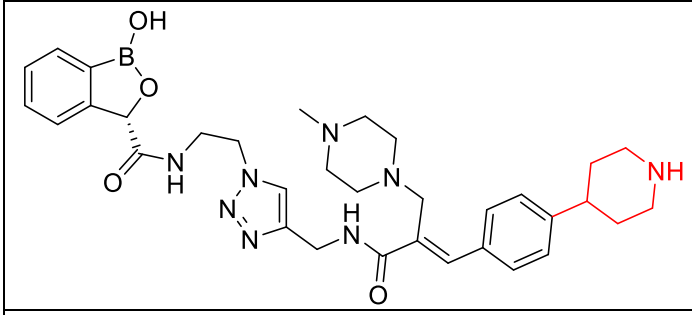
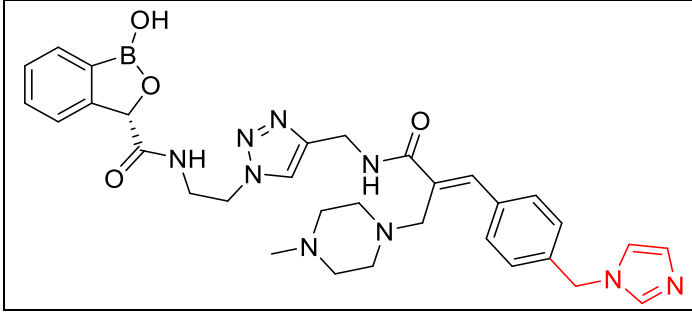
Structure	MM-GBSA ΔG_{Bind}
	-74.60
	-69.35

Table 102

Hits from enumeration at all modification sites of LL3_6.

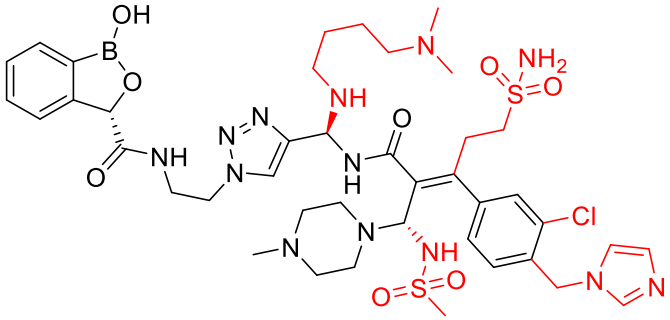
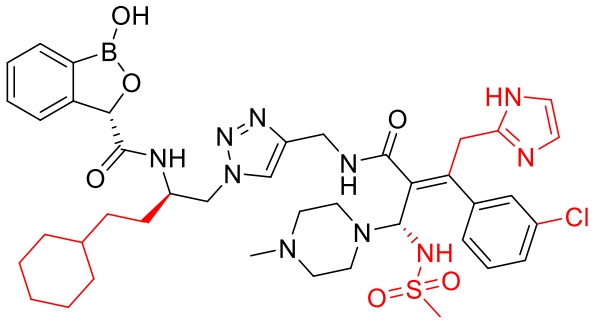
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-87.245	-9.053
	-82.812	-4.707

Table 102 (continued)

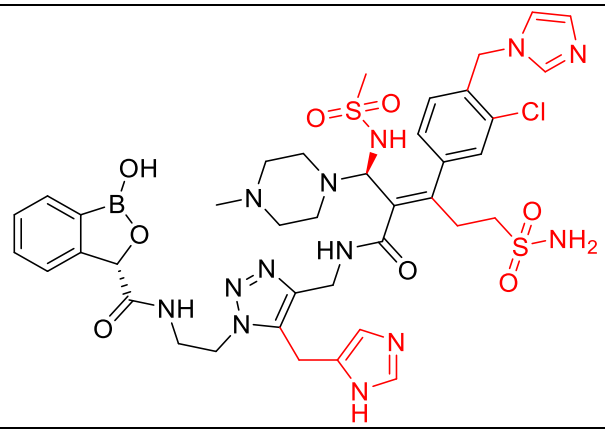
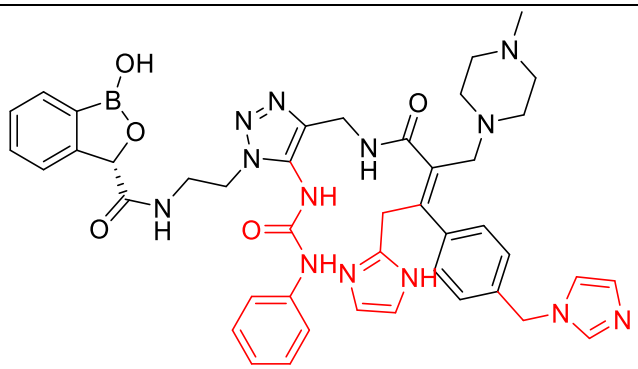
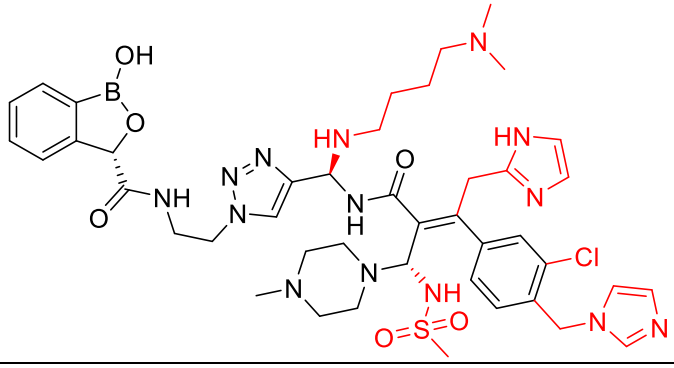
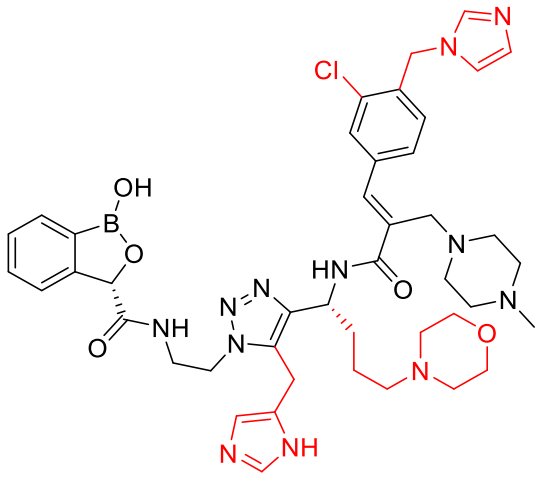
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-82.475	-6.543
	-82.203	-10.462
	-80.186	-11.144

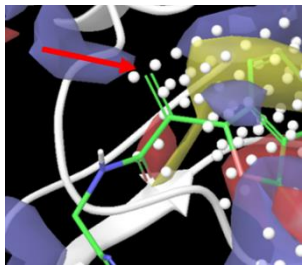
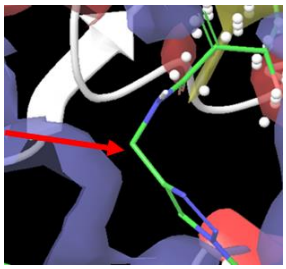
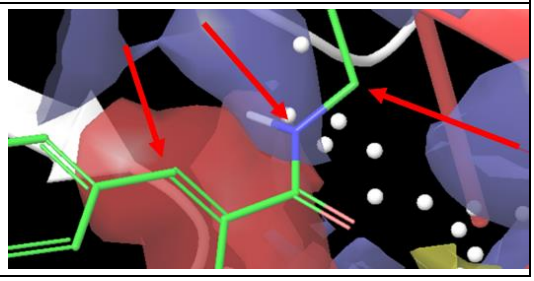
Table 102 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-77.346	-5.635

2.9.5.17 LL3_9.

Table 103

The surfaces generated by binding site analysis at the LeuRS-LL3_9 binding site and arrows point to modification site of R_1, R_2, R_3, R_4 and R_5.

R_1	R_2	R_3, R_4, and R_5
		

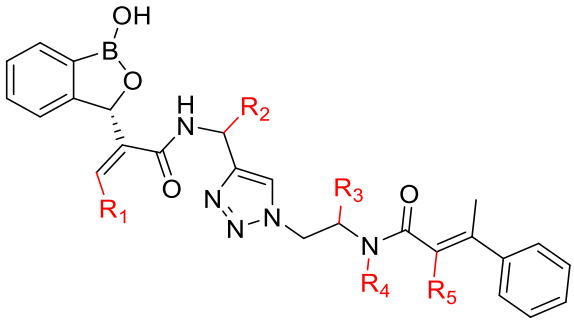


Figure 28. 2D structure of LL3_9 illustrating all the prospective modification sites.

Table 104

Hits from single site enumeration at R_1 of LL3_9.

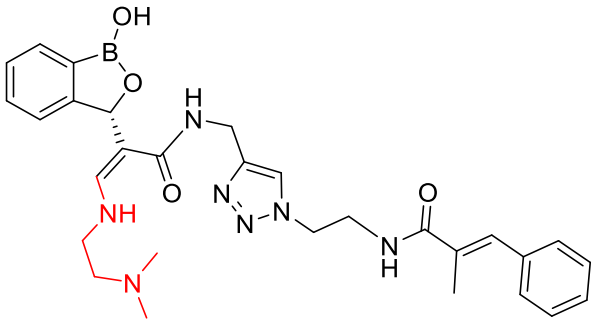
Structure	MM-GBSA ΔG_{Bind}
	-69.20

Table 104 (continued)

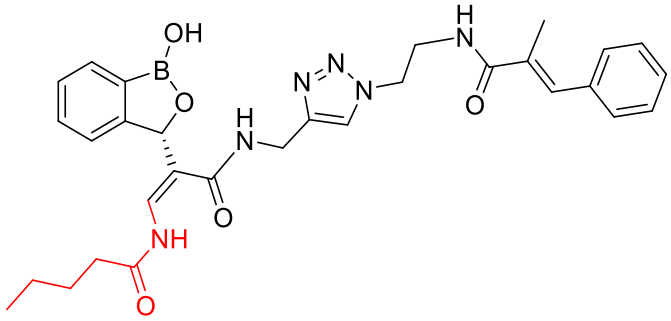
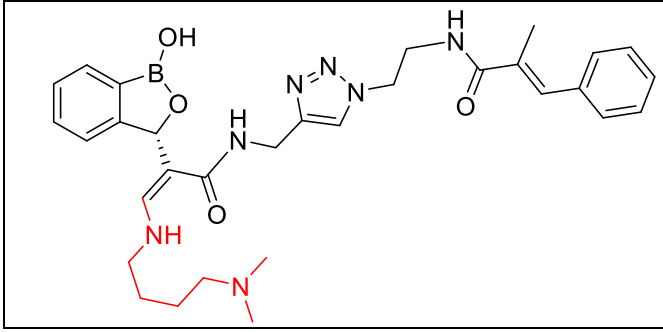
Structure	MM-GBSA ΔG_{Bind}
	-69.27
	-74.13

Table 105

Hits from single site enumeration at R_2 of LL3_9.

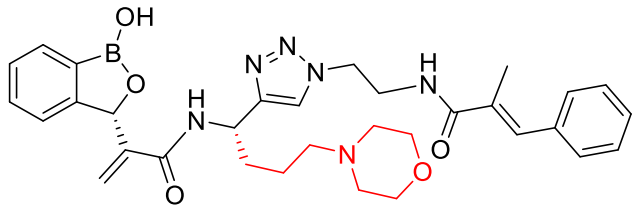
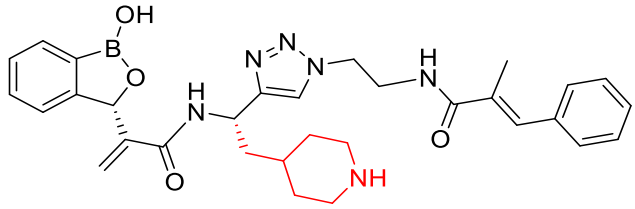
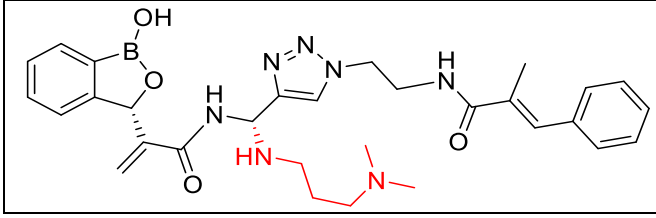
Structure	MM-GBSA ΔG_{Bind}
	-71.32
	-77.16
	-69.44

Table 105 (continued)

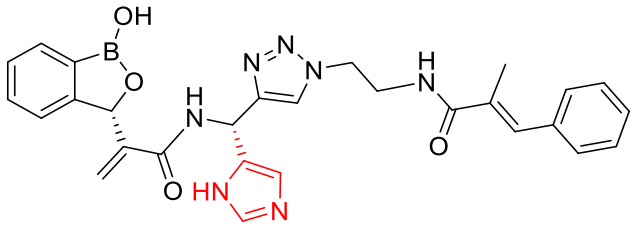
Structure	MM-GBSA ΔG_{Bind}
	-101.38

Table 106

Hits from single site enumeration at R_3 of LL3_9.

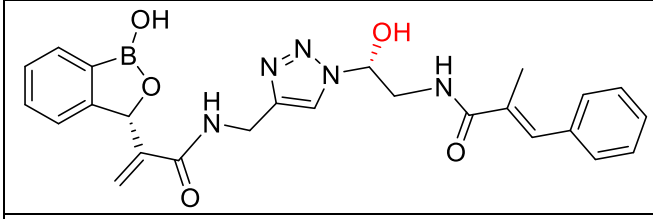
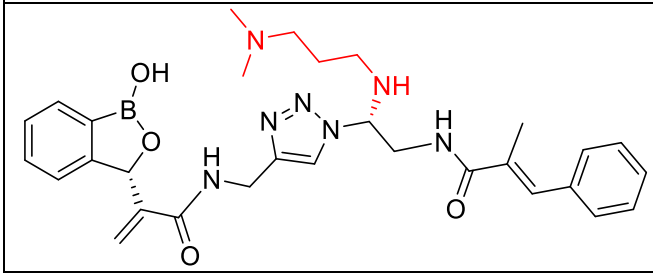
Structure	MM-GBSA ΔG_{Bind}
	-64.91
	-67.21

Table 107

Hits from single site enumeration at R_4 of LL3_9.

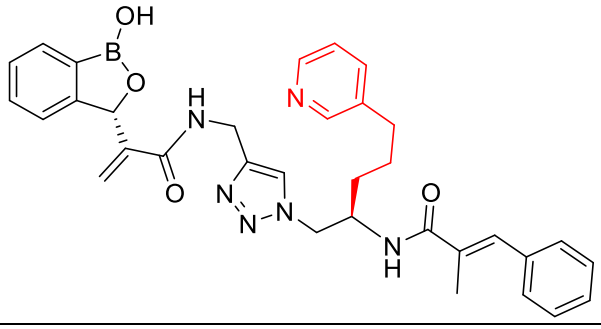
Structure	MM-GBSA ΔG_{Bind}
	-65.45

Table 107(continued)

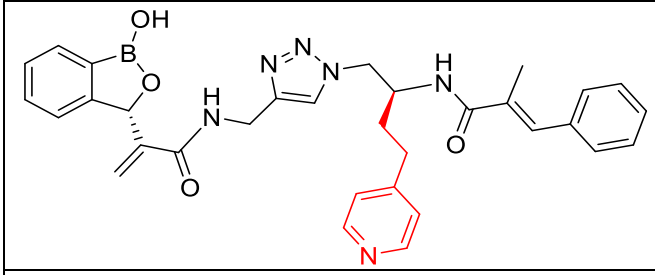
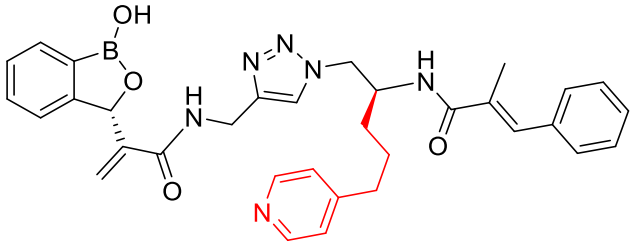
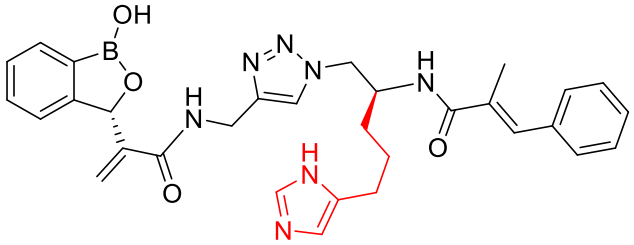
Structure	MM-GBSA ΔG_{Bind}
	-91.55
	-90.82
	-81.91

Table 108

Hits from single site enumeration at R_5 of LL3_9.

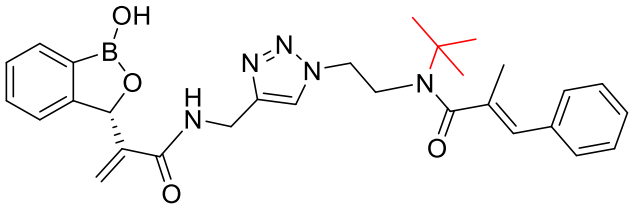
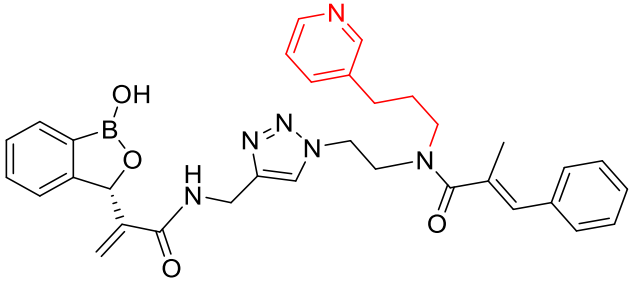
Structure	MM-GBSA ΔG_{Bind}
	-61.94
	-62.581

Table 108 (continued)

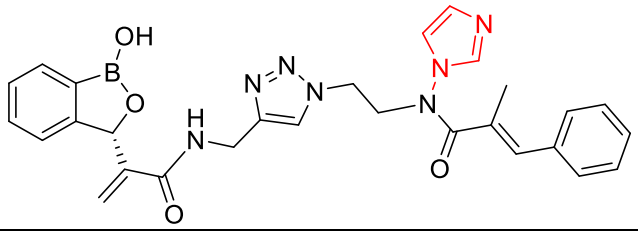
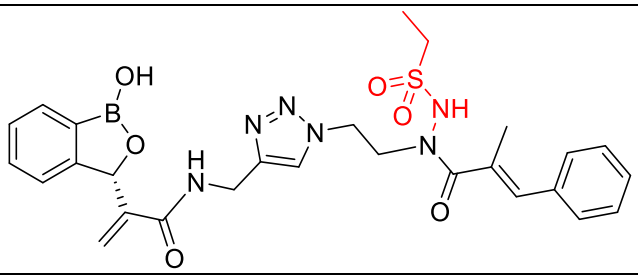
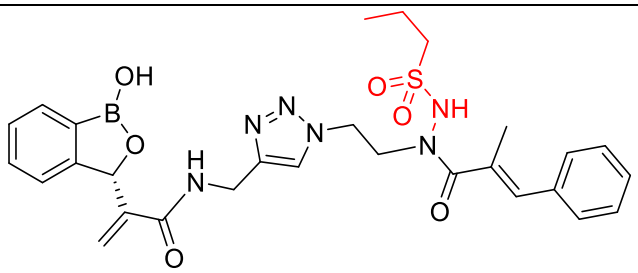
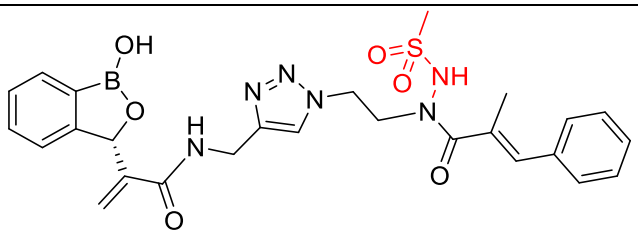
Structure	MM-GBSA ΔG_{Bind}
	-64.493
	-61.884
	-70.463
	-67.986

Table 109

Hits from enumeration at all modification sites of LL3_9.

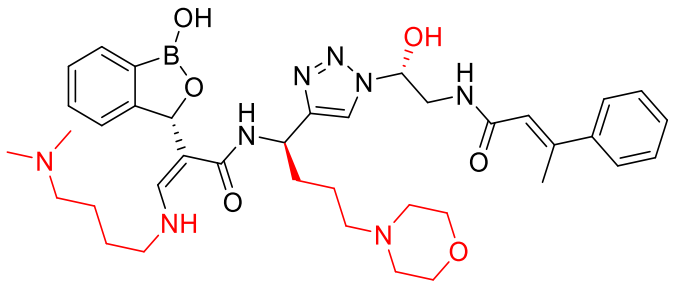
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-148.153	-10.79

Table 109 (continued)

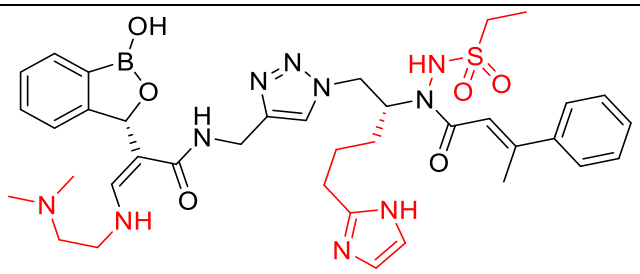
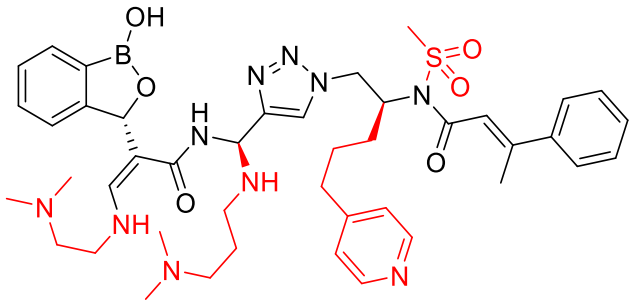
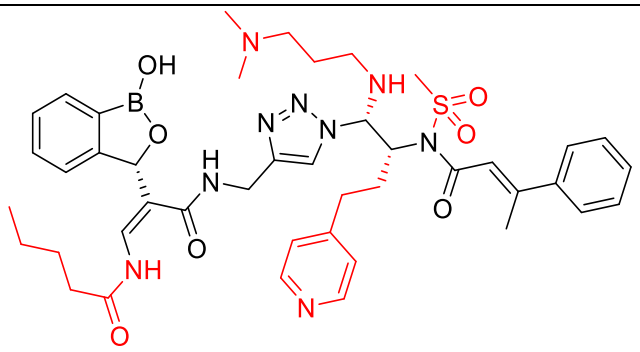
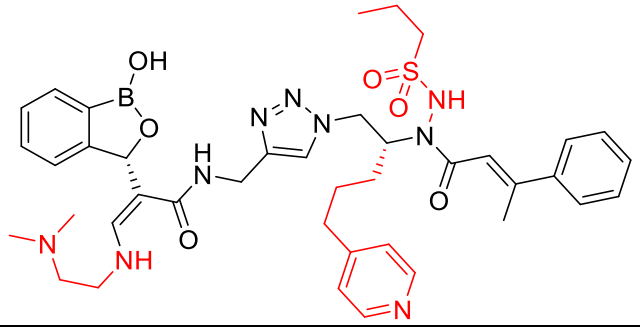
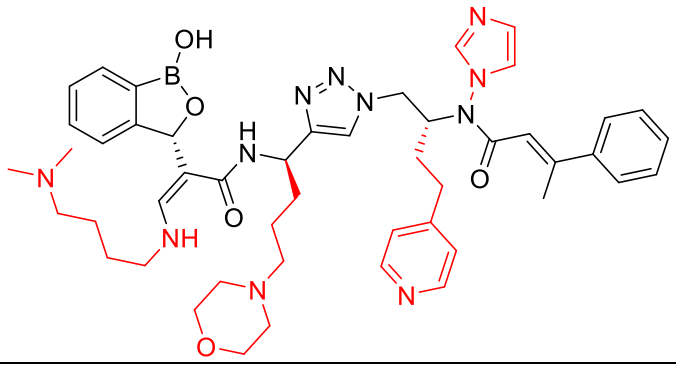
Structure	MM-GBSA ΔG_{Bind}	Docking score
	-114.819	-6.665
	-112.359	-7.908
	-111.764	-6.42
	-107.239	-4.223

Table 109 (continued)

Structure	MM-GBSA ΔG_{Bind}	Docking score
	-101.668	-7.933

Chapter 3

Probing the Binding Pathway of BRACO19 to a Parallel-Stranded Human Telomeric G-Quadruplex Using Molecular Dynamics Binding Simulation with AMBER DNA OL15 and Ligand GAFF2 Force Fields

3.1 Abstract

Human telomeric DNA G-quadruplex has been identified as a good therapeutic target in cancer treatment. G-quadruplex specific ligands that stabilize the G-quadruplex, have great potential to be developed as anticancer agents. Two crystal structures (an apo form of parallel stranded human telomeric G-quadruplex and its holo form in complex with BRACO19, a potent G-quadruple ligand) have been solved, yet the binding mechanism and pathway remains to be elusive. In this study, we simulated the binding of a free BRACO19 molecule to the apo form of the G-quadruplex using the latest AMBER DNA (OL15) and ligand GAFF2 force field. Three binding modes have been identified: top stacking, bottom intercalation and groove binding. Bottom intercalation (51% of the population) resembles the bottom binding pose in the complex crystal structure very well. The groove binding mode is less stable than the bottom binding mode, and is likely to be an intermediate state leading to bottom binding mode. A flip-insertion mechanism was observed in the bottom intercalation mode, during which the flipping out of the bases made space for ligand insertion, followed by bases flipping back to increase the stability of the complex. In addition to reproducing correct base-flipping behavior for some loop residues upon the ligand binding, the direct alignment type of ATAT-tetrad was observed in our simulations for the first time. These successes provide an initial support for using this force field combination of OL15 and GAFF2 force fields to study quadruplex/ligand interactions.

3.2 Introduction

In a guanidine-rich sequence, the formation of eight Hoogsteen H-bonds between four guanine bases instead of the typical Watson–Crick H-bonds observed in duplex DNA, leads to a square-planar configuration known as G-tetrad. And multiple G-tetrads further stack together to form a G-quadruplex. G-quadruplex can be formed by one, two or more strands of DNA or RNA and can fold into diverse topologies.(Burge, Parkinson, Hazel, Todd, & Neidle, 2006) The electron dense void generated by the oxygens of the adjacent guanidine bases are typically filled by a monovalent cation stabilizing the whole G-quadruplex structure. And as the K^+ and Na^+ are the pronounced cations, the G-quadruplexes with these cations are physiologically favored. Accounting to the better co-ordination of K^+ with eight oxygens of four guanidine bases, it is preferred over Na^+.(Burge et al., 2006; Collie, Sparapani, Parkinson, & Neidle, 2011) Computational tools have identified over 350,000 putative G-quadruplex sequences in the human genome, both the promoter regions of genes as well as within telomeres. (Huppert & Balasubramanian, 2005, 2007) Evidence supporting G-quadruplex formation in human cells has been reported in various studies.(Biffi, Tannahill, McCafferty, & Balasubramanian, 2013; Di Antonio, Rodriguez, & Balasubramanian, 2012; Hänsel et al., 2009; Hänsel et al., 2011b; Hänsel, Löhr, Trantirek, & Dötsch, 2013) In particularly, G-quadruplexes are over-represented specifically in areas of DNA damage in cancer cells and happen to appear more frequently in tumors than in normal tissues.(Cree & Kennedy, 2014; Duchler, 2012; Onel, Lin, & Yang, 2014; Shalaby et al., 2013) For that reason, G-quadruplexes are becoming important pharmacological targets for developing cancer therapeutics.(Balasubramanian, Hurley, &

Neidle, 2011; Cree & Kennedy, 2014; Duchler, 2012; Onel et al., 2014; Shalaby et al., 2013)

The first therapeutically important G-quadruplex formation was observed in the 3'-end overhang of human telomeric DNA(Doluca, Withers, & Filichev, 2013b). The telomeric overhang with a length of 100-200 nucleotides, containing repeats of the sequence d(TTAGGG), is capped by Shelterin complexes.(Chung et al., 2013b; de Lange, 2005b; Moyzis et al., 1988b; Wright, Tesmer, Huffman, Levene, & Shay, 1997a) After each cell replication, the telomere truncates by 50-200 base pairs and when the telomere is exhausted and Hayflick limit is reached, cell senescence and apoptosis are triggered(Harley, Futcher, & Greider, 1990b; Zakian, 1995b). In cancer cells, a reverse transcriptase called telomerase which is overexpressed in 80-85% of tumor cells, adds nucleotides to the telomere thus immortalizing the cells.(Greider & Blackburn, 1989a; Moorhouse et al., 2006b) It has been reported that the telomere cannot be hybridized by telomerase when the 3' overhang folds into a G-quadruplex (Zahler, Williamson, Cech, & Prescott, 1991b), leading to the inhibition of telomerase and thus cell apoptosis. In addition, the telomeric G-quadruplex adopted by the guanidine-rich 3' overhang prevents the binding of telomere protection proteins, which causes chromosomal fusions and stimulate cell apoptosis.(Denchi & de Lange, 2007b; Doluca et al., 2013b) Therefore, the G-quadruplex ligands that stablize the G-quadruplex are considered as promising anti-cancer agents and are under intensive development.(Hänsel et al., 2011b) Tricyclic aromatic chromophore based G-quadruplex binding molecules has been identified. The activity of these molecules was optimized by substituting side chains with amido-alkylamino character. A study by Read et al. on telomeric inhibitors makes the best case for BRACO19

reporting better proportion between IC50 of 10-13 µM against various ovarian tumor cell lines and telomeric inhibition EC50 of 95 nM compared to other molecules of the same class.

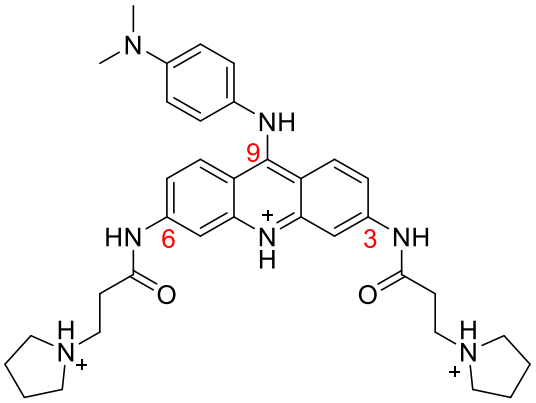


Figure 29. Chemical structure of BRACO19 (3^{+}).

BRACO19 (Figure 29), a computationally designed G-quadruplex ligand targeting the parallel-stranded G-quadruplex binding site(Yang & Okamoto, 2010b), inhibits telomerase, causes telomere shortening and also produces end-to-end chromosomal fusions in cancer cells. (Incles et al., 2004) It shows significant in-vivo anticancer activity in various tumor cell lines (Table 110)(Akagi & Kimoto, 1976; Alizadehnohi, Nabiuni, Nazari, Safaeinejad, & Irian, 2012; Brandes & Hermonat, 1983; Burger et al., 2005; Chen, Drabkowski, Hay, Macy, & Peterson Jr, 1987; Fang & Aust, 1997; Gunaratnam et al., 2007; Harrison et al., 2004; Kellner, Wierda, Shpall, Keating, & McNiece, 2016a; Landers, Cassel, & George, 1997; Mickey et al., 1977; Morimoto, Safrit, & Bonavida, 1991b; Nichols et al., 1977; Olopade et al., 1992b; Rankin, Faller, & Spanjaard, 2008a; G. T. Zhou et al., 2016).

Table 110

In vivo activity of BRACO19 against various cancer cell lines.

Cell lines	**Tissue type**	**IC_{50}**
MCF7	Breast cancer (human) (Brandes & Hermonat, 1983)	2.5 μM (Gunaratnam et al., 2007)
A549	Lung cancer (human) (Fang & Aust, 1997)	2.4 μM
DU145	Prostate cancer (human) (Mickey et al., 1977)	2.3 μM
HT-29	Colon cancer (human) (Chen et al., 1987)	2.7 μM
HGC-27	Gastric carcinoma (Akagi & Kimoto, 1976)	2.6 μM
A2780	Ovarian cancer (human) (Alizadehnohi et al., 2012)	2.5 μM
WI-38	Lung fibroblast (human) (Landers et al., 1997)	10.7 μM (Gunaratnam et al., 2007)
IMR90	Lung fibroblast (human) (Nichols et al., 1977)	>25 μM
U87	Glioblastoma (human) (Olopade et al., 1992a)	1.45 μM (G. T. Zhou et al., 2016)
U251	Glioblastoma (human)	1.55 μM
SHG-44	Glioma (human)	2.5 μM
UXF1138L	Uterus carcinoma (human)	2.5μM (Burger et al., 2005)
CH1	Lymphoma (mouse)	10.1μM (Harrison et al., 2004)
SKOV3	Ovarian cancer (human) (Morimoto, Safrit, & Bonavida, 1991a)	13.0μM
CLL	Chronic lymphocytic leukemia (Kellner, Wierda, Shpall, Keating, & McNiece, 2016b)	80μM (Rankin, Faller, & Spanjaard, 2008b)
AML	Acute myeloid leukemia	80μM (Rankin et al., 2008a)
--	Prolymphocytic leukemia	80μM (Rankin et al., 2008a)

In addition, BRACO19 also demonstrates broad anti-viral activity by stabilizing the G-quadruplexes found in pro-viral DNA.(Perrone et al., 2014) The crystal structure of a

parallel telomeric G-quadruplex d(TAGGGTTAGGGT)$_2$ with and without BRACO19 (PDB ID: 3CE5 and 1K8P, respectively) have been identified (Figure 30).

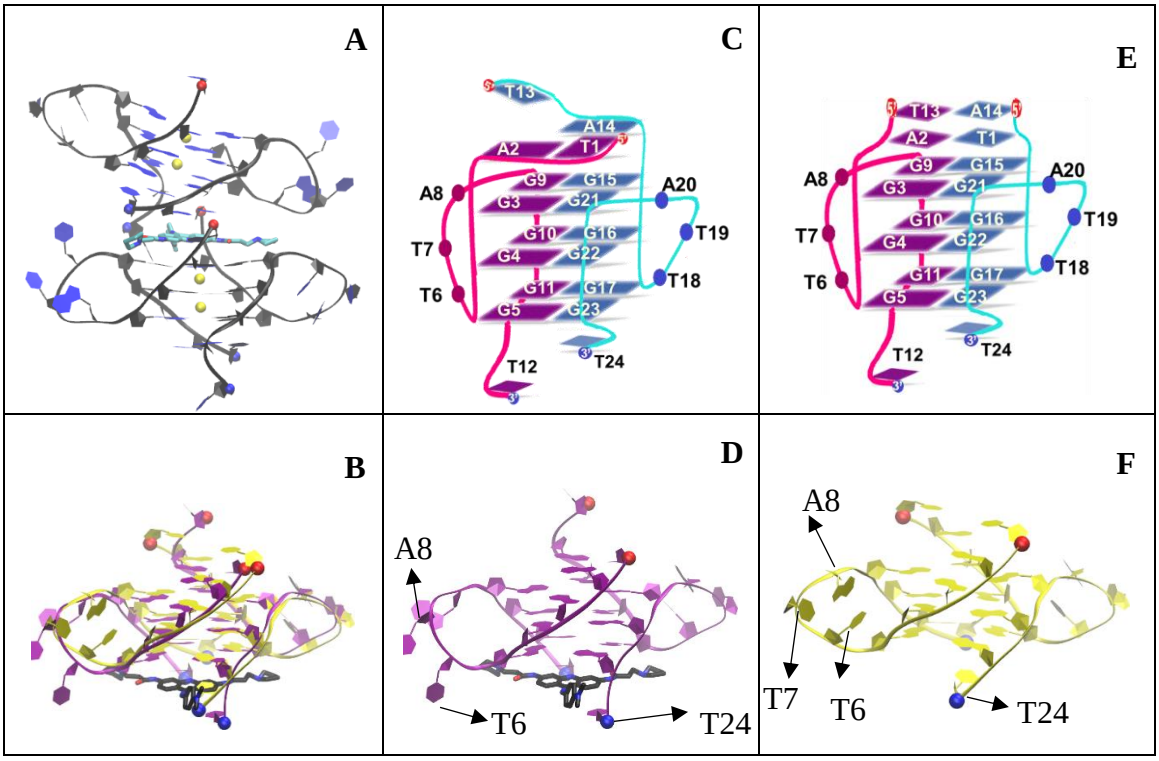


Figure 30. Comparison between apo and holo crystal structure of a parallel telomeric DNA G-quadruplex. (**A**) The holo structure of the telomeric DNA G-quadruplex in complex with BRACO19 (PDB: 3CE5). (**B**) Superimposition of the apo and holo form. (**C**) Cartoon representation of the holo form highlighting the four layers formed by DNA bases. (**D**) The holo structure of the telomeric DNA G-quadruplex in complex with BRACO19 at the bottom (PDB: 3CE5). (**E**) Cartoon representation of the apo form highlighting the five layers formed by DNA bases. (**F**) The apo form of the telomeric DNA G-quadruplex (PDB: 1K8P).

In the holo form, BRACO19 molecule binds at the interface of two parallel folded G-quadruplexes, sandwiched between a G-tetrad and a AT tetrad (Figure 31), where ATA is from the bottom G-quadruplex and T is from the top G-quadruplex.

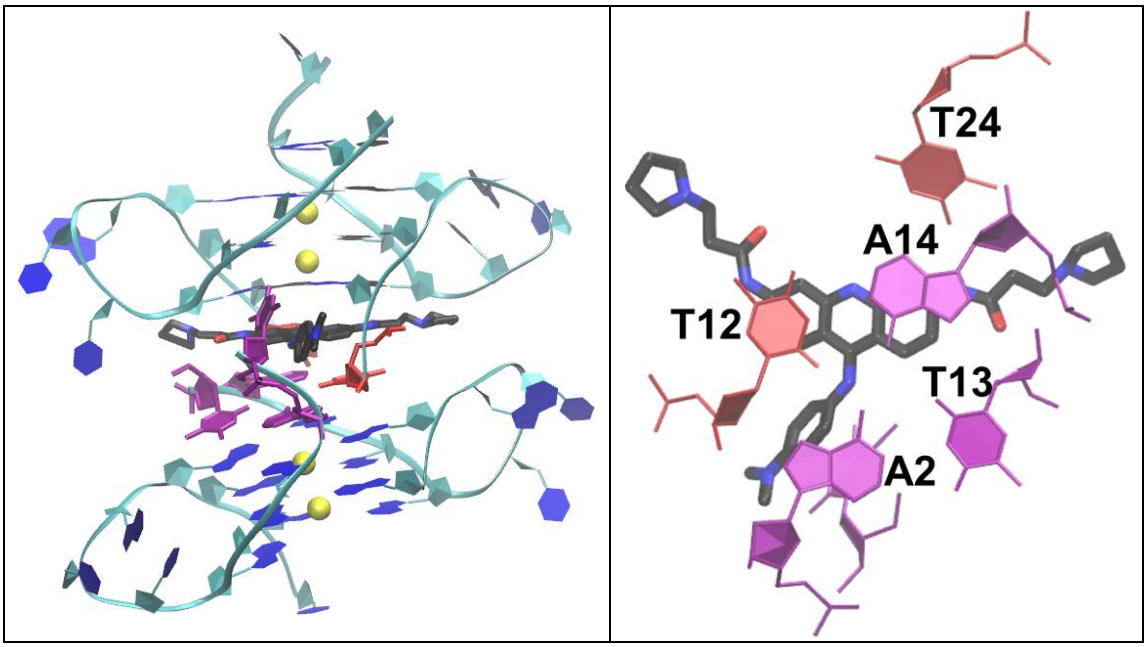


Figure 31. Quadruplex-ligand-quadruplex interface in crystal structure.

The comparison of the holo form with the apo form indicates that the binding of BRACO19 induces some local conformational changes in the G-quadruplex. First, the bases of T6, T7 and A8 flip out in the holo form with respect to the apo form, probably facilitating the insertion of the 3-pyrrolodino-propionamido branch of BRACO19. Second, the two adenine bases (residues 2 and 14) and the two thymine bases (residues 1 and 13) are paired in the apo form whereas in the holo form, the residues 1, 2 and 14 in are paired leaving the thymine residue, 13, unpaired. These local conformational adjustments clearly indicate that the intercalation of BRACO19 into the G-quadruplex follows an induced-fit binding mechanism rather than lock-key. Yet, the binding pathway and detailed mechanism remain elusive. The induced fit binding mechanism proposes that the initial weak binding interactions between ligand and receptor induce conformation changes in the receptor and ligand; and these changes in turn facilitate better binding affinity and specificity. Therefore understanding the binding pathway of BRACO19 to telomeric DNA G-quadruplex is

essential in designing more potent drugs. Furthermore, this intercalation mode with these subtle local conformational changes from the experimental structure also provides an excellent test to check whether molecular dynamics (MD) simulation based on the latest force fields is accurate enough to reproduce this binding mode.

MD stability simulations with various force fields have been widely used in studying G-quadruplexes in complex with BRACO19 and other ligands. Moore et al modelled 22mer parallel G-quadruplex with BRACO19 analogs; and conducted MD simulations to probe the qualitative structure-activity relationships (Moore et al., 2006a) using the AMBER parm99 force field.(Cornell et al., 1996; Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983b; J. M. Wang, Cieplak, & Kollman, 2000; J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, & D. A. Case, 2004) Hou et al conducted stability simulations using the AMBER parm99 force field(Duan et al., 2003; J. M. Wang et al., 2004) on G-quadruplex in complex with BRACO19 and 5 other ligands; and revealed that the H-bonds in the G-quadruplex to be major contributors for the stability of the G-quadruplex and ligand-quadruplex complex.(J. Q. Hou et al., 2010a) Dhamodharan et al docked bis-quinolinium and bis-pyridinium derivatives of 1,8-naphthyridine to an antiparallel G-quadruplex and consequently, conducted MD simulations; and reported that end-stacking was the favored binding mode.(Dhamodharan, Harikrishna, Jagadeeswaran, Halder, & Pradeepkumar, 2012b) Jain et al docked dimeric 1,3-phenylene-bis(piperazinyl benzimidazole)s to a 22mer parallel G-quadruplex, ran MD simulations and reported that both end-stacking and groove-binding were favored.(Jain, Paul, Maji, Muniyappa, & Bhattacharya, 2012a) Using docking and MD simulations based on AMBER parm99(Akhshi, Mosey, & Wu, 2012; Biffi, Tannahill, Miller, Howat, & Balasubramanian,

2014; Di Leva, Novellino, Cavalli, Parrinello, & Limongelli, 2014) force field, Ungvarsky et al successfully characterized the binding pose of a set of novel BRACO19 derivatives to the human telomeric parallel G-quadruplex.(Ungvarsky et al., 2014a) Zhou et al employed steered molecular dynamics and umbrella simulations using the charmm27 force field(MacKerell et al., 1998) to understand the ligand unbinding from human telomeric antiparallel G-quadruplex.(J. K. Zhou, Yang, & Sheu, 2015a) Recently, Diveshkumar et al identified indolyl, methylene-indanone scaffolds, by docking and conducting MD simulations using AMBER DNA parm99 force field with the updates of parmbsc0 and χOL4 refinement(Cheatham, Cieplak, & Kollman, 1999; Krepl et al., 2012a) on various G-quadruplexes (PDB IDs: 2L7V, 2O3M, 1KF1, 143D, and 2MB3), which specifically binds to parallel promoter G-quadruplexes rather than telomeric DNA G-quadruplex or duplex DNA(Diveshkumar et al., 2016a). However, these stability simulations don't provide detailed information on the binding pathway.

A recent AMBER DNA force field OL15 has been developed to include the corrections on several backbone torsional angle parameters (i.e. β_{OL1}(Zgarbová et al., 2015), $\varepsilon\zeta_{OL1}$(Zgarbová et al., 2013)and χ_{OL4}(Krepl et al., 2012a) to ff99bsc0(Galindo-Murillo et al., 2016)). These corrections are expected to improve the backbone sub-state description in G-quadruplexes and Z-DNA. So far, the tests on a DNA force field is mainly limited to long stability simulations on various DNA systems including B-DNA, Z-DNA, duplexes, triplexes, G-quadruplexes as well as unfolding simulations of DNA duplex.(Galindo-Murillo et al., 2016; Ivani et al., 2016; Sponer, Cang, & Cheatham, 2012; Zgarbová et al., 2013; Zgarbová et al., 2015) The performance of this DNA force field coupled with a recently updated AMBER GAFF2(J. M. Wang et al., 2004) ligand force

field on the binding simulations of ligand to G-quadruplexes is yet to be validated. During which the local unfolding and refolding of the bases particularly, base flipping is expected and thus provide a good test for the backbone torsional angle parameters of the improved DNA force field OL15. In this study, MD free binding simulations of unbound ligand-DNA complex were utilized to probe the binding pathway and mechanism of BRACO19 to the human telomeric parallel G-quadruplex DNA and stability simulations of two crystal binding poses were used to generate reference structures under physiological solution conditions and to understand the limitations of X-ray crystal packing. While the crystal bottom pose was stable, the crystal top pose altered significantly. In our free binding simulations, the three major binding modes were observed: top stacking, bottom intercalation and groove binding modes. The most abundant mode, the bottom intercalation mode, resembles the MD relaxed crystal pose well. Encouragingly, the local conformation adjustments were observed in the simulated structures. For the bottom intercalation mode, these conformational changes are consistent with the crystal pose in terms of the backbone torsion angles. These provide an initial evidence to the correctness of the torsion parameter corrections made to the OL15 force field. The dynamic and energetic properties of the three major binding modes were thoroughly studied, providing vivid examples of induced-fit binding mechanism.

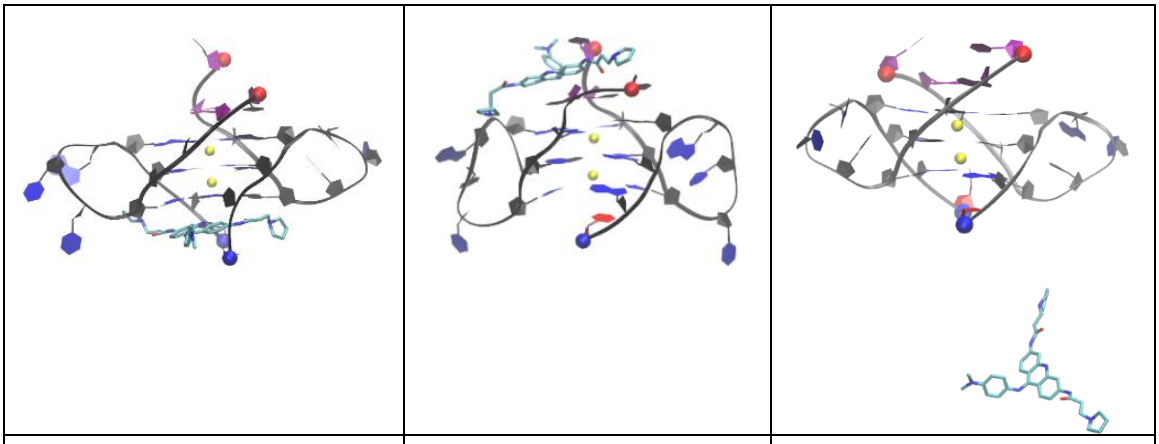
3.3 Methods

Table 111

Molecular dynamics simulations.

DNA	No. of Ligands	No. of Water molecules	K^+ ions	Box Size (Å)[1]*	Runs	Drug Initial Pose	NPT eq. (ns)	NVT (ns)
N/A	1	1491	3 Cl^-	41.5	1	N/A	1	499
G-Quad(1K8P)[2]	0	5141/4639	20/22	62.5/60.9	2x1	N/A	1	499
G-Quad(3CE5)	1	5114	17	62.5	1	top pose	1	999
G-Quad(3CE5)	1	5075	19	62.5	1	bottom pose	1	999
G-Quad(1K8P)	1	7621	19	70.0	10	Free	1	499

3.3.1 Simulation systems.



A **B** **C**

Figure 32. The initial configuration of the simulation system (DNA Quadruplex + Unbound BRACO19). 5' and 3' of the telomeric G-quadruplex DNA are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow.

[1] Triclinic box equivalent to the true truncated octahedral box
[2] 2 DNA structures: 1K8P top /1K8P bottom G-quadruplex with 2 missing residues (cf. Simulation system)

A DNA-ligand system was constructed, using the X-ray solved human telomeric DNA G-quadruplex (Figure 32C, PDB ID: 1K8P(Parkinson, Lee, & Neidle, 2002)) with an unbound BRACO19 10 Å away from the G-quadruplex, to simulate 10 simulation runs. This 10 Å distance was to ensure that there were at least three layers of water molecules separating ligand and DNA, and thus enabling the simulations to start from an unbound state to probe the binding pathways and mechanisms. The four non-standard brominated Uracil residues were replaced by standard Thymine residues in the original PDB structure. Two bound DNA-ligand systems, one characterizing bottom intercalation mode (Figure 32A, BRACO19 stacked below the G-tetrad formed by residues G5, G11, G17 and G23) and other characterizing top stacking mode (Figure 32B, BRACO19 stacked above the ATA formed by residues A2, T13 and A14) were constructed from the bound X-ray solved human telomeric DNA G-quadruplex (PDB ID: 3CE5) and simulated. Note that the crystal structure characterizing top stacking mode is missing two terminal residues, G23 and T24, in the second chain. A water box of a truncated octahedron, 10 Å water buffer and K^+ to act as counter ions, was used to solvate the unbound and bound systems. For the unbound system, the water buffer starts from the outmost atom of the ligand to the box surface. A TIP3P model (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983a) represented the water molecules, a K^+ model recently developed by Cheatham group and a refined version of the AMBER OL15 with the addition of corrections βOL1, εζOL1 and χOL4 to ff99bsc0 (Galindo-Murillo et al., 2016) represented the DNA G-quadruplex(Joung & Cheatham, 2008a). The standard AMBER protocol was used to generate the partial charges of BRACO19 with 3^+ charge at physiological pH=7; HF/6-31G* level was used to obtain the electrostatic potential after the geometrical optimization and the RESP (Restrained Electro-

Static Potential) method was then used to generate the partial charges(Bayly, Cieplak, Cornell, & Kollman, 1993b) and AMBER GAFF2(J. M. Wang et al., 2004) force field provided the other parameters.

3.3.2 Simulation protocols. The AMBER 16 package(Case et al., 2016) was used to conduct 12 simulation runs for the unbound DNA-ligand system, 1 simulation run for BRACO19 only and 1 simulation for G-Quadruplex only (Table 111). After the potential energy of the system was minimized, 14 independent simulation runs were conducted with different initial velocities, which were assigned based on random seeds. For the free binding system, an extra 500 ps pre-run at high temperature (500 K) was carried out to randomize the position and orientation of the free ligand(Lei, Wang, & Wu, 2012a), while the receptor was fixed. A short 1.0 ns molecular dynamics in the NPT ensemble mode (constant pressure and temperature) was used to relax the system density with cartesian restraints (1.0 kcal/mol/Å) on the ligand and the G-quadruplex and then 499.0 ns dynamics for unbound systems and 999.0 ns dynamics for bound systems in the equivalent NVT ensemble mode (constant volume and temperature) was run at 300 K. 2.0 fs time step was enabled in the simulations by applying SHAKE(Ryckaert, Ciccotti, & Berendsen, 1977a) to constrain all bonds connecting hydrogen atoms. The long-range electrostatic interactions were treated with the particle-mesh Ewald method(Essmann et al., 1995a) under periodic boundary conditions (charge grid spacing of ~1.0 Å, the fourth order of the B-spline charge interpolation; and direct sum tolerance of 10^{-5}). For the short-range non-bonded interactions, the cutoff distance was 10 Å; and the long-range van der Waals interactions were based on a uniform density approximation. A two-stage RESPA approach(Procacci & Berne, 1994a) was used in calculating the non-bonded forces, where the frequency of

updating short range forces was once per time step and long range forces was twice per time step . The Langevin thermostat with a coupling constant of 2.0 ps was used to control the temperature. The trajectories were saved at 100.0 ps intervals for analysis.

3.3.3 Convergence of simulations. The root mean square deviation (RMSD) of DNA backbone and ligand-DNA atom contacts using a cutoff of 3.0 Å were used to monitor the convergence of the simulations. Attaining a steady bound state and the stability of the bound state was indicated by the flat and small RMSD of 2.5 Å (Figure 60) and the stable contact number (Figure 61) in the last 200 ns. A complex with the number of atom contacts greater than 40 is defined as a stable complex. A good sampling of the binding sites is indicated by the free drugs are binding to different sites as shown in the last snapshots for the ten runs (Figure 33).

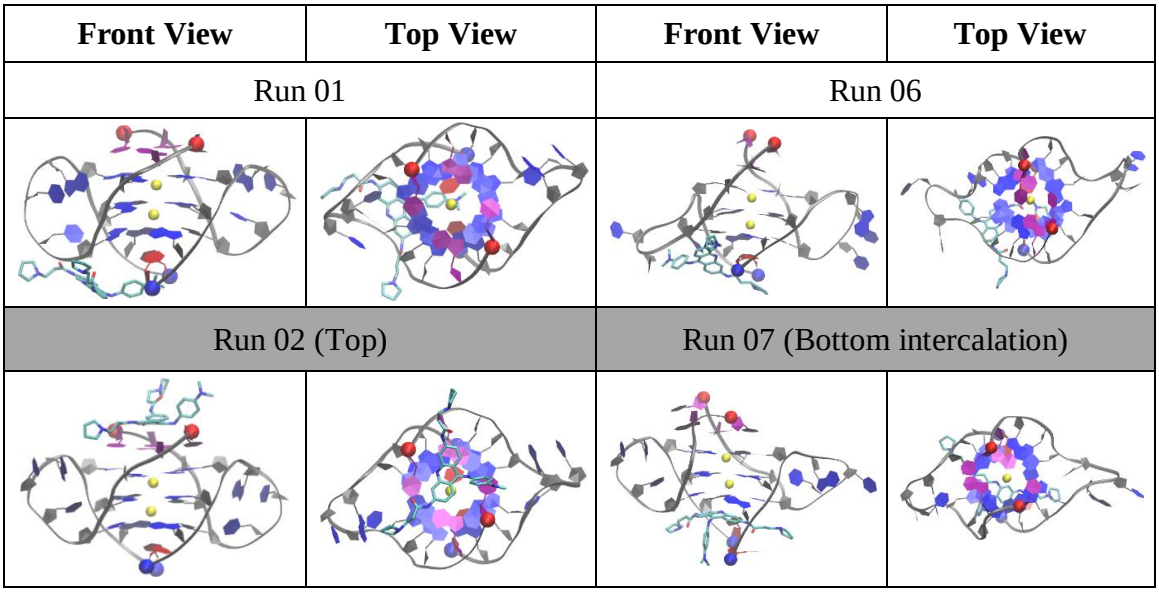


Figure 33. Last snapshots of the ten quadruplex-BRACO19 simulations. 5' and 3' of the telomeric G-quadruplex DNA are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow.

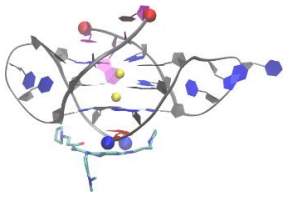
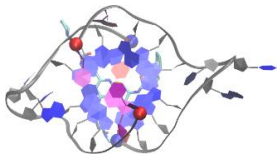
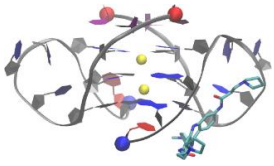
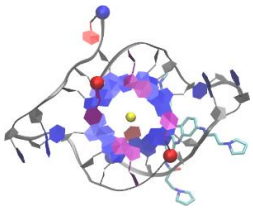
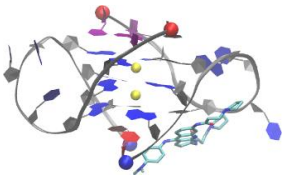
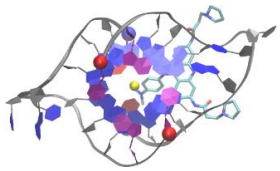
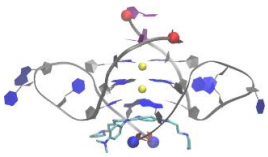
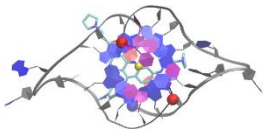
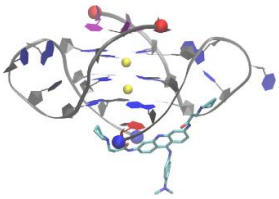
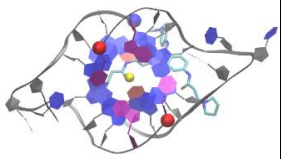
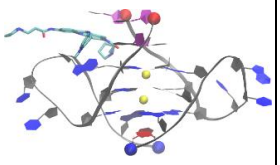
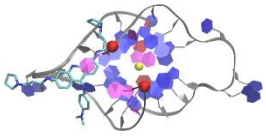
Front View	Top View	Front View	Top View
Run 03		Run 08	
			
Run 04		Run 09 (Bottom intercalation)	
			
Run 05		Run 10 (Groove)	
			

Figure 33 (continued)

3.3.4 Binding mode identification. DNA backbone of the stable complexes from the trajectories were aligned by least square fitting as the DNA backbone remained relatively stable. Daura algorithm(Daura et al., 1999a) was used to cluster the aligned complexes into different structural families based on the ligand's 2 Å pair-wise RMSD cutoff without fit. Every structural family is represented by the centroid structure. The centroid structure of the populated structural families (>1% of total structure population) are shown in Table 112. Based on visual inspection, these centroid structures were further merged into three major binding modes: top stacking, bottom stacking and groove binding modes.

Table 112

Representative structures of the most populated complex structure families (population ≥ 1 %) from the clustering analysis of the combined binding trajectories. 5' and 3' of the telomeric G-quadruplex DNA are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red.

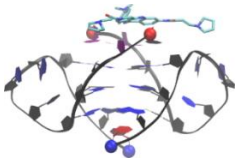
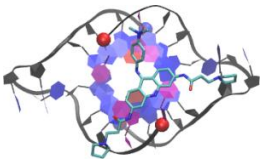
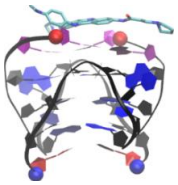
Binding model	Top Stacking
Cluster ID	A1
Representative Structure (Front View)	
Representative Structure (Top View)	
Representative Structure (Side View)	
Population	11.9%

Table 112 (continued)

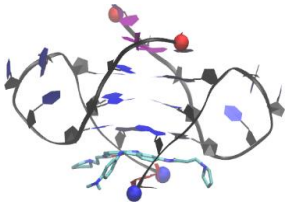
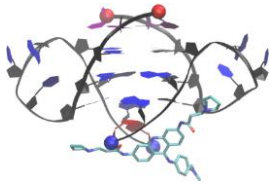
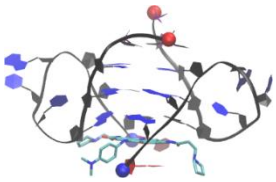
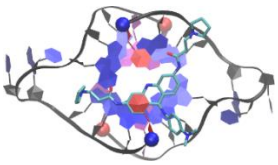
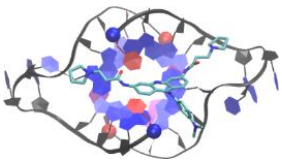
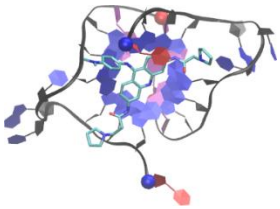
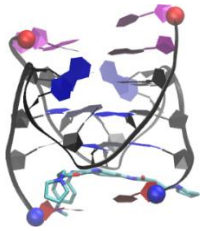
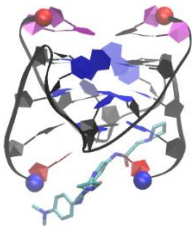
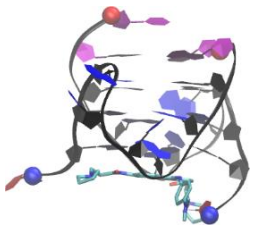
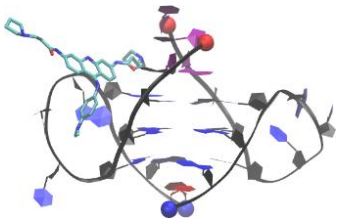
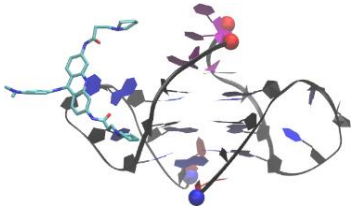
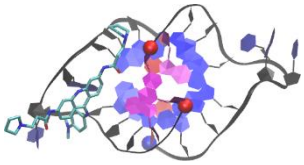
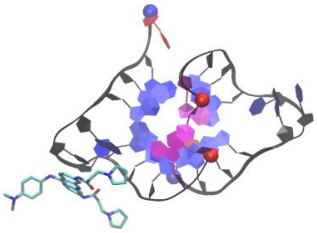
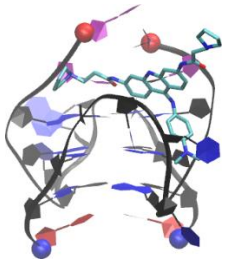
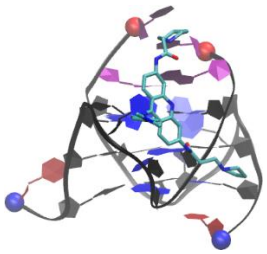
Binding model	End Intercalation (Bottom)		
Cluster ID	B1	B2	B3
Representative Structure (Front View)			
Representative Structure (Bottom View)			
Representative Structure (Side View)			
Population	44%	5.3%	1.8%

Table 112 (continued)

Binding model	Groove Binding	
Cluster ID	C1	C2
Representative Structure (Front View)		
Representative Structure (Lateral View)		
Representative Structure (Side View)		
Population	8.3%	1.3%

3.3.5 Parameters for characterizing DNA-drug binding pathway. The DNA-drug binding process was characterized by five order parameters: MM-PBSA binding energy (ΔE), center-to-center distance (Å), K^+-K^+ distance, ligand RMSD, drug-base dihedral angle, and hydrogen bond analysis. In order to avoid large energy fluctuation of explicit solvent, the energetics of the bound complexes were analyzed using MM-PBSA(Tan, Tan, & Luo, 2007) (Molecular Mechanics Poisson-Boltzmann/Surface Area) module in the AMBER package (PB1 model with mBondi radii set, salt concentration of 0.2 M, and surface tension of 0.0378 kcal/mol/Å^2 and offset of -0.5692 kcal/mol). And

MM-GBSA(Gregory D. Hawkins, Christopher J. Cramer, & Donald G. Truhlar, 1995; Hawkins, Cramer, & Truhlar, 1996; Tsui & Case, 2000) (Molecular Mechanics Generalized-Born/Surface Area) module in AMBER package (GB1 model with mBondi radii set, at various ionic strengths of 0.0 M, 0.05 M, 0.1 M, 0.15 M and 0.2 M, and surface tension of 0.00720 kcal/mol/$Å^2$) was also used to characterize the binding affinities of different binding modes. An evaluation study by Case et al. suggested that the GB1 model performs better than GB2-OBC1 and GB5-OBC2.(Gaillard & Case, 2011) To include the conformation energy change, the MM-PBSA/MM-GBSA binding energy for a system was calculated from three simulations(Kelly Mulholland, Siddiquei, & Wu, 2017): ligand only, DNA only and DNA-ligand complex. The center-center distance was defined as the distance between the center of the DNA G-quadruplex and BRACO19. The K^+-K^+ distance is defined as the distance between the K^+ ions in the G-quadruplex DNA. Receptor and ligand RMSD for bottom and groove binding mode were calculated with reference to the last snapshot of the simulation of bound crystal complex characterizing bottom intercalation mode (PDB ID: 3CE5); and receptor and ligand RMSD for top stacking was with reference to the last snapshot of the simulation of bound crystal complex (PDB ID: 3CE5) characterizing top stacking mode. The dihedral angle was defined as the angle between the plane of the top layer of the G-tetrads of the DNA G-quadruplex and the BRACO19's center, the plane of the acridine ring. The geometric definition of H-bonds: distance cutoff between the donor and acceptor was 3.5 Å and the angle cutoff between the same with -H was 120°. The H-bonds were defined and calculated for the five base layers over the course of the trajectories: the first/AT-tetrad (T1, A2, T13 and A14), the second/top layer of the G-tetrads (G3, G9, G15 and G16), the third/middle layer of the G-

tetrads (G4, G10, G16 and G22) the fourth/bottom layer of the G-tetrads and the fifth/T-dyad (T12 and T24) illustrated in (Figure 71). The standard backbone dihedral angles (α, β, γ, δ, ε and ζ) around the covalent bonds of the deoxyribose and χ about the glycosidic bond were defined to characterize the conformational changes. The defined dihedral angles are depicted in Figure 34.

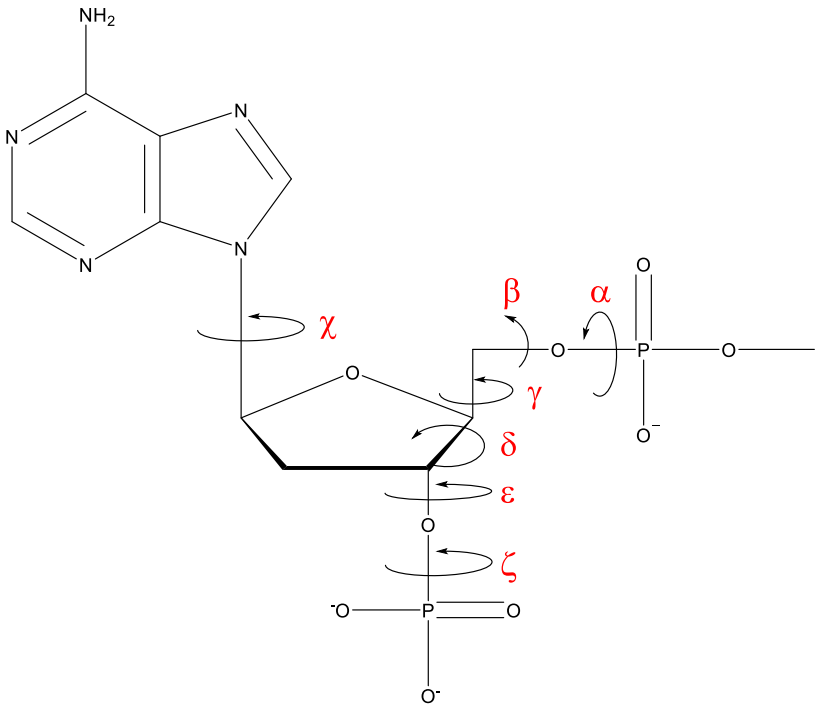


Figure 34. Backbone Torsion Angles of DNA.

3.4 Results

Two bound DNA-ligand systems, one characterizing bottom intercalation mode and other characterizing top stacking mode, were simulated to characterize the stability of the binding modes observed in the X-ray solved human telomeric DNA G-quadruplex (Figure 30, PDB ID: 3CE5).

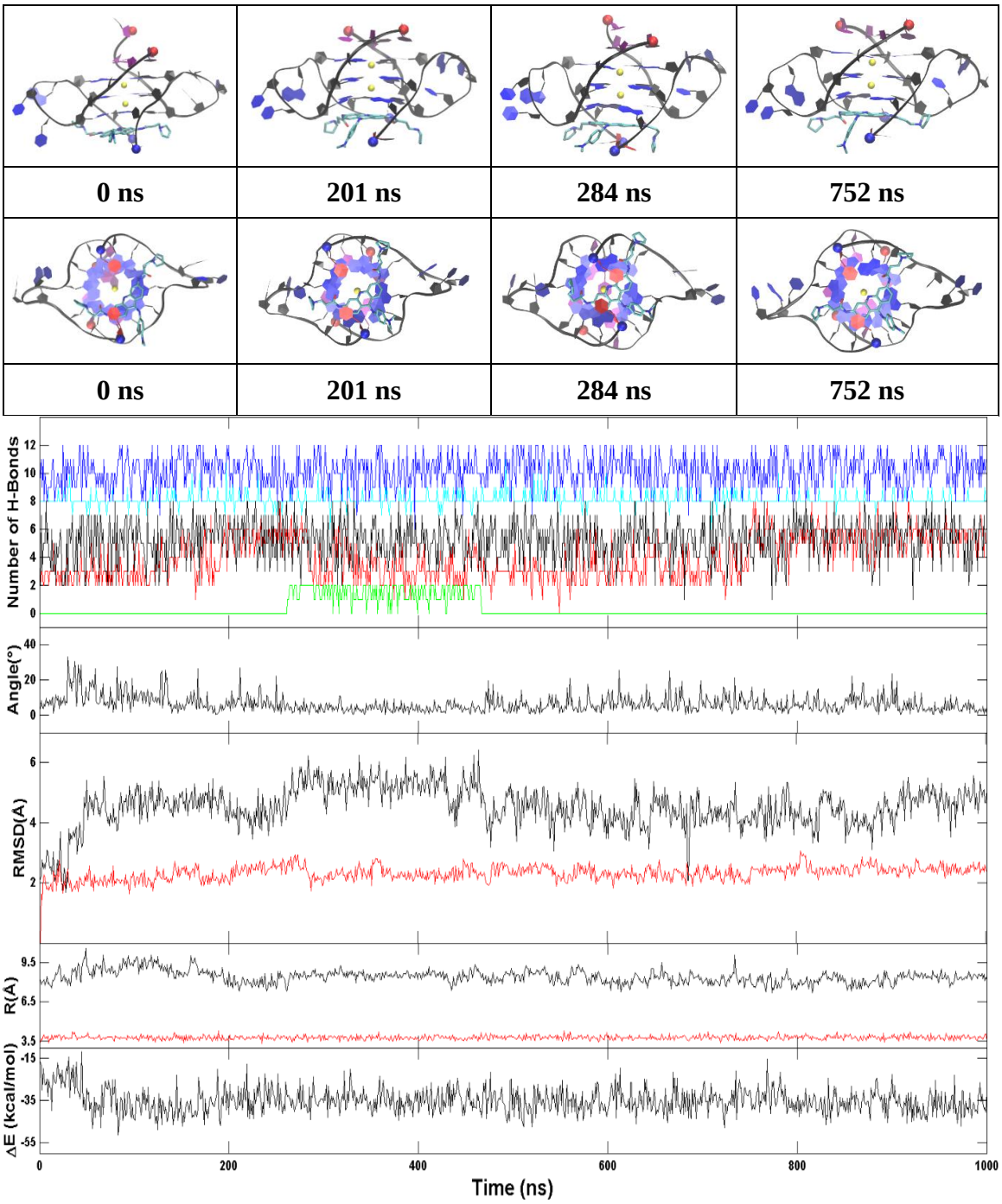


Figure 35. Stability simulation of the bottom binding mode in the crystal structure (PDB: 3CE5). **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 71), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-PBSA binding energy (ΔE) (cf. methods section for definition).

3.4.1 The crystal bottom intercalation mode of BRACO19 was stable. The simulation of DNA-ligand system characterizing bottom intercalation mode illustrated a stable binding mode (Figure 35) as indicated by the small receptor and ligand RMSD and drug-base dihedral angle in the order parameter plot. The planarity observed in the crystal structure (Figure 37A) is lost and conformational changes are observed in 3-pyrrolodino-propionamido branches of BRACO19. This loss of planarity is explained by the absence of stacking between two tetrads, G-tetrad and an ATAT-tetrad, as observed in the crystal structure (Figure 31). The ligand interactions of BRACO19 with the G-quadruplex in crystal structure (bottom pose) and MD relaxed crystal structure (bottom pose) are compared (Figure 70).

And the final MD relaxed crystal structure illustrated (i) the acridine ring stacked on residue, G5 (ii) the protonated N in the acridine ring formed a H-bond with T12, (iii) the 3-pyrrolodino-propionamido side chain at the 6^{th} position formed a H-bond with G17 and (iv) the other 3-pyrrolodino-propionamido side chain at the 3^{rd} position formed a H-bond with G11 (Figure 70 (B)) whereas in the crystal structure the acridine is stacked on residues G5 and G23 and the N from the amide of the 3-pyrrolodino-propionamido side chain on 3^{rd} position formed a H-bond with T12 (Figure 70 (A)). Formation of TT-dyad and ATAT-tetrad are also observed (Figure 71) and illustrated in the order parameter plot by increase in the number of H-bonds (Figure 35). This explains that its formation was blocked in the crystal structure due to crystal packing constraints.

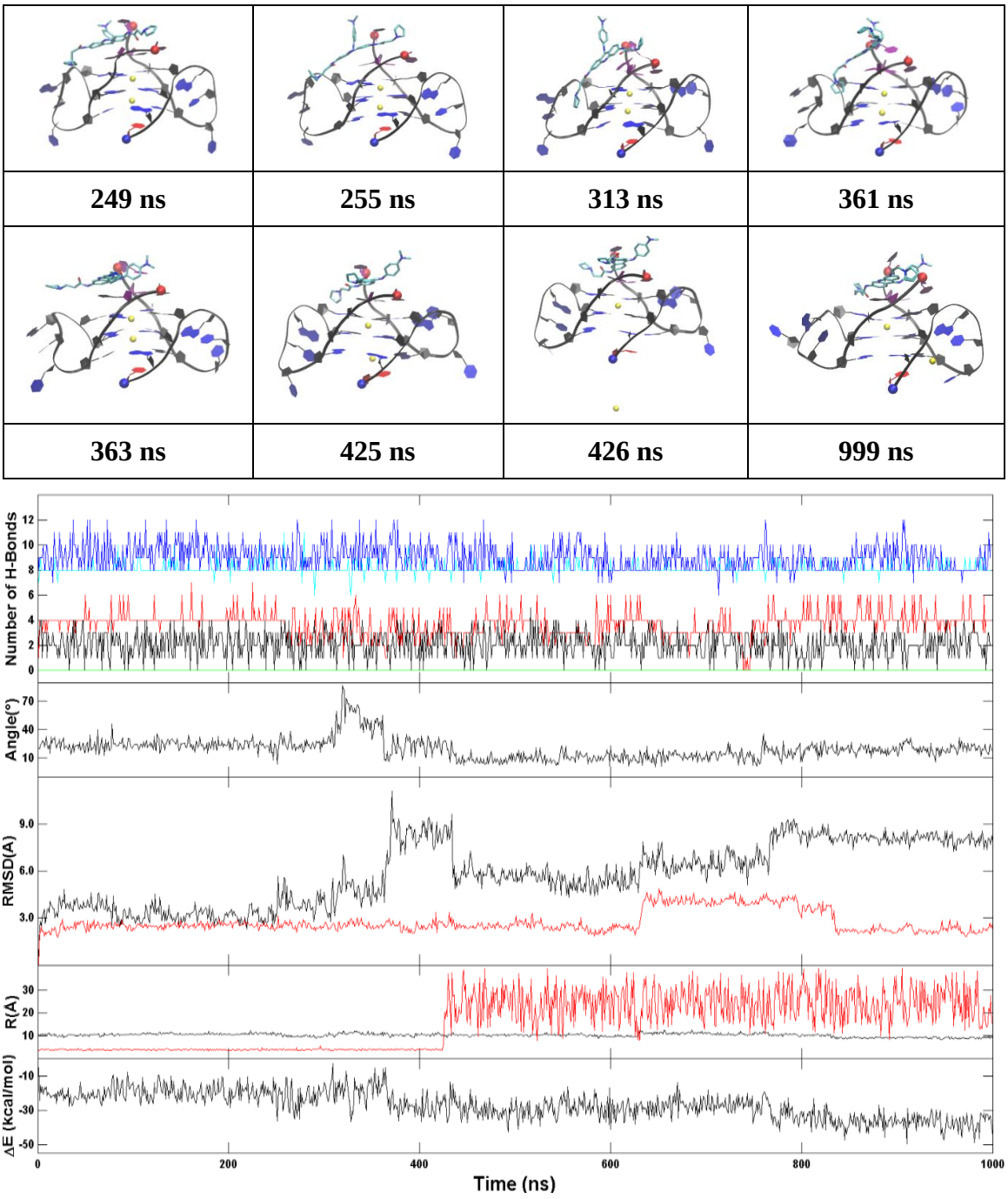


Figure 36. Stability simulation of the top binding mode in the crystal structure (PDB: 3CE5). **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 71), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-PBSA binding energy (ΔE) (cf. methods section for definition).

3.4.2 The crystal top mode of BRACO19 changed significantly, caused by lack of crystal packing constraints. The simulation of DNA-ligand system characterizing top stacking mode illustrated an unstable binding mode. The ligand moves away from the G-quadruplex and binds again although not in the same orientation as illustrated in Figure 36. This is demonstrated by the ligand RMSD and drug-base dihedral angle. Concluding that the top stacking mode is due to crystal packing constraints. The ligand interactions of BRACO19 with the G-quadruplex in crystal structure (top pose) and MD simulated crystal structure (top pose) are compared (Figure 70). And the final MD simulated crystal structure illustrated (i) acridine ring stacked on residue A14, in the 3-pyrrolodino-propionamido side chain at the 6^{th} position, (ii) the pyrrolodino ring formed a H-bond with A2 and (iii) the O from the amide formed H-bond with A14 and in other 3-pyrrolodino-propionamido side chain at the 3^{rd} position, (iv) the pyrrolodino ring formed a H-bond with A14 and (v) the protonated N in the acridine ring was also interacting with A14 (Figure 70 (E)) whereas in the crystal structure the acridine is stacked on A14 and the side chain at the 6^{th} position formed a salt bridge with G21 (Figure 70 (D)). The K^+ ion moves out of the G-quadruplex (Figure 36), this might be caused by the two missing terminal residues.

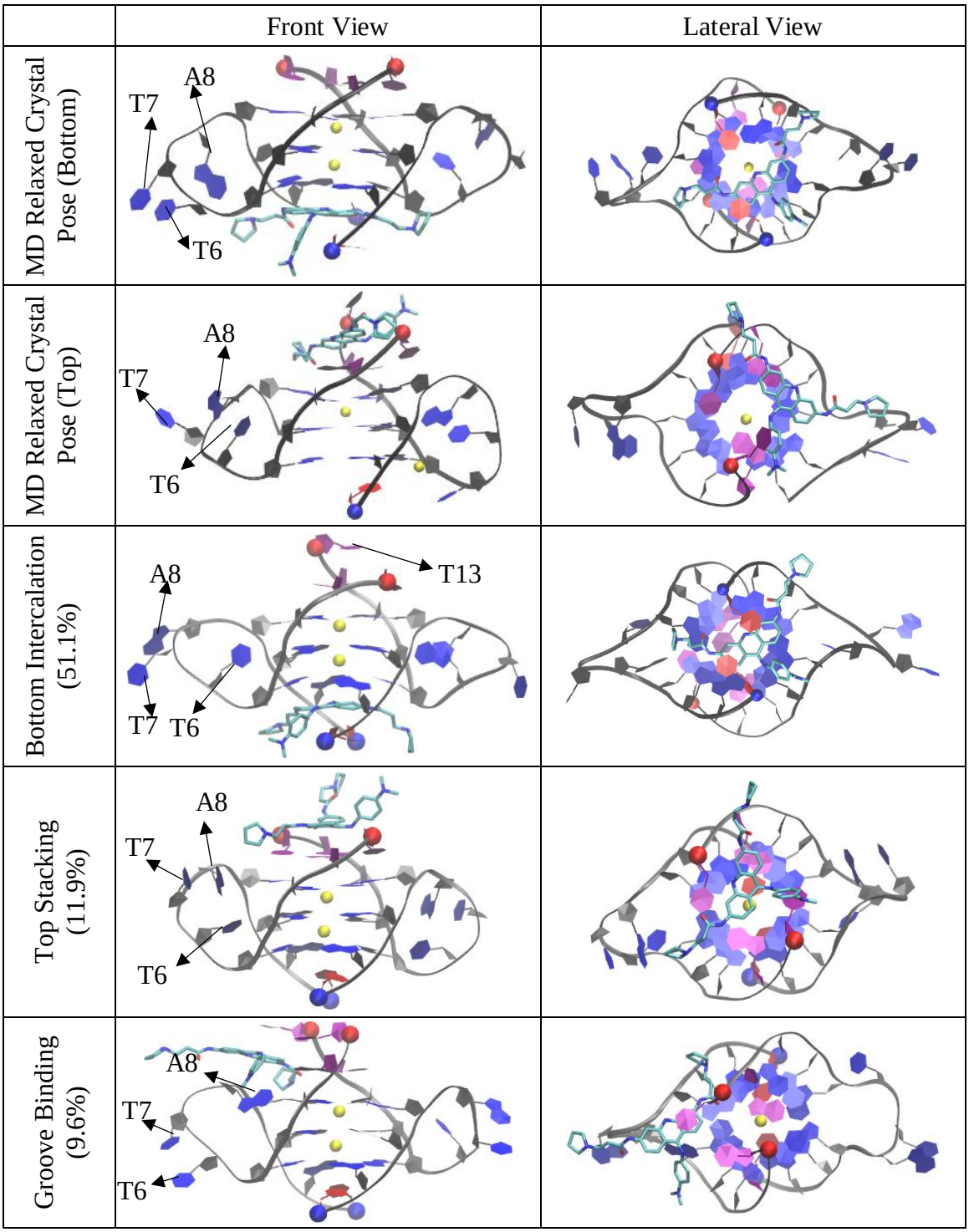


Figure 37. The MD relaxed crystal structure of **(A)** bottom intercalation mode and **(B)** top stacking mode (PDB ID: 3CE5) and the major binding modes, **(C)** bottom intercalation mode **(D)** top binding mode and **(E)** groove binding mode from the binding simulations. BRACO19 is shown in licorice; 5' and 3' of the telomeric G-quadruplex DNA are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red for lateral view and the K^+ ions are represented in yellow.

3.4.3 Three drug binding modes were observed in free binding DNA-ligand simulations. 10 simulation runs (500 ns of each) were executed to study the binding pathway of BRACO19 to the telomeric DNA G-quadruplex. The convergence of the binding simulations was confirmed (see the method section). Starting at an unbounded state, the ligand was observed to bind to the top and bottom of the G-quadruplex in seven runs (run 2, 3, 4, 5, 6, 7 and 9, Figure 33) and to the groove/side of the G-quadruplex in the remaining three runs (run 1, 8 and 10, Figure 33) at 500 ns. Six structural families with over 1% population were obtained from the clustering analysis (Table 112). Three binding modes were identified from these six structural families (Figure 37): top stacking (11.9 % of the total population), bottom intercalation (51.1%) and groove binding (9.6%). Encouragingly, the bottom pose is very similar to the ligand pose in the crystal complex structure (Figure 37A, 37C). In the bottom-intercalation mode, BRACO19 is sandwiched between the bottom G-tetrad and T-dyad. This pose is very consistent with the bottom intercalation pose of BRACO19 in the crystal structure (Figure 37A, 37C): a) the similar position and orientation of BRACO19; b), the intercalation between G-tetrad and T-dyad; c) flipping out of the bases A8 and T7. Subtle differences were also observed: a). a higher planarity of BRACO19 in the crystal structure; b) flipping out of T6 in the crystal structure; c). pairing between T12 and T24 in the MD structure.

The high planarity of BRACO19 in the crystal structure might be accounted by the fact that it is stacked between an ATAT-tetrad and a G-tetrad; and the paring between T12 and T24 is blocked by the formation ATAT-tetrad in the crystal structure (Figure 31). This ATAT-tetrad comprises a thymine residue from the top G-quadruplex and two adenine bases and one thymine base from the bottom G-quadruplex. Nonetheless, the ligand RMSD

between the MD simulated structure and the crystal structure for this pose is ~1 Å, indicating the high accuracy of the MD prediction. In the top binding pose, BRACO19 stacks on the top of the ATAT-tetrad, which is formed on the top of the first G-tetrad (Figure 37D). In addition, the pairing between T12 and T24 and the flipping in of the bases T6, T7 and A8 are observed in the top binding pose. In the groove binding mode, interaction of BRACO19 with the groove, pairing between T12 and T24 and flipping out of the bases T6, T7 and A8 are observed (Figure 37E). The groove binding pose appears to be the intermediate binding pose of the end binding mode based on our binding pathway (will be further discussed later). To characterize the conformational changes in the DNA backbone upon binding, the dihedral angles were calculated for each residue in every binding mode and compared with the simulated crystal poses. For the stability simulations of the X-ray poses and the free binding simulations, the residues in three G-tetrads show minimum fluctuations whereas the residues in terminal layers, ATAT-tetrad and T-dyad larger fluctuations were observed. Encouragingly, the dihedral angles of the residues in the representative structures of free binding simulations are consistent with most of the residues in the crystal structures of stability simulations. This data supports the torsional parameter corrections (βOL1, εζOL1 and χOL4) to the OL15 DNA force field.

3.4.4 The stable G-Quadruplex backbone scaffold is maintained as indicated by the flat and small RMSD of ~2.5 Å in all trajectories (Figure 60). For the five base layers in the G-quadruplex (Figure 71), while the middle three G4 are the stable ones (maintained in all the representative structures), the first and fifth layers are less stable and modulated by ligand binding. Among the three G-tetrad layers, the most stable G-tetrad is found to be the middle G-tetrad with ~10 H-bonds, followed by the top G-tetrad with ~8

H-bonds and the bottom G-tetrad with ~5 H-bonds in all trajectories. The detailed account of H-bonds maintained in the representative structure of different binding poses with respect to the experimentally solved X-ray crystal structure is illustrated in Figure 71.

The binding pathway of the three binding modes was characterized by calculating the five order parameters as described in the methods section. A representative trajectory for each mode is shown in Figure 38, 40 and 42.

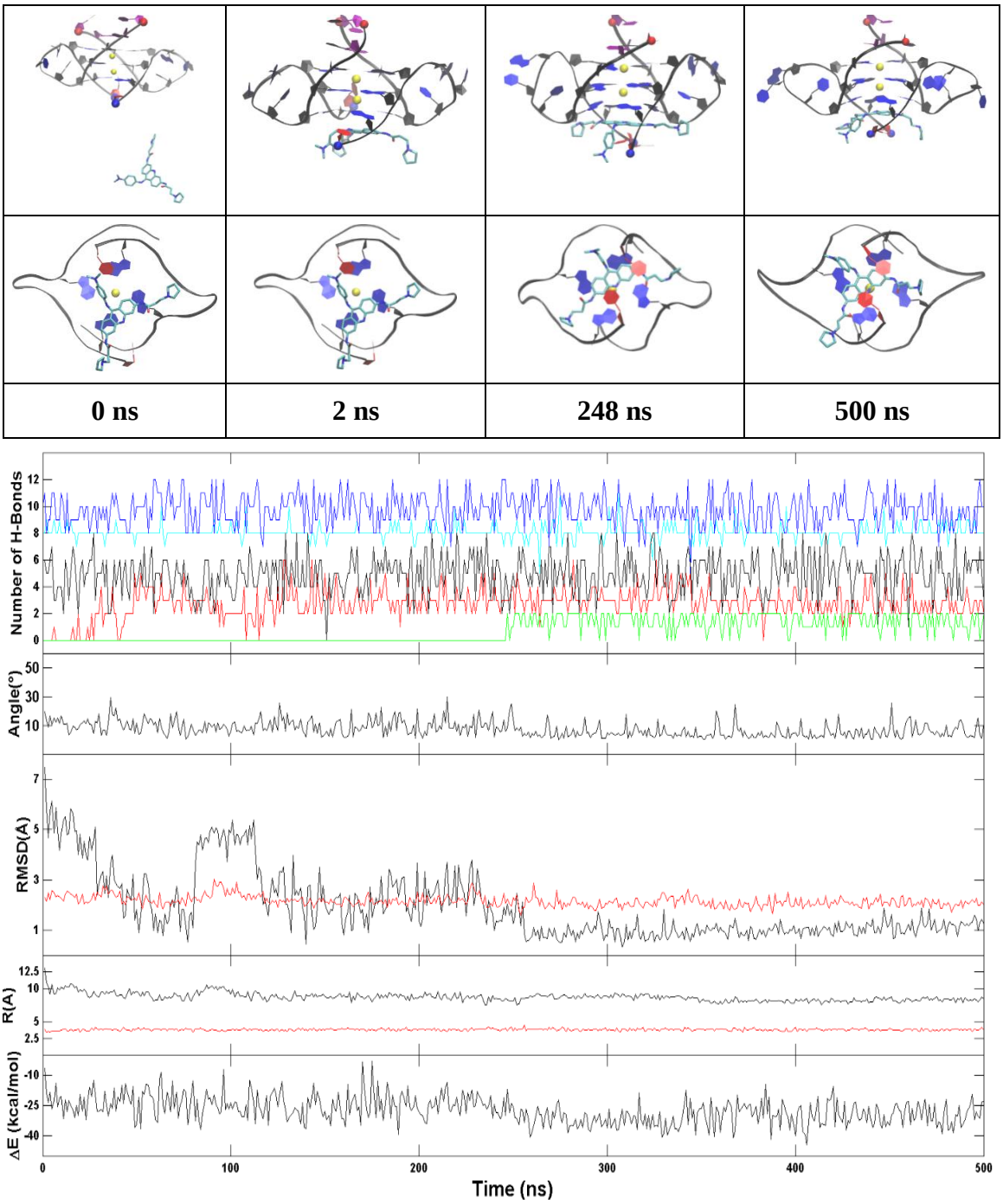


Figure 38. A representative trajectory of the bottom intercalation mode. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 71), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-PBSA binding energy (ΔE) (cf. methods section for definition).

3.4.5 Bottom intercalation of BRACO19 follows the flip-insertion mechanism.

In the representative trajectory of the bottom intercalation mode (Figure 38), the first interaction between the G-quadruplex and BRACO19 occurred at about 2 ns from the starting unbound state. Interestingly, successful intercalation of BRACO19 was observed at 248 ns as the bases of T12 and T24 paired over BRACO19 and the bases, A8 and T7 flipped out. It was noted that the initial contact with the G-quadruplex is concurrent with flipping out of the residues, T12 and T24 making space for BRACO19 to interact with G-tetrad. Once BRACO19 stacked below the bottom G-tetrad, the bases of T12 and T24 flipped back and paired to form two hydrogen bonds (Figure 71). As a result, the ligand is sandwiched between the bottom G-tetrad and the T-dyad of the G-quadruplex, leading to a stable intercalation mode with stable order parameters (ligand RMSD of ~1 Å and MM-PBSA binding energy of ~-30 kcal/mol). Clearly, the drug binding facilitates the formation of T-dyad (Figure 38). We name this intercalation as "flip-insertion" mechanism. This mechanism is also observed in the other representative trajectory of bottom intercalation mode. The ligand interactions of BRACO19 with the G-quadruplex in MD relaxed crystal structure (bottom pose) and the representative structure of bottom intercalating trajectory from the free binding simulation are compared (Figure 70). And the final MD simulated crystal structure illustrated (i) the acridine ring stacked on residue G5, (ii) the protonated N in the acridine ring formed a H-bond with T12, (iii) the 3-pyrrolodino-propionamido side chain at the 6th position formed a H-bond with G17 and (iv) the other 3-pyrrolodino-propionamido side chain at the 3rd position formed a H-bond with G11 (Figure 70 (B)) whereas in the representative structure of bottom intercalating trajectory from the free binding simulation (i) the acridine is stacked on G23, (ii) one pyrrolidine ring formed H-

bond with T18, (iii) other formed H-bond with T12 and G11 and (iv) the side chain on 9th position is stacked on G5 (Figure 70 (C)).

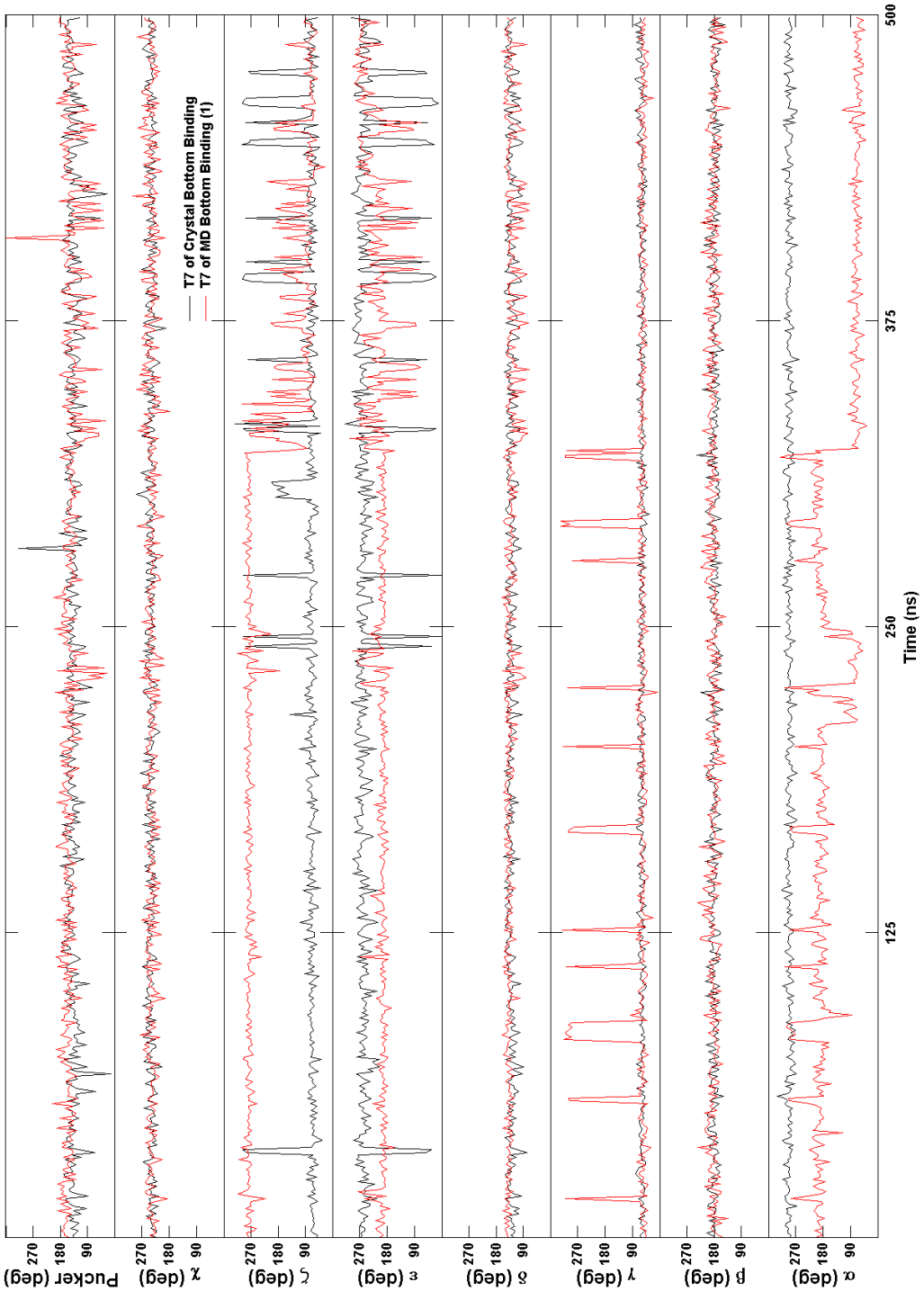


Figure 39. Comparison for the backbone torsion angles of residue T7 between the free ligand binding simulation (red) and the stability simulation of the crystal pose (black) for the bottom intercalation mode.

Another notable feature in this trajectory is the flipping out of the bases A8 and T7 after the intercalation of BRACO19. This flipping out of the bases is also observed in the experimental X-ray solved crystal structure of G-quadruplex in complex with BRACO19, PDB ID: 3CE5 (Figure 30). These conformational changes are characterized by calculating the backbone dihedral angles of each base and comparing with its corresponding base in the stability simulation of the crystal bottom mode. Figure 39 features the dihedral angles of residue T7 in both representative trajectories of free binding and stability simulations characterizing the bottom intercalation mode. Starting at the different values, the torsional angles ε and ζ of residues from MD simulated crystal structure and free binding simulations converged. The changes in the torsional angles are concurrent with the flipping out of residue T7 from 323 ns through the rest of the trajectory with fluctuations. Therefore, correct torsional parameters appear to be critical for reproducing base flipping shown in the crystal structures.

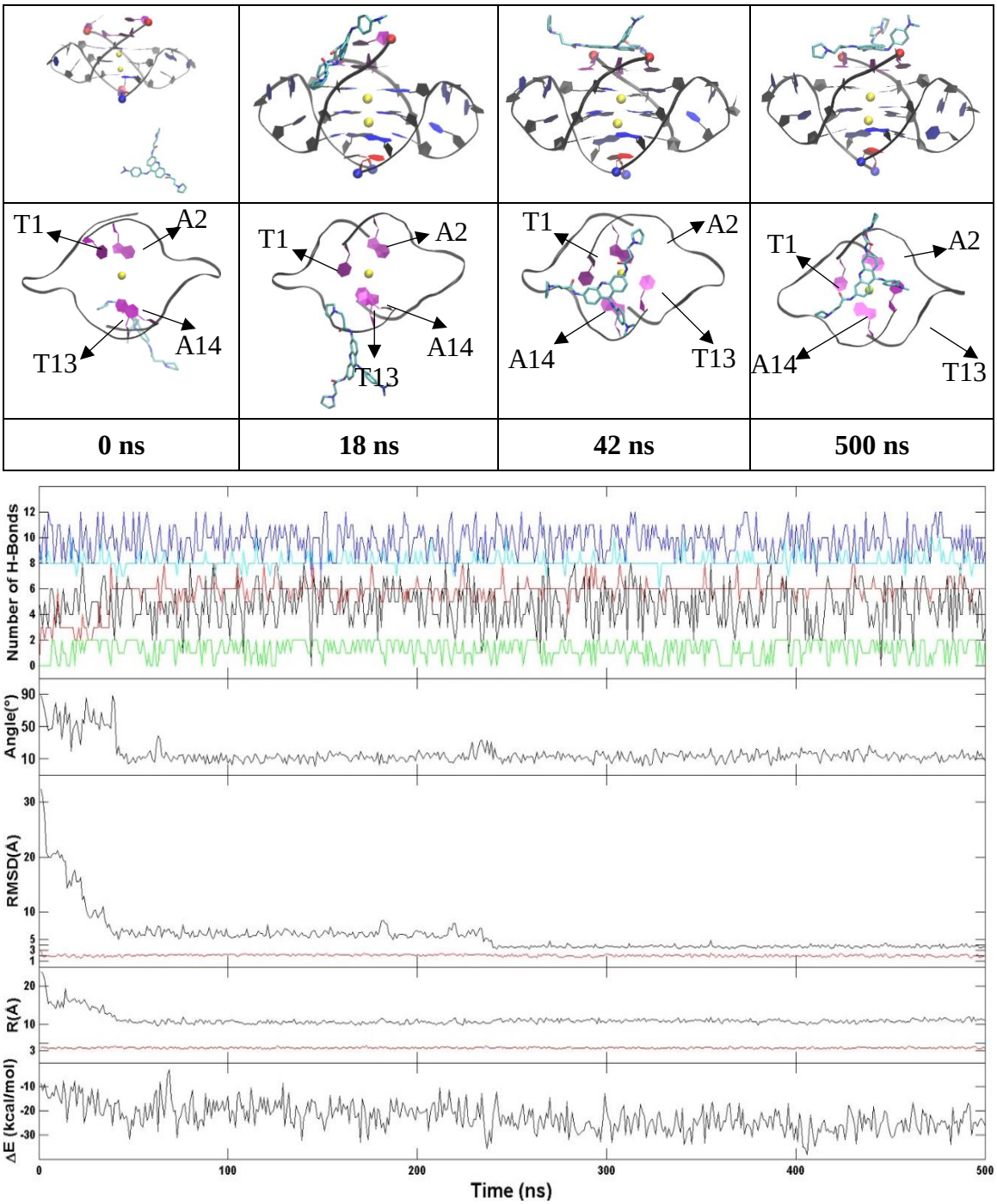


Figure 40. A representative trajectory of the top stacking mode. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K+ ions are represented in yellow. **Bottom:** A representative trajectory of the top stacking mode. An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 71), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-PBSA binding energy (ΔE) (cf. methods section for definition).

3.4.6 Top stacking of BRACO19 leads to the formation of ATAT-tetrad in a trajectory. The representative trajectory for the top stacking mode is shown in Figure 40 with the 500-ns order parameter plot. In this trajectory, the first interaction of BRACO19 with the top of the G-quadruplex occurred at about 18 ns from the starting unbound state. From about 18-42 ns of the trajectory, BRACO19 flipped orientation and the residues T1, A2, T13, A14 formed an ATAT-tetrad and remained stable through the rest of the simulation (ligand RMSD of ~3.5 Å and MM-PBSA binding energy of ~-25 kcal/mol). The formation of ATAT-tetrad(Zhang et al., 2001) by residues T1, A2, T13, A14 in the first layer is indicated by the fact that the number of H-bonds increases to ~6 from initial ~3 in the first layer at 42 ns.

The ligand interactions of BRACO19 with the G-quadruplex in MD simulated crystal structure (top pose) and the representative structure of top stacking trajectory of free binding simulation are compared (Figure 70). And the final MD simulated crystal structure illustrated (i) acridine ring stacked on residue A14, in the 3-pyrrolodino-propionamido side chain at the 6th position, (ii) the pyrrolodino ring formed a H-bond with A2 and (iii) the O from the amide formed H-bond with A14 and in other 3-pyrrolodino-propionamido side chain at the 3rd position, (iv) the pyrrolodino ring formed a H-bond with A14 and (v) the protonated N in the acridine ring was also interacting with A14 (Figure 70 (E)) whereas in the representative structure of top stacking trajectory of free binding simulation the N in the acridine formed a H-bond with T13 and (ii) N from the amide of side chain on 6th position formed a H-bond with A14 (Figure 70 (G)).

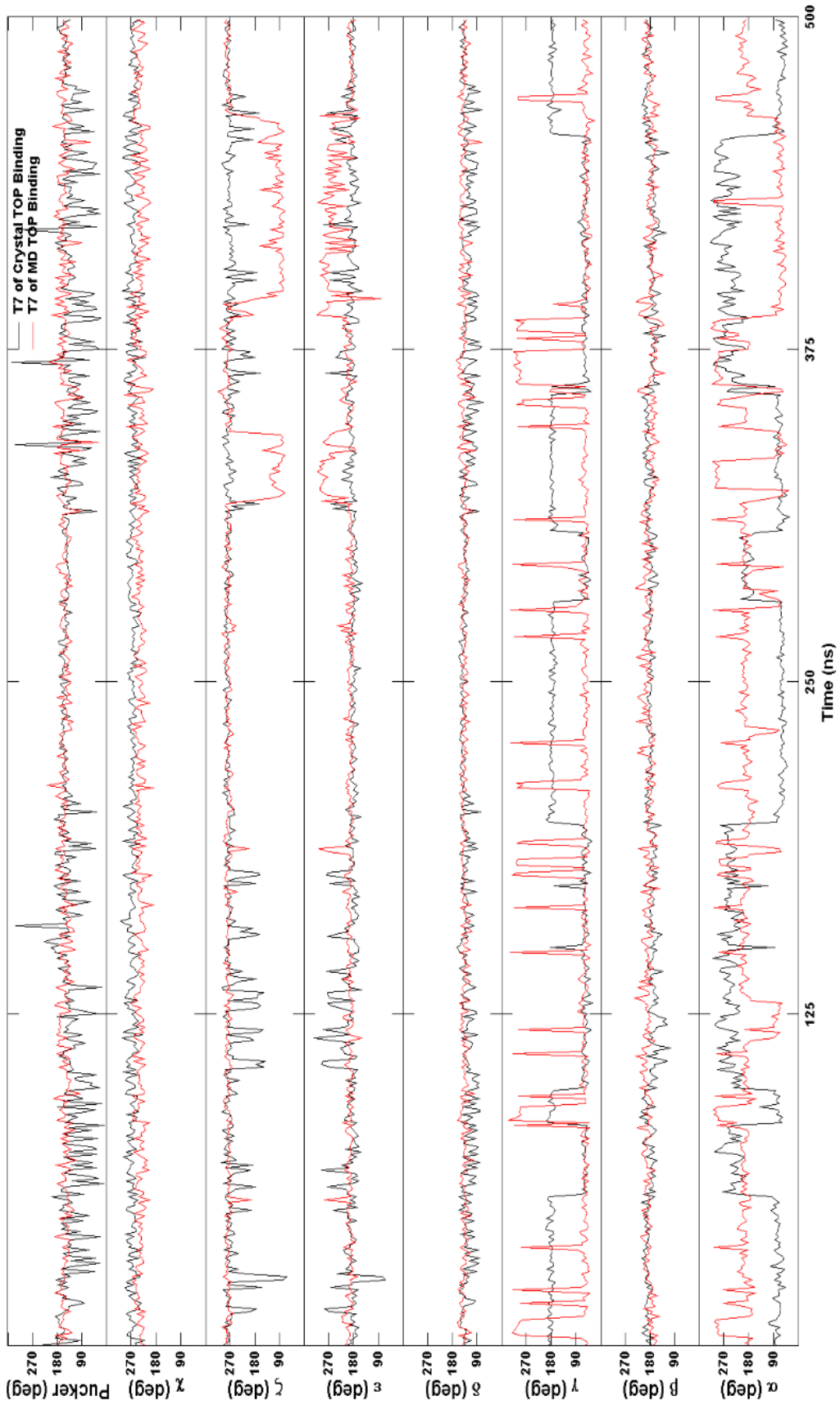


Figure 41. Comparison for the backbone torsion angles of residue T7 between the free ligand binding simulation (red) and the stability simulation of the crystal pose (black) for the top stacking mode.

Occasional flipping out of the bases in the loops is also observed in this trajectory. Figure 41 features the dihedral angles of residue T7 in both stability simulation of the crystal top mode and the representative trajectory of the free binding simulations charactering top stacking mode. In the stability simulation trajectory, the residue T7 is flipped out from 74 ns to 95 ns, from 110 to 129 ns and from 258 to 333 ns with fluctuations mainly characterized by α and γ. In the free binding simulation trajectory, the residue T7 is flipped out from 322 ns to 345 ns and from 395 ns to 462 ns and is mainly characterized by α, ε and ζ angles.

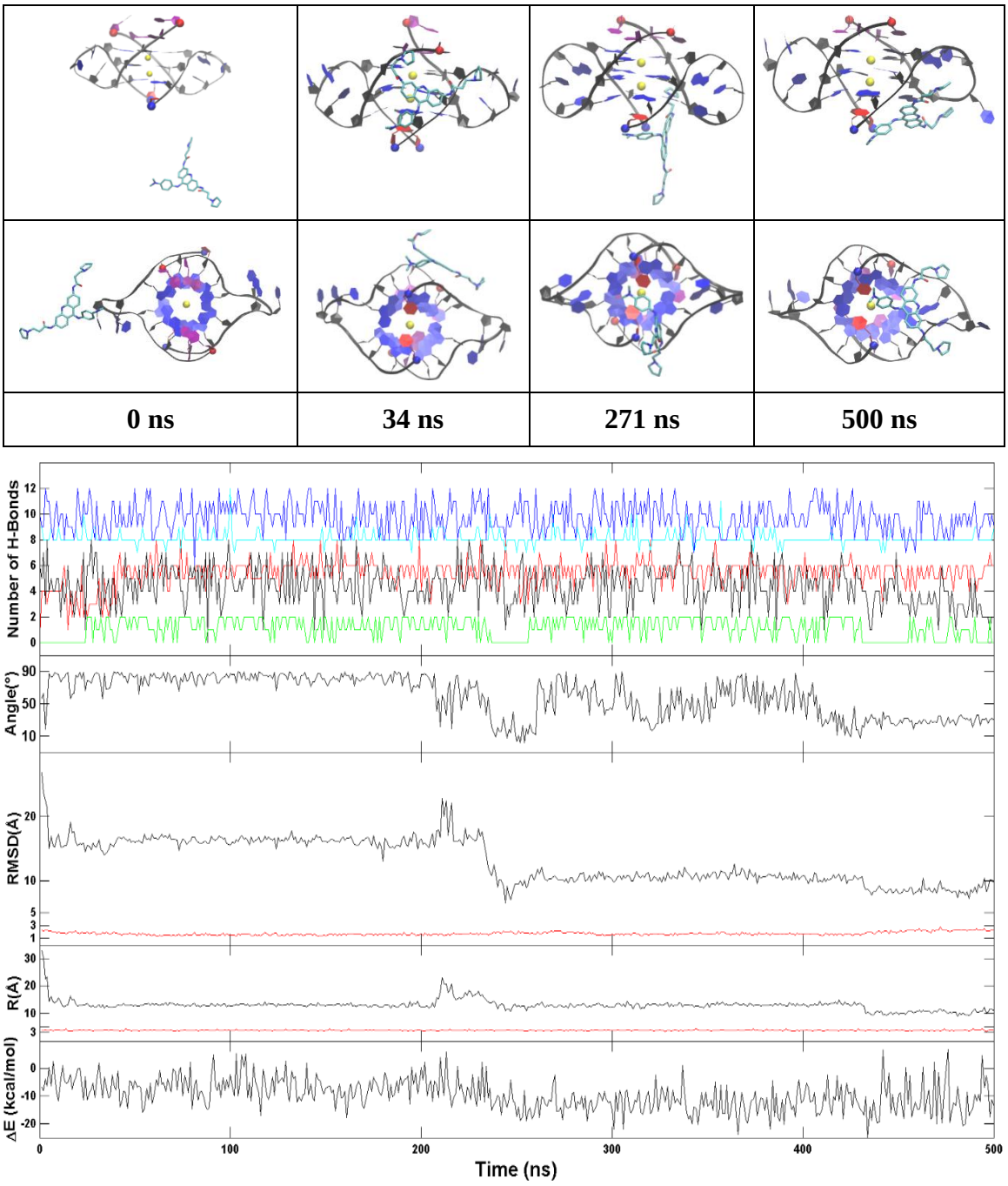


Figure 42. A representative trajectory of the groove binding mode. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 71), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-PBSA binding energy (ΔE) (cf. methods section for definition).

3.4.7 Groove binding of BRACO19 might be an intermediate state for the end binding pose. In the representative trajectory of the groove binding mode, BRACO19 made initial contact with the telomeric G-quadruplex at 34 ns (Figure 42). At 20 ns, two additional hydrogen bonds formed in the first base layer, indicating the formation of ATAT-tetrad. At ~271 ns, the ligand moved to the groove, but rather than remaining stable, it kept adjusting pose toward the G-quadruplex bottom end. This also was reflected in the large fluctuation of the five order parameters during 200-450 ns. This system reached a steady state at 450 ns, showing minor fluctuations through the rest of the trajectory. The terminal MM-PBSA binding energy is ~20 kcal/mol, which is much lower than that of the top stacking (-25 kcal/mol) and the bottom binding mode (~30 kcal/mol). Therefore, this system can further be converted into an end binding pose. In fact, the conversion from a groove binding pose to a bottom binding pose was observed at 236 ns in this trajectory indicated by ligand RMSD and drug-base dihedral angle in the order parameter plot and the conversion to a top binding pose was observed in another representative groove binding trajectory. BRACO19 initially bound to the side of the telomeric G-quadruplex at 9 ns, slightly moved down and bound to a groove at 86 ns, but moved out and bound to the top at 345 ns, and then remained in this binding mode through the rest of the trajectory. The van der Waals energy difference between groove binding pose and end binding pose might be the driving force. The ligand interactions of BRACO19 with the G-quadruplex in the representative structure of the groove binding trajectory and the representative structure of the bottom trajectory are compared (Figure 70).

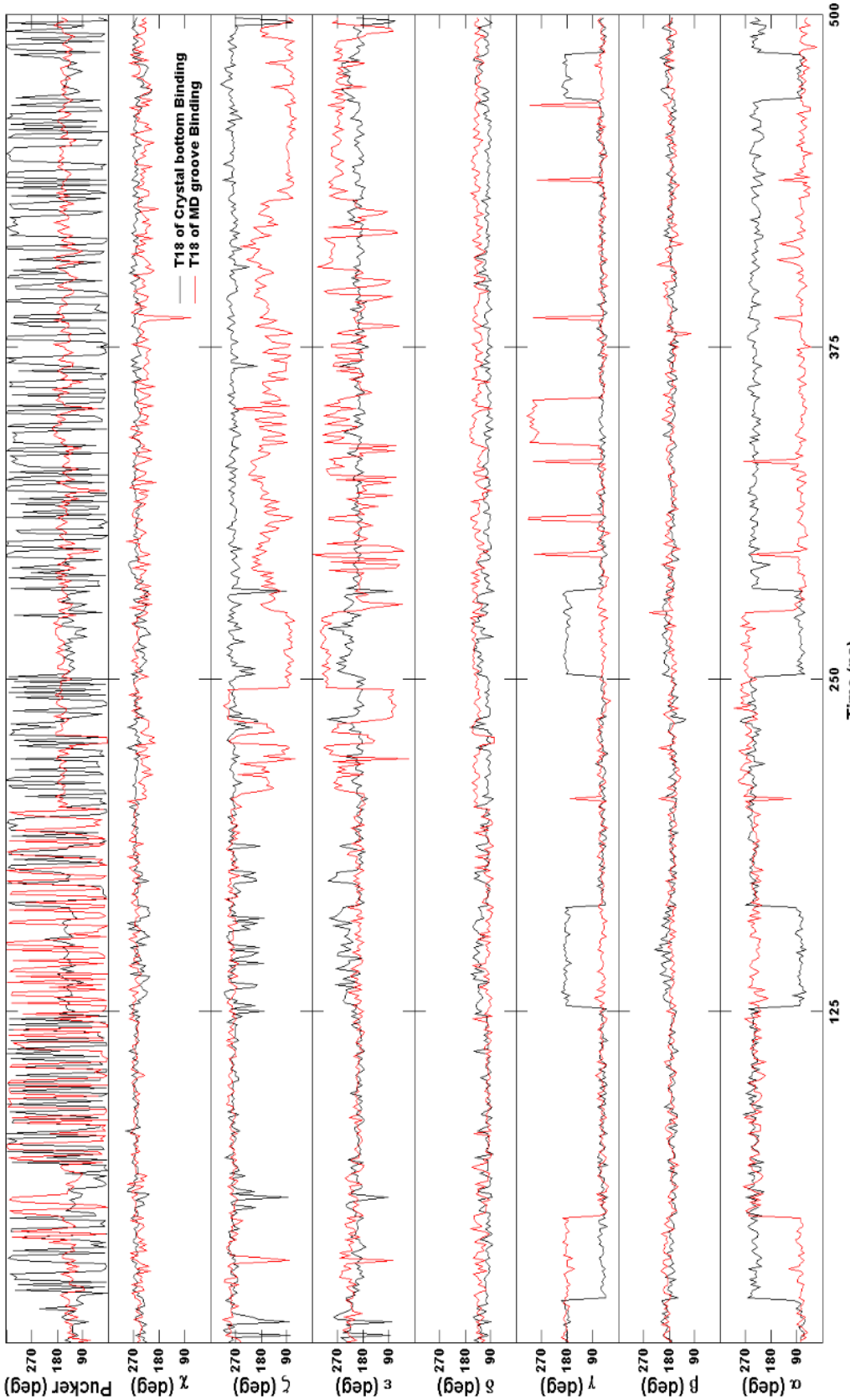


Figure 43. Comparison for the backbone torsion angles of residue T18 between the free ligand binding simulation (red) of the groove binding mode and the stability simulation of the crystal pose (black) of the bottom binding mode.

In the representative structure of the bottom intercalating trajectory (i) the acridine is stacked on G23, (ii) one pyrrolidine ring formed H-bond with T18, (iii) other formed H-bond with T12 and G11 and (iv) the side chain on 9th position is stacked on G5 (Figure 70 (C)) whereas in the representative structure of the groove binding trajectory the acridine is stacked on G17, (ii) pyrrolidine ring of side chain on 3rd position formed a H-bond with G17, (iii) N from the amide of side chain on 6th position formed a H-bond with G23 and (iv) the secondary amine from the side chain on 9th position formed a H-bond with T12 (Figure 70 (G)).

Occasional flipping out of the bases in the loops is also observed in this trajectory. Figure 43 features the dihedral angles of residue T18 in representative trajectories of stability simulations charactering crystal bottom intercalation mode and free binding simulations charactering groove binding mode. In the stability simulation, the residue T18 does not flip out at all and this is characterized by the minimum fluctuation of ε and ζ. In the free binding trajectory, the residue T18 is flipped out at 276 ns remains as such through the rest of the trajectory and is mainly characterized by α, ε and ζ angles.

Table 113

MM-PBSA energy of BRACO19 to the G-Quadruplex DNA in different binding modes.

Binding Pose	[3]Δ^{VDW}	[4]Δ^{SUR}	[5]$\Delta^{PB\text{-}ELE}$	[6]Δ^{CONF}	[7]Δ^{TOT}	[8]$\Delta\Delta^{E}$
Top Stacking	-60.1±6.2	29.4±3.3	11.2±5.5	-5.6±6.7	-25.1±6.7	4.2
Groove Binding	-42.7±5.3	22.2±3.0	6.5±6.1	-14.9±4.8	-26.1±6.6	3.2
Bottom Intercalation	-84.8±4.2	36.6±1.8	18.0±5.0	0.9±2.7	-29.3±5.7	0

[3] VDW = Change of VDW energy in gas phase upon complex formation (Units: kcal/mol)
[4] SUR = Change of energy due to surface area change upon complex formation (Units: kcal/mol)
[5] PB-ELE = Change of PB reaction field energy + Elec. energy upon complex formation (Units: kcal/mol)
[6] CONF = Change of energy due to conformational changes (Units: kcal/mol)
[7] TOT = (ΔVDW +ΔSUR + ΔPB-ELE + ΔCONF) Change of potential energy upon complex formation
[8] ΔE = Difference to the most favorable binding mode

3.4.8 MM-PBSA binding energy data ranks the bottom intercalation mode as the most stable mode among the three binding modes. The relative stability of the three binding modes was examined by conducting MM-PBSA binding energy calculations on each mode. The best binding energy toward the G-quadruplex (-29.3±5.7 kcal/mol) is observed in the bottom intercalation, making this the most favorable binding pose. The -25.1±6.7 kcal/mol and -26.1±6.6 kcal/mol binding energies of top stacking and groove binding respectively are quite comparable. To further understand the nature of binding, the binding energy was fragmented into van der Waals (VDW) interaction, hydrophobic interaction (SUR), and electrostatic interaction (PB-ELE) (Table 113). As demonstrated in the Table 113, most of the binding interactions are contributed by van der Waals interactions. As expected, the most favorable VDW energy was demonstrated by the bottom stacking pose (-84.8±4.2 kcal/mol), which is 24.7 kcal/mol more favorable than that of top stacking pose and 42.1 kcal/mol more favorable than that of the groove binding pose. This high VDW interactions of the bottom intercalation mode can be accounted by the interaction of BRACO19 with a G-tetrad and a T-dyad whereas, the top stacking mode and groove binding modes are only interacting with an ATAT-tetrad and the residues in the loop, respectively. MM-GBSA binding energies were also calculated for each binding pose of free binding simulations at various ionic strengths. The binding energies decreased as the ionic strength increased indicating a stronger screening effect. But, the relative binding energies of three modes at different ionic strength are quite similar, supporting the use of MM-GBSA in ranking poses ($\Delta\Delta E$=0.0, 4.6 and 6.5 kcal/mol for bottom, top and groove binding modes). Clearly, both MM-PBSA and MM-GBSA data indicates the most stable pose is the bottom binding mode and the groove binding mode is less stable, which

was observed to convert into the end binding mode. Thus, the groove binding is likely an intermediate state of end binding mode.

3.5 Discussion

G-quadruplex DNA has been gaining increasing attention as a promising target for cancer therapeutics(Biffi et al., 2014). The binding of BRACO19 to the G-quadruplex depends on its ability to π-π stack onto the G-quadruplex which is facilitated by its planarity. Molecular dynamics binding simulations and MM-PBSA binding energy calculations were utilized in this study, to analyze the binding pathway of BRACO19 to a telomeric DNA G-quadruplex. To tackle the fore-mentioned question, whether the latest AMBER DNA force field (OL15) is accurate enough to probe the binding pose of BRACO19 to G-quadruplex, the complex obtained from the simulation of the apo form and a free BRACO19 was compared to the simulated structure of the experimental X-ray crystal structure of the bound form (PDB ID: 3CE5). Comparing with an NMR solved structure would have been an ideal practice, but due to the unavailability of such NMR structure, X-ray solved crystal structure was used. Because of the crystal packing constraints, the potential AT-tetrad that could be formed by the bases T1, A2, T13 and A14 was disrupted and instead the AT tetrad was formed by A2, T13, A14 and the T24 (Figure 31) of different chains, thus disrupting the pairing of T24 with T12 of its respective chain. In the previous X-ray study,(Chung et al., 2013a) only two binding modes were observed, in which BRACO19 intercalates at the bottom of the G-tetrad and stack on the top of the ATAT-tetrad (Figure 30A). Stability simulations were conducted on these two crystal binding poses and illustrated the loss of planarity of BRACO19 in the bottom mode and instability of top stacking mode. In addition, the ATAT-tetrad was observed in the crystal

bottom pose, which is consistent with the bottom binding trajectory of our free binding simulations. Clearly, the lack of the crystal packing constraints led to these changes. Therefore, the relaxation by MD simulations generated better solution reference structures for our free binding simulations.

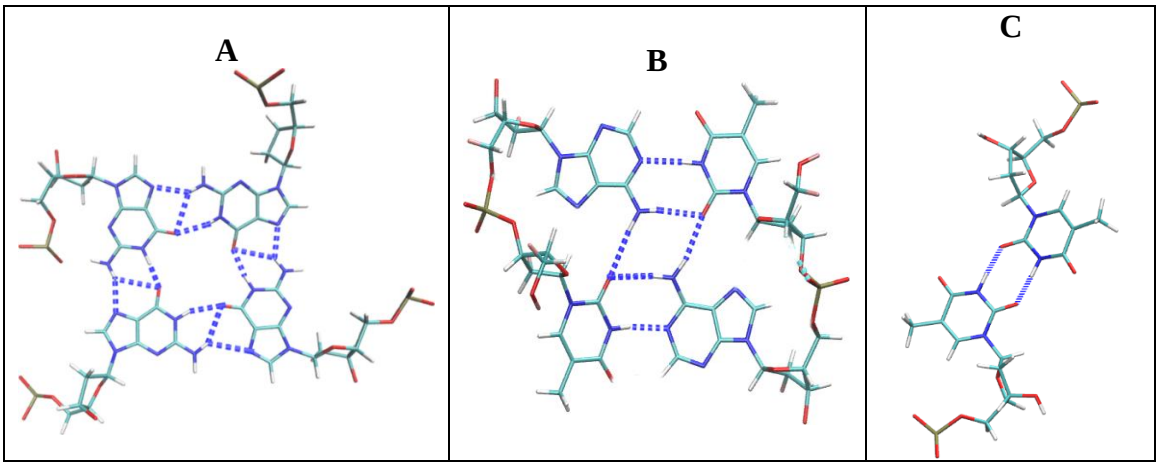


Figure 44. H-bond network indicating maximum number of H-bonds possible in G-tetrad (A), ATAT-tetrad (B) and T-dyad (C) respectively.

Our free binding simulations found that BRACO19 binds to the same telomeric DNA G-quadruplex structure in three different modes: Top stacking (11.9% total population), bottom intercalation (51.1% total population), and in the groove (9.6% total population). The bottom intercalation and top stacking mode resemble the BRACO19 binding pose in the crystal structure. Our MM-PBSA/MM-GBSA and trajectory analysis show the bottom intercalation mode to be more stable than top stacking mode or groove binding mode. The apo form of the G-quadruplex in the crystal (Figure 30E &F) is known to have three G-tetrads whereas the top stacking and groove binding mode from the simulations were observed to have three G-tetrads, one ATAT-tetrad and one T-dyad. In

the ATAT-tetrad every adenine and thymine forms 6 H-bonds with adjacent thymine and adenine respectively (Figure 44) and in T-dyad, the thymine bases share 2 H-bonds with each other. The detailed analysis of number of H-bonds in each predefined layer is illustrated in Figure 71. Interestingly, the direct alignment type of an intra-quadruplex ATAT-tetrad observed in our simulation is different from the slipped alignment type of ATAT-tetrad identified by the previous NMR study of d(GAGCAGGT) sequence in 1M NaCl solution under which it forms a head-to-head dimeric quadruplex containing sequentially stacked GCGC, GGGG and slipped ATAT tetrads.(Zhang et al., 2001) This intra-quadruplex ATAT-tetrad is also different from the inter-quadruplex ATAT-tetrad that was observed in the crystal structure 3CE5 (Figure 31), although both share the similar geometry and H-bond pattern. Therefore, further experimental evidence is required to prove our prediction of the direct intra-quadruplex ATAT-tetrad. In the groove binding mode, T6 flipped out to facilitate BRACO19 insertion. A notable speculation in this study is that the groove binding mode is likely to be an intermediate stage in the process of achieving the final stable end stacking mode, which has been observed in our early study of binding of RHPS4 to human telomeric G-quadruplex(Kelly Mulholland et al., 2017) and in a study of binding of BRACO19 to a single stranded parallel telomeric G-quadruplex (to be published).

The fore mentioned conformational differences between the apo and holo form of the G-quadruplex (i.e. the bottom intercalation mode) in the crystal structure were partially reproduced by the simulations. First, as to the flipping out of T6, T7 and A8 in the crystal complex structure (Figure 37C and 37E), the flipping out of the latter two was observed in the simulations (Figure 37C). Second, as to the formation of the first base layer by T1, A2

and A14 in the crystal structure, this was observed in both bottom intercalation trajectories (Figure 38). In addition, further optimization of the bottom intercalation was observed in the simulation: while the T12 and T24 are not yet paired in the X-ray crystal structure (Figure 31), this T-dyad was formed in the simulation. As a result, BRACO19 is intercalated between the G-tetrad formed by residues 5, 11, 17 and 23 and the T-dyad by residues 12 and 24. The intercalation between G-tetrad and T-dyad increased the ligand binding energy to the G-quadruplex.

The planarity of BRACO19 central rings permits the compound to stack on top and intercalate at the bottom of telomeric G-quadruplex. For this reason, planarity is a critical feature to be considered in developing G-quadruplex specific ligands. A planar scaffold not only increases the binding selectivity, but also boosts the intercalation thus increasing overall binding affinity. These findings may aid future attempts at creating a promising telomeric G-quadruplex stabilizer with large central rings.

The dihedral angles of the 3 G-tetrads in both stability simulation and free binding simulations have low fluctuations and are consistent through the binding process. The dihedral angles of the terminal layers, ATAT-tetrad and T-dyad demonstrated more fluctuations. The torsion angle analysis indicated that the conformational changes are characterized mainly by α, γ, ε and ζ and in some cases, changes in χ dihedral angle. And the recent corrections (εζOL1, εζOL1and χOL4) in the AMBER OL15 DNA force field appears to provide a more balanced dihedral angle sampling which contributed to a good agreement to the experimental structures.

3.6 Conclusions

Computational methods are getting more and more significant in drug discovery as they provide detailed structural information. Molecular dynamics binding simulations, MM-PBSA and MM-GBSA binding energy calculations were utilized in this study, to characterize binding modes of BRACO19 to a telomeric DNA parallel G-quadruplex at physiological solution conditions and validate the utilization of latest AMBER DNA force field (OL15) with recent corrections βOL1, εζOL1 and χOL4 coupled with GAFF2 ligand force field in studying G-quadruplex in complex with a ligand. Three binding modes have been identified: top stacking, bottom intercalation and groove binding. Bottom intercalation and top stacking resembles very well the binding pose in the X-ray solved crystal structure of the same telomeric G-quadruplex with BRACO19. The groove binding mode is likely to be an intermediate state leading to the end binding mode. A flip-insertion mechanism was observed in the bottom intercalation mode, during which the flipping out of the bases make space for ligand insertion, followed by the flipping back of the bases to increase the stability of the complex. Formation of an intra-quadruplex ATAT-tetrad has been observed for the first time. Torsion angle analysis indicated good sampling of dihedral angles and a good agreement with the experimental structures.

Chapter 4

Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent

4.1 Abstract

High efficacy is displayed by the potently binding human telomeric DNA G-quadruplex drug, BRACO19, in inhibition of tumor cell growth. Although, the improvement of its' 62-fold preferential binding affinity towards DNA G-quadruplex over DNA duplex remains to be a challenge. The crystal structure of BRACO19 in complex with neither single-stranded telomeric DNA G-quadruplexes nor B-DNA duplex is available yet. Consequently, the characteristic binding nature of BRACO19 and these DNA forms remains elusive. In this study, the binding pathway of BRACO19 is characterized by simulating 200 ns MD binding simulations with a free ligand (BRACO19) to a DNA duplex and three different topological folds of the human telomeric DNA G-quadruplex (parallel, antiparallel and hybrid). Groove binding mode was found to be the most stable binding mode for the duplex and top stacking mode for parallel G-quadruplex, antiparallel and hybrid G-quadruplexes. The non-existential binding selectivity of BRACO19 can be accounted to the similar binding affinities of groove binding to both the duplex and the G-quadruplex. For that reason, a modification should be induced such that this prospective ligand destabilizes binding to the duplex form but stabilizes the G-quadruplex binding. Such modification can improve this mere 62-fold binding selectivity toward the G-quadruplex. Furthermore, the groove binding mode was found to be an intermediate stage of the top stacking mode.

4.2 Introduction

The formation of the first therapeutically important G-quadruplex was observed in the single stranded overhang of human telomeric DNA.(J. Debray et al., 2009; Doluca, Withers, & Filichev, 2013a) The single stranded 3' overhang (100-200 nucleotides) is the termini of the human telomeric DNA which contains numerous repeats of d(TTAGGG) sequences and is capped by Shelterin complexes.(Chung et al., 2013c; de Lange, 2005a; Moyzis et al., 1988a; Wright, Tesmer, Huffman, Levene, & Shay, 1997b) Shelterin complexes provides protection against nuclease attacks, chromosomal end-to-end fusion and gene erosion at cell divisions.(Palm & de Lange, 2008) After each cell replication, the telomere truncates by 50-200 base pairs and when the telomere is exhausted and Hayflick limit is reached, cell senescence and apoptosis are triggered.(Harley, Futcher, & Greider, 1990a; Zakian, 1995a).

In cancer cells, a reverse transcriptase called telomerase adds nucleotides to the telomere thus immortalizing the cells. (Greider & Blackburn, 1989b; Moorhouse et al., 2006a) Telomerase is found to be overexpressed in 80-85% of tumor cells. It can be logically concluded that telomerase inhibition is a valid therapeutic approach in cancer treatment. But the challenges with this approach are (i) there is a time delay in which the telomere length needs to be established for the ultimate apoptosis trigger(Asai et al., 2003; Harley et al., 1990a; Shay & Wright, 2006) and (ii) studies suggest an alternate mechanism for telomerase maintenance might be activated upon telomerase inhibition.(Bechter, Zou, Walker, Wright, & Shay, 2004; Dunham, Neumann, Fasching, & Reddel, 2000; Hu et al., 2012) It has been reported that the telomere cannot be hybridized by telomerase when the single stranded 3' overhang folds into a G-quadruplex.(Zahler, Williamson, Cech, &

Prescott, 1991a) Consequently, stabilizing the telomeric G-quadruplex adopted by guanidine-rich single stranded 3' overhang which will be perceived as DNA damage and stimulates cell apoptosis. (Denchi & de Lange, 2007a; Doluca et al., 2013a)

G-quadruplexes can be formed from a single or double stranded DNA duplex. It has been experimentally established that the telomeric sequences can fold into four topologies in dilute solutions; hybrid [3+1] (PDB IDs: 2HY9 and 2JPZ), parallel (PDB ID: 1KF1), one 2-tetrad antiparallel and one 3- tetrad antiparallel (PDB ID: 143D) folds. And this folding depend on sequence, ions and presence of small molecules.(Hänsel et al., 2011a) Traditional studies suggest that the polymorphism is lost in 40% PEG or 50% ethanol solutions, in other words dehydrated solutions, parallel stranded conformation prevails. Concluding that parallel G-quadruplex is biologically relevant. Many studies were reported to develop lead compounds targeting them. Hansel et al suggested that parallel G-quadruplex might not be the most prevalent form and other topologies need to be studied to understand and design lead compounds with better binding affinities and selectivity.(Hänsel et al., 2011a)

BRACO19, tri-substituted acridine shown in figure 30, was logically designed with computer modelling by understanding the structural requirements of the parallel-stranded G-quadruplex binding site.(Yang & Okamoto, 2010a) BRACO19 has been reported to inhibit telomerase causing telomere shortening(Incles et al., 2004) and its experimental in-vivo activity has been reported (Table 110). It was also reported that BRACO19 demonstrated broad anti-viral activity by stabilizing the G-quadruplexes found in pro-viral DNA.(Perrone et al., 2014) Lack of selectivity towards G-quadruplex over duplex DNA is one of the reasons BRACO19 has never been approved.(Yang & Okamoto, 2010a) To

achieve higher selectivity (in the order of 10^5), better understanding of characteristic binding of BRACO19 with DNA G-quadruplex and duplex DNA is required.

The only available crystal structure with BRACO19 (PDB ID: 3CE5) is double stranded with a parallel G-quadruplex arrangement. The G-quadruplex asymmetrically interacts with the ligand via π–π interactions with the guanine bases, stacking the K^+ in-line. Computational studies reported that homologous variation of the side chains decreases the binding affinity (Campbell, Parkinson, Reszka, & Neidle, 2008) although these studies might be irrelevant as the telomeric overhang that folds into a G-quadruplex is single strand DNA. It is also to be noted that not many studies could be found on other scaffolds i.e., anti-parallel and hybrid.

Debray et al synthesized and evaluated fused bis-pyrimidinoacridines, pentacyclic analogs of BRACO19 in order to understand the interactions of these analogs with the G-quadruplex. The analogs were docked onto DNA G-quadruplex (PDB ID: 22AG), and DNA-duplex (PDB ID: DS17) and parallel G-quadruplex with BRACO19 (PDB ID: 3CE5) in the binding site.(Julien Debray et al., 2009) Xue et al synthesized and docked Neomycin-Perylene conjugate onto the antiparallel G-quadruplex. Their docking data indicated that perylene moiety stacked onto the DNA bases and the two neomycin units occupied two G-quadruplex grooves.(Xue, Ranjan, & Arya, 2011) Long et al introduced a peptidyl group on benzo-furo-quinoline derivatives; and their combined experimental and molecular docking data using parallel G-quadruplex suggest that the peptidyl group increased their selectivity significantly towards telomeric DNA quadruplex over duplex DNA.(Long et al., 2012) Alcaro et al identified and characterized novel G-quadruplex binders by docking-based virtual screening using the three known folds of DNA-quadruplex; hybrid [3+1],

parallel and antiparallel folds.(Alcaro et al., 2013) Their docking data showed that most of the ligands stacked at the bottom of 1KF1 and 2HY9, but in the case of 143D and 2JPZ most of the ligands docked laterally. Multiple computational methods including pharmacophore modeling, shape-based modeling and docking were employed on DNA G-quadruplexes (PDB IDs: 3SC8, 3UYH, 3CE5 and 3R6R) and lead candidates with promising potency were identified.(Kaserer et al., 2016) Nonetheless, it is well known that the docking with a rigid receptor might lead to incorrect binding modes and poor docking scores, therefore eliminating a prospective lead compound.(Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005)

G-quadruplexes in complex with BRACO19 and various ligands have been widely studied using molecular dynamic (MD) simulations. Moore et al conducted MD simulations to investigate the structure-activity relationships of tri-substituted acridines analogs (BRACO19 analogs) and a modelled 22mer parallel G-quadruplex.(Moore et al., 2006b) Hou et al revealed H-bonds to be the major contributors for stability of the G-quadruplex and ligand-quadruplex complex by conducting stability simulations on G-quadruplex-ligand complexes involving BRACO19 and 5 other ligands, known for affinity towards DNA G-quadruplex.(J. Q. Hou et al., 2010b) Dhamodharan et al advised end-stacking to be the favored binding mode after docking bis-quinolinium and bis-pyridinium derivatives of 1,8-naphthyridine onto antiparallel G-quadruplex and consequently, conducting MD simulations. (Dhamodharan, Harikrishna, Jagadeeswaran, Halder, & Pradeepkumar, 2012a) However, Jain et al reported that both end-stacking and groove-binding were favored after docking dimeric 1,3-phenylene-bis(piperazinyl benzimidazole)s to 22mer parallel G-quadruplex followed by MD simulations.(Jain, Paul,

Maji, Muniyappa, & Bhattacharya, 2012b) Ungvarsky et al characterized the binding poses of a novel set of BRACO19 derivatives to the human telomeric parallel G-quadruplex by successfully employing docking and MD simulations.(Ungvarsky et al., 2014b) Zhou et al attempted to understand the ligand unbinding from G-quadruplex using steered molecular dynamics and umbrella simulations.(J. K. Zhou, Yang, & Sheu, 2015b). Recently Diveshkumar et al conducted, by docking and MD simulation studies on various G-quadruplexes (PDB IDs: 2L7V, 2O3M, 1KF1, 143D, and 2MB3) and identified indolyl, methylene-indanone scaffolds which demonstrate selectivity towards parallel promoter G-quadruplexes over telomeric DNA quadruplex or duplex DNA.(Diveshkumar et al., 2016b)

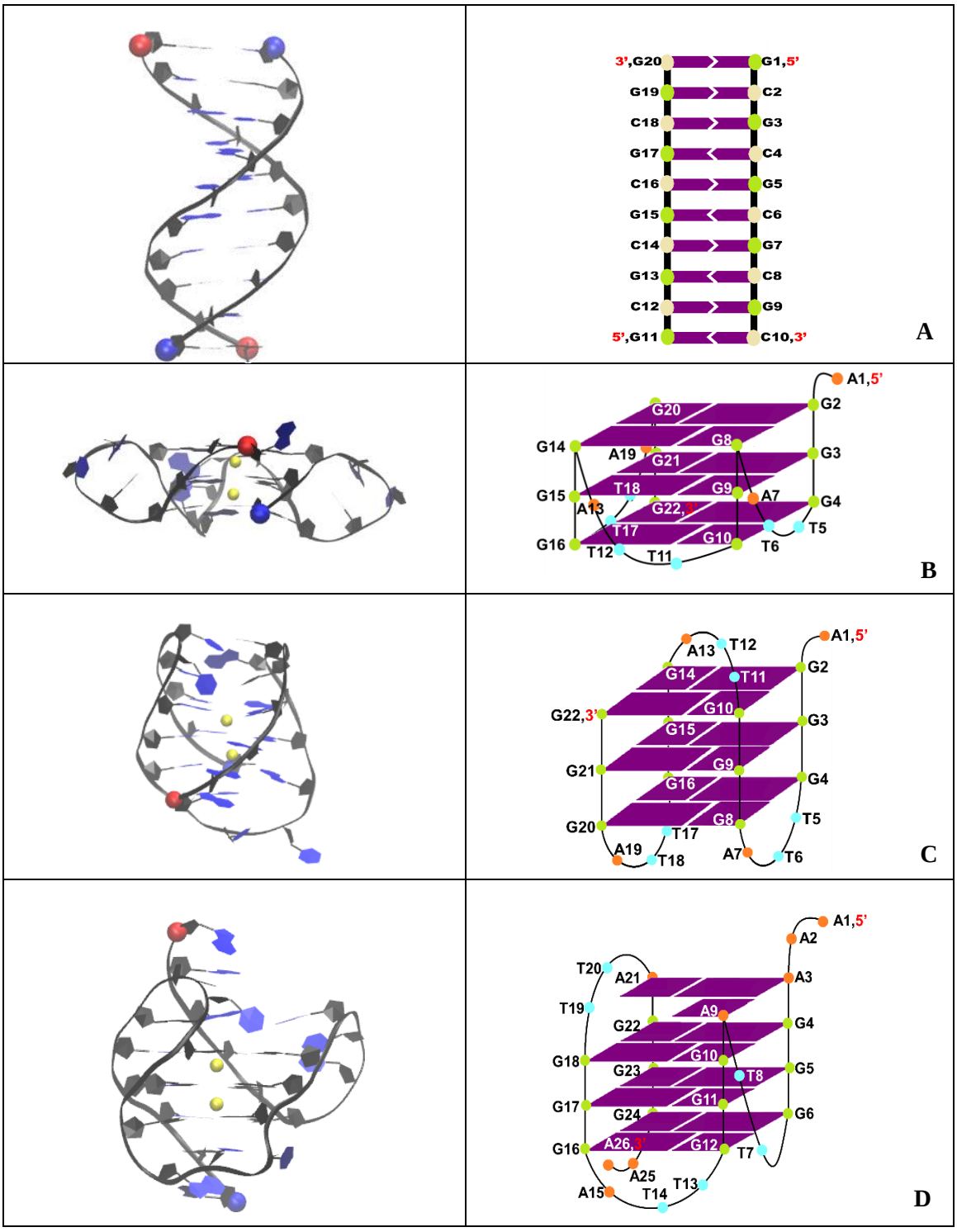


Figure 45. Structure of human telomeric DNA duplex (**A**), human telomeric parallel DNA quadruplex (PDB ID: 1KF1) (**B**), human telomeric antiparallel DNA quadruplex (PDB ID: 143D) (**C**), and human telomeric hybrid DNA quadruplex (PDB ID: 2HY9) (**D**). 5' and 3' of the DNA chain are indicated by a red and blue ball, respectively.

In this study, the binding pathway of BRACO19 to parallel, anti-parallel, hybrid DNA G-quadruplex and duplex DNA is characterized by conducting molecular dynamics binding simulations. In this study, the binding pathway of BRACO19 to parallel, anti-parallel, hybrid DNA G-quadruplex and duplex DNA is characterized by conducting molecular dynamics binding simulations. MD free ligand binding simulations in which no constraints are placed on the relative position of the ligand were utilized to probe the binding pathway and mechanism of BRACO19 to the human telomeric parallel G-quadruplex DNA. Major binding poses, (top binding, end stacking, bottom binding and groove binding) were identified and detailed binding pathways were characterized. The dynamic and energetic properties of the three major binding modes were thoroughly studied, providing vivid examples of induced-fit binding mechanism. The similar binding energy of the groove binding pose to the duplex and the G-quadruplexes may be responsible for the low selectivity of BRACO19. The binding pathway of BRACO19 to various G-quadruplexes is characterized using torsion angle parameters. This analysis indicated good sampling of dihedral angles and a good agreement with the experimental structures.

4.3 Methods

4.3.1 Simulation systems.

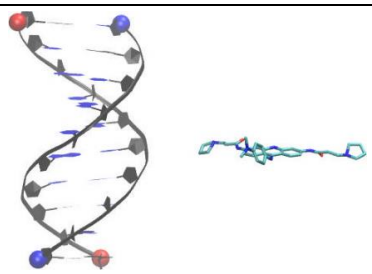
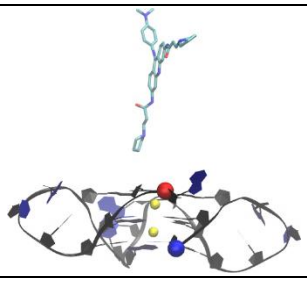
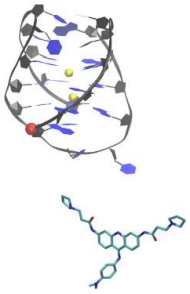
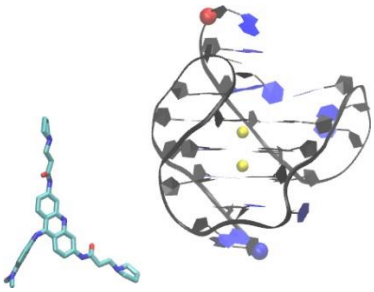
DNA Duplex + Unbound BRACO19	Parallel DNA G-Quadruplex + Unbound BRACO19
	
Anti-parallel DNA Quadruplex + Unbound BRACO19	Hybrid DNA Quadruplex + Unbound BRACO19
	

Figure 46. Initial configuration of the simulation systems. 5' and 3' of the telomeric DNA are indicated by a red and blue ball, respectively and the K+ ions are represented in yellow.

A total of 4 DNA-ligand systems were constructed: B-DNA duplex structure of $d([GC]_{10})_2$, X-ray crystal structure of the parallel telomeric DNA G-quadruplex, NMR solved anti-parallel telomeric DNA G-quadruplex and NMR-solved (3+1) hybrid telomeric DNA G-quadruplex (figure 46). One B-DNA duplex structure of $d([GC]_{10})_2$, built using Maestro program, one X-ray solved human telomeric parallel G-quadruplex and two NMR solved human telomeric G-quadruplex structure were each used to construct four unbound DNA-ligand systems with a BRACO19 molecule that was 10 Å away from the DNA (Figure 46). A water box of truncated octahedron with 10 Å water buffer was used to

solvate the unbound system. And it was neutralized by K^+. The DNA structures were represented by a refined version of the AMBER DNA OL15 (i.e. parm99bsc0(Pérez et al., 2007) $+\chi_{OL4}$ (Krepl et al., 2012b)+ ε/ζ_{OL1}(Zgarbova et al., 2013)+ β_{OL1}(Zgarbova et al., 2015) updates), water was represented by TIP3P model(Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983c) and the K^+ ions were represented by the K^+ model developed by Cheatham group.(Joung & Cheatham, 2008b) The standard AMBER protocol was used to obtain the force field for the BRACO19 molecule: after the geometry optimization of the BRACO19 at the HF/6-31G* level, the molecular electrostatic potential (MEP) of the BRACO19 molecule was calculated at the same theory level; then the partial charges of BRACO19 atoms was determined by MEP using Restrained Electrostatic Potential/RESP method with two stage fitting;(Bayly, Cieplak, Cornell, & Kollman, 1993a) and the AMBER GAFF2(Case et al., 2016) force field provided the rest of the force field parameters. The nucleic acid simulations have been widely practiced in AMBER DNA force fields.(Cosconati et al., 2010; Fadrna et al., 2009; Lavery et al., 2010; A. Mukherjee, Lavery, Bagchi, & Hynes, 2008) In our studies, the binding pathway of doxorubicin(Lei, Wang, & Wu, 2012b) and telomestatin(K. Mulholland & Wu, 2016), anti-cancer drugs, to the B-DNA fragment(Lei et al., 2012b) and to the human telomeric hybrid G-quadruplex(K. Mulholland & Wu, 2016), respectively have been simulated.

4.3.2 Simulation protocols. The ten production runs for all systems were conducted using the AMBER 16 simulation package.(Case et al., 2016) The detailed protocol followed our previous studies.(Lei et al., 2012b; K. Mulholland & Wu, 2016) After minimizing the energy, the Maxwell-Boltzmann distribution was followed in using different random seeds to assign different initial velocities to the atoms of the system.

Better sampling of binding poses and pathway was enabled by multiple independent simulations. To equilibrate the system density, a 1.0 μs production run at 300 K which included a short 1.0 ns MD simulation in the NPT ensemble mode (constant pressure and temperature), where the DNA and ligand were subjected to Cartesian restraints (1.0 kcal/mol/Å), and 200.0 ns (500 ns for one trajectory of parallel G-quadruplex) MD simulation in the NVT ensemble mode (constant volume and temperature). All bonds connecting hydrogen atoms were constrained by SHAKE(Ryckaert, Ciccotti, & Berendsen, 1977b) which enabled a 2.0 fs time step in the simulations. Long-range electrostatic interactions under periodic boundary conditions were treated using the particle-mesh Ewald method(Essmann et al., 1995b) (the fourth order of the B-spline charge interpolation, charge grid spacing of ~1.0 Å; and direct sum tolerance of 10^{-5}). The cutoff distance for short-range non-bonded interactions was 10 Å, with the long-range van der Waals interactions based on a uniform density approximation. To reduce the computation cost, a two-stage RESPA approach(Procacci & Berne, 1994b) was used to calculate non-bonded forces where the short range forces were updated every step and the long range forces were updated every two steps. The Langevin thermostat with a coupling constant of 2.0 ps was used to control the temperature. The trajectories were saved at 50.0 ps intervals for analysis.

4.3.3 Convergence of simulations. The initial structure was used as a reference to calculate the root mean square deviation (RMSD) of DNA backbone. The stability of the DNA structures was indicated by the flat and small RMSDs (Figure 62, 64, 66 and 68). An atom-to-atom distance cutoff of 3.0 Å was used to calculate atom contacts between the DNA structure and the BRACO19. The stable contact number indicated the steady state of

the simulation systems (Figure 63, 65, 67 and 69). A complex with the number of atom contacts greater than 10 was defined as a stable complex.

4.3.4 Binding mode identification. Accounting to the stability of the DNA backbone in the binding process, the DNA backbone of the stable complexes was aligned by a least square fitting. Daura algorithm (Daura et al., 1999b) was used to cluster the aligned complexes into different structural families based on the 2 Å pair-wise RMSD cutoff of the BRACO19 only without ligand fit. The centroid structure was defined as a structure with the largest number of neighbors in the structural family. And this structural family was represented by this centroid structure. Based on visual inspection, super-families corresponding to major binding modes were formed by merging the centroid structures.

4.3.5 Order parameters. The DNA-drug binding process was characterized by using five order parameters: hydrogen bond analysis, drug-base dihedral angle, ligand RMSD, center-to-center and K^+-K^+ distance (R) and MM-GBSA binding energy (ΔE). A hydrogen bond was defined by 3.5Å distance cutoff between H-bond donor and H-bond acceptor and 120° donor-H-acceptor angle cutoff. The hydrogen bonds were calculated for the top/first, middle/second and bottom/third base layers. For the duplex, the three base layers were defined based on the drug insertion position. For the three G-quadruplexes, the three G-tetrads were defined so that 5' is close to the first G-tetrad. The dihedral angle between the plane of the stable G-tetrad layer of the DNA that is close to drug binding site and the BRACO19's ring plane was defined as the dihedral angle. After aligning the DNA, the ligand RMSD was calculated with reference to the first frame of the trajectory. The length from the DNA center to the drug molecule center was defined as the center-to-center

distance (R). The distance between the K^+ ions present in the DNA G-quadruplex was defined as K^+-K^+ distance. The energetics of the bound complexes were analyzed using MM-GBSA(P. A. Kollman et al., 2000) (Molecular Mechanics Generalized Born-Surface Area) module in the AMBER package (GB1 model with salt concentration of 0.2 M, mBondi radii set, and surface tension of 0.0072 kcal/$Å^2$) to avoid the large energy fluctuation of the explicit solvent.

It was reported that even when considering the relative solvation free energy, good predictions can be made for charged molecules by the GB models on the hydration free energy.(Kongsted, Soderhjelm, & Ryde, 2009a) Under this assumption, in this study, ions were removed from charged DNA systems. This was already validated in our previous study, in which this MMGBSA protocol successfully assessed the binding energy of doxorubicin, an anti-cancer drug, to a B-DNA fragment (d(CGATCG)2).(Lei, Wang, & Wu, 2012c) Under comparable entropic terms, the relative binding free energy estimated by the MMGBSA binding energies can be used to rank drugs or their binding poses if a single molecule is considered.(Kongsted, Soderhjelm, & Ryde, 2009b) It has been established by systematic benchmarking studies up to 1864 crystal complexes that ranking of the ligand binding affinity can be achieved by relative MM-GBSA binding energy calculations.(Hou, Wang, Li, & Wang, 2010, 2011; P. Kollman et al., 2000; Sun, Li, Tian, Xub, & Hou, 2014; Xu, Sun, Li, Wang, & Hou, 2013) The standard backbone dihedral angles (α, β, γ, δ, ε and ζ) around the covalent bonds of the deoxyribose and χ about the glycosidic bond were defined (figure 34) to characterize the conformational changes.

4.4 Results

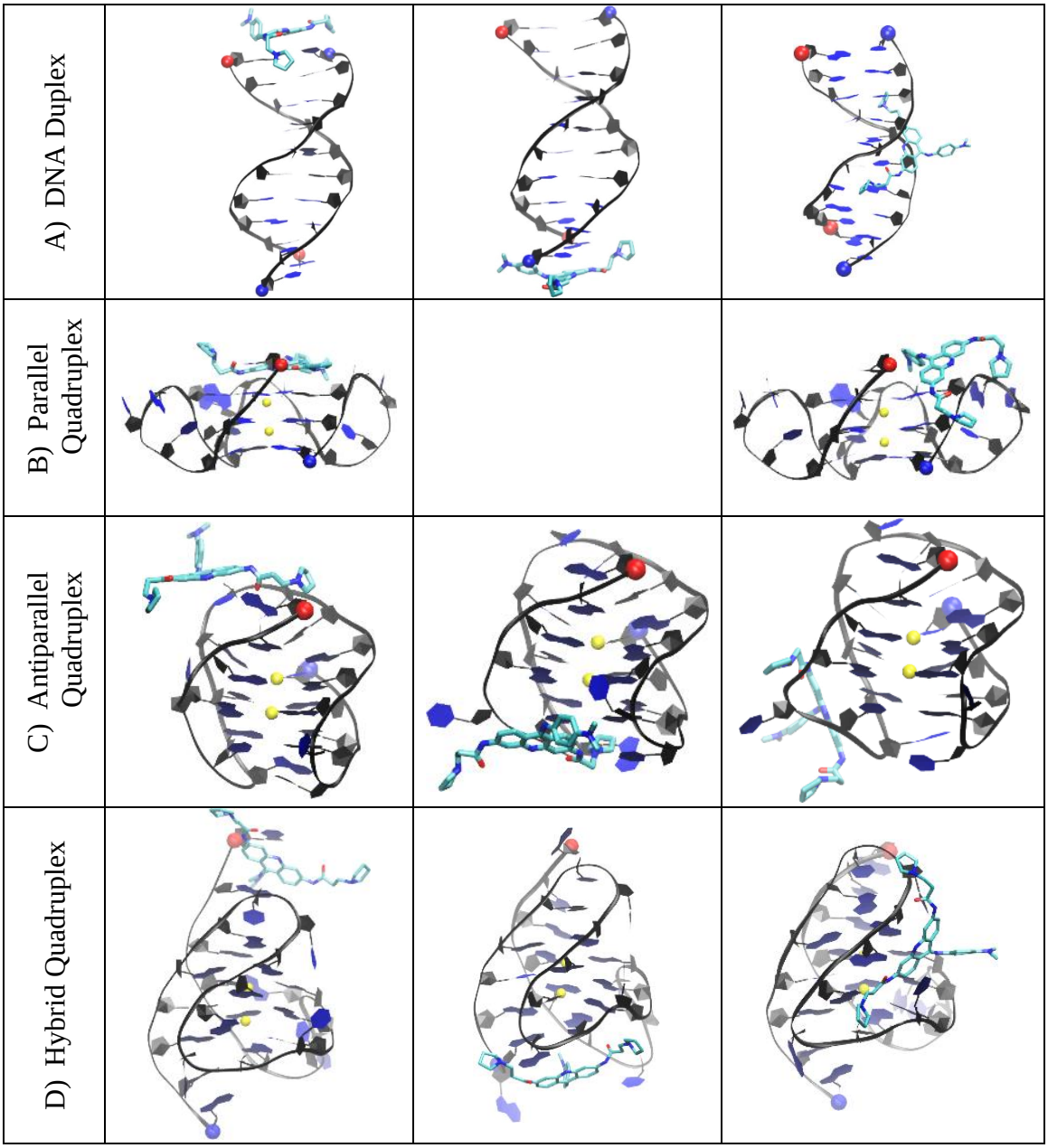


Figure 47. Simulated structures of human telomeric DNA duplex (A), human telomeric parallel DNA quadruplex (PDB ID: 1KF1) (B), human telomeric antiparallel DNA quadruplex (PDB ID: 143D) (C), and human telomeric hybrid DNA quadruplex (PDB ID: 2HY9) (D) in complex with BRACO19. A-D: Top pose (left), Bottom (middle) and groove (right) 5' and 3' of the telomeric DNA are indicated by a red and blue ball, respectively and the K+ ions are represented in yellow.

4.4.1 Multiple drug binding modes were observed in binding simulations. Starting from an unbound state, ten production runs for all four systems were simulated. The convergence of the binding simulations was confirmed (see the method section). The last snapshots of all the simulated trajectories of duplex are listed in figure 48 and they indicate the stability of the structures; the base pairing was maintained.

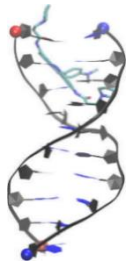
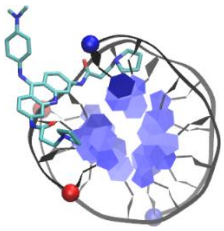
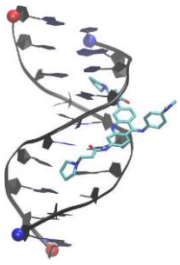
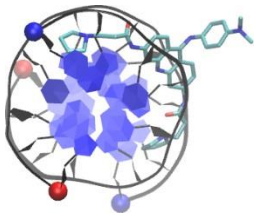
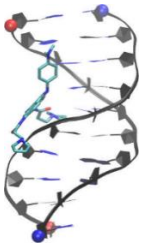
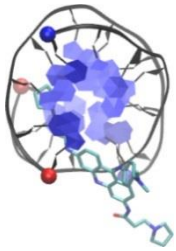
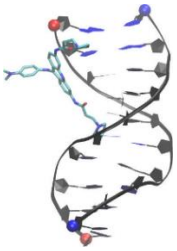
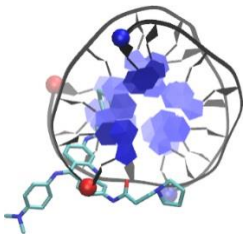
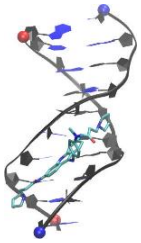
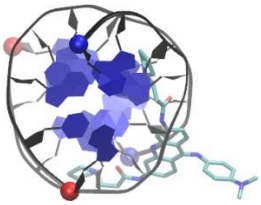
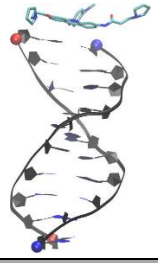
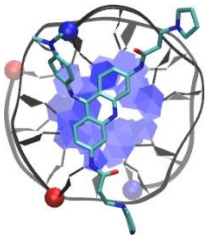
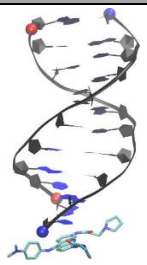
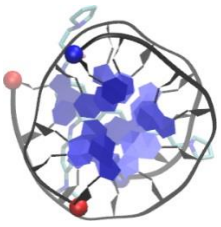
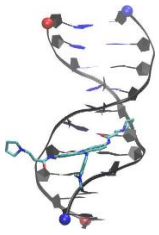
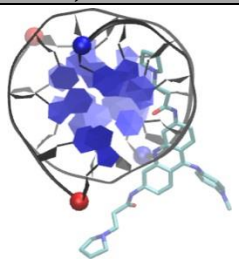
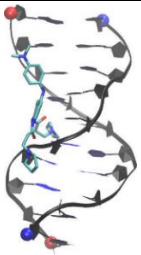
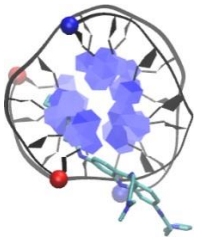
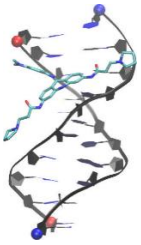
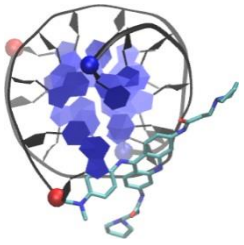
Front View	Description	Front View	Description
Run 01 (Groove)		Run 06 (Groove)	
			
Run 02 (Groove)		Run 07 (Groove)	
			
Run 03 (Groove)		Run 08 (Top)	
			
Run 04 (Bottom)		Run 09 (Groove)	
			
Run 05 (Groove)		Run 10 (Groove)	
			

Figure 48. Last snapshots of 10 DNA duplex and BRACO19 simulations. 5' and 3' of the telomeric DNA are indicated by a red and blue ball.

And the last snapshots of all the simulated trajectories of G-quadruplexes are listed in figure 49, 50 and 51. They indicate the stability of the structures; the G-tetrads were maintained and the K^+ ions retained their position in almost all the trajectories. It is to be noted that in figure 49, run 07 and 09 the K^+ ion moved out of the quadruplex and this disrupted the G-quadruplex. This will be discussed later.

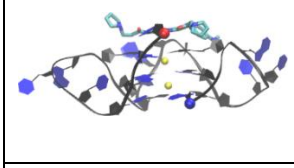
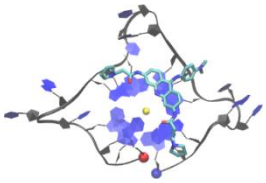
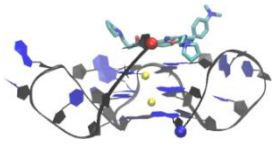
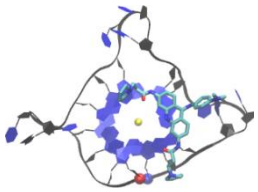
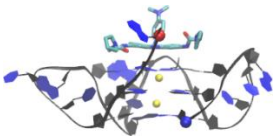
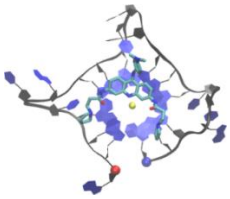
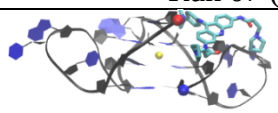
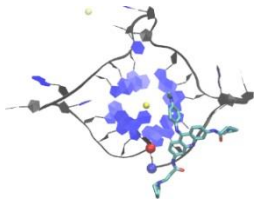
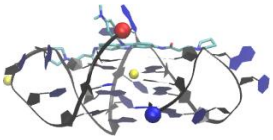
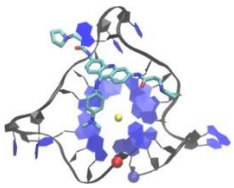
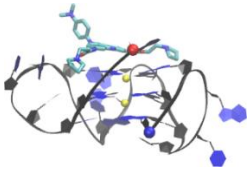
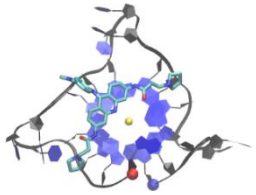
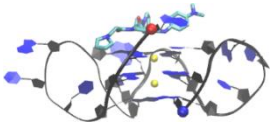
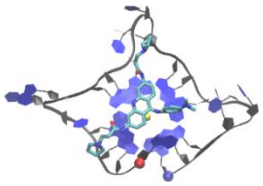
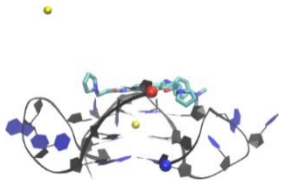
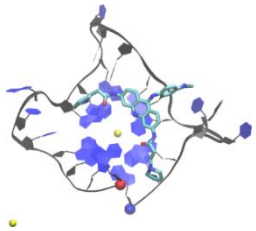
Front View	Description	Front View	Description
Run 01 (Top)		Run 06 (Top)	
			
Run 02 (Top)		Run 07 (Groove)	
			
Run 03 (Top)		Run 08 (Top)	
			
Run 04 (Top)		Run 09 (Top)	
			

Figure 49. Last snapshots of 10 parallel telomeric DNA G-quadruplex and BRACO19 simulations. 5' and 3' of the telomeric DNA G-quadruplex are indicated by a red and blue ball, respectively.

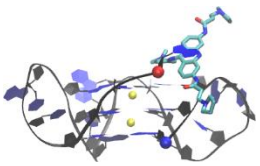
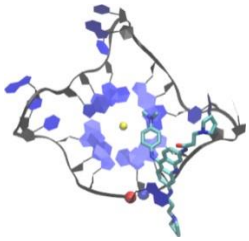
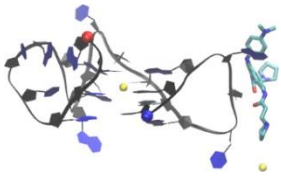
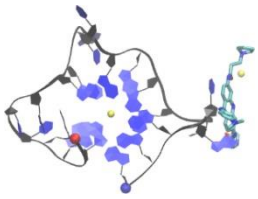
Run 05 (Groove)		Run 10 (Groove)	
			

Figure 49 (continued)

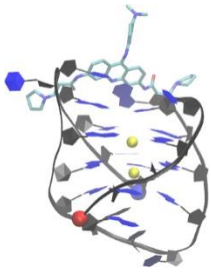
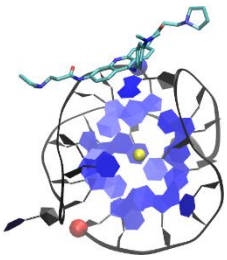
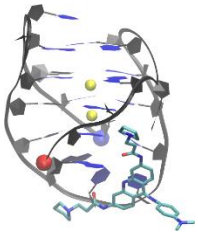
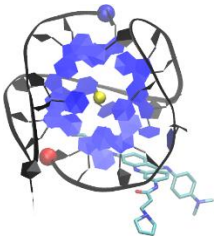
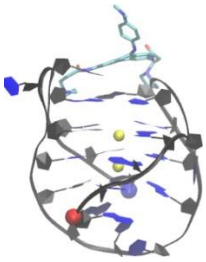
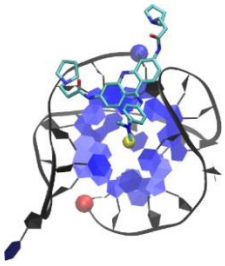
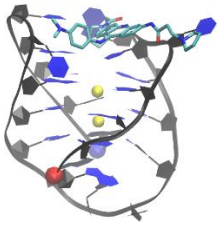
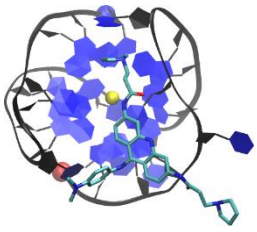
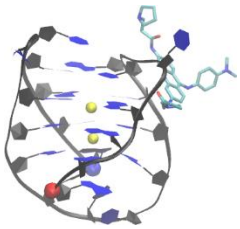
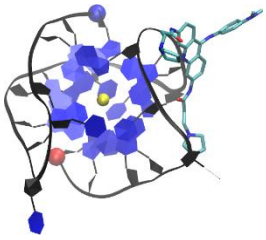
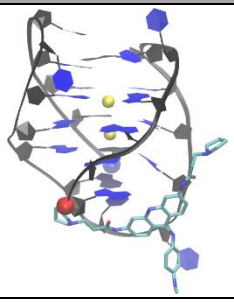
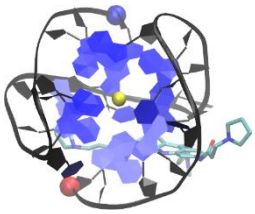
Front View	Description	Front View	Description
Run 01 (Side)		Run 06 (Side groove)	
			
Run 02 (Top)		Run 07 (Top)	
			
Run 03 (Side groove)		Run 08(Side groove)	
			

Figure 50. Last snapshots of 10 anti-parallel telomeric DNA G-quadruplex and BRACO19 simulations. 5' and 3' of the telomeric DNA G-quadruplex are indicated by a red and blue ball, respectively.

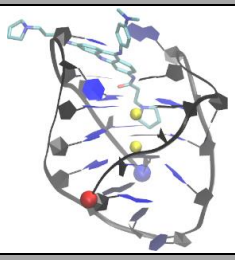
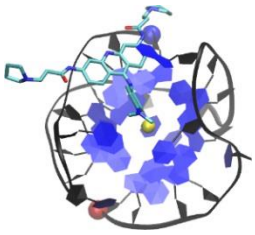
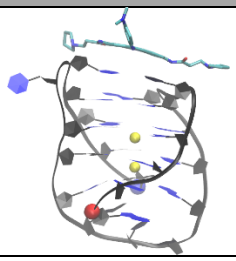
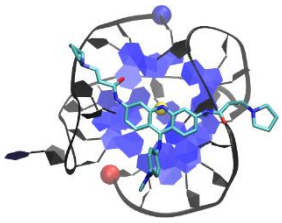
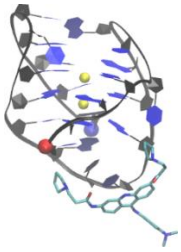
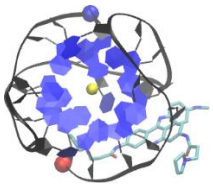
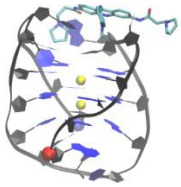
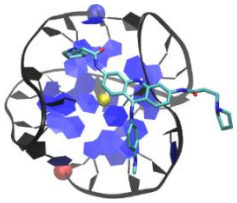
Run 04 (Top)		Run 09 (Top)	
			
Run 05 (Side groove)		Run 10 (Top)	
			

Figure 50 (continued)

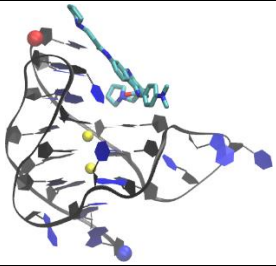
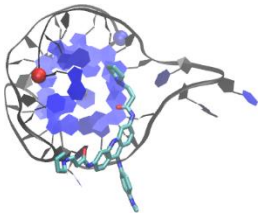
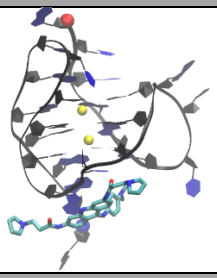
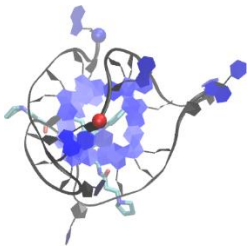
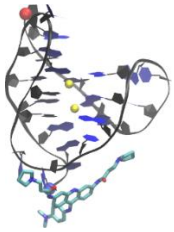
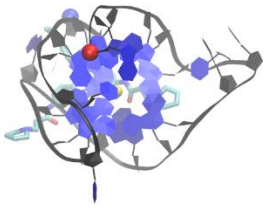
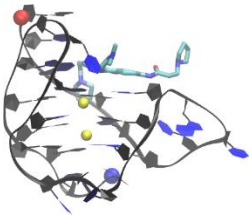
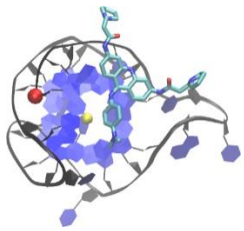
Front View	Description	Front View	Description
Run 01 (Top)		Run 06 (Bottom)	
			
Run 02 (Bottom)		Run 07 (Top)	
			

Figure 51. Last snapshots of 10 hybrid telomeric DNA G-quadruplex and BRACO19 simulations. 5' and 3' of the telomeric DNA G-quadruplex are indicated by a red and blue ball, respectively.

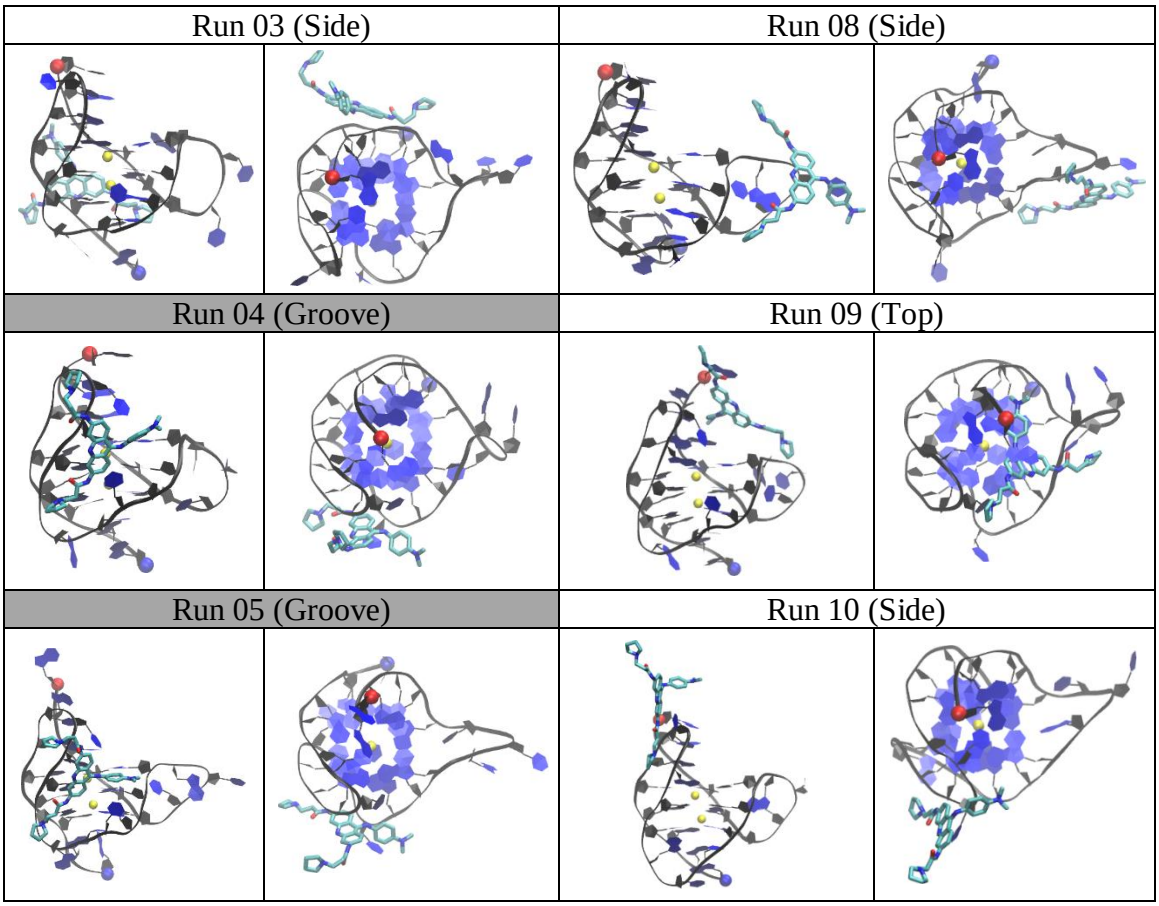


Figure 51 (continued)

Multiple binding sites were observed in the ten duplex DNA-BRACO19 trajectories. The clustering analysis described in the methods section was employed to categorize the stable complexes, extracted from these trajectories, into structural families. By setting a threshold of 1% population, 14 structural families of complexes were identified. These 8 structural families were further merged into three binding modes: groove binding, top stacking and bottom stacking. Binding to the groove of the duplex accounted for 81% of the total population. Additionally, end stacking to the top of the duplex accounted for 4% and end stacking to the bottom of the duplex made up 2% of the total population. Two binding modes were observed in the ten parallel G-quadruplex DNA-

BRACO19 trajectories. The same clustering analysis was employed to categorize the stable complexes, extracted from these trajectories, into 10 structural families. Two binding modes were observed: top stacking and groove binding. Top stacking to the parallel G-quadruplex DNA accounted for 56% and groove binding for 41% of the total population. Multiple binding sites were observed in the ten antiparallel G-quadruplex DNA-BRACO19 trajectories. The same clustering analysis was employed to categorize the stable complexes, extracted from these trajectories, into 9 structural families. Three binding modes were observed: top, bottom and groove binding. Bottom binding to the antiparallel G-quadruplex DNA accounted for 47%, top binding for 33% and groove binding for 21% of the total population. Multiple binding sites were observed in the ten hybrid G-quadruplex DNA-BRACO19 trajectories. The same clustering analysis was employed to categorize the stable complexes, extracted from these trajectories, into 12 structural families. Three binding modes were observed: top, groove and bottom binding. Groove binding to the hybrid G-quadruplex DNA accounted for 74%, Top binding for 19%, and bottom binding for 9% of the total population.

4.4.2 VDW interaction contributes most to the total binding energy, ranking the binding poses for each DNA-ligand system.

Table 114

MM-GBSA binding energy (kcal/mol) of BRACO19 to human telomeric DNA duplex and Quadruplexes.

System	Pose	[9]ΔE_{VDW}	[10]ΔE_{SUR}	[11]ΔE_{GBELE}	[12]ΔE_{CONF}	[13]ΔE_{TOT}	[14]$\Delta\Delta E_{TT}$
DNA Duplex	[15]TS	-28.9±4.0	-2.3±0.5	-5.2±3.1	2.7±1.2	-33.7±5.3	28.6
	[16]BB	-28.4±4.1	-2.2±0.5	-4.8±3.2	0.8±3.8	-34.6±5.7	27.7
	[17]GB	-57.9±9.5	-5.2±0.7	-3.9±4.5	5.2±2.8	-61.7±8.0	0.6
Parallel Quad	TS	-63.1±5.7	-5.2±0.6	-1.9±4.4	7.9±5.2	-62.3±4.5	0
	GB	-37.0±6.4	-3.1±0.4	-8.9±4.5	11.4±4.5	-37.6±7.2	24.7
Anti-Parallel Quad	TS	-41.5±11.4	-4.0±1.1	-8.5±4.3	0.1±4.6	-53.9±5.8	8.4
	BB	-29.1±9.0	-2.5±0.8	-9.2±3.0	-2.1±4.4	-42.8±4.1	19.5
	GB	-43.0±6.0	-3.4±0.5	-7.5±2.6	10.9±2.3	-43.1±7.2	19.2
Hybrid Quad	TS	-44.2±11.4	-4.3±1.0	-12.1±5.0	20.0±9.2	-40.5±5.4	21.8
	BB	-25.7±5.8	-2.8±0.7	-16.3±6.0	15.8±8.5	-29±12.9	33.3
	GB	-40.5±6.6	-4.0±0.5	-14.9±5.2	23.7±3.6	-35.7±5.1	26.6

MM-GBSA binding energy calculations were carried out as depicted in methods section to examine the relative binding affinities major binding modes of BRACO19 with respect to DNA and summarized in Table 114. Of the three binding modes of BRACO19 to the DNA duplex, the best binding energy was in the groove binding mode (-69.5±8.0 kcal/mol), followed by the top stacking mode (-34.8±5.3 kcal/mol).

[9] Change of van der Waals energy in gas phase upon complex formation
[10] Change of surface area term change upon complex formation
[11] Change of GBELE generalized Born term + gas phase electrostatic energy upon complex formation
[12] Change of conformational energy upon complex formation
[13] Change of total potential energy in water upon complex formation (VDW+SUR+GBELE+CONF)
[14] Change in binding energy with a reference to top stacking parallel G-quadruplex
[15] Top Stacking
[16] Bottom Binding
[17] Groove Binding

VDW packing responsible for the VDW energy contribution governed the binding energy order of the three modes. The VDW contribution in the groove mode (-59.8±5.5 kcal/mol) points out that in the groove binding mode, one side of BRACO19 was still exposed to solvent. BRACO19 bound to parallel G-quadruplex DNA in two binding poses. Top stacking (-72.1±4.5kcal/mol) was the most, making the groove binding (-40.3±7.2 kcal/mol) being the least stable of the two. BRACO19 bound to antiparallel G-quadruplex DNA in three binding poses. Top binding (-60.9±5.8 kcal/mol) was the most stable of the three with groove binding exhibiting a binding energy of only -57.2±7.2kcal/mol. Bottom binding was the lowest with a binding energy of -45.2±4.1kcal/mol. BRACO19 bound to hybrid G-quadruplex DNA in three binding poses as well. Top binding (-63.4±5.4 kcal/mol) was the most stable of the three, followed by groove binding (-56.8±5.1 kcal/mol) and bottom binding (-55.3±12.9 kcal/mol).

4.4.3 BRACO19 binds to the duplex DNA, without inducing structural changes.

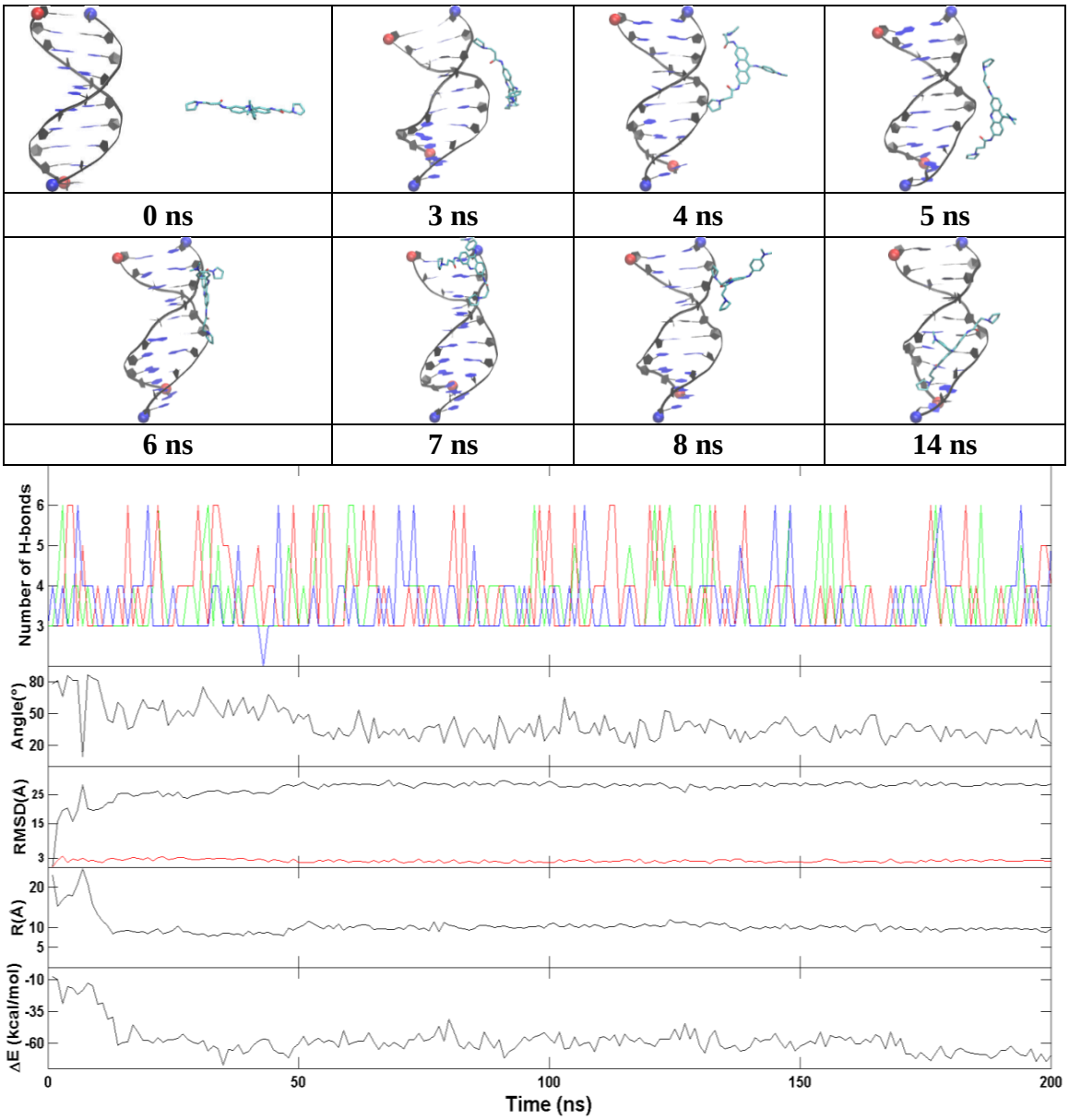


Figure 52. A representative groove binding trajectory of the DNA duplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first base pair (green), second base pair (red) and third base pair (blue) layers of the DNA structure (figure 45), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

The representative trajectory for the groove binding of BRACO19 to the duplex DNA are characterized in Figure 52. In all ten binding trajectories, the DNA showed low structural fluctuation with RMSD of 2.4 Å (Figure 62) and the hydrogen bonds between the base pairs were maintained. In the representative trajectory of BRACO19 binding to the groove of the human telomeric duplex DNA in figure 52, an initial interaction was observed as early as 3 ns and the final binding pose was achieved at an astounding 14 ns and was maintained throughout the remainder of the trajectory. The limited fluctuation in the five order parameters explains the limited structural dynamics. The other representative trajectories of BRACO19 top stacking, groove binding, bottom stacking also exhibited rapid binding and limited dynamics, binding to the complex at 8 ns and 19 ns respectively.

4.4.4 Groove binding might be an intermediate state for the top stacking mode

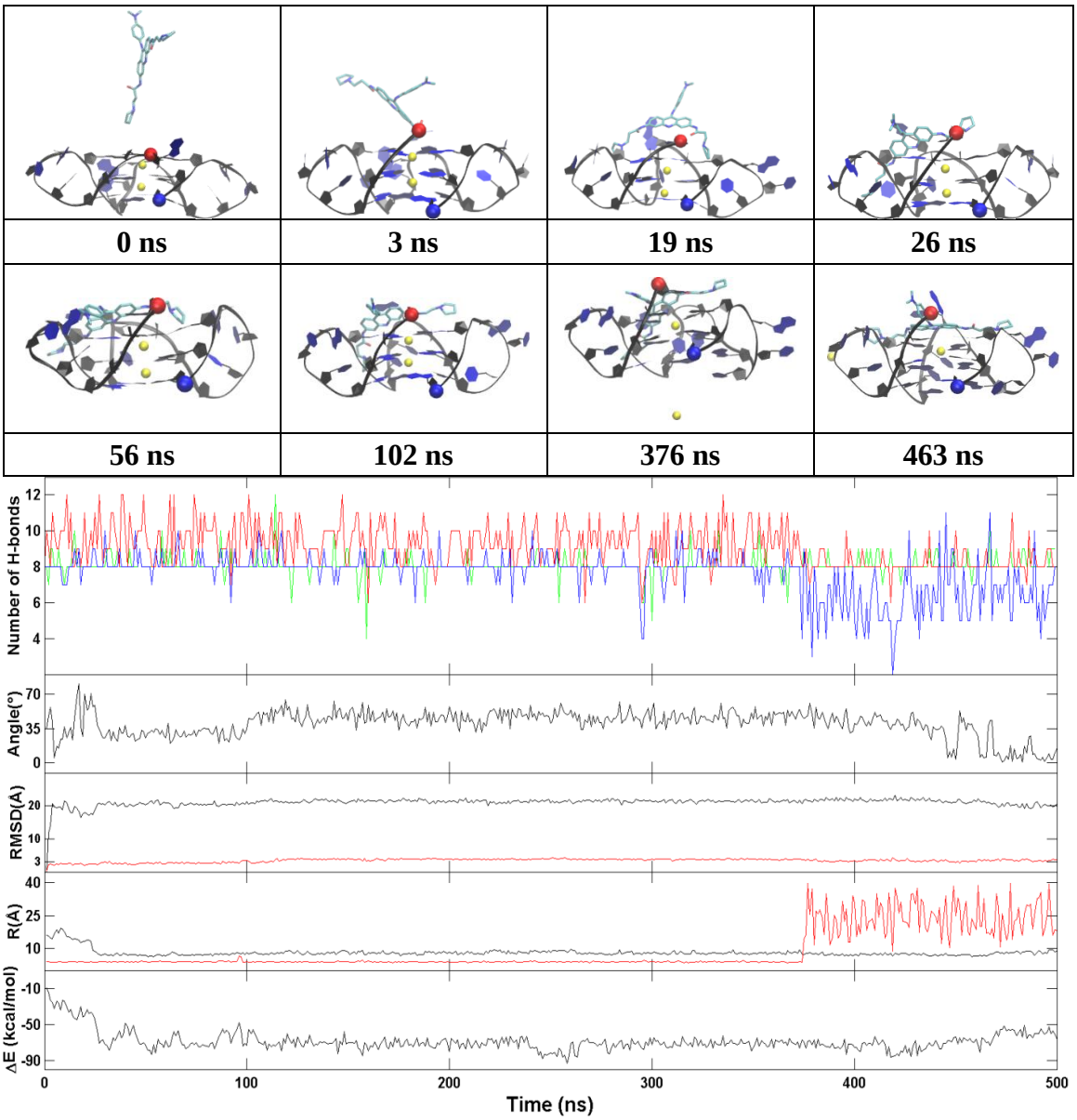


Figure 53. A representative top stacking trajectory of the parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first G4 (green), second G4 (red) and third G4 (blue) layers of the DNA structure (Figure 45), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

The representative trajectory for the top stacking mode of BRACO19 to the parallel human telomeric G-quadruplex DNA are characterized in figure 53. In all ten binding trajectories, the DNA showed low structural fluctuation with RMSD of 2.4 Å (Figure 64) and the hydrogen bonds in the three G-tetrads were maintained. In the representative trajectory of BRACO19 binding to the top of the human telomeric parallel G-quadruplex DNA in figure 53, an initial interaction to the complex at 8 ns and attaining the stable groove binding pose at 19 ns. However, on further simulation to 500 ns, at exactly 463 ns BRACO19 was stacked on top of the parallel G-quadruplex. It can be inferred that groove binding is an intermediate state for top stacking mode. This further simulation also showed that the potassium ion from the G-quadruplex moved out followed by the disruption of the third G-tetrad layer of the G- quadruplex inferring that the K+ ions are essential for the stability of the G-quadruplex. The binding energy for top stacking fluctuated between -60 and -75 kcal/mol while bottom stacking varied between -35 and -45 kcal/mol after attaining the steady binding pose.

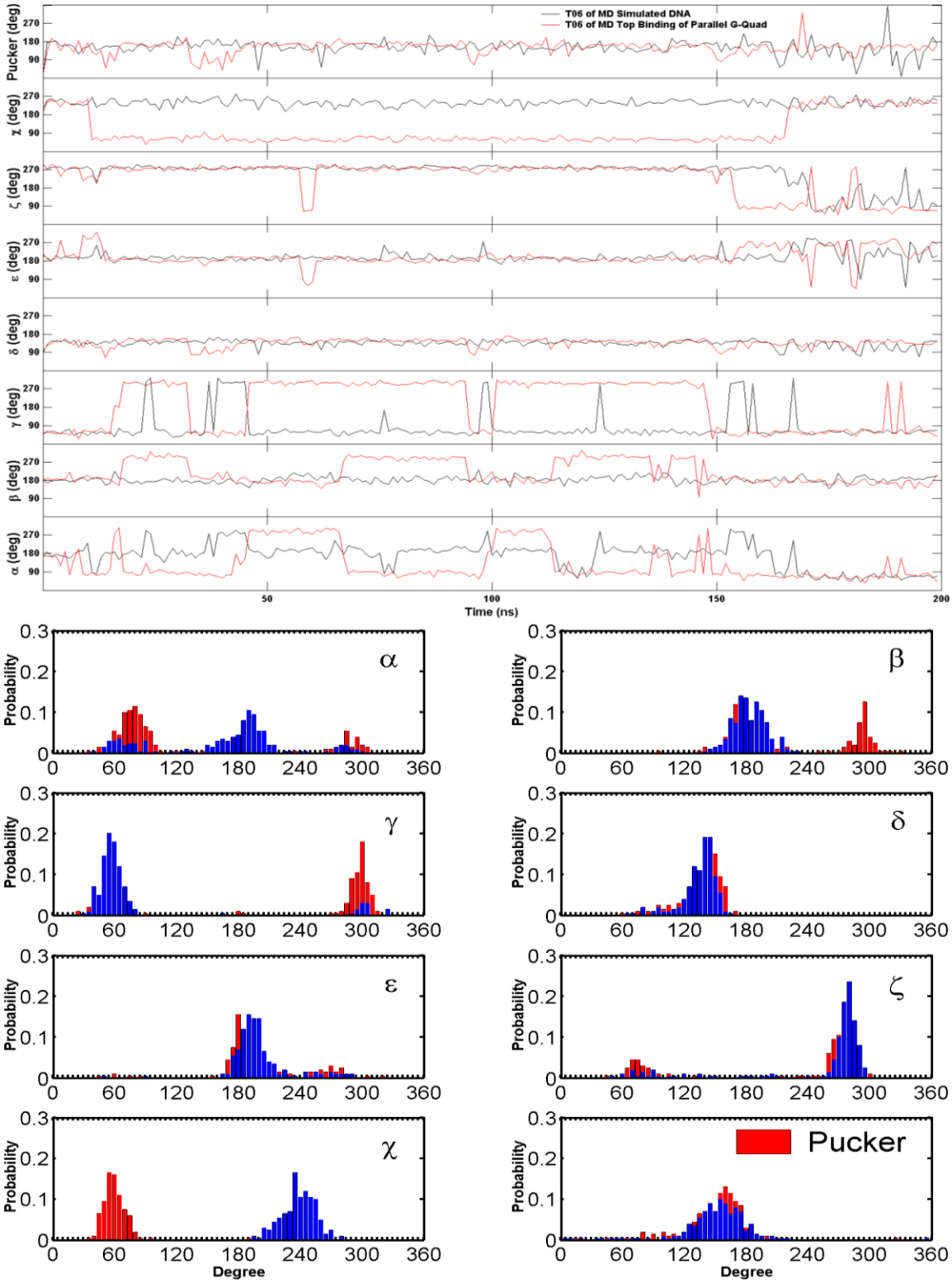


Figure 54. Comparison for the backbone torsion angles of residue T6 between the free ligand binding simulation (red) of the top stacking mode of the parallel G-quadruplex and the stability simulation of the crystal structure (black) of the parallel G-quadruplex. **Top:** Time series, **Bottom:** Histograms.

Dihedral angles of all DNA bases in the simulated parallel G-quadruplex were analyzed. The dihedral angles of the G-tetrads in free ligand binding simulations indicate low fluctuations and are consistent through the binding process. Major fluctuations were observed in the terminal residues, T6 in particular is discussed here as it demonstrates highest fluctuation. T6 flipped out at 15 ns and flipped back at 45 ns, flipped out at 69 ns and flipped in at 100 ns and it finally flipped out at 114 ns and remained same throughout the rest of the trajectory. This flipping out of the base is mainly characterized by α, β, γ and χ (figure 54).

4.4.5 BRACO19 binds to the anti-parallel G-Quadruplex, without inducing structural changes.

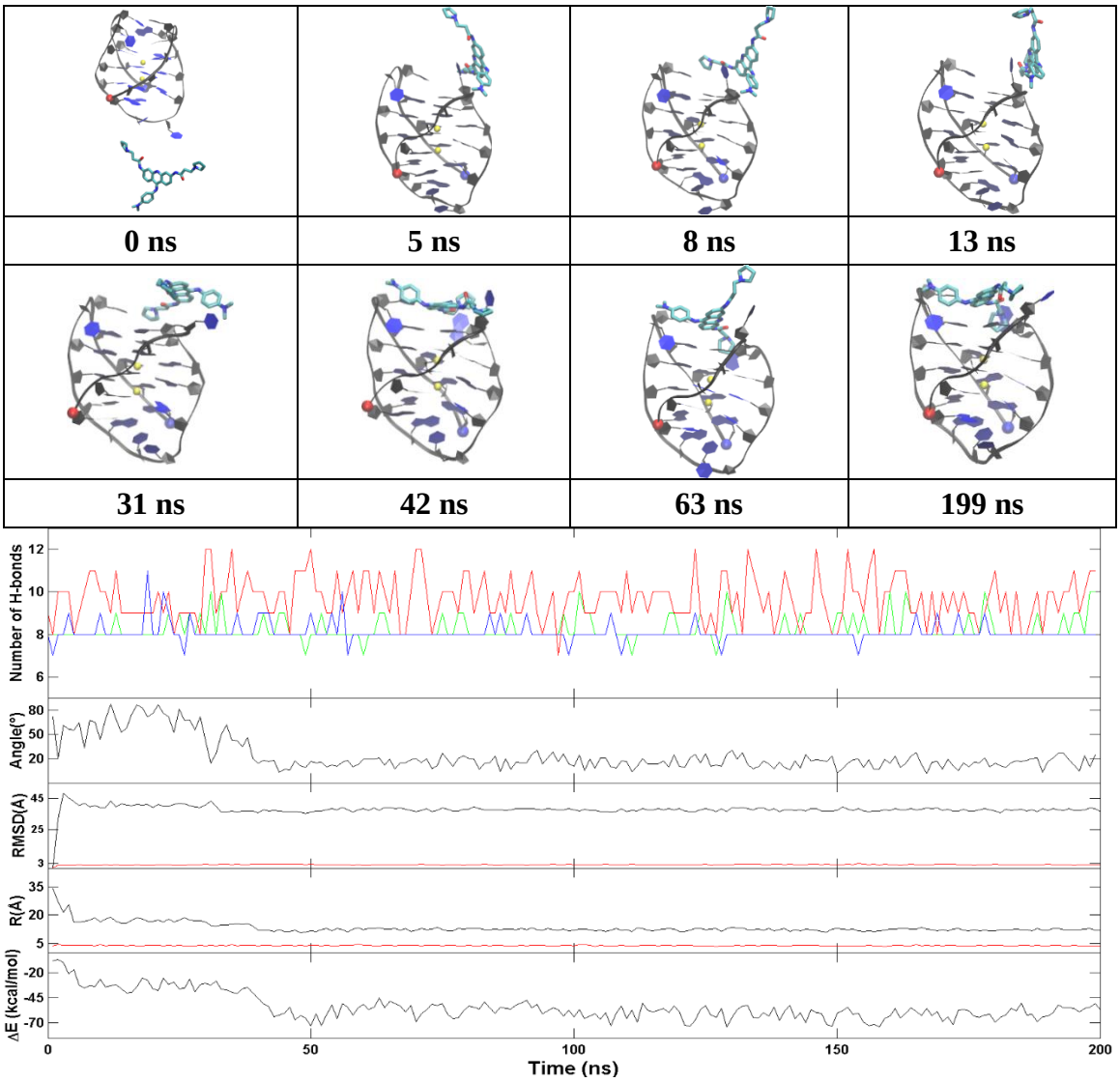


Figure 55. A representative top stacking trajectory of the anti-parallel G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 45), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

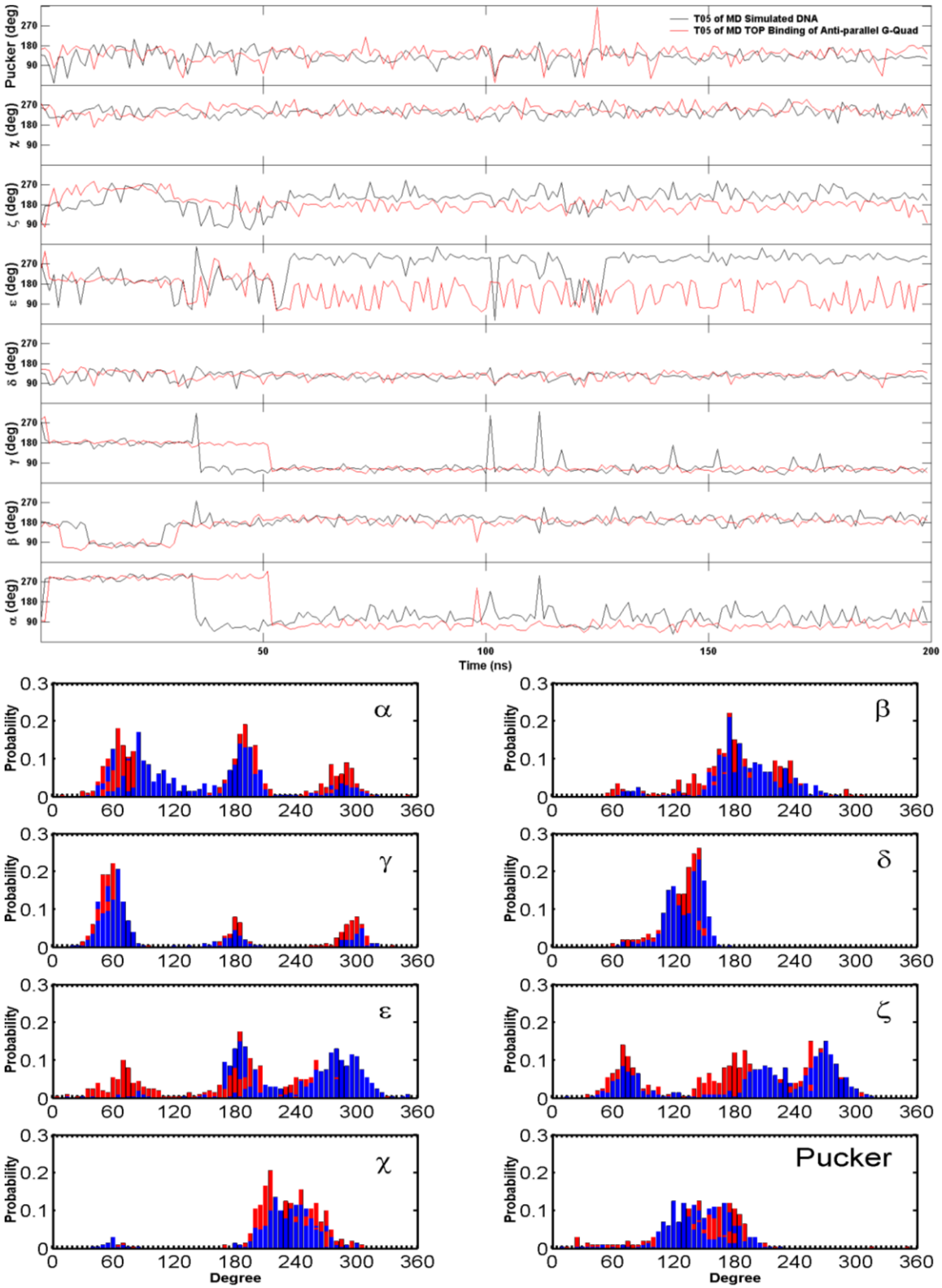


Figure 56. Comparison for the backbone torsion angles of residue T05 between the free ligand binding simulation (red) of the top stacking trajectory of the anti-parallel G-quadruplex and the stability simulation of the crystal structure (black) of the anti-parallel G-quadruplex. **Top:** Time series, **Bottom:** Histograms.

The representative trajectory for the top stacking mode of BRACO19 to the antiparallel human telomeric G-quadruplex DNA are characterized in figure 55. In all ten binding trajectories, the DNA showed high structural fluctuation in four trajectories with RMSD of 3.2 Å (Figure 66), the hydrogen bonds in the three G-tetrads were maintained and the distance between K^+ ions remained stable in all trajectories. The representative trajectories of top stacking of BRACO19 with the human telomeric antiparallel G-quadruplex DNA, showed an initial interaction at an early 5 ns and the final binding pose was achieved at an astonishing 42 ns and was maintained throughout the rest of the trajectory. The limited structural dynamics was explained by the limited fluctuation in the five order parameters. The representative trajectories of the groove binding and bottom binding are similar to the top binding trajectory with a rapid binding and limited fluctuation of order parameters. Early interaction at 1 and 5 ns respectively and attainment of final binding pose by 16 and 55 ns respectively. The other representative trajectories of BRACO19 top stacking, groove binding and bottom stacking also exhibited rapid binding and limited dynamics, binding to the complex at 19, 5 and 2 ns respectively and attaining the final binding pose at 117, 143 and 107 ns respectively and maintained it throughout the rest of the trajectories. The binding energy for top stacking and groove binding fluctuated between -55 and -65 kcal/mol while bottom stacking varied between -40 and -50 kcal/mol after attaining the steady binding pose.

Dihedral angles of all DNA bases in the simulated anti-parallel G-quadruplex were analyzed. The dihedral angles of the G-tetrads in free ligand binding simulations indicate low fluctuations and are consistent through the binding process. Major fluctuations were observed in the terminal residues, T5 in particular is discussed here as it demonstrates

highest fluctuation. Through the binding process BRACO19 majorly interacted with T5, it opened up as BRACO19 approached and at 29 ns, flipped out to let BRACO19 in, flipped back at 40 ns and it stayed open afterward while interacting with BRACO19. This flipping out of the base is mainly characterized by ε and ζ (figure 56).

4.4.6 BRACO19 binds to the hybrid telomeric G-Quadruplex DNA, without inducing structural fluctuation. The representative trajectory for the top stacking of BRACO19 with respect to the hybrid human telomeric G-quadruplex DNA are characterized in figure 57. In all ten binding trajectories, the DNA showed high structural fluctuation in five trajectories with RMSD of 2.9 Å (Figure 68), the hydrogen bonds in the three G-tetrads were maintained and the distance between K^+ ions remained stable in all trajectories. The representative trajectory of BRACO19 top stacking onto the hybrid G-quadruplex DNA showed an initial interaction at 3 ns and the final binding pose was attained as early as 30 ns and was maintained throughout the rest of the trajectory. The limited structural dynamics was explained by the limited fluctuation in the five order parameters. The representative trajectories of the groove binding and bottom binding are similar to the top binding trajectory with a rapid binding and limited fluctuation of order parameters. Early interaction at 2 and 9 ns respectively and final binding pose was attained by 13 and 51 ns respectively. The other representative trajectories of BRACO19 top stacking, groove binding and bottom stacking also exhibited rapid binding and limited dynamics. The binding energy for all binding modes varied between -55 and -65 kcal/mol after attaining the steady binding pose.

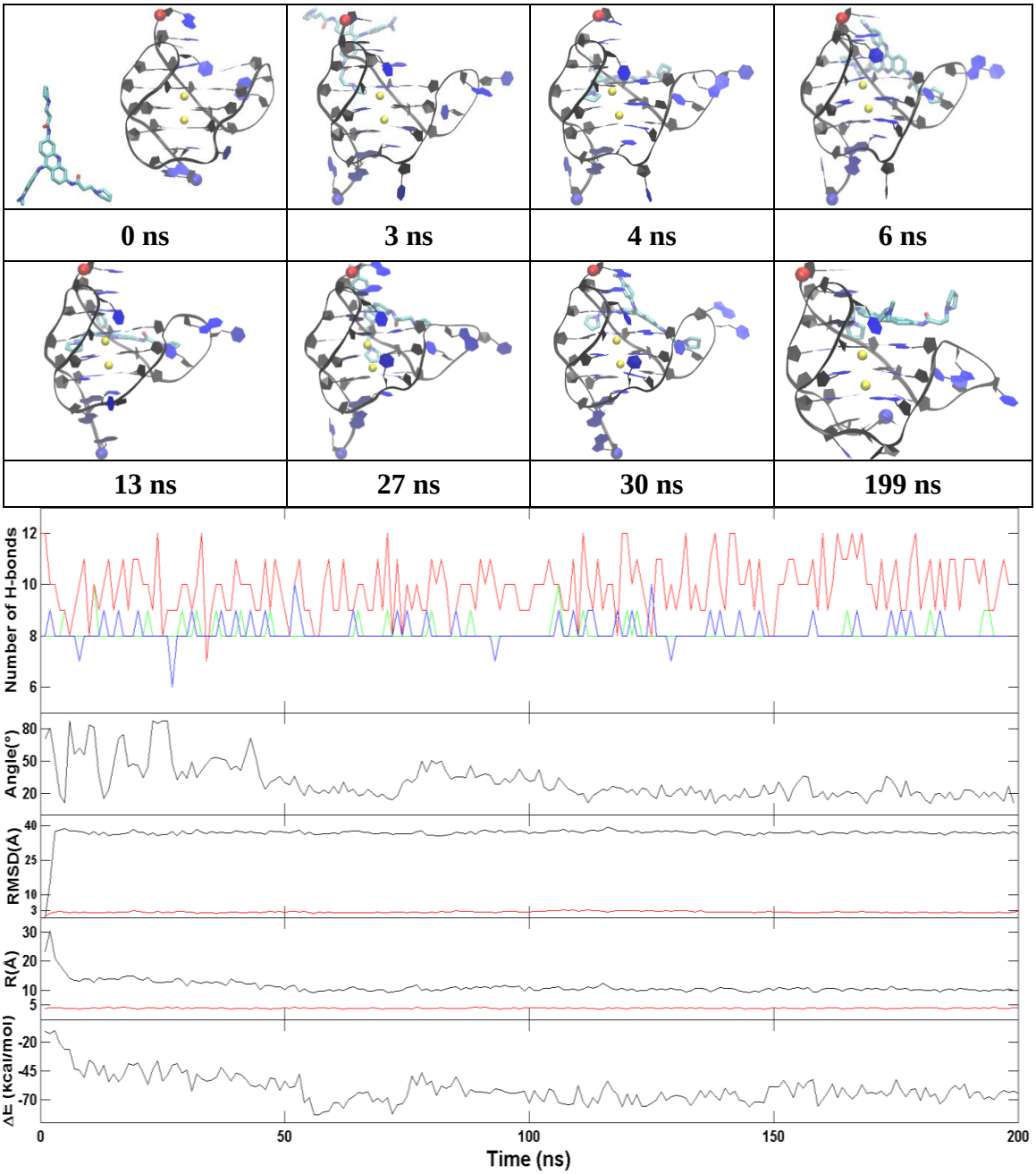


Figure 57. A representative top binding trajectory of the hybrid G-quadruplex. **Top:** Representative structures with time annotation. 5' and 3' are indicated by a red and blue ball, respectively. Residues 1, 2, 13, 14 are indicated in purple and residues 12, 24 are indicated in red and the K^+ ions are represented in yellow. **Bottom:** An order parameter plot depicting number of hydrogen bonds present in first (red), second G4 (cyan), third G4 (blue), fourth G4 (black) and fifth (green) layers of the DNA structure (Figure 45), the drug-base dihedral angle, receptor (red) and ligand (black) RMSD relative to the original crystal pose, center-to-center distance (R/black) and K^+-K^+ distance (R/red) and MM-GBSA binding energy (ΔE) (cf. methods section for definition).

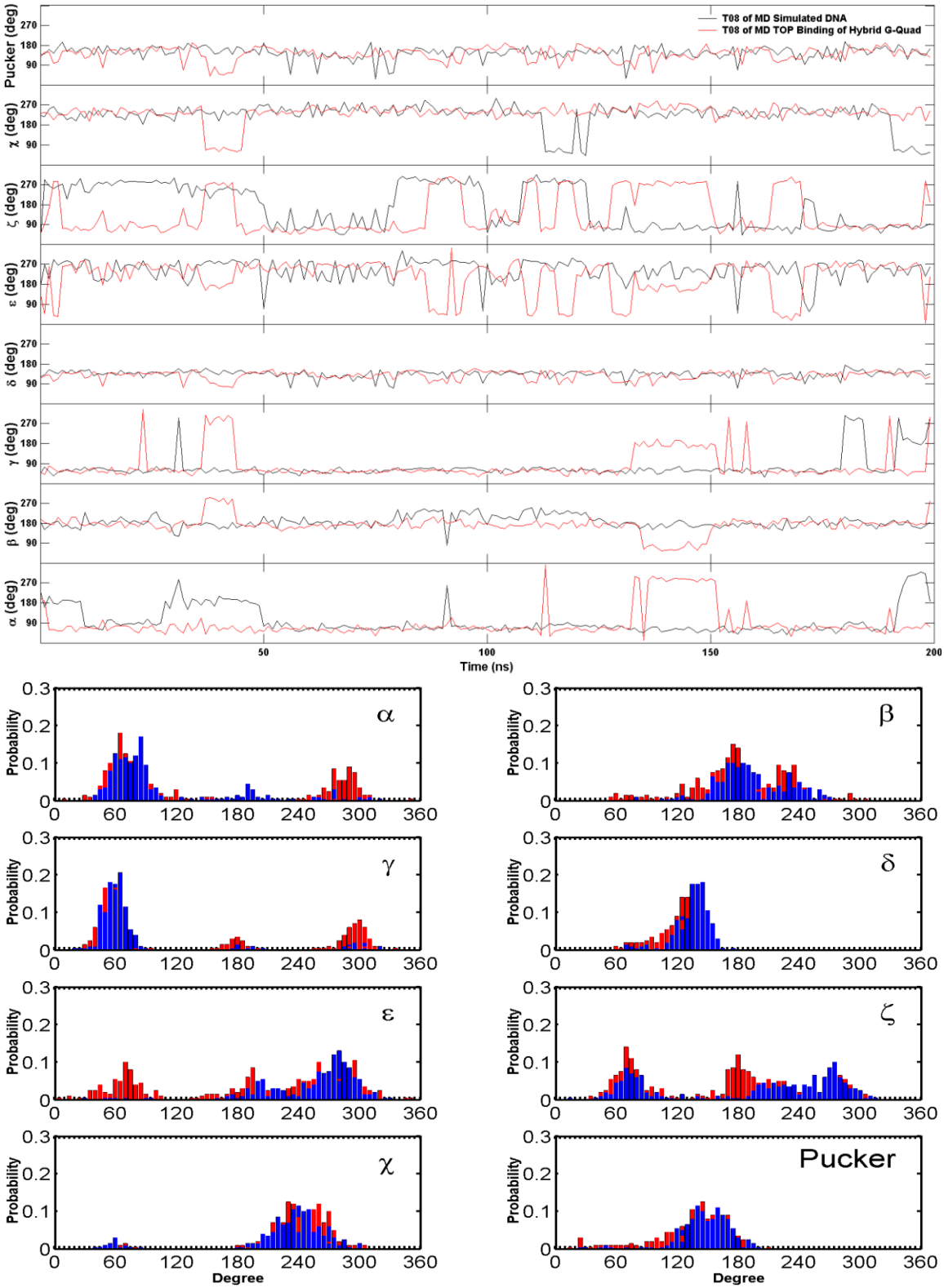


Figure 58. Comparison for the backbone torsion angles of residue T8 between the free ligand binding simulation (red) of the top binding trajectory of the hybrid G-quadruplex and the stability simulation of the crystal structure (black) of the hybrid G-quadruplex. **Top:** Time series, **Bottom:** Histograms.

Dihedral angles of all DNA bases in the simulated hybrid G-quadruplex were analyzed. The dihedral angles of the G-tetrads in free ligand binding simulations indicate low fluctuations and are consistent through the binding process. Major fluctuations were observed in the terminal residues, T8 in particular is discussed here as it demonstrates highest fluctuation. T8 flipped out upon simulation and remained flipped through the rest of the simulation except for a few ns after 38 ns and 143 ns. This flipping out of the base is mainly characterized by α, δ, ε and ζ (figure 58).

4.5 Discussion

After the recent discovery of the greater existence of G-quadruplex in malignant tumors than in normal tissues interest in G-quadruplex DNA as a promising target for cancer therapeutics has increased. BRACO19, one of the most effective G-quadruplex binding ligands, is a promising anticancer drug candidate, yet its low preferential binding affinity (about ~62-fold) to the telomeric single-stranded G-quadruplex DNA over duplex DNA remains to be enhanced. For better molecular insights, the binding of BRACO19 to a duplex 20mer DNA ($d([GC]_{10})_2$) and to the parallel, antiparallel and hybrid telomeric G-quadruplexes was investigated in this study using binding molecular dynamics simulations with a free ligand.

Out of various binding modes for each system, the MM-GBSA binding energy calculations showed that the most stable binding pose was the groove binding mode for the duplex and the top stacking mode for the parallel G-quadruplex, the antiparallel G-quadruplex and the hybrid G-quadruplex (figure 45). The order of the relative binding energy of BRACO19 in these most stable poses are as follows: -72.1±4.5 kcal/mol; the top stacking to the parallel G-quadruplex (ΔΔE=0 kcal/mol) > -69.5±8.0 kcal/mol; the groove

binding to the duplex DNA (2.6 kcal/mol) > -63.4±5.4 kcal/mol; the top stacking to the hybrid G-quadruplex (8.7 kcal/mol) > -60.9±5.8 kcal/mol; the top stacking to antiparallel G-quadruplex (11.2 kcal/mol).

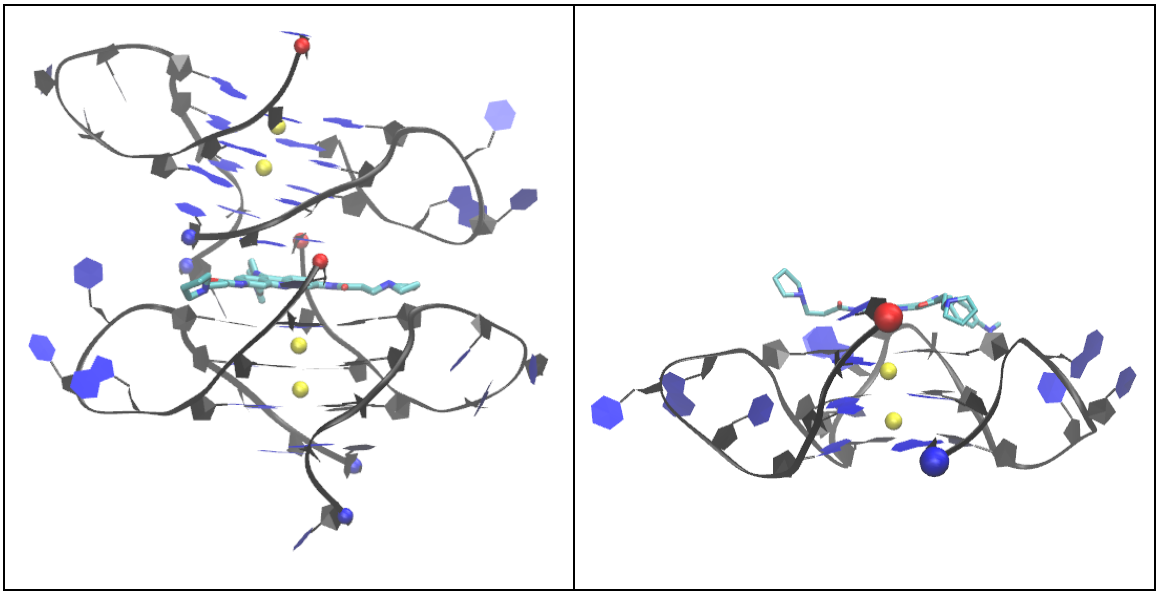


Figure 59. (A) The experimental binding mode (PDB ID: 3CE5) of double stranded parallel telomeric DNA G-quadruplex (B) Major binding pose of simulated single stranded parallel telomeric DNA G-quadruplex.

The combination of the long time (1 μs) stability and the large magnitude of these binding energies suggests an enthalpy driven binding is likely and the contribution of entropy to the binding free energy to be of minor importance. Analysis by breaking down the binding energy indicated that the VDW term makes the biggest contribution to the total binding energy (Table 114). This indication suggests introducing target or drug specific packing optimization as a prospect for further stabilization of the G-quadruplex. If these binding modes have comparable entropic energies and the parallel G-quadruplex is the major telomeric G-quadruplex species, then our relative binding energy signifies that

BRACO19 binds preferentially to the telomeric G-quadruplexes than to the DNA duplex. This qualitatively explains the experimental observation of preferential binding affinity difference of BRACO19 on the two DNA forms. The emphasis is being directed to the fact that the binding energies of the groove binding mode of the duplex and top stacking mode of the G-quadruplex are comparable. This rationalizes the lack of binding selectivity of BRACO19 to the two DNA forms. For that reason, it can be suggested that a ligand modification that destabilizes the duplex groove binding mode but stabilizes the G-quadruplex top stacking mode will enhances the binding selectivity of the ligand. For example, adding a planar ring fragment to the acridine would facilitate the top stacking rather than groove binding and increase the van deer Waals interactions there for increasing selectivity and binding affinity of the prospective drug towards the G-quadruplex. This suggestion is consistent with the original SAR data in the development of BRACO19 from prototype BSU6048 in which the addition of the ring at position 9 (makings of BRACO19) increased the drug selectivity from 10-fold to 62 fold towards human telomeric G-quadruplexes over duplex DNA. (Harrison, Gowan, Kelland, & Neidle, 1999; White et al., 2007; Yang & Okamoto, 2010a) And it is also to be noted that the sidechains on 3 and 6 contribute to the groove binding of both DNA duplex and G-quadruplex which could be the reason behind low selectivity. So, suggestions can be made to reduce the length of these side chains. These side chains exist in protonated form at physiological pH however, Table 114 indicates that the contribution of electrostatic interactions to the binding affinity is very low and therefore modifications can be suggested to the substituents at 3rd and 6th position of the acridine. Modifications such as loss of positive charge which would increase the

hydrophobicity which could in fact increase the van der Waals interactions and reduction of the length of the side chains.

The most stable binding mode of BRACO19 to the DNA duplex is the groove binding and the single stranded G-quadruplexes is top stacking mode, which is evidently similar to the binding pose in the only X-ray solved crystal structure of a double stranded G-quadruplex in complex of BRACO19 (Figure 70). The plane of BRACO19 is parallel to the plane of G-tetrads. However, molecular details are different. In the groove binding mode of duplex and the top stacking mode of parallel G-quadruplex, antiparallel and hybrid G-quadruplex only one side of BRACO19 molecule interacted with the DNA. Lastly, the groove binding mode was observed to be an intermediate stage of top stacking mode. The dihedral angles of the 3 G-tetrads in free ligand binding simulations indicate low fluctuations and are consistent through the binding process. The torsion angle analysis indicated that the conformational changes are characterized mainly by α, β, γ, ε, and ζ and in some cases changes in χ dihedral angle. Significant overlap between in the histogram of free ligand binding simulation and crystal pose simulation indicate good prediction of the torsion angle from MD simulation.

4.6 Conclusion

The detailed structural knowledge of the intramolecular human telomeric G-quadruplexes in complex with a ligand is required for the rational design of human telomeric G-quadruplex binding drugs. In this study, molecular dynamics binding simulations were used to probe and understand the binding nature of BRACO19, a potent human telomeric G-quadruplex drug, to a B-DNA duplex and the three scaffolds of a single stranded human telomeric G-quadruplex. The most stable binding mode indicated by the MM-GBSA binding energy analysis for the duplex DNA is the groove binding mode and top stacking for parallel G-quadruplex, antiparallel and hybrid G-quadruplexes. The similar binding affinity of BRACO19's groove binding mode with respect to both the duplex and the G-quadruplexes explains its lack of preferential binding selectivity. Therefore, a ligand modification that destabilizes the duplex groove binding mode but stabilizes the G-quadruplex top stacking mode will improve the binding selectivity of the ligand. Our study presents a successful example of the ability of molecular dynamic simulations with the latest AMBER force field to facilitate detailed structural and dynamic information which will further decipher the binding nature of DNA ligands.

Publications Note

The study illustrated in Chapter 2, CADD in Boron therapeutics is under preparation for publication

The study illustrated in Chapter 3, probing the binding mechanism of BRACO19 and human telomeric DNA G-quadruplex is accepted for publication by Journal of Chemical Information and Modelling

The study illustrated in Chapter 4, Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent is under preparation for submission to Physical Chemistry and Chemical Physics.

References

Adamczyk-Woźniak, A., Komarovska-Porokhnyavets, O., Misterkiewicz, B., Novikov, V. P., & Sporzyński, A. (2012). Biological activity of selected boronic acids and their derivatives. *Applied Organometallic Chemistry, 26*(7), 390-393. doi: 10.1002/aoc.2880

Adams, J., Behnke, M., Chen, S., Cruickshank, A. A., Dick, L. R., Grenier, L., . . . Stein, R. L. (1998). Potent and selective inhibitors of the proteasome: Dipeptidyl boronic acids. *Bioorganic & Medicinal Chemistry Letters, 8*(4), 333-338. doi: http://dx.doi.org/10.1016/S0960-894X(98)00029-8

Ajay, & Murcko, M. A. (1995). Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes. *Journal of Medicinal Chemistry, 38*(26), 4953-4967. doi: 10.1021/jm00026a001

Akagi, T., & Kimoto, T. (1976). Human Cell Line (Hgc-27) Derived from Metastatic Lymph-Node of Gastric Cancer. *Acta Med. Okayama, 30*, 215-219.

Akhshi, P., Mosey, N. J., & Wu, G. (2012). Free-Energy Landscapes of Ion Movement Through a G-Quadruplex DNA Channel. *Angew. Chem., 51*, 2850-2854. doi: 10.1002/anie.201107700

Alcaro, S., Musetti, C., Distinto, S., Casatti, M., Zagotto, G., Artese, A., . . . Sissi, C. (2013). Identification and Characterization of New DNA G-Quadruplex Binders Selected by a Combination of Ligand and Structure-Based Virtual Screening Approaches. *Journal of Medicinal Chemistry, 56*(3), 843-855. doi: 10.1021/jm3013486

Alizadehnohi, M., Nabiuni, M., Nazari, Z., Safaeinejad, Z., & Irian, S. (2012). The Synergistic Cytotoxic Effect of Cisplatin and Honey Bee Venom on Human Ovarian Cancer Cell Line A2780cp. *J. Venom Res., 3*, 22-27.

Asai, A., Oshima, Y., Yamamoto, Y., Uochi, T.-a., Kusaka, H., Akinaga, S., . . . Gryaznov, S. (2003). A Novel Telomerase Template Antagonist (GRN163) as a Potential Anticancer Agent. *Cancer Research, 63*(14), 3931.

Baker, S. J., Ding, C. Z., Akama, T., Zhang, Y.-K., Hernandez, V., & Xia, Y. (2009). Therapeutic potential of boron-containing compounds. *Future Medicinal Chemistry, 1*(7), 1275-1288. doi: 10.4155/fmc.09.71

Baker, S. J., Tomsho, J. W., & Benkovic, S. J. (2011). Boron-containing inhibitors of synthetases. *Chemical Society Reviews, 40*(8), 4279-4285. doi: 10.1039/C0CS00131G

Baker, S. J., Zhang, Y.-K., Akama, T., Lau, A., Zhou, H., Hernandez, V., . . . Plattner, J. J. (2006). Discovery of a New Boron-Containing Antifungal Agent, 5-Fluoro-1,3-dihydro-1-hydroxy-2,1- benzoxaborole (AN2690), for the Potential Treatment of Onychomycosis. *Journal of Medicinal Chemistry, 49*(15), 4447-4450. doi: 10.1021/jm0603724

Balasubramanian, S., Hurley, L. H., & Neidle, S. (2011). Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discovery, 10*, 261-275. doi: 10.1038/nrd3428

Banik, S. D., & Nandi, N. (2013). Chirality and Protein Biosynthesis. In P. Cintas (Ed.), *Biochirality: Origins, Evolution and Molecular Recognition* (pp. 255-305). Berlin, Heidelberg: Springer Berlin Heidelberg.

Barnard, J. M., & Downs, G. M. (1992). Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures. *Journal of Chemical Information and Computer Sciences, 32*(6), 644-649. doi: 10.1021/ci00010a010

Bayly, C. I., Cieplak, P., Cornell, W. D., & Kollman, P. A. (1993a). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem., 97*(40), 10269-10280.

Bechter, O. E., Zou, Y., Walker, W., Wright, W. E., & Shay, J. W. (2004). Telomeric Recombination in Mismatch Repair Deficient Human Colon Cancer Cells after Telomerase Inhibition. *Cancer Research, 64*(10), 3444.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*(1), 235-242. doi: 10.1093/nar/28.1.235

Beveridge, D. L., & DiCapua, F. M. (1989). Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annual Review of Biophysics and Biophysical Chemistry, 18*(1), 431-492. doi: 10.1146/annurev.bb.18.060189.002243

Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.*, *5*, 182-186. doi: 10.1038/nchem.1548

Böhm, H.-J., & Stahl, M. (2003). The Use of Scoring Functions in Drug Discovery Applications *Reviews in Computational Chemistry* (pp. 41-87): John Wiley & Sons, Inc.

Brandes, L. J., & Hermonat, M. W. (1983). Receptor Status and Subsequent Sensitivity of Subclones of MCF-7 Human Breast Cancer Cells Surviving Exposure to Diethylstilbestrol. *Cancer Res.*, *43*, 2831-2835.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*, *4*. doi: 10.1002/jcc.540040211

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, *21*(3), 167-195. doi: 10.1002/prot.340210302

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.*, *34*, 5402-5415. doi: 10.1093/nar/gkl655

Burger, A. M., Dai, F. P., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., & Neidle, S. (2005). The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.*, *65*, 1489-1496. doi: 10.1158/0008-5472.can-04-2910

Campbell, N. H., Parkinson, G. N., Reszka, A. P., & Neidle, S. (2008). Structural basis of DNA quadruplex recognition by an acridine drug. *Journal of the American Chemical Society*, *130*(21), 6722-+. doi: 10.1021/ja8016973

Case, D. A., Betz, R. M., Botello-Smith, W., Cerutti, D. S., Cheatham, I., T.E. , Darden, T. A., . . . Madej, B. (2016). AMBER 2016. University of California, San Francisco.

Chandonia, J.-M., & Brenner, S. E. (2005). Implications of structural genomics target selection strategies: Pfam5000, whole genome, and random approaches. *Proteins: Structure, Function, and Bioinformatics, 58*(1), 166-179. doi: 10.1002/prot.20298

Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *Journal of Medicinal Chemistry, 42*(25), 5100-5109. doi: 10.1021/jm990352k

Cheatham, T. E., 3rd, Cieplak, P., & Kollman, P. A. (1999). A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn., 16*, 845-862. doi: 10.1080/07391102.1999.10508297

Chen, T. R., Drabkowski, D., Hay, R. J., Macy, M., & Peterson Jr, W. (1987). WiDr is a Derivative of Another Colon Adenocarcinoma Cell Line, HT-29. *Cancer Genet. Cytogenet., 27*, 125-134. doi: http://dx.doi.org/10.1016/0165-4608(87)90267-6

Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., . . . van Gunsteren, W. F. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem, 26*. doi: 10.1002/jcc.20303

Chung, W. J., Heddi, B., Tera, M., Iida, K., Nagasawa, K., & Phan, A. T. (2013a). Solution Structure of an Intramolecular (3 + 1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. *J Am Chem Soc, 135*(36), 13495-13501. doi: 10.1021/ja405843r

Cleaves, H. J. (2011). Molecular Recognition. In M. Gargaud, R. Amils, J. C. Quintanilla, H. J. Cleaves, W. M. Irvine, D. L. Pinti & M. Viso (Eds.), *Encyclopedia of Astrobiology* (pp. 1079-1080). Berlin, Heidelberg: Springer Berlin Heidelberg.

Collie, G. W., Sparapani, S., Parkinson, G. N., & Neidle, S. (2011). Structural Basis of Telomeric RNA Quadruplex-Acridine Ligand Recognition. *J. Am. Chem. Soc., 133*, 2721-2728. doi: 10.1021/ja109767y

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., . . . Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc, 117*. doi: 10.1021/ja00124a002

Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., . . . Randazzo, A. (2010). Structural and Conformational Requisites in DNA Quadruplex Groove Binding: Another Piece to the Puzzle. *J. Am. Chem. Soc., 132*(18), 6425-6433.

Cree, S. L., & Kennedy, M. A. (2014). Relevance of G-Quadruplex Structures to Pharmacogenetics. *Front. Pharmacol., 5*, 1-8. doi: 16010.3389/fphar.2014.00160

Csermely, P., Palotai, R., & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences, 35*(10), 539-546. doi: http://dx.doi.org/10.1016/j.tibs.2010.04.009

Cusack, S., Yaremchuk, A., & Tukalo, M. (2000). The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *The EMBO Journal, 19*(10), 2351-2361. doi: 10.1093/emboj/19.10.2351

Das, B. C., Thapa, P., Karki, R., Schinke, C., Das, S., Kambhampati, S., . . . Evans, T. (2013). Boron chemicals in diagnosis and therapeutics. *Future Medicinal Chemistry, 5*(6), 653-676. doi: 10.4155/fmc.13.38

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., & Mark, A. E. (1999a). Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., 38*, 236-240.

de Lange, T. (2005a). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev, 19*(18), 2100-2110. doi: 10.1101/gad.1346005

de Ruiter, A., & Oostenbrink, C. (2011). Free energy calculations of protein–ligand interactions. *Current Opinion in Chemical Biology, 15*(4), 547-552. doi: http://dx.doi.org/10.1016/j.cbpa.2011.05.021

Debray, J., Zeghida, W., Jourdan, M., Monchaud, D., Dheu-Andries, M.-L., Dumy, P., . . . Demeunynck, M. (2009). Synthesis and evaluation of fused bispyrimidinoacridines as novel pentacyclic analogues of quadruplex-binder BRACO-19. *Organic & Biomolecular Chemistry, 7*(24), 5219-5228. doi: 10.1039/b912716j

Demchenko, A. P. (2001). Recognition between flexible protein molecules: induced and assisted folding†. *Journal of Molecular Recognition, 14*(1), 42-61. doi: 10.1002/1099-1352(200101/02)14:1<42::AID-JMR518>3.0.CO;2-8

Denchi, E. L., & de Lange, T. (2007a). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature, 448*(7157), 1068-1071. doi: http://www.nature.com/nature/journal/v448/n7157/suppinfo/nature06065_S1.html

Dhamodharan, V., Harikrishna, S., Jagadeeswaran, C., Halder, K., & Pradeepkumar, P. I. (2012a). Selective G-quadruplex DNA Stabilizing Agents Based on Bisquinolinium and Bispyridinium Derivatives of 1,8-Naphthyridine. *Journal of Organic Chemistry, 77*(1), 229-242. doi: 10.1021/jo201816g

Di Antonio, M., Rodriguez, R., & Balasubramanian, S. (2012). Experimental Approaches To Identify Cellular G-Quadruplex Structures and Functions. *Methods (Amsterdam, Neth.), 57*, 84-92. doi: 10.1016/j.ymeth.2012.01.008

Di Leva, F. S., Novellino, E., Cavalli, A., Parrinello, M., & Limongelli, V. (2014). Mechanistic Insight into Ligand Binding to G-Quadruplex DNA. *Nucleic Acids Res., 42*, 5447-5455. doi: 10.1093/nar/gku247

Diveshkumar, K. V., Sakrikar, S., Rosu, F., Harikrishna, S., Gabelica, V., & Pradeepkumar, P. I. (2016a). Specific Stabilization of c-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleneindanone Scaffolds. *Biochemistry, 55*, 3571-3585. doi: 10.1021/acs.biochem.6b00120

Doluca, O., Withers, J. M., & Filichev, V. V. (2013b). Molecular Engineering of Guanine-Rich Sequences: Z-DNA, DNA Triplexes, and G-Quadruplexes. *Chem. Rev. (Washington, DC, U. S.), 113*, 3044-3083. doi: 10.1021/cr300225q

Drake, L. A., Patrick, D. L., Fleckman, P., André, J., Baran, R., Haneke, E., . . . Tosti, A. (1999). The impact of onychomycosis on quality of life: Development of an international onychomycosis-specific questionnaire to measure patient quality of life. *Journal of the American Academy of Dermatology, 41*(2), 189-196. doi: http://dx.doi.org/10.1016/S0190-9622(99)70047-2

Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., . . . Liu, S.-Q. (2016). Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences, 17*(2), 144. doi: 10.3390/ijms17020144

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., . . . Kollman, P. (2003). A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem., 24*, 1999-2012. doi: 10.1002/jcc.10349

Duchler, M. (2012). G-Quadruplexes: Targets and Tools in Anticancer Drug Design. *J. Drug Targeting, 20*, 389-400. doi: 10.3109/1061186X.2012.669384

Dunham, M. A., Neumann, A. A., Fasching, C. L., & Reddel, R. R. (2000). Telomere maintenance by recombination in human cells. *Nat Genet, 26*(4), 447-450. doi: 10.1038/82586

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). 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, 11*(5), 425-445.

Elewski, B. E. (2000). Onychomycosis. Treatment, quality of life, and economic issues. *Am J Clin Dermatol, 1*(1), 19-26.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T. A., Lee, H., & Pedersen, L. G. (1995a). A Smooth Particle Mesh Ewald Method. *J. Phys. Chem., 103*, 8577-8593.

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des, 15*(5), 411-428.

Fadrna, E., Spackova, N. a., Sarzynska, J., Koca, J., Orozco, M., Cheatham, T. E., III;, . . . Sponer, J. (2009). Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *J. Chem. Theory Comput., 5*(9), 2514-2530

Fang, R., & Aust, A. E. (1997). Induction of Ferritin Synthesis in Human Lung Epithelial Cells Treated with Crocidolite Asbestos. *Arch. Biochem. Biophys., 340*, 369-375. doi: http://dx.doi.org/10.1006/abbi.1997.9892

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges, 27*. doi: 10.1002/cber.18940270364

Foloppe, N., & Hubbard, R. (2006). Towards Predictive Ligand Design With Free-Energy Based Computational Methods? *Current Medicinal Chemistry, 13*(29), 3583-3608. doi: http://dx.doi.org/10.2174/092986706779026165

Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science, 254*(5038), 1598.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., . . . Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739-1749. doi: 10.1021/jm0306430

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., . . . Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177-6196. doi: 10.1021/jm051256o

Fukunaga, R., & Yokoyama, S. (2005). Crystal Structure of Leucyl-tRNA Synthetase from the Archaeon Pyrococcus horikoshii Reveals a Novel Editing Domain Orientation. *Journal of Molecular Biology*, *346*(1), 57-71. doi: http://dx.doi.org/10.1016/j.jmb.2004.11.060

Gaillard, T., & Case, D. A. (2011). Evaluation of DNA Force Fields in Implicit Solvation. *J. Chem. Theory Comput.*, *7*, 3181-3198. doi: 10.1021/ct200384r

Galindo-Murillo, R., Robertson, J. C., Zgarbová, M., Šponer, J., Otyepka, M., Jurečka, P., & Cheatham, T. E. (2016). Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.*, *12*, 4114-4127. doi: 10.1021/acs.jctc.6b00186

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449-461. doi: 10.1517/17460441.2015.1032936

Gilson, M. K., & Zhou, H.-X. (2007). Calculation of Protein-Ligand Binding Affinities. *Annual Review of Biophysics and Biomolecular Structure*, *36*(1), 21-42. doi: 10.1146/annurev.biophys.36.040306.132550

Gohlke, H., & Klebe, G. (2002). Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angewandte Chemie International Edition*, *41*(15), 2644-2676. doi: 10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O

Greider, C. W., & Blackburn, E. H. (1989a). A Telomeric Sequence in the RNA of Tetrahymena Telomerase Required for Telomere Repeat Synthesis. *Nature (London, U. K.)*, *337*, 331-337.

Grinter, S., & Zou, X. (2014). Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules*, *19*(7), 10150.

Gunaratnam, M., Greciano, O., Martins, C., Reszka, A. P., Schultes, C. M., Morjani, H., . . . Neidle, S. (2007). Mechanism of Acridine-Based Telomerase Inhibition and Telomere Shortening. *Biochem. Pharmacol. (Amsterdam, Neth.), 74*, 679-689. doi: 10.1016/j.bcp.2007.06.011

Guo, M., & Schimmel, P. (2013). Essential Non-Translational Functions of tRNA Synthetases. *Nature chemical biology*, *9*(3), 145-153. doi: 10.1038/nchembio.1158

Gupta, A. K., & Simpson, F. C. (2012). New therapeutic options for onychomycosis. *Expert Opinion on Pharmacotherapy*, *13*(8), 1131-1142. doi: 10.1517/14656566.2012.681779

Gupta, A. P., Verma, G. L., & Ikram, S. (2000). Studies on a new heteropolyacid-based inorganic ion exchanger; zirconium(IV) selenomolybdate. *Reactive and Functional Polymers*, *43*(1), 33-41. doi: http://dx.doi.org/10.1016/S1381-5148(98)00091-1

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, *47*(7), 1750-1759. doi: 10.1021/jm030644s

Hänsel, R., Foldynová-Trantírková, S., Löhr, F., Buck, J., Bongartz, E., Bamberg, E., . . . Trantírek, L. (2009). Evaluation of Parameters Critical for Observing Nucleic Acids Inside Living Xenopus Laevis Oocytes by In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.*, *131*, 15761-15768. doi: 10.1021/ja9052027

Hänsel, R., Löhr, F., Foldynová-Trantírková, S., Bamberg, E., Trantírek, L., & Dötsch, V. (2011a). The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Research*, *39*(13), 5768-5775. doi: 10.1093/nar/gkr174

Hänsel, R., Löhr, F., Trantirek, L., & Dötsch, V. (2013). High-Resolution Insight into G-Overhang Architecture. *J. Am. Chem. Soc.*, *135*, 2816-2824. doi: 10.1021/ja312403b

Hansson, T., Marelius, J., & Åqvist, J. (1998). Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des, 12*(1), 27-35. doi: 10.1023/A:1007930623000

Harley, C. B., Futcher, A. B., & Greider, C. W. (1990a). Telomeres shorten during ageing of human fibroblasts. *Nature, 345*(6274), 458-460.

Harrison, R. J., Gowan, S. M., Kelland, L. R., & Neidle, S. (1999). Human telomerase inhibition by substituted acridine derivatives. *Bioorg Med Chem Lett, 9*(17), 2463-2468.

Harrison, R. J., Reszka, A. P., Haider, S. M., Romagnoli, B., Morrell, J., Read, M. A., . . . Neidle, S. (2004). Evaluation of by Disubstituted Acridone Derivatives as Telomerase Inhibitors: The Importance of G-Quadruplex Binding. *Bioorg. Med. Chem. Lett., 14*, 5845-5849. doi: 10.1016/j.bmcl.2004.09.037

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100-108. doi: 10.2307/2346830

Hawkins, G. D., Cramer, C. J., & Truhlar, D. G. (1995). Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chemical Physics Letters, 246*(1-2), 122-129. doi: Doi 10.1016/0009-2614(95)01082-K

Hawkins, G. D., Cramer, C. J., & Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem. B., 100*, 19824-19839. doi: 10.1021/jp961710n

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature, 450*(7172), 964-972.

Hernandez, V., Crépin, T., Palencia, A., Cusack, S., Akama, T., Baker, S. J., . . . Plattner, J. J. (2013). Discovery of a Novel Class of Boron-Based Antibacterials with Activity against Gram-Negative Bacteria. *Antimicrobial Agents and Chemotherapy, 57*(3), 1394-1403. doi: 10.1128/aac.02058-12

Hong, K. W., Ibba, M., Weygand-Durasevic, I., Rogers, M. J., Thomann, H. U., & Söll, D. (1996). Transfer RNA-dependent cognate amino acid recognition by an aminoacyl-tRNA synthetase. *The EMBO Journal, 15*(8), 1983-1991.

Hou, J. Q., Chen, S. B., Tan, J. H., Ou, T. M., Luo, H. B., Li, D., . . . Huang, Z. S. (2010a). New Insights into the Structures of Ligand-Quadruplex Complexes from Molecular Dynamics Simulations. *J. Phys. Chem. B., 114*, 15301-15310. doi: 10.1021/jp106683n

Hou, T., Wang, J., Li, Y., & Wang, W. (2010). Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *J. Comput. Chem., 32*, 866-877.

Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model., 51*, 69-82.

Hu, J., Hwang, S. S., Liesa, M., Gan, B., Sahin, E., Jaskelioff, M., . . . DePinho, R. A. (2012). Anti-telomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. *Cell, 148*(4), 651-663. doi: 10.1016/j.cell.2011.12.028

Huang, N., Kalyanaraman, C., Irwin, J. J., & Jacobson, M. P. (2006). Physics-Based Scoring of Protein−Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *Journal of Chemical Information and Modeling, 46*(1), 243-253. doi: 10.1021/ci0502855

Huang, S.-Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics, 12*(40), 12899-12908. doi: 10.1039/C0CP00151A

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res., 33*, 2908-2916. doi: 10.1093/nar/gki609

Huppert, J. L., & Balasubramanian, S. (2007). G-Quadruplexes in Promoters Throughout the Human Genome. *Nucleic Acids Res., 35*, 406-413.

Ibba, M., & Söll, D. (2000). Aminoacyl-tRNA Synthesis. *Annual Review of Biochemistry, 69*(1), 617-650. doi: 10.1146/annurev.biochem.69.1.617

Incles, C. M., Schultes, C. M., Kempski, H., Koehler, H., Kelland, L. R., & Neidle, S. (2004). A G-Quadruplex Telomere Targeting Agent Produces p16-Associated Senescence and Chromosomal Fusions in Human Prostate Cancer Cells. *Mol. Cancer Ther., 3*, 1201-1206.

Ivani, I., Dans, P. D., Noy, A., Perez, A., Faustino, I., Hospital, A., . . . Orozco, M. (2016). parmBSC1: A Refined Force Field for DNA Simulations. *Nat. Methods, 13*, 55-58. doi: 10.1038/nmeth.3658

http://www.nature.com/nmeth/journal/v13/n1/abs/nmeth.3658.html#supplementary-information

Jacobson, M. P., Friesner, R. A., Xiang, Z., & Honig, B. (2002). On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of Molecular Biology, 320*(3), 597-608. doi: http://dx.doi.org/10.1016/S0022-2836(02)00470-9

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics, 55*(2), 351-367. doi: 10.1002/prot.10613

Jain, A. K., Paul, A., Maji, B., Muniyappa, K., & Bhattacharya, S. (2012a). Dimeric 1,3-Phenylene-bis(piperazinyl benzimidazole)s: Synthesis and Structure–Activity Investigations on their Binding with Human Telomeric G-Quadruplex DNA and Telomerase Inhibition Properties. *J. Med. Chem., 55*, 2981-2993. doi: 10.1021/jm200860b

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking1. *Journal of Molecular Biology, 267*(3), 727-748. doi: https://doi.org/10.1006/jmbi.1996.0897

Jones, J. E. (1924). On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London. Series A, 106*(738), 463-477. doi: 10.1098/rspa.1924.0082

Joo, K., Lee, J., & Lee, J. (2012). Methods for Accurate Homology Modeling by Global Optimization. In A. J. W. Orry & R. Abagyan (Eds.), *Homology Modeling: Methods and Protocols* (pp. 175-188). Totowa, NJ: Humana Press.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983a). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys., 79*, 926-935. doi: Doi 10.1063/1.445869

Joseph M. Hayes, G. A. (2012). MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies, Molecular Dynamics - Studies of Synthetic and Biological Macromolecules.

Joung, I. S., & Cheatham, T. E. (2008a). Determination of Alkali and Halide Monovalent ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B., 112*, 9020-9041. doi: 10.1021/jp8001614

Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., . . . Schulten, K. (1999). NAMD2: greater scalability for parallel molecular dynamics. *J Comput Phys, 151*. doi: 10.1006/jcph.1999.6201

Karpen, M. E., Tobias, D. J., & Brooks, C. L. (1993). Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry, 32*(2), 412-420. doi: 10.1021/bi00053a005

Kaserer, T., Rigo, R., Schuster, P., Alcaro, S., Sissi, C., & Schuster, D. (2016). Optimized Virtual Screening Workflow for the Identification of Novel G-Quadruplex Ligands. *Journal of Chemical Information and Modeling, 56*(3), 484-500. doi: 10.1021/acs.jcim.5b00658

Kellner, J., Wierda, W., Shpall, E., Keating, M., & McNiece, I. (2016a). Isolation of a Novel Chronic Lymphocytic Leukemic (CLL) Cell Line and Development of an In Vivo Mouse Model of CLL. *Leuk. Res., 40*, 54-59. doi: http://dx.doi.org/10.1016/j.leukres.2015.10.008

Kellner, J., Wierda, W., Shpall, E., Keating, M., & McNiece, I. (2016b). Isolation of a novel chronic lymphocytic leukemic (CLL) cell line and development of an in vivo mouse model of CLL. *Leukemia Research, 40*, 54-59. doi: http://dx.doi.org/10.1016/j.leukres.2015.10.008

Khan, S., Sharma, A., Jamwal, A., Sharma, V., Pole, A. K., Thakur, K. K., & Sharma, A. (2011). Uneven spread of cis- and trans-editing aminoacyl-tRNA synthetase domains within translational compartments of P. falciparum. *1*, 188. doi: 10.1038/srep00188

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov, 3*(11), 935-949. doi: http://www.nature.com/nrd/journal/v3/n11/suppinfo/nrd1549_S1.html

Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews, 93*(7), 2395-2417. doi: 10.1021/cr00023a004

https://www.nature.com/articles/srep00188#supplementary-information

Kollman, P., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., . . . Cheatham, T. (2000). Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res., 33*, 889-897.

Kongsted, J., Soderhjelm, P., & Ryde, U. (2009a). How accurate are continuum solvation models for drug-like molecules? *J. Comput. Aided Mol. Des., 23*(7), 395-409.

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 44*(2), 98-104. doi: 10.1073/pnas.44.2.98

Krepl, M., Zgarbova, M., Stadlbauer, P., Otyepka, M., Banas, P., Koca, J., . . . Sponer, J. (2012a). Reference Simulations of Noncanonical Nucleic Acids with Different Chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA and Z-DNA. *J. Chem. Theory Comput., 8*, 2506-2520. doi: 10.1021/ct300275s

Kumar, S., Ma, B., Tsai, C.-J., Wolfson, H., & Nussinov, R. (1999). Folding funnels and conformational transitions via hinge-bending motions. *Cell Biochemistry and Biophysics, 31*(2), 141-164. doi: 10.1007/BF02738169

Lamb, M. L., & Jorgensen, W. L. (1997). Computational approaches to molecular recognition. *Current Opinion in Chemical Biology, 1*(4), 449-457. doi: http://dx.doi.org/10.1016/S1367-5931(97)80038-5

Landers, J. E., Cassel, S. L., & George, D. L. (1997). Translational Enhancement of Oncogene Expression in Human Tumor Cells Containing a Stabilized Wild-Type p53 Protein. *Cancer Res., 57*, 3562-3568.

Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T. E. I., . . . Sponer, J. (2010). A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA *Nucleic Acids Res., 38*(1), 299-313.

Lei, H., Wang, X., & Wu, C. (2012a). Early Stage Intercalation of Doxorubicin to DNA Fragments Observed in Molecular Dynamics Binding Simulations. *J. Mol. Graphics Modell., 38*, 279-289. doi: 10.1016/j.jmgm.2012.05.006

Leśnikowski, Z. J. (2016). Challenges and Opportunities for the Application of Boron Clusters in Drug Design. *Journal of Medicinal Chemistry, 59*(17), 7738-7758. doi: 10.1021/acs.jmedchem.5b01932

Li, H., Xie, Y., Liu, C., & Liu, S. (2014). Physicochemical bases for protein folding, dynamics, and protein-ligand binding. *Science China Life Sciences, 57*(3), 287-302. doi: 10.1007/s11427-014-4617-2

Lincecum, T. L., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B. S., . . . Cusack, S. (2003). Structural and Mechanistic Basis of Pre- and Posttransfer Editing by Leucyl-tRNA Synthetase. *Molecular Cell, 11*(4), 951-963. doi: http://dx.doi.org/10.1016/S1097-2765(03)00098-4

Liu, J., & Wang, R. (2015). Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling, 55*(3), 475-482. doi: 10.1021/ci500731a

Long, Y., Li, Z., Tan, J.-H., Ou, T.-M., Li, D., Gu, L.-Q., & Huang, Z.-s. (2012). Benzofuroquinoline Derivatives Had Remarkable Improvement of their Selectivity for Telomeric G-Quadruplex DNA over Duplex DNA upon Introduction of Peptidyl Group. *Bioconjugate Chemistry, 23*(9), 1821-1831. doi: 10.1021/bc300123m

Lyne, P. D. (2002). Structure-based virtual screening: an overview. *Drug Discovery Today, 7*(20), 1047-1055. doi: http://dx.doi.org/10.1016/S1359-6446(02)02483-2

Ma, B., Kumar, S., Tsai, C. J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng, 12*(9), 713-720. doi: 10.1093/protein/12.9.713

Ma, B., Shatsky, M., Wolfson, H. J., & Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci, 11*. doi: 10.1110/ps.21302

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., . . . Karplus, M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B., 102*, 3586-3616. doi: 10.1021/jp973084f

Maestro© 2014 Schrödinger, LLC. Manuals.

Magrane, M., & Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database: The Journal of Biological Databases and Curation, 2011*, bar009. doi: 10.1093/database/bar009

Manly, C. J., Chandrasekhar, J., Ochterski, J. W., Hammer, J. D., & Warfield, B. B. (2008). Strategies and tactics for optimizing the Hit-to-Lead process and beyond—A computational chemistry perspective. *Drug Discovery Today, 13*(3), 99-109. doi: http://dx.doi.org/10.1016/j.drudis.2007.10.019

Martínez-Aguirre, M. A., Villamil-Ramos, R., Guerrero-Alvarez, J. A., & Yatsimirsky, A. K. (2013). Substituent Effects and pH Profiles for Stability Constants of Arylboronic Acid Diol Esters. *The Journal of Organic Chemistry, 78*(10), 4674-4684. doi: 10.1021/jo400617j

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature, 267*. doi: 10.1038/267585a0

Mickey, D. D., Stone, K. R., Wunderli, H., Mickey, G. H., Vollmer, R. T., & Paulson, D. F. (1977). Heterotransplantation of a Human Prostatic Adenocarcinoma Cell Line in Nude Mice. *Cancer Res., 37*, 4049.

Miller, D. W., & Dill, K. A. (1997). Ligand binding to proteins: The binding landscape model. *Protein Science, 6*(10), 2166-2179. doi: 10.1002/pro.5560061011

Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: Successes and challenges. *Current Pharmaceutical Design, 11*(3), 323-333. doi: Doi 10.2174/1381612053382106

Moore, M. J. B., Schultes, C. M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F. A., . . . Neidle, S. (2006a). Trisubstituted Acridines as G-quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation. *J. Med. Chem., 49*, 582-599. doi: 10.1021/jm050555a

Moorhouse, A. D., Santos, A. M., Gunaratnam, M., Moore, M., Neidle, S., & Moses, J. E. (2006a). Stabilization of G-quadruplex DNA by highly selective ligands via click chemistry. *Journal of the American Chemical Society, 128*(50), 15972-15973. doi: 10.1021/ja0661919

Morimoto, H., Safrit, J. T., & Bonavida, B. (1991a). Synergistic effect of tumor necrosis factor-alpha- and diphtheria toxin-mediated cytotoxicity in sensitive and resistant human ovarian tumor cell lines. *The Journal of Immunology, 147*(8), 2609.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785-2791. doi: 10.1002/jcc.21256

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., . . . Wu, J. R. (1988a). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America, 85*(18), 6622-6626.

Muegge, I. (2006). PMF Scoring Revisited. *Journal of Medicinal Chemistry, 49*(20), 5895-5902. doi: 10.1021/jm050038s

Mukherjee, A., Lavery, R., Bagchi, B., & Hynes, J. T. (2008). On the Molecular Mechanism of Drug Intercalation into DNA: A Simulation Study of the Intercalation Pathway, Free Energy, and DNA Structural Changes. *J. Am. Chem. Soc., 130*(30), 9747-9755.

Mukherjee, S., Balius, T. E., & Rizzo, R. C. (2010). Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *Journal of Chemical Information and Modeling, 50*(11), 1986-2000. doi: 10.1021/ci1001982

Mulholland, K., Siddiquei, F., & Wu, C. (2017). Binding Modes and Pathway of RHPS4 to Human Telomeric G-quadruplex and Duplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent *Phys. Chem. Chem. Phys.*(19), 18685-18694 doi: DOI: 10.1039/C7CP03313C

Mulholland, K., & Wu, C. (2016). Binding of Telomestatin to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *J. Chem. Inf. Model, 56*(10), 2093-2102. doi: 10.1021/acs.jcim.6b00473

Murray, C. W., Auton, T. R., & Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *12 5*, 503-519.

Nichols, W. W., Murphy, D. G., Cristofalo, V. J., Toji, L. H., Greene, A. E., & Dwight, S. A. (1977). Characterization of a New Human Diploid Cell Strain, IMR-90. *Am. Assoc. Adv. Sci., 196*, 60.

Olopade, O. I., Jenkins, R. B., Ransom, D. T., Malik, K., Pomykala, H., Nobori, T., . . . Diaz, M. O. (1992a). Molecular Analysis of Deletions of the Short Arm of Chromosome 9 in Human Gliomas. *Cancer Research, 52*(9), 2523.

Onel, B., Lin, C., & Yang, D. (2014). DNA G-Quadruplex and Its Potential as Anticancer Drug Target. *Sci. China: Chem., 57*, 1605-1614. doi: 10.1007/s11426-014-5235-3

Palm, W., & de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet, 42*, 301-334. doi: 10.1146/annurev.genet.41.110306.130350

Parkinson, G. N., Lee, M. P. H., & Neidle, S. (2002). Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature (London, U. K.), 417*, 876-880. doi: http://www.nature.com/nature/journal/v417/n6891/suppinfo/nature755_S1.html

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E. r., Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J., 92*(11), 3817–3829.

Perozzo, R., Folkers, G., & Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res, 24*(1-2), 1-52.

Perrone, R., Butovskaya, E., Daelemans, D., Palù, G., Pannecouque, C., & Richter, S. N. (2014). Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *J. Antimicrob. Chemother., 69*, 3248-3258. doi: 10.1093/jac/dku280

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem, 26*. doi: 10.1002/jcc.20289

Pirzadeh, P., Beaudoin, E. N., & Kusalik, P. G. (2012). Interfacial Free Energy: An Entropy Portent to Energy Changes. *Crystal Growth & Design, 12*(1), 124-128. doi: 10.1021/cg200861e

Procacci, P., & Berne, B. J. (1994a). Multiple Time-Scale Methods for Constant-Pressure Molecular-Dynamics Simulations of Molecular-Systems. *Molec. Phys., 83*, 255-272.

Rankin, A. M., Faller, D. V., & Spanjaard, R. A. (2008a). Telomerase Inhibitors and 'T-oligo' as Cancer Therapeutics: Contrasting Molecular Mechanisms of Cytotoxicity. *Anti-Cancer Drugs, 19*, 329-338.

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology, 261*(3), 470-489. doi: http://dx.doi.org/10.1006/jmbi.1996.0477

Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., & Honig, B. (2002). Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *Journal of Computational Chemistry, 23*(1), 128-137. doi: 10.1002/jcc.1161

Rock, F. L., Mao, W., Yaremchuk, A., Tukalo, M., Crépin, T., Zhou, H., . . . Alley, M. R. K. (2007). An Antifungal Agent Inhibits an Aminoacyl-tRNA Synthetase by Trapping tRNA in the Editing Site. *Science, 316*(5832), 1759.

Ryckaert, J., Ciccotti, G., & Berendsen, H. J. C. (1977a). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comp. Phys., 23*, 327-341.

Schimmel, P., & Schmidt, E. (1995). Making connections: RNA-dependent amino acid recognition. *Trends in Biochemical Sciences, 20*(1), 1-2. doi: http://dx.doi.org/10.1016/S0968-0004(00)88937-9

Seebacher, C., Brasch, J., Abeck, D., Cornely, O., Effendy, I., Ginter-Hanselmayer, G., . . . Tietz, H. J. (2007). Onychomycosis. *Mycoses, 50*(4), 321-327. doi: 10.1111/j.1439-0507.2006.01351.x

Sergio, F. S., Nuno, M. F. S. A. C., Pedro, A. F., & Maria Joao, R. (2010). Virtual Screening in Drug Design and Development. *Combinatorial Chemistry & High Throughput Screening, 13*(5), 442-453. doi: http://dx.doi.org/10.2174/138620710791293001

Shalaby, T., Fiaschetti, G., Nagasawa, K., Shin-ya, K., Baumgartner, M., & Grotzer, M. (2013). G-quadruplexes as Potential Therapeutic Targets for Embryonal Tumors. *Molecules, 18*, 12500-12537. doi: 10.3390/molecules181012500

Shao, J., Tanner, S. W., Thompson, N., & Cheatham, T. E. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation, 3*(6), 2312-2334. doi: 10.1021/ct700119m

Shay, J. W., & Wright, W. E. (2006). Telomerase therapeutics for cancer: challenges and new directions. *Nat Rev Drug Discov, 5*(7), 577-584.

Shenkin, P. S., & McDonald, D. Q. (1994). Cluster analysis of molecular conformations. *Journal of Computational Chemistry, 15*(8), 899-916. doi: 10.1002/jcc.540150811

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature, 432*(7019), 862-865. doi: 10.1038/nature03197

Singh, N., & Warshel, A. (2010). Absolute binding free energy calculations: On the accuracy of computational scoring of protein–ligand interactions. *Proteins: Structure, Function, and Bioinformatics, 78*(7), 1705-1723. doi: 10.1002/prot.22687

Sousa, S. F., Ribeiro, A. J. M., Coimbra, J. T. S., Neves, R. P. P., Martins, S. A., Moorthy, N. S. H. N., . . . Ramos, M. J. (2013). Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Current Medicinal Chemistry, 20*(18), 2296-2314. doi: http://dx.doi.org/10.2174/0929867311320180002

Sponer, J., Cang, X., & Cheatham, T. E., 3rd. (2012). Molecular Dynamics Simulations of G-DNA and Perspectives on the Simulation of Nucleic Acid Structures. *Methods (Amsterdam, Neth.), 57*, 25-39. doi: 10.1016/j.ymeth.2012.04.005

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *Journal of the American Chemical Society, 120*(37), 9401-9409. doi: Doi 10.1021/Ja981844+

Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society, 112*(16), 6127-6129. doi: Doi 10.1021/Ja00172a038

Suman, P., Patel, B. P., Kasibotla, A. V., Solano, L. N., & Jonnalagadda, S. C. (2015). Synthesis and evaluation of functionalized aminobenzoboroxoles as potential anti-cancer agents. *Journal of Organometallic Chemistry, 798, Part 1*, 125-131. doi: https://doi.org/10.1016/j.jorganchem.2015.06.021

Sun, H., Li, Y., Tian, S., Xub, L., & Hou, T. (2014). Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols Using PDBbind Data Set. *PCCP, 16*, 16719-16729.

Tan, C., Tan, Y.-H., & Luo, R. (2007). Implicit Nonpolar Solvent Models. *J. Phys. Chem. B., 111*, 12263-12274. doi: 10.1021/jp073399n

Tang, Y. (2010). Examination of Molecular Recognition in Protein-Ligand Interactions. *(2010). All Theses and Dissertations (ETDs). 341*. doi: https://doi.org/10.7936/K74747VH

Thomas, J., Jacobson, G. A., Narkowicz, C. K., Peterson, G. M., Burnet, H., & Sharpe, C. (2010). Toenail onychomycosis: an important global disease burden. *Journal of Clinical Pharmacy and Therapeutics, 35*(5), 497-519. doi: 10.1111/j.1365-2710.2009.01107.x

Thomas, S., & Andreas, L. (2010). Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Current Medicinal Chemistry, 17*(8), 767-785. doi: http://dx.doi.org/10.2174/092986710790514453

Tibi, S. (2006). Al-Razi and Islamic medicine in the 9th century. *Journal of the Royal Society of Medicine, 99*(4), 206-207.

Tobi, D., & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proceedings of the National Academy of Sciences of the United States of America, 102*(52), 18908-18913.

Tomsho, J. W., & Benkovic, S. J. (2012). Elucidation of the Mechanism of the Reaction between Phenylboronic Acid and a Model Diol, Alizarin Red S. *The Journal of Organic Chemistry, 77*(5), 2098-2106. doi: 10.1021/jo202250d

Tsai, C.-J., Kumar, S., Ma, B., & Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Science, 8*(6), 1181-1190. doi: 10.1110/ps.8.6.1181

Tsui, V., & Case, D. A. (2000). Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers, 56*, 275-291. doi: 10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E

Ungvarsky, J., Plsikova, J., Janovec, L., Koval, J., Mikes, J., Mikesova, L., . . . Kozurkova, M. (2014a). Novel Trisubstituted Acridines as Human Telomeric Quadruplex Binding Ligands. *Bioorg. Chem., 57*, 13-29. doi: 10.1016/j.bioorg.2014.07.010

van Gunsteren, W. F., Dolenc, J., & Mark, A. E. (2008). Molecular simulation as an aid to experimentalists. *Curr Opin Struc Biol, 18*. doi: 10.1016/j.sbi.2007.12.007

Verkhivker, G., Appelt, K., Freer, S. T., & Villafranca, J. E. (1995). Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Engineering, Design and Selection, 8*(7), 677-691. doi: 10.1093/protein/8.7.677

Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., . . . Rose, P. W. (2000). Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des, 14*(8), 731-751.

Vitkup, D., Melamud, E., Moult, J., & Sander, C. (2001). Completeness in structural genomics. *Nat Struct Mol Biol, 8*(6), 559-566.

Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. *Indian Journal of Pharmaceutical Sciences, 74*(1), 1-17. doi: 10.4103/0250-474X.102537

Wang, J., Morin, P., Wang, W., & Kollman, P. A. (2001). Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *Journal of the American Chemical Society, 123*(22), 5221-5230. doi: 10.1021/ja003834q

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry, 25*(9), 1157-1174. doi: 10.1002/jcc.20035

Wang, J. M., Cieplak, P., & Kollman, P. A. (2000). How Well does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem., 21*, 1049-1074. doi: Doi 10.1002/1096-987x(200009)21:12<1049::Aid-Jcc3>3.0.Co;2-F

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des, 16*(1), 11-26. doi: 10.1023/A:1016357811882

Westbrook, J., Feng, Z., Chen, L., Yang, H., & Berman, H. M. (2003). The Protein Data Bank and structural genomics. *Nucleic Acids Research, 31*(1), 489-491.

Westerberg, D. P., & Voyack, M. J. (2013). Onychomycosis: Current trends in diagnosis and treatment. *Am Fam Physician, 88*(11), 762-770.

White, E. W., Tanious, F., Ismail, M. A., Reszka, A. P., Neidle, S., Boykin, D. W., & Wilson, W. D. (2007). Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge. *Biophysical Chemistry, 126*(1–3), 140-153. doi: http://dx.doi.org/10.1016/j.bpc.2006.06.006

Withana-Gamage, T. S., Hegedus, D. D., Qiu, X., & Wanasundara, J. P. D. (2011). In Silico Homology Modeling To Predict Functional Properties of Cruciferin. *Journal of Agricultural and Food Chemistry, 59*(24), 12925-12938. doi: 10.1021/jf201979a

Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D., & Shay, J. W. (1997a). Normal Human Chromosomes have Long G-Rich Telomeric Overhangs at One End. *Genes Dev., 11*, 2801-2809.

Xu, L., Sun, H., Li, Y., Wang, J., & Hou, T. (2013). Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B., 117*, 8408-8421.

Xue, L. A., Ranjan, N., & Arya, D. P. (2011). Synthesis and Spectroscopic Studies of the Aminoglycoside (Neomycin)-Perylene Conjugate Binding to Human Telomeric DNA. *Biochemistry, 50*(14), 2838-2849. doi: 10.1021/bi1017304

Yamaguchi, T., Blázquez, B., Hesek, D., Lee, M., Llarrull, L. I., Boggess, B., . . . Mobashery, S. (2012). Inhibitors for Bacterial Cell-Wall Recycling. *ACS Medicinal Chemistry Letters, 3*(3), 238-242. doi: 10.1021/ml2002746

Yang, D., & Okamoto, K. (2010a). Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Medicinal Chemistry, 2*(4), 619-646.

Zahler, A. M., Williamson, J. R., Cech, T. R., & Prescott, D. M. (1991a). Inhibition of telomerase by G-quartet DMA structures. *Nature, 350*(6320), 718-720.

Zakian, V. A. (1995a). Telomeres: Beginning to Understand the End. *Science, 270*(5242), 1601.

Zgarbova, M., Luque, F. J., Sponer, J., Cheatham, T. E., Otyepka, M., & Jurecka, P. (2013). Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *Journal of Chemical Theory and Computation, 9*(5), 2339-2354. doi: 10.1021/ct400154j

Zgarbova, M., Sponer, J., Otyepka, M., Cheatham, T. E., Galindo-Murillo, R., & Jurecka, P. (2015). Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of Chemical Theory and Computation, 11*(12), 5723-5736. doi: 10.1021/acs.jctc.5b00716

Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E., & Patel, D. J. (2001). Dimeric DNA Quadruplex Containing Major Groove-Aligned A·T·A·T and G·C·G·C Tetrads Stabilized by Inter-Subunit Watson-Crick A·T and G·C Pairs1. *J. Mol. Biol., 312*, 1073-1088. doi: http://dx.doi.org/10.1006/jmbi.2001.5002

Zhou, G. T., Liu, X. R., Li, Y. Q., Xu, S. B., Ma, C. Y., Wu, X. M., . . . Chen, Y. (2016). Telomere Targeting with a Novel G-Quadruplex-Interactive Ligand BRACO-19 Induces T-Loop Disassembly and Telomerase Displacement in Human Glioblastoma Cells. *Oncotarget, 7*, 14925-14939.

Zhou, J. K., Yang, D. Y., & Sheu, S. Y. (2015a). The Molecular Mechanism of Ligand Unbinding from the Human Telomeric G-Quadruplex by Steered Molecular Dynamics and Umbrella Sampling Simulations. *Phys. Chem. Chem. Phys., 17*, 12857-12869. doi: 10.1039/c5cp00378d

Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S., & Harder, E. (2014). Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring. *Journal of Chemical Information and Modeling, 54*(7), 1932-1940. doi: 10.1021/ci500118s

Appendix A

RMSD and Contact Plots from Chapter 3

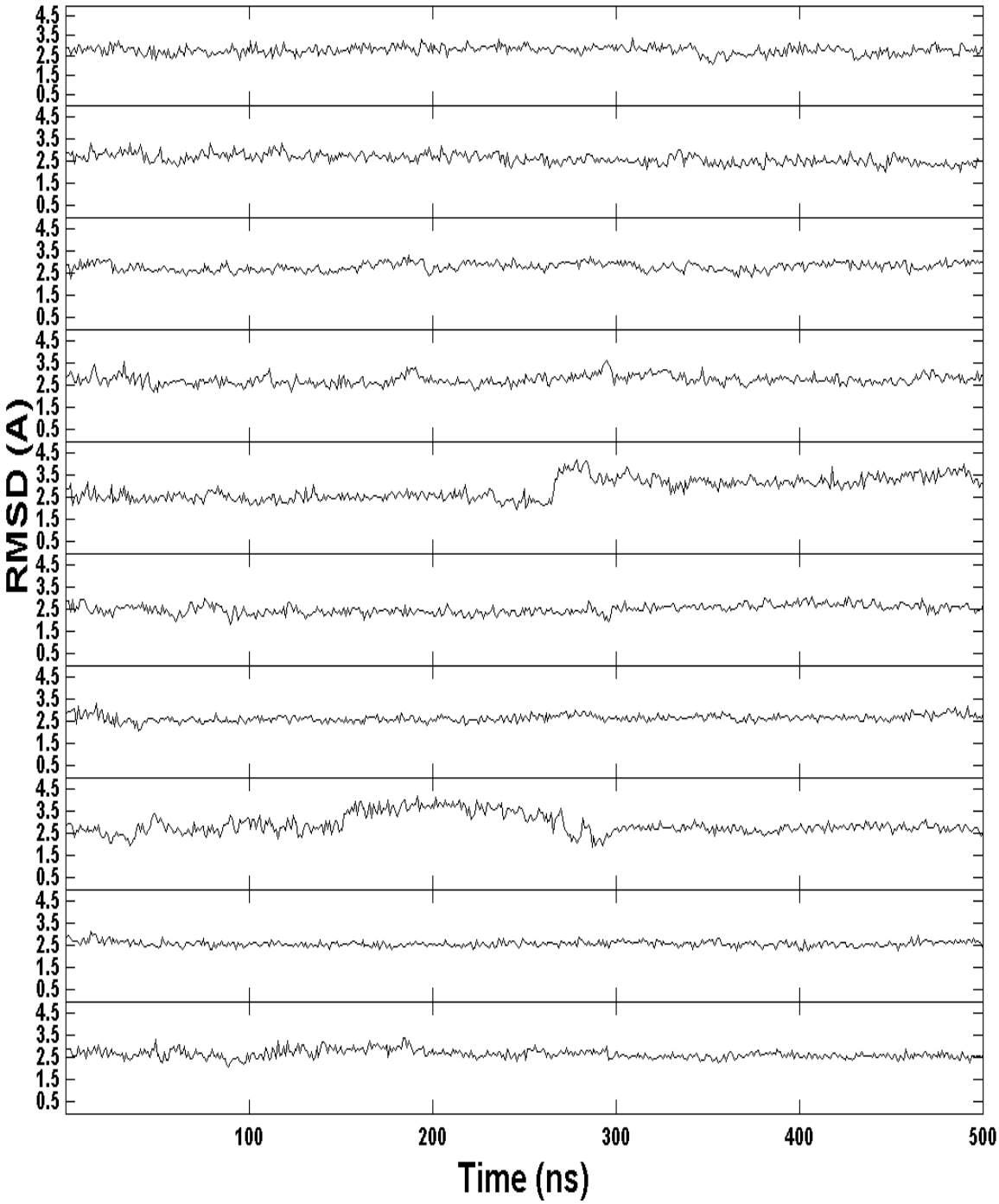


Figure 60. RMSD plot for each trajectory of parallel DNA G-quadruplex and BRACO19 system in ten runs

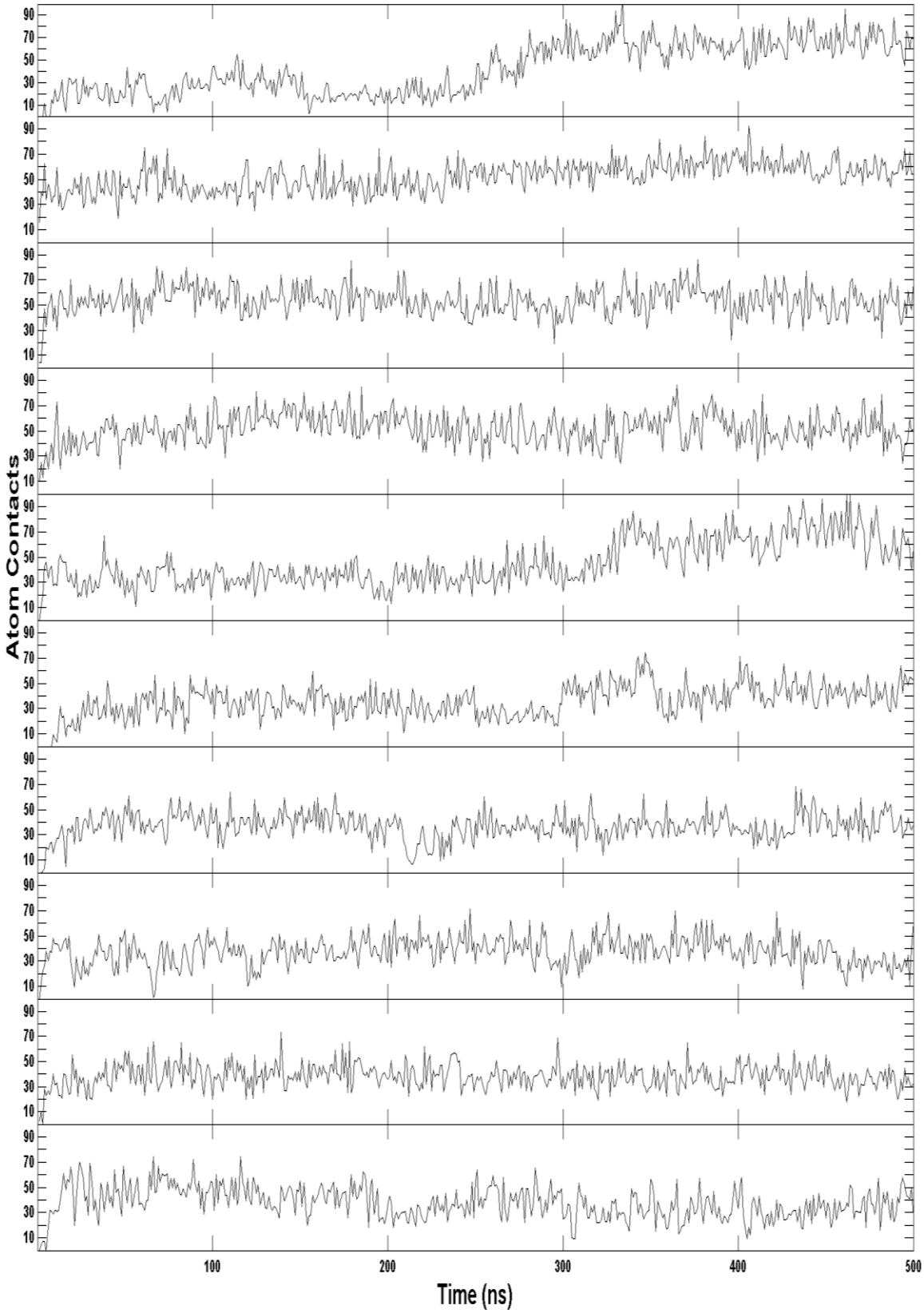


Figure 61. The contact number between parallel DNA G-quadruplex and BRACO19 for each trajectory in ten runs.

Appendix B

RMSD and Contact Plots from Chapter 4

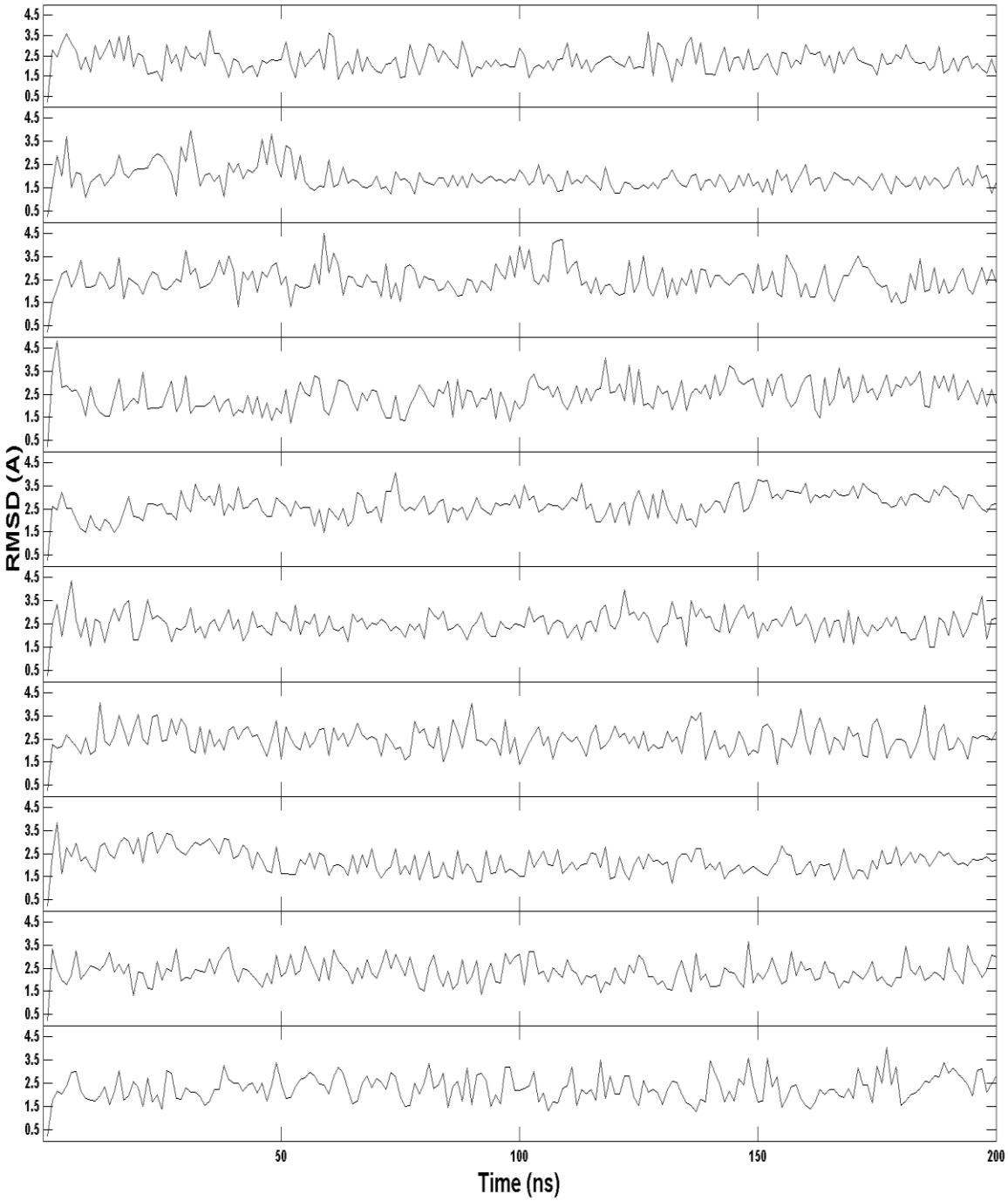


Figure 62. RMSD plot for each trajectory of DNA duplex and BRACO19 system in ten runs

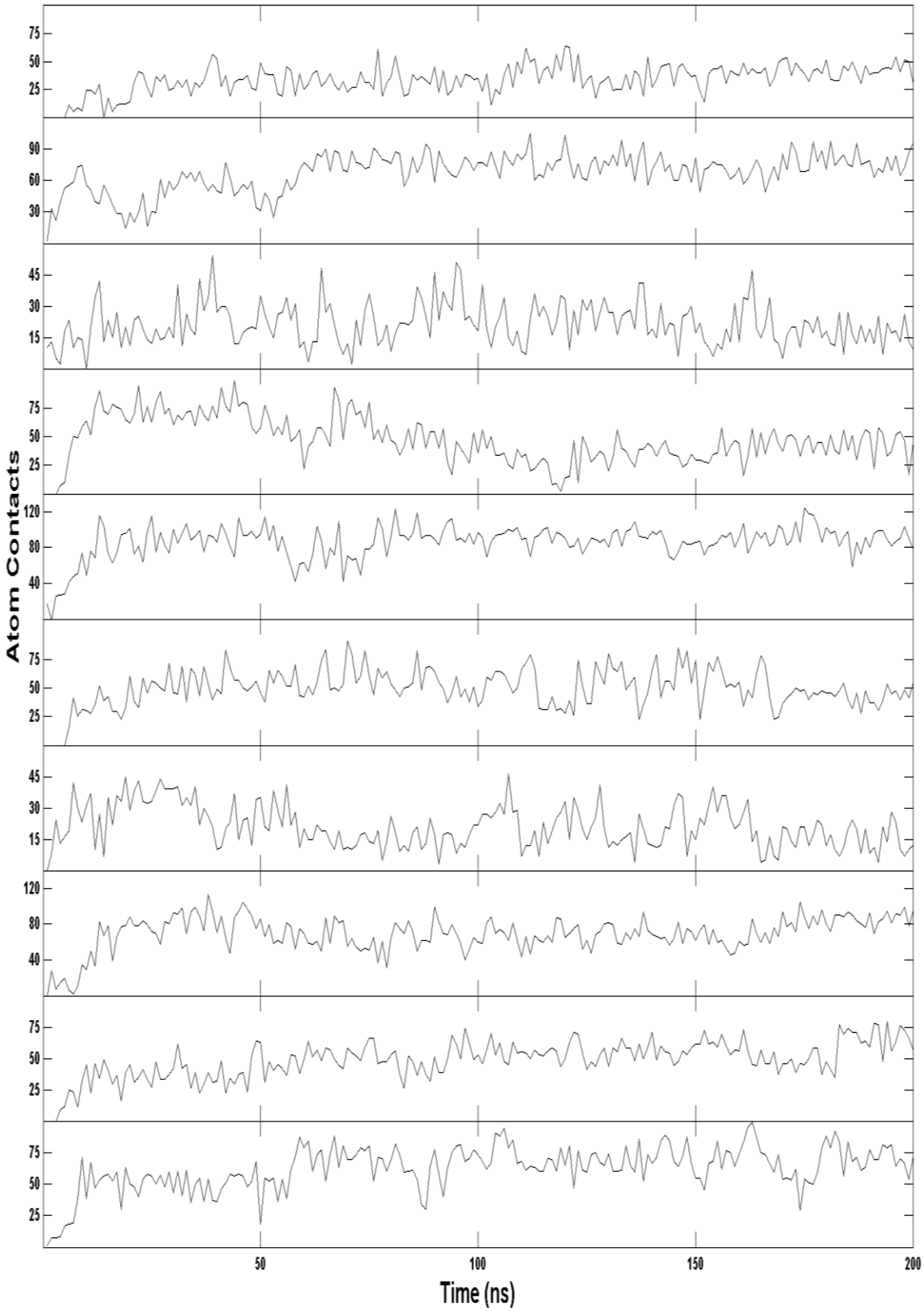


Figure 63. The contact number between DNA duplex and BRACO19 for each trajectory in ten runs.

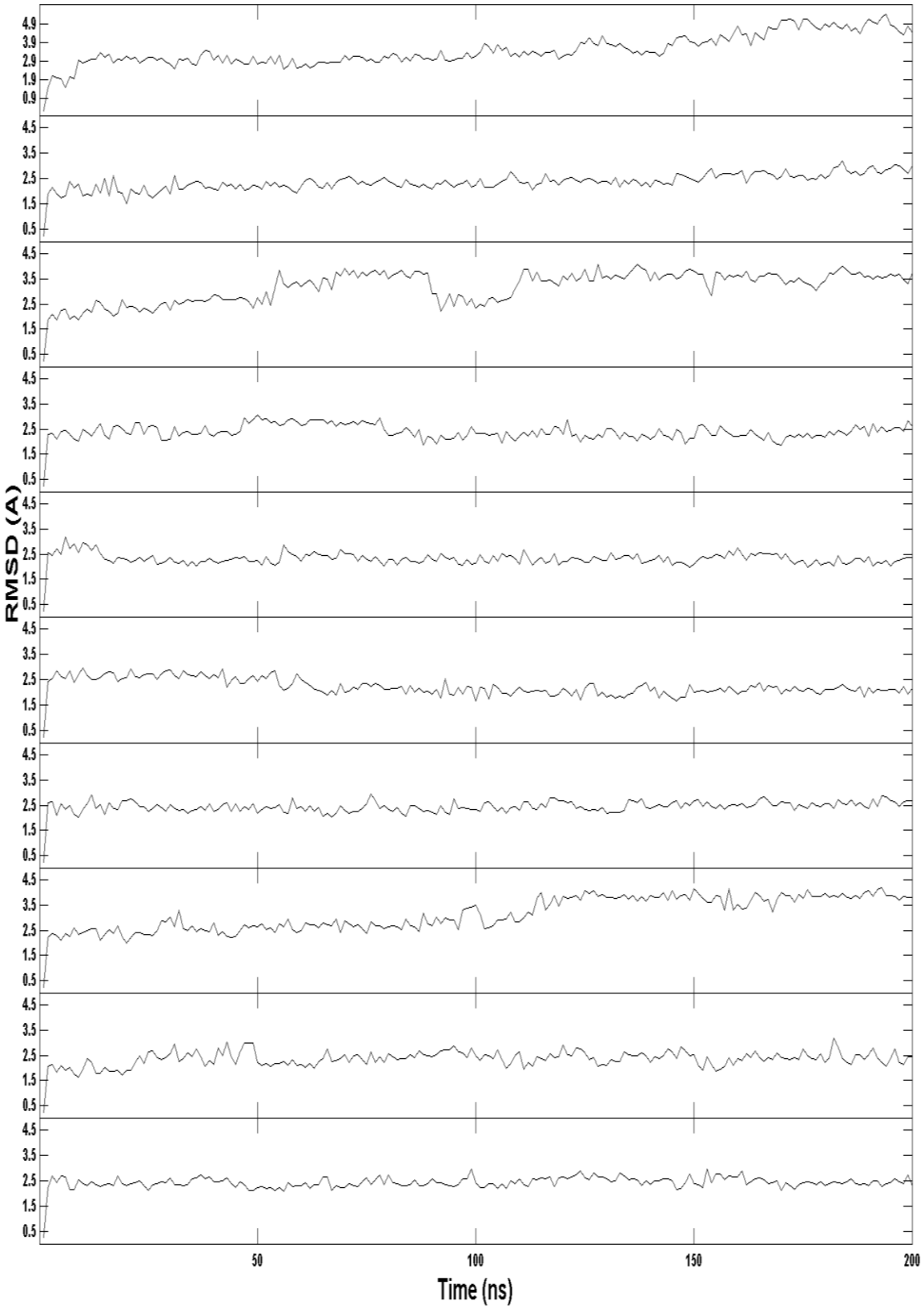


Figure 64. RMSD plot for each trajectory of parallel DNA G-quadruplex and BRACO19 system in ten runs

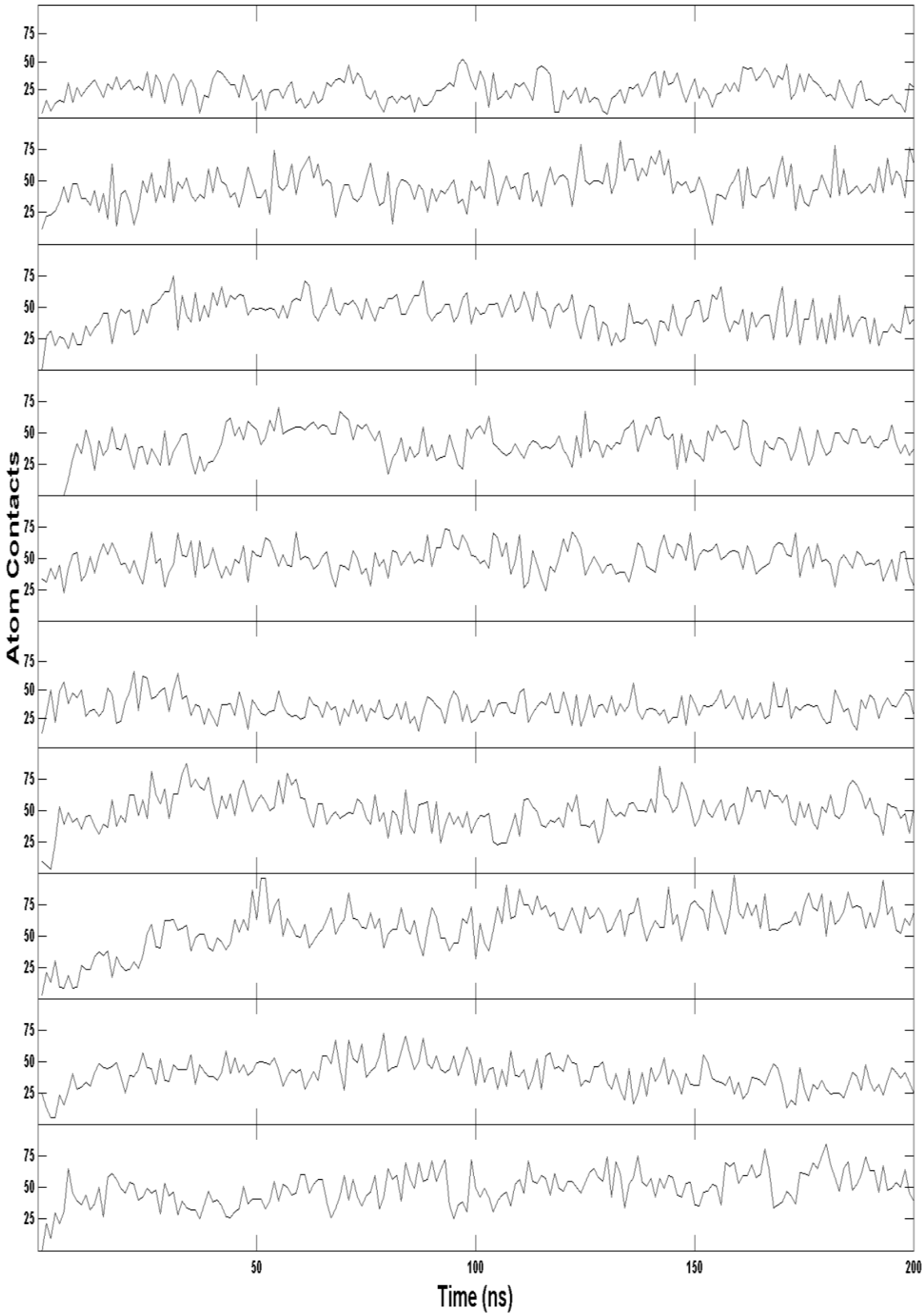


Figure 65. The contact number between parallel DNA G-quadruplex and BRACO19 for each trajectory in ten runs.

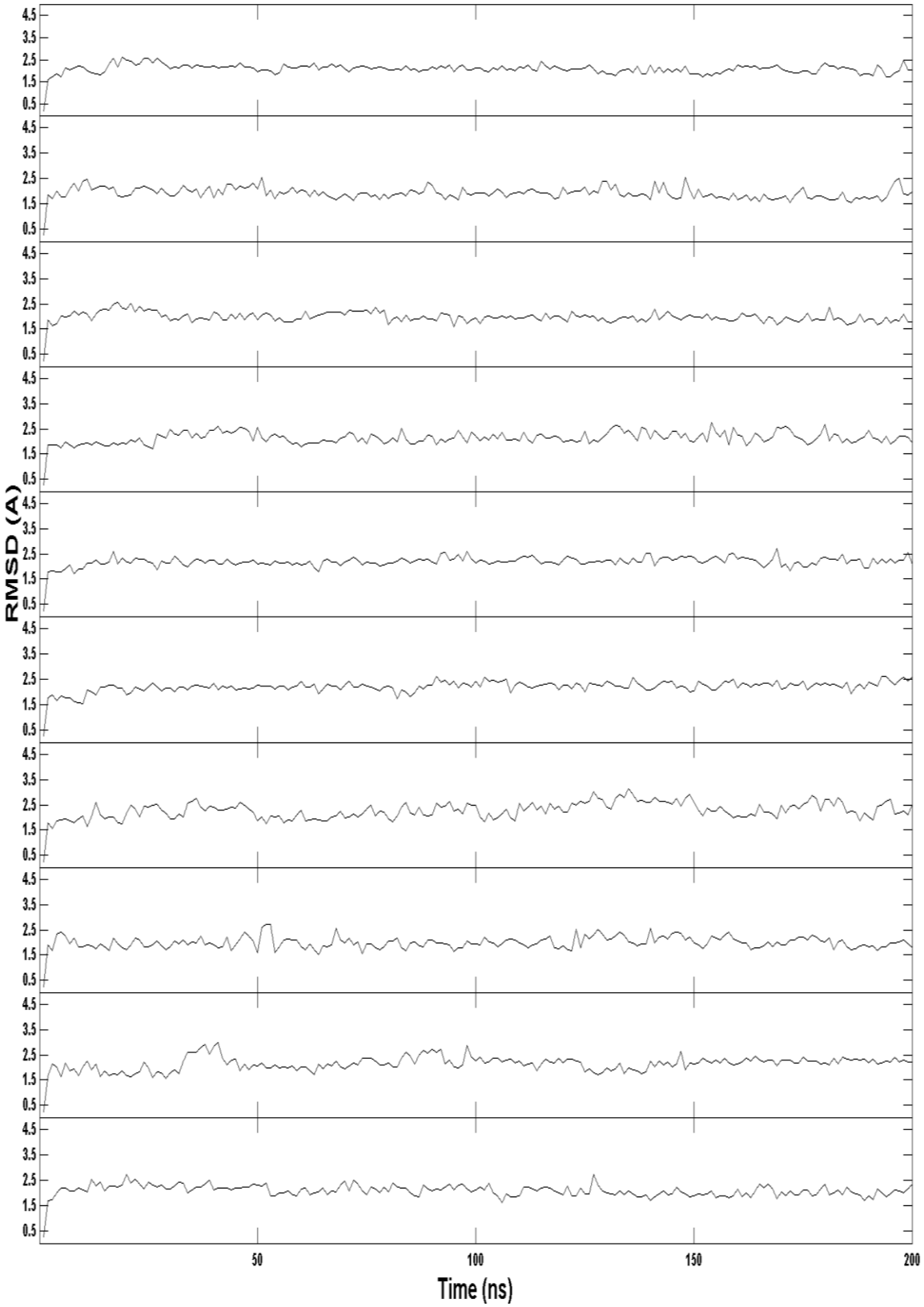


Figure 66. RMSD plot for each trajectory of anti-parallel DNA G-quadruplex and BRACO19 system in ten runs

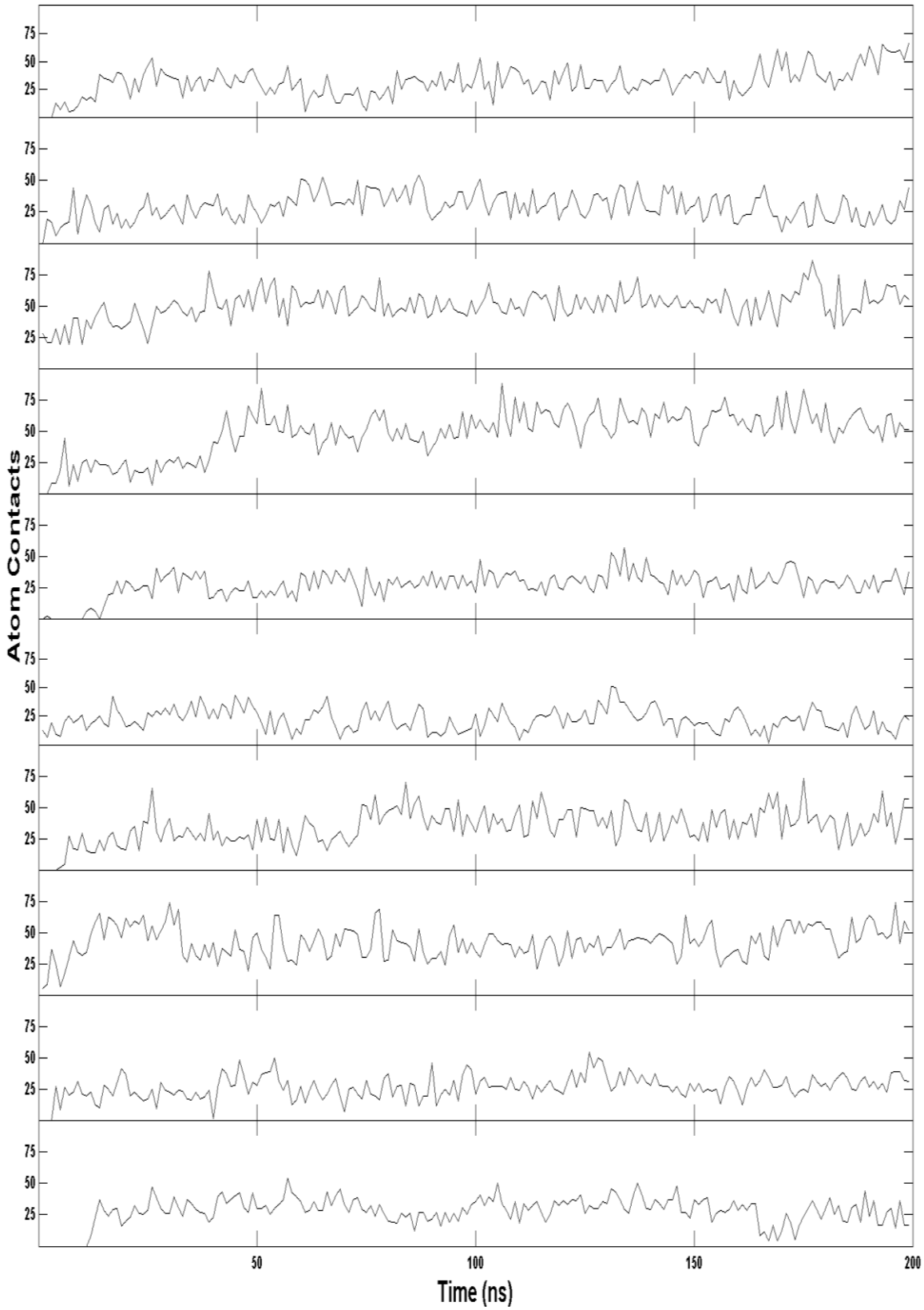


Figure 67. The contact number between anti-parallel DNA G-quadruplex and BRACO19 for each trajectory in ten runs.

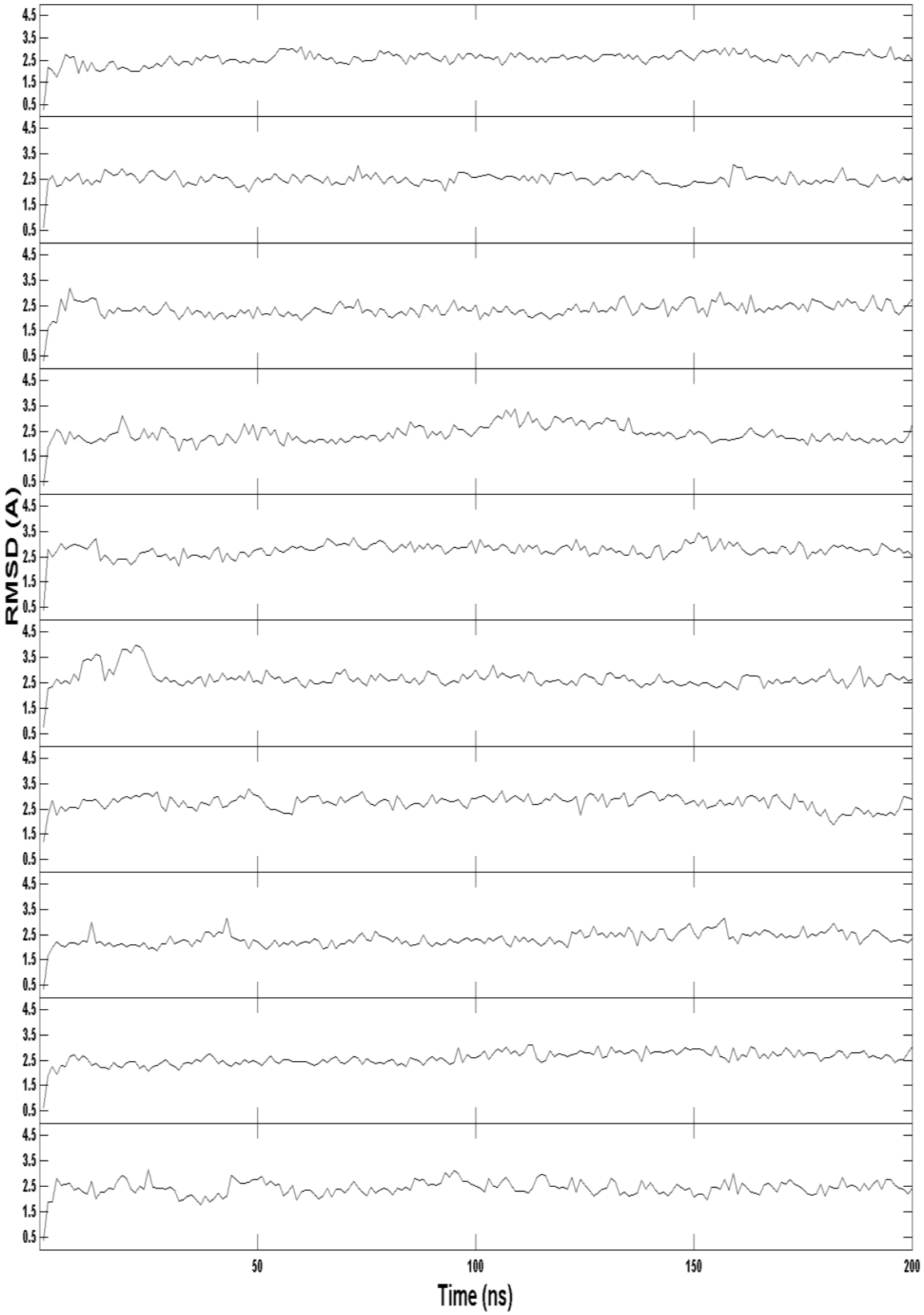


Figure 68. RMSD plot for each trajectory of hybrid DNA G-quadruplex and BRACO19 system in ten runs

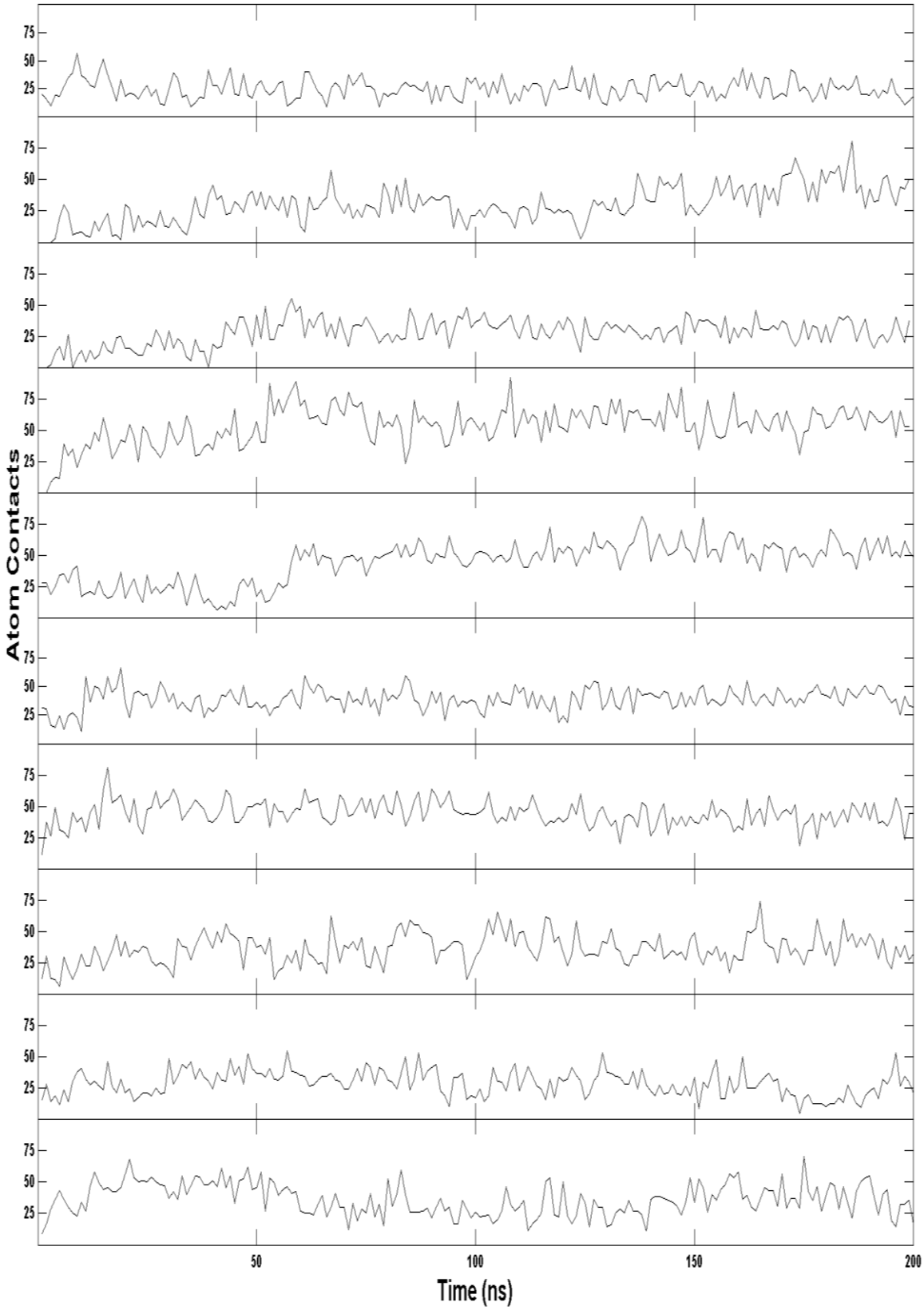


Figure 69. The contact number between hybrid DNA G-quadruplex and BRACO19 for each trajectory in ten runs.

Appendix C

2D Interactions and H-bond Network from Chapter 3

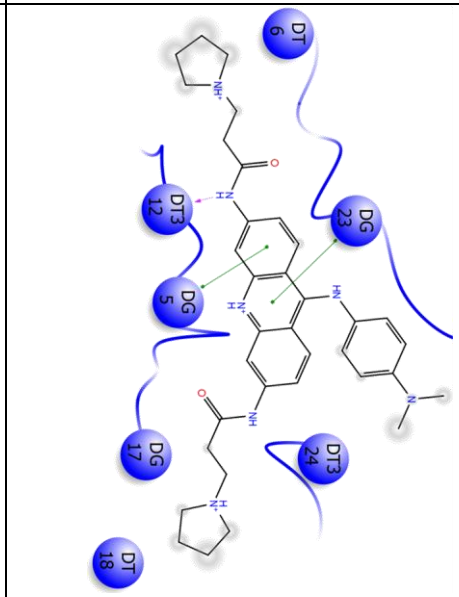
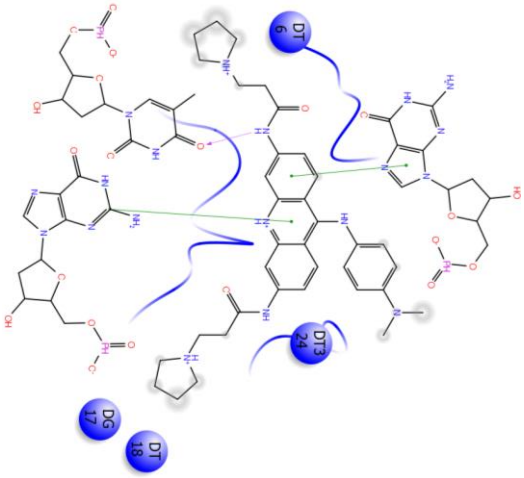
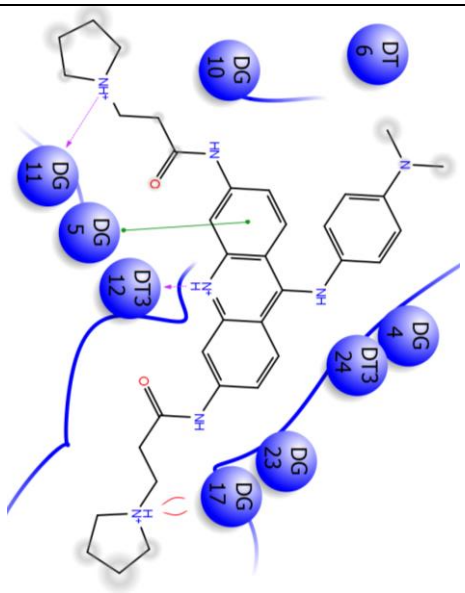
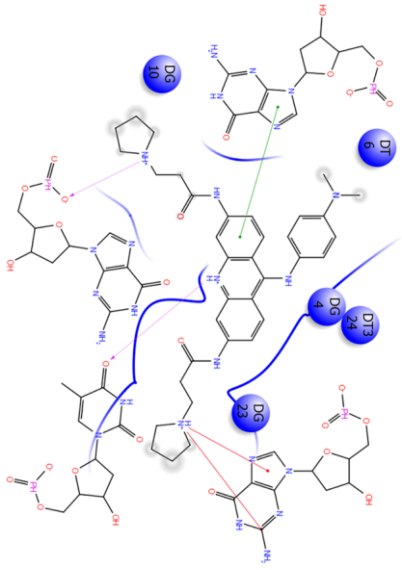
	2D interactions	2D interactions with expanded interacting residues
(A) Crystal Bottom Pose		
(B) MD Relaxed Crystal Bottom Pose		

Figure 70. 2D ligand-DNA interactions of BRACO19 in complex structures of (A) crystal bottom pose, (B) MD relaxed crystal bottom pose, (C) bottom binding pose from free binding simulations, (D) crystal top binding pose, (E) MD relaxed crystal top pose (F) top binding pose from free binding simulations and (G) groove binding pose from free binding simulations.

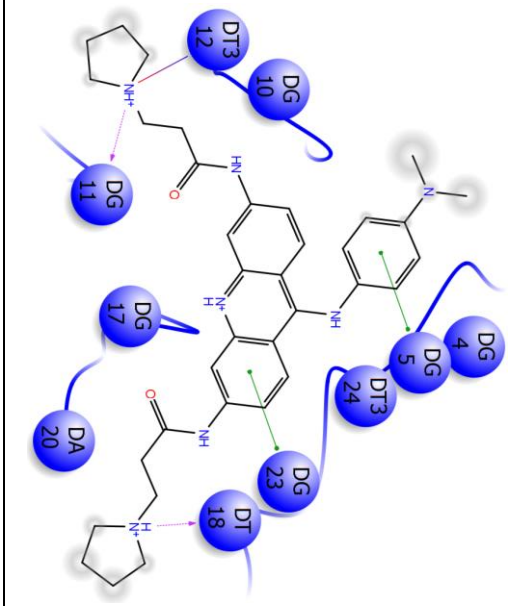
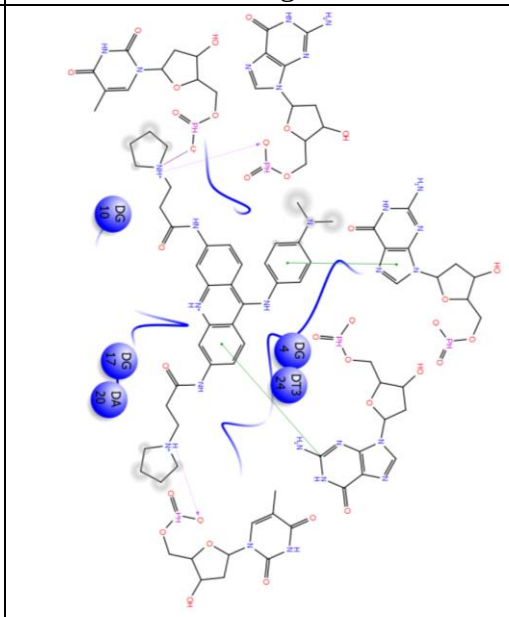
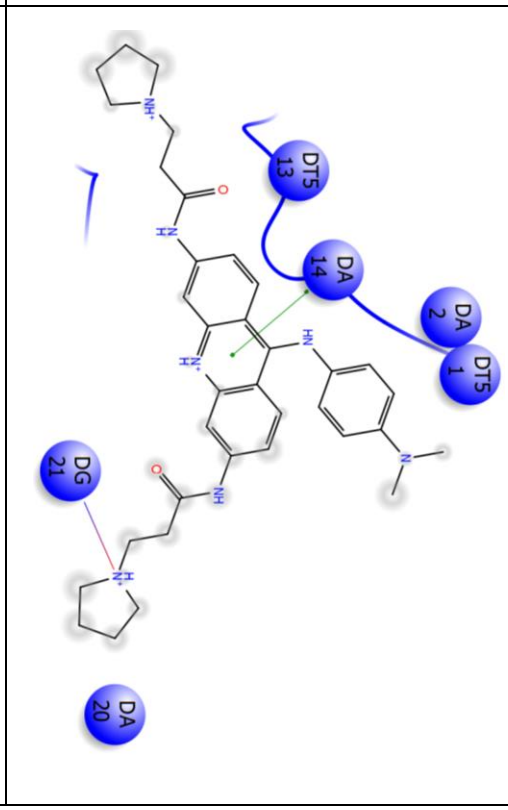
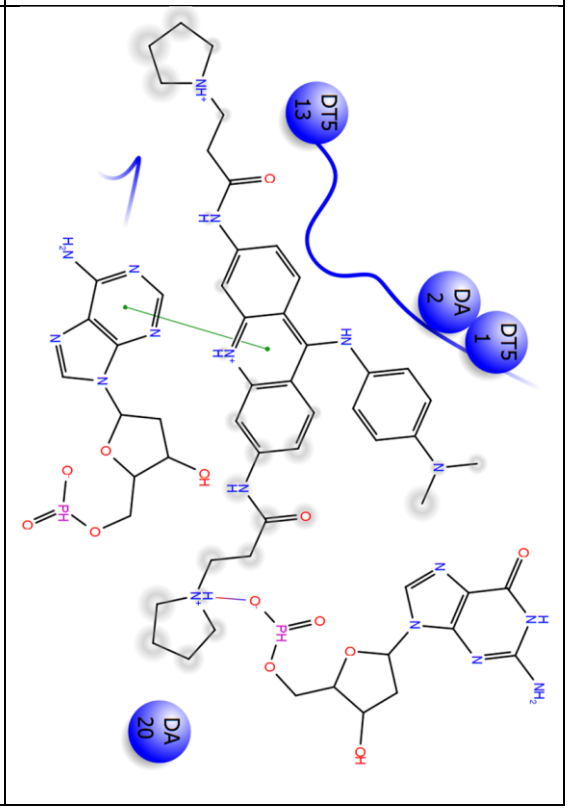
	2D interactions	2D interactions with expanded interacting residues
(C) Free MD bottom Binding Pose		
(D) Crystal top binding Pose		

Figure 70 (continued)

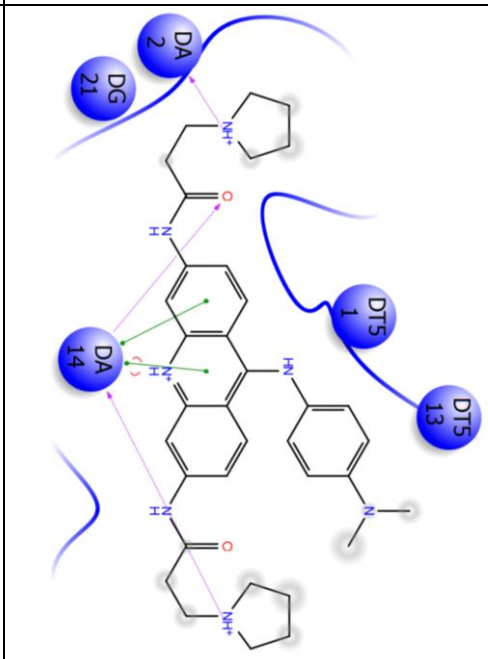
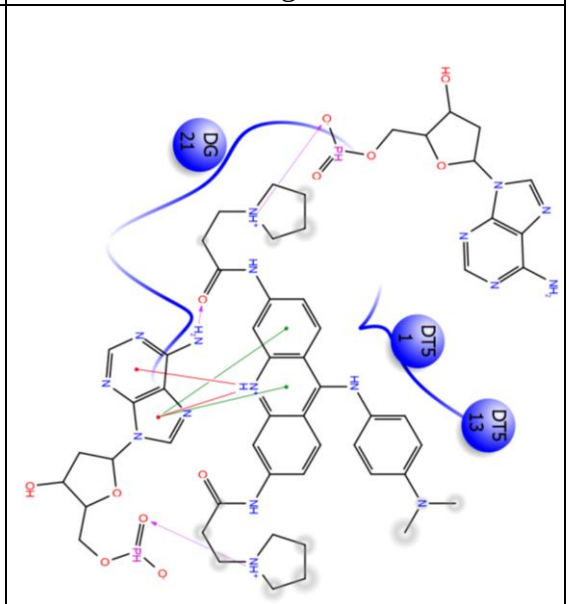
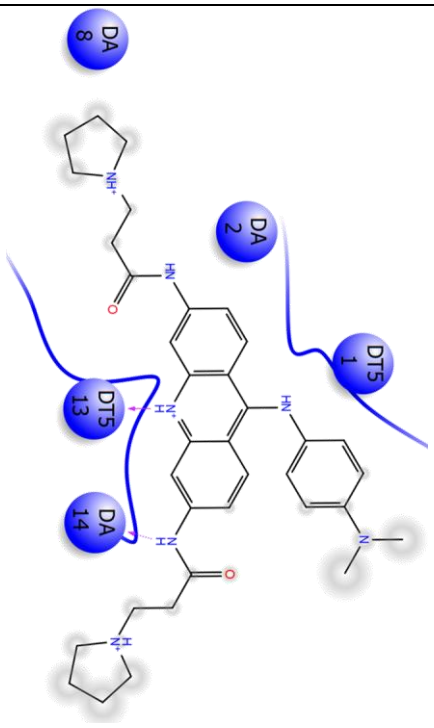
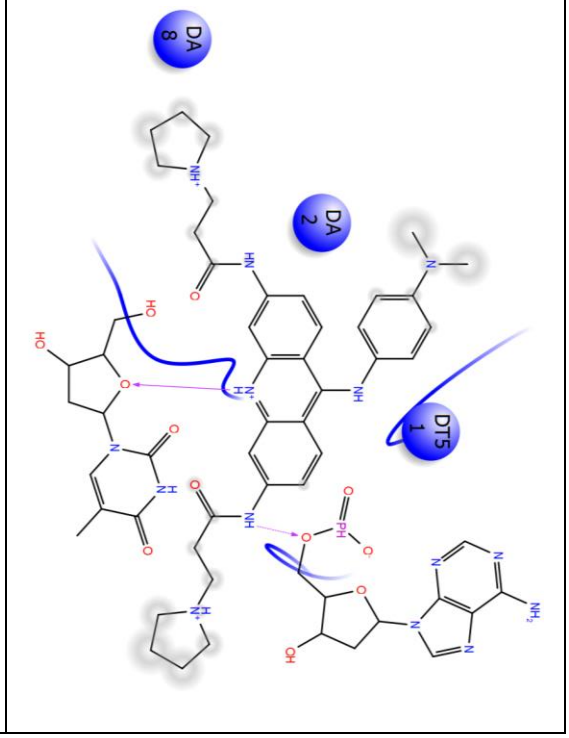
	2D interactions	2D interactions with expanded interacting residues
(E) MD Relaxed Crystal top Pose		
(F) Free MD top Binding Pose		

Figure 70 (continued)

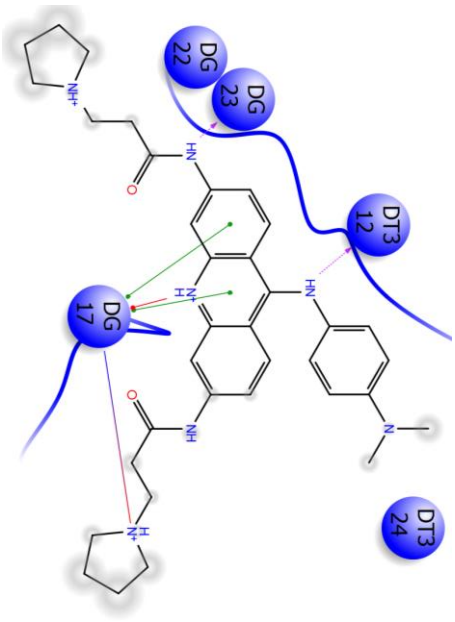
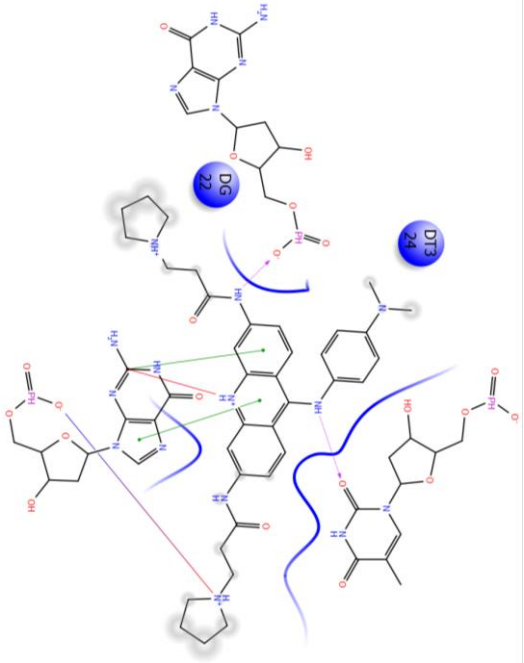
	2D interactions	2D interactions with expanded interacting residues
(G) Free MD Groove Binding Pose		

Figure 70 (continued)

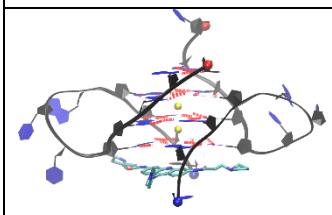
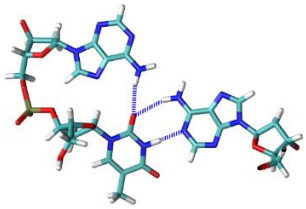
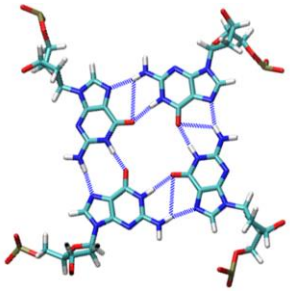
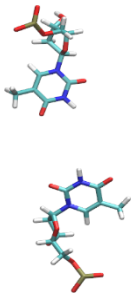
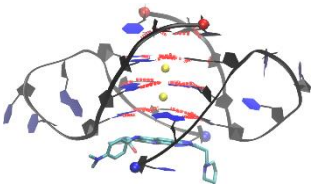
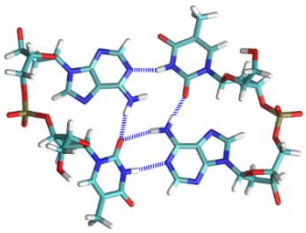
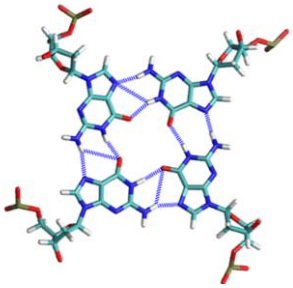
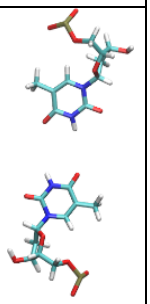
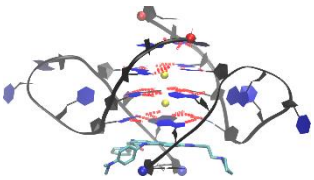
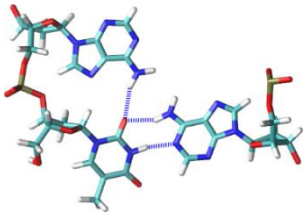
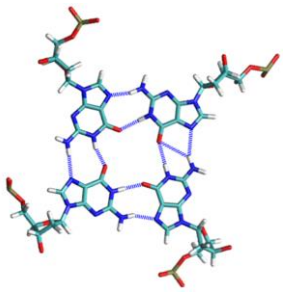
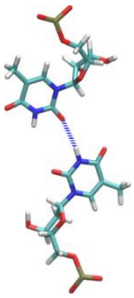
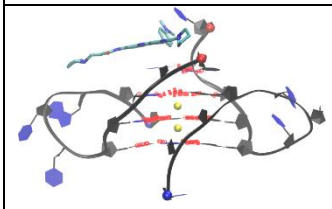
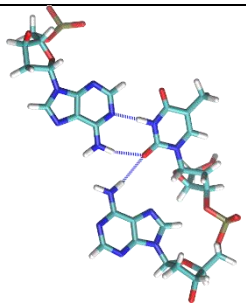
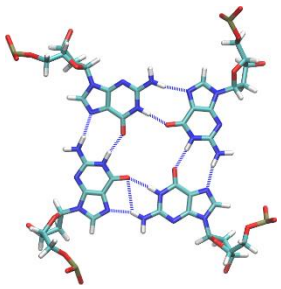
H-bonds	AT-tetrad	G-tetrad (1)	T-dyad
 **(A)** X – Ray Bottom Pose			
 **(B)** MD Relaxed Crystal Bottom Pose			
 **(C)** Free MD bottom Binding Pose			
 **(D)** Crystal top binding Pose			

Figure 71. H-bond network in the layers formed by DNA residues in the representative structure of (A) Crystal Bottom Pose, (B) MD Simulated Crystal Bottom Pose, (C) Bottom Binding Pose from Free MD Binding Simulations, (D) Crystal top binding Pose, (E) MD Simulated Crystal top Pose (F) top Binding Pose from Free MD Binding Simulations and (G) Groove Binding Pose from Free MD Binding Simulations.

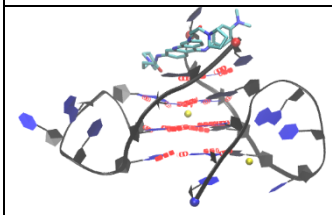
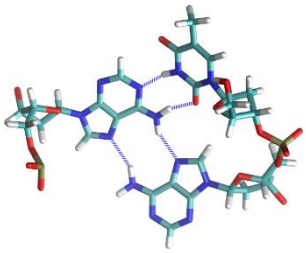
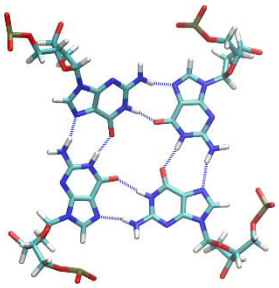
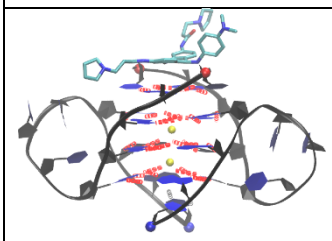
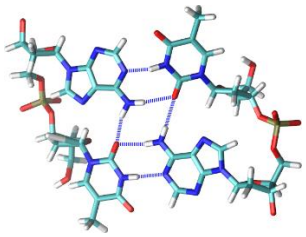
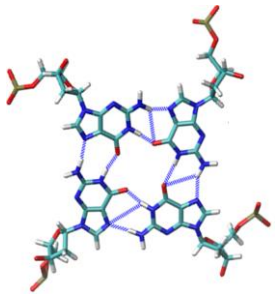
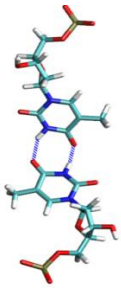
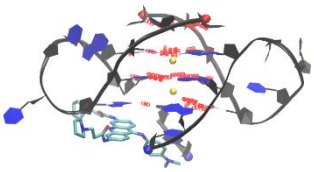
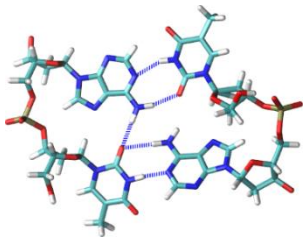
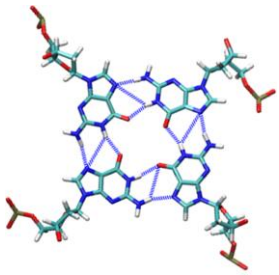
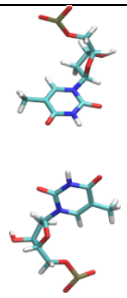
H-bonds	AT-tetrad	G-tetrad (1)	T-dyad
 **(E)** MD Relaxed Crystal top Pose			
 **(F)** Free MD top Binding Pose			
 **(G)** Free MD Groove Binding Pose			

Figure 71 (continued)