



THESIS APPROVAL
GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Chemistry)

DEGREE

Chemistry

FIELD

Chemistry

DEPARTMENT

TITLE: Molecular Dynamics Simulations of Wild Type, Y181C Mutant HIV-1 RT/Nevirapine and Minor Groove Binder, Thiazotropsin A, Complex: The Energetic and Interactions by Means of MM-PBSA Calculations

NAME: Mr. Witcha Treesuwan

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Associate Professor Supa Hannongbua, Dr.rer.nat.)

COMMITTEE MEMBER

(Assistant Professor Waraporn Parasuk, Dr.rer.nat.)

COMMITTEE MEMBER

(Mr. Chak Sangma, Ph.D.)

COMMITTEE MEMBER

(Professor Simon Mackay, Ph.D.)

DEPARTMENT HEAD

(Assistant Professor Noojaree Prasitpan, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

MOLECULAR DYNAMICS SIMULATIONS OF WILD TYPE, Y181C MUTANT HIV-1 RT/NEVIRAPINE AND MINOR GROOVE BINDER, THIAZOTROPSIN A, COMPLEX: THE ENERGETIC AND INTERACTIONS BY MEANS OF MM-PBSA CALCULATIONS

WITCHA TREESUWAN

**A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Chemistry)
Graduate School, Kasetsart University**

2009

Witcha Treesuwan 2009: Molecular Dynamics Simulations of Wild Type, Y181C Mutant HIV-1 RT/Nevirapine and Minor Groove Binder, Thiazotropsin A, Complex: The Energetic and Interactions by Means of MM-PBSA Calculations. Doctor of Philosophy (Chemistry), Major Field: Chemistry, Department of Chemistry. Thesis Advisor: Associate Professor Supa Hannongbua, Dr.rer.nat. 156 pages.

The combination of Molecular Dynamics simulations/MM-PBSA calculations were used to explore thermodynamic properties of two systems; the HIV-1 Reverse Transcriptase (RT) complex to nevirapine and the minor groove binder, thiazotropsin A, complex to DNA, which are comparable to the available experimental data. First, the simulations of wild type and Y181C HIV-1 RT/nevirapine complexes revealed the characteristic hydrogen bonds from the bridge water molecule (WAT1075), which was the key in the stabilizing of the bound complex. Improvement of binding energies calculations was observed when an explicit solvent, WAT1075, was included in the MM-PBSA calculations. Binding energies of -37.65 and -29.82 kcal/mol found in the wild type and Y181C HIV-1 RT, respectively. The attractive interactions via the bridge water brought His235 and Leu234 became major contributions. However, the presence of WAT1075 in the Y181C RT complex presented the weaker hydrogen bond distance formation, lack of attractive force to nevirapine and lack of binding efficiency led to the fail of nevirapine against the Y181C HIV-1 RT. Quantitative understanding of the role of bridge water will help to develop and design for novel HIV-1 RT inhibitors active against mutant enzyme. Second, the simulations of thiazotropsin A and DNA dodecamer, d(GCGACTAGTCGC)₂, at 2:1 ratio were performed using several combination of parameter sets and simulation protocols. Evaluation of the model revealed that combination of ligand charges from HF/6-31G*, polarizable force field for DNA and loop protocol equilibration reproduced the best binding energy of -10.06 kcal/mol, compared to the experimental data of -10.0 kcal/mol from the Isothermal Titration Calorimetry (ITC). The reproducible energy was observed only when the isolate trajectories were used in the MM-PBSA calculations. The major and minor interactions also revealed the recognition pattern of thiazotropsin A to the floor of the DNA minor groove. High correlation between protocols was observed, but not models and parameter sets. These will be used as a benchmark for the side-by-side simulations in term of the Quantitative Protocol Activities Relationship (QPAR) in the future.

Student's signature

Thesis Advisor's signature

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to a number of people who giving me the guidance, help and support to reach my goal of this thesis. First of all, most of credits in this thesis should justifiably go to my advisor, Associate Professor Dr. Supa Hannongbua and Professor Simon Mackay, for valuable guidance, continuous support, kindness and encouragement throughout the course of my graduate. I would like to thank my committees, Dr. Chak Sangma, Associate Professor Dr. Supanna Techasakul, Assistance Professor Dr. Waraporn Parasuk and Assistance Professor Dr. Luckhana Lawtrakul for their helpful comments and suggestions. Furthermore, thanks Dr. Nahoum Anthony and Guillaume Huchet for guiding me producing good researches. Thanks my friend who helped me when I feel hopeless. My colleagues at Laboratory for Computational and Applied Chemistry (LCAC) at Kasetsart University are grateful for their providing helpful assistance and sharing useful ideas.

I would like to acknowledge financial supports from the Royal Golden Jubilee Ph.D. Program (3C.KU/47/B.1) for a scholarship. Partial supporting by the University of Strathclyde, the Thailand Research Fund (RSA5080005) and National Nanotechnology Center (NANOTEC), Ministry of Science and Technology, Thailand, through its program of Center of Excellence Network is grateful.

The gorgeous supply of facilities at the Laboratory for Computational and Applied Chemistry (LCAC), Kasetsart University and the Pharmaceutical Department, University of Strathclyde are very appreciated for computing time and research facilities. The computer resource supported from the High Performance Computing and Networking Center (HPCNC) at Kasetsart University. A generous support by the Postgraduate Education and Research Programs in Petroleum, Petrochemical Technology and Advanced Materials is thankful.

Finally, I would like to express my most profound gratitude to my family who have been giving me love, sustained, encouraged and supported me since the beginning of my educations.

Witcha Treesuwan

April, 2008

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	i
LIST OF TABLES	iii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	xiv
CHAPTER I: MOLECULAR DYNAMICS SIMULATIONS OF WILD TYPE, Y181C MUTANT HIV-1 RT/NEVIRAPINE AND MINOR GROOVE BINDER, THIAZOTROPSIN A, COMPLEX: THE ENERGETIC AND INTERACTIONS BY MEANS OF MM-PBSA CALCULATIONS	
INTRODUCTION	1
METHODS OF CALCULATIONS	9
Theoretical background of Molecular Dynamics (MD) simulations	9
Theoretical background of the Molecular Mechanics and Poisson Boltzmann Surface Area (MM-PBSA) calculations	11
Starting geometries for the models of HIV-1 RT/Nevirapine Complexes	14
Molecular Dynamics Simulations of HIV-1 RT/Nevirapine Complexes	15
Computational Analysis of HIV-1 RT/Nevirapine Complexes	18
RESULTS AND DISCUSSIONS	22
Equilibration of HIV-1 RT/Nevirapine Complexes	22
Binding Energy Calculations of HIV-1 RT/Nevirapine Complexes	29
Folded Resistance Calculations of HIV-1 RT/Nevirapine Complexes	41
Decomposition Energy Calculations of HIV-1 RT/Nevirapine Complexes	42
CONCLUSIONS	47

TABLE OF CONTENTS (continued)

CHAPTER II: MOLECULAR DYNAMICS SIMULATIONS OF WILD TYPE, Y181C MUTANT HIV-1 RT/NEVIRAPINE AND MINOR GROOVE BINDER, THIAZOTROPSIN A, COMPLEX: THE ENERGETIC AND INTERACTIONS BY MEANS OF MM-PBSA CALCULATIONS	
INTRODUCTION	48
METHODS OF CALCULATIONS	59
Force Field and Parameters for Thiazotropsin A	59
Force Field for DNA Sequence	60
Starting Models of the DNA/Thiazotropsin A Complex	60
Setup the Simulations of the DNA/Thiazotropsin A Complex System	61
Computational Calculations of the DNA/Thiazotropsin A Complex	63
Computational Analysis of the DNA/Thiazotropsin A Complex	74
RESULTS AND DISCUSSIONS	76
Comparison of the Applied Force Fields on Thiazotropsin A	76
The Energy Equilibration of the DNA/Thiazotropsin A Complexes	77
Structural Equilibration of the DNA/Thiazotropsin A Complexes	83
The CPU timing in the simulations of the DNA/Thiazotropsin A Complexes	98
Binding Free Energy Calculations of the DNA/Thiazotropsin A Complexes	100
Hydrogen Bond Interactions in the DNA/Thiazotropsin A Complexes	127
CONCLUSIONS	132
LITERATURE CITED	136

LIST OF TABLES

Chapter I

Table		Page
1	Series of minimization steps use for both wild type and Y181C HIV-1 RT complex with nevirapine. Restraint weight applied on each part (kcal/mol)	16
2	Series of simulation steps use for Y181C HIV-1 RT complex with nevirapine. Restraint weight applied on each part (kcal/mol)	17
3	The experimental binding affinity between both wild type and Y181C HIV-1 RT with nevirapine (μM)	29
4	MM-PBSA energies (kcal/mol) of the HIV-1 RT complex with nevirapine calculated from the complex, enzyme, nevirapine and solvent water (WAT)	34
5	The closest water molecule within 7 Å from nevirapine inside the binding pocket of the wild type HIV-1 RT	35
6	Closest water molecule within 7 Å from nevirapine inside the binding pocket of the Y181C HIV-1 RT	36
7	MM-PBSA energies (kcal/mol) of the wild type HIV-1 RT complex with nevirapine calculated by include different numbers of solvent molecules	37
8	MM-PBSA energies (kcal/mol) of the Y181C HIV-1 RT complex with nevirapine calculated by including different numbers of solvent molecules	37
9	Binding energy of wild type and Y181C HIV-1 RT with nevirapine from experimental and calculations (kcal/mol)	38
10	Binding entropy of wild type and Y181C HIV-1 RT with nevirapine from calculations (kcal/mol)	40
11	Relative folded resistance energies ($\Delta\Delta\text{G}$) in kcal/mol for Y181C HIV-1 RT mutations normalized to nevirapine	42

LIST OF TABLES (Continued)

Table		Page
12	Interaction energies (IE, kcal/mol) of nevirapine with individual residues	44
13	Decomposition energies (kcal/mol) of bridging WAT1075 with amino acid residues and nevirapine	46
Chapter II		
Table		
14	The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 1	64
15	The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 2	65
16	The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 3	66
17	The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 4, the straight equilibration	68
18	The stepwise to apply the restraint force on DNA and thiazotropin A in the simulations system type 4, the loop equilibration	70
19	The stepwise to apply the restraint force on DNA and thiazotropin A in the simulations system type 5	71
20	Summary of combination systems in the MD1, B3LYP/6-31G** and force field version 1 for thiazotropsin A, performed in the simulations (✓)	72
21	Summary of combination systems in the MD2, B3LYP/6-31G** and force field version 2 for thiazotropsin A, performed in the simulations (✓)	72

LIST OF TABLES (Continued)

Table		Page
22	Summary of combination systems in the MD3, HF/6-31G* and force field version 1 for thiazotropsin A, performed in the simulations (✓)	73
23	Summary of combination systems in the MD4, HF/6-31G* and force field version 2 for thiazotropsin A, performed in the simulations (✓)	73
24	Summary of the performed simulations systems 5 (✓)	73
25	The combination systems type sand model types	79
26	The total CPU time usage (hours) for combination systems types and model types with ff03 force field	99
27	The total CPU time usage (hours) for combination systems types and model types with ff02 force field	99
28	The CPU time usage (seconds) per 100 steps in the production MD simulations from combination systems types and model types with ff03 force field	99
29	The CPU time usage (seconds) per 100 steps in the production MD simulations from combination systems types and model types with ff02 force field	100
30	Binding energy (kcal/mol) results: comparison between single CPU (Vader) and Opteron cluster (Alien)	101
31	Effect of the number of snapshots on the binding free energy (kcal/mol), taken from simulation system 2 combined with MD1 and ff03 force field	102
32	Binding energy (kcal/mol) results for invacuo systems with restrain DNA using optimized structure of thiazotropsin A as a starting	103

LIST OF TABLES (Continued)

Table		Page
33	Binding energy (kcal/mol) results for invacuo systems with restrained DNA using NMR structure of thiazotropsn A as a starting with RESP charges	104
34	Binding energy (kcal/mol) results for invacuo systems with restraints on DNA	106
35	Binding energy (kcal/mol) results for invacuo systems with restraints on DNA and thiazotropsin A	106
36	Binding energy (kcal/mol) results for invacuo systems with higher restraints on DNA and thiazotropsin A	107
37	Binding energy (kcal/mol) results for invacuo systems including counter ions with restraints on DNA	108
38	Binding energy (kcal/mol) of neutralized solvated systems in implicit and explicit solvent with restraints on DNA and ff03 force field	109
39	Binding energy (kcal/mol) of neutralized solvated systems in implicit and explicit solvent with restraints on DNA and ff02 force field	110
40	Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B1 with ff03 force field	112
41	Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B1 with ff02 force field	112
42	Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B2 with ff03 force field	114

LIST OF TABLES (Continued)

Table		Page
43	Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B2 with ff02 force field	114
44	Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using the AMBER tutorial equilibration protocol	122
45	The binding free energies (kcal/mol) from <i>single</i> and <i>separate</i> trajectories calculations with non-polarizable force field FF03	123
46	The binding free energies (kcal/mol) from <i>single</i> and <i>separate</i> trajectories calculations with polarizable force field FF02	124
47	Summary of the binding free energies (kcal/mol) from various models, protocols and force field types	125
48	Summary of the hydrogen bonds in the 2:1 complex formed between thiazotropsin A and d(CGACTAGTCG) ₂ based on ¹ H NMR chemical shift, labile proton exchange characteristics, and solution structure information	128

LIST OF FIGURES

Chapter I

Figure		Page
1	The structure of HIV-1 RT polymerization complex with DNA. Subdomains in p66 are colorized by finger (blue), palm (orange), thumb (green), connection (purple) and RNaseH (red). P51 subunit is shown in light blue color. The DNA duplex is rendered as arrow and ladder. In the red circle located the active site of the HIV-1 RT	2
2	Structures of the anti-HIV nucleoside analogues (NRTIs)	4
3	Structures of non-nucleoside reverse transcriptase inhibitor (NNRTIs)	4
4	The schematic of thermodynamic cycle	12
5	The diagram shows the preparing of the cap water model of the Y181C HIV-1 RT	15
6	Total energy, potential energy and kinetic energy of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine	23
7	Temperature of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine	24
8	Root-mean-square displacement (RMSD) without <i>fit function</i> of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine. The pocket 10 Å of enzyme complex shows in blue, the pocket 10 Å of enzyme alone shows in pink, nevirapine shows in green and the WAT1075 shows in brown color	27

LIST OF FIGURES (Continued)

Figure		Page
9	The root-mean-square displacement (RMSD) with <i>fit function</i> of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine. The pocket 10 Å of enzyme complex shows in blue, the pocket 10 Å of enzyme alone shows in pink, nevirapine shows in green and the WAT1075 shows in brown color	28
10	Distance between water molecules and the center of mass of nevirapine along with the simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine	31
11	Distance of hydrogen-bonded bridge at 3 ns referred by heteroatomic distance. In a) wild type HIV-1 RT complex and b) Y181C HIV-1 RT complex	33
12	Plot of interaction energies (kcal/mol) of nevirapine with individual residues. The calculations from ONIOM method with the wild type RT is shown in light blue circle, decomposition method with wild type RT is shown in red diamond, decomposition method with Y181C HIV-1 RT is shown in black triangle, the MFCC method with wild type RT is shown in green square, and the MFCC method with Y181C HIV-1 RT is shown in dark blue triangle	45
Chapter II		
Figure		
13	The molecular structure of DNA	48
14	The back bond DNA strand components; a) DNA nucleotide, b) the four bases of DNA showing their complementary binding properties	50
15	A intercalator bound and orientated inside the DNA sequence	51

LIST OF FIGURES (Continued)

Figure		Page
16	The structures of distamycin, netropsin, and several other minor groove ligands are shown. All have in common a planar aromatic core and positively charged end groups	53
17	Structures are shown of distamycin bound in 1:1 (left) and 2:1 (right) complexes. The DNA is shown in gray, while the distamycin molecules in the minor groove are colored according to atom type. The backbone atoms are shaded darker gray to highlight the change in the minor groove width between the two forms of complex	54
18	(I) Representation of the binding to DNA of netropsin; (II) proposed representation of a model lexitropsin molecule with guanine residues in DNA. Heavy arrows are hydrogen bonds, from donor to acceptor. Dashed lines mark close van der Waals nonbonded contacts between DNA and drug	57
19	Compound 1 , thiazole lexitropsins (thiazotropsin A), and the self complementary DNA duplex 2	57
20	The starting structure of Thiazotropsin A used in the simulations. a) From NMR. b) From B3LYP/6-31G** optimization	61
21	The setup steps of the DNA/Thiazotropsin A simulations system. A) Starting solute. B) Solvation in a box of water. C) Neutralized by adding counter ions, Cl ⁻	62
22	Superimposition of thiazotropsin a) B3LYP/6-31G** Optimized (dark blue) and minimized from force field version 1 (light blue). b) B3LYP/6-31G** Optimized (dark blue) and minimized from force field version 2 (purple)	76
23	The total energy from opt. and NMR starting	79

LIST OF FIGURES (Continued)

Figure		Page
24	The total energy from ff03 and ff02 force field	80
25	The total energy from various systems; a) the overview, b) focusing only in the invacuo simulations group and c) focusing only in the solvation simulations group	80
26	The total energy from various topology parameters; a) the overview and b) at higher resolutions	82
27	The total energy from simulations system 5	83
28	The RMSD fluctuation in simulations system 4 (B1) with the optimized and NMR starting structure. a) complex, b) DNA and c) thiazotropsin A	84
29	The RMSD fluctuation in simulations system 3 (the restraint simulations) with the ff03 and ff02 force field for DNA. a) complex, b) DNA and c) thiazotropsin A	86
30	The RMSD fluctuation in simulations system 4 (B1) (the free simulations) with the ff03 and ff02 force field for DNA. a) complex, b) DNA and c) thiazotropsin A	88
31	The RMSD fluctuation from simulations system 3 (the restraint simulations) with various topology parameters (MD1, MD2, MD3 and MD4) usage for thiazotropsin A. a) complex, b) DNA and c) thiazotropsin A	91
32	The RMSD fluctuation from simulations system 4 (the free simulations) with various topology parameters (MD1, MD2, MD3 and MD4) usage for thiazotropsin A. a) complex, b) DNA and c) thiazotropsin A	93
33	The RMSD fluctuation from various simulations system (system 1, 2, 3, 4 (B1) and 4 (B2)) by using the same MD2 parameter set. a) complex, b) DNA and c) thiazotropsin A	95
34	The RMSD fluctuation from the simulations system 5	97

LIST OF FIGURES (Continued)

Figure		Page
35	Superimposition of DNA and thiazotropsin A from difference simulations in solution	97
36	The 3-Dimensional drawing of nucleic acid called Calladine-Drew style, showing an ideal B-DNA and A-DNA form	115
37	The Calladine-Drew projections base-centered of the starting structure Thiazotropsin A/ds(CGCACTAGTGCG) ₂ complex from NMR refinement. a) top view. b) side view	115
38	The Calladine-Drew projections base-centered from top view of the simulations structure Thiazotropsin A / ds(CGCACTAGTGCG) ₂ complex taken the last configure from all solvation trajectories with ff03 force field	116
39	The Calladine-Drew projections base-centered from top view of the simulations structure Thiazotropsin A / ds(CGCACTAGTGCG) ₂ complex taken the last configure from all solvation trajectories with ff02 force field	117
40	The Calladine-Drew projections base-centered from side view of the simulations structure Thiazotropsin A / ds(CGCACTAGTGCG) ₂ complex taken the last configure from all solvation trajectories with ff03 force field	118
41	The Calladine-Drew projections base-centered from side view of the simulations structure Thiazotropsin A / ds(CGCACTAGTGCG) ₂ complex taken the last configure from all solvation trajectories with ff02 force field	120
42	The correlation of model usage corresponding to various protocols for the <i>single</i> trajectory calculations	126
43	The correlation of protocols usage corresponding to various models for the <i>single</i> trajectory calculations	127

LIST OF FIGURES (Continued)

Figure		Page
44	The correlation of protocols usage corresponding to various models for the <i>separate</i> trajectories	127
45	Interactions labeling between thiazotropsin A and DNA base-pairs correspond to NMR chemical shift assignment	128
46	Five hydrogen bonds distance between heteroatom observed in the solvate simulations from MD1, MD2, MD3 and MD4 within non-polarizable (ff03) force field. The simulations system 3 labeled in red, system 4 (B1) is in black and system 4 (B2) is in blue line. a) H2--T ⁵ O2, b) H9--A ⁶ N3, c) H16--G ⁷ N3, d) N21--G ⁷ H22, e) H26--T ⁸ O2	130
47	Five hydrogen bonds distance between heteroatom observed in the solvate simulations from MD1, MD2, MD3 and MD4 within non-polarizable (ff02) force field. The simulations system 3 labeled in red, system 4 (B1) is in black and system 4 (B2) is in blue line. a) H2--T ⁵ O2, b) H9--A ⁶ N3, c) H16--G ⁷ N3, d) N21--G ⁷ H22, e) H26--T ⁸ O2	131

LIST OF ABBREVIATIONS

AIDS	=	Acquired Immune Deficiency Syndrome
BP	=	Binding Pocket
Cys (C)	=	Cysteine
DNA	=	Deoxyribonucleic Acid
dsDNA	=	double-stranded DNA
HIV-1	=	Human Immunodeficiency Virus Type 1
His (H)	=	Histidine
ITC	=	Isothermal Titration Calorimetry
K _i	=	Inhibitor Binding Constant
Leu (L)	=	Leucine
MD	=	Molecular Dynamic Simulations
MM	=	Molecular Mechanics
MM-PBSA	=	Molecular Mechanics-Poisson Boltzmann Surface Areas
Nev	=	Nevirapine
NNRTIs	=	Non-Nucleoside Reverse Transcriptase Inhibitors
NRTIs	=	Nucleoside Reverse Transcriptase Inhibitors
QM	=	Quantum Mechanics
REC	=	Receptor
RMSD	=	Root Mean Squares Deviation
Tyr (Y)	=	Tyrosine
WAT	=	Water molecule

CHAPTER I: MOLECULAR DYNAMICS SIMULATIONS OF WILD TYPE, Y181C MUTANT HIV-1 RT/NEVIRAPINE AND MINOR GROOVE BINDER, THIAZOTROPSIN A, COMPLEX: THE ENERGETIC AND INTERACTIONS BY MEANS OF MM-PBSA CALCULATIONS

INTRODUCTION

HIV-1 Reverse Transcriptase (RT) is a versatile enzyme that synthesizes double-stranded proviral DNA from single stranded viral RNA (Katz *et al.*, 2001; Whitcomb and Hughes, 1992). This cause the HIV-1 RT is a key enzyme in the replication cycle of the human immunodeficiency virus type 1 (HIV-1). In the process, RT catalyzes both RNAand DNA-dependent DNA polymerization and RNase H hydrolysis. This essential step in the retroviral life cycle is targeted by a variety of drugs in clinical use to combat AIDS (Acquired Immuno Deficiency Syndromes). The native enzyme in the virion is a heterodimer of p66 and p51 subunits (Veronese *et al.*, 1986). The p66 subunit has polymerase which consists of fingers, palm and thumb subdomains, named according to their resemblance to a right hand, as well as connection and ribonuclease H (RNaseH) subdomains. The p51 subunit is derived from p66 by proteolytic removal of the C-terminal RNase H domain. A dimeric enzyme form is requisite for enzyme activity. The p66/p51 heterodimer and p66/p66 homodimer have both polymerase and RNase H activities (Restle *et al.*, 1990); the p51/p51 homodimer has only polymerase activity (Bavand *et al.*, 1993). The heterodimer has an asymmetric structure with a single polymerase active site on the p66 subunit (Le Grice *et al.*, 1991). In addition to RT, several cis-acting sequences on the nucleic acid orchestrate key steps in HIV-1 replication. These include the primer binding site, long terminal repeats, 3' and central polypurine tracts, and central termination sequence (CTS). The CTS is a 29 base-pairs (bp) sequence with weak and strong termination sites for (+) strand DNA synthesis (Charneau *et al.*, 1994). The structural detail of the HIV-1 RT complex to the DNA duplex is shown in figure 1.

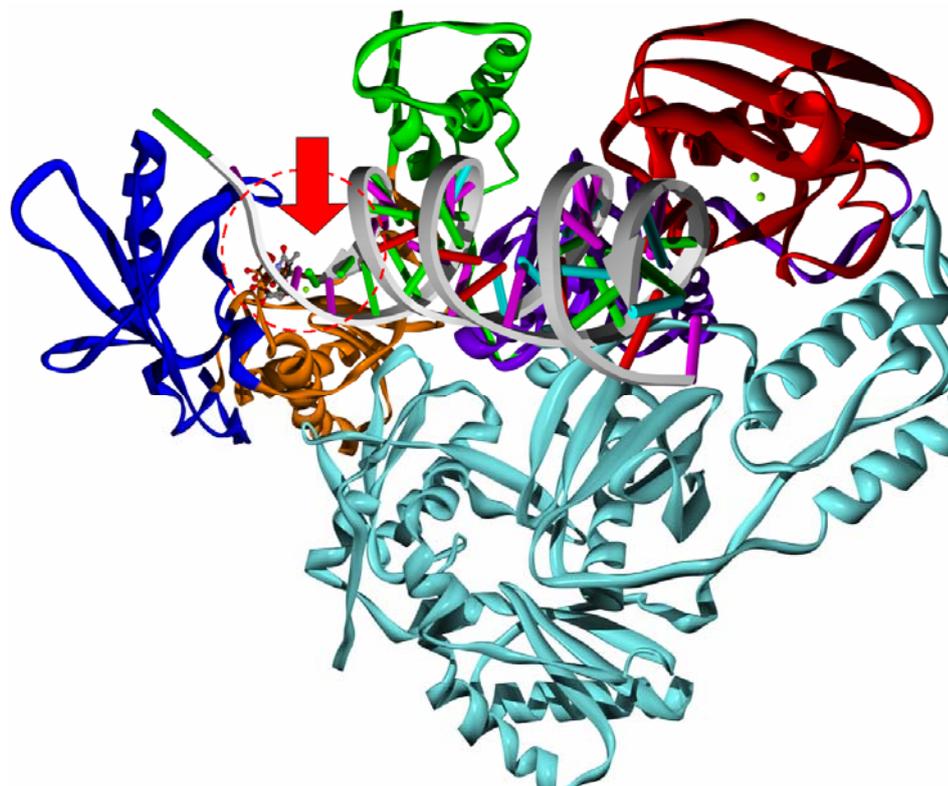


Figure 1 The structure of HIV-1 RT polymerization complex with DNA.

Subdomains in p66 are colorized by finger (blue), palm (orange), thumb (green), connection (purple) and RNaseH (red). P51 subunit is shown in light blue color. The DNA duplex is rendered as arrow and ladder. In the red circle located the active site of the HIV-1 RT.

Several RT inhibitors have been developed and approved by the FDA and are in clinical use (Koup *et al.*, 1991; Richman *et al.*, 1991). The two main classes of RT inhibitors are identified as nucleoside analogues (NRTIs) and non-nucleoside inhibitors (NNRTIs) (Katz and Skalka, 1994; Mitsuya *et al.*, 1990; Rizzo *et al.*, 2001; Wang *et al.*, 2001). The first class is nucleoside analogues (NRTIs) such as AZT (3'-azido-deoxythymidine), ddi (dideoxyxymosine) and ddC (dideoxycytidine) which are competitive inhibitors of the nucleotide substrate and bind to the polymerase active site upon metabolic activation. After incorporation in the DNA strand instead of dNTP, they cause premature termination of the newly synthesized chain. In addition, NRTIs are also acting on other host DNA polymerases, which explains their toxicity (De Clercq, 1995b; Maga *et al.*, 1997). The structures of the NRTIs are listed and

shown in figure 2. Another class is non-nucleoside inhibitors (NNRTIs) such as HEPT (1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine) (Baba *et al.*, 1989), TIBO (tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one) (Pauwels *et al.*, 1990), nevirapine (Dipyridodiazepinones) (Merluzzi *et al.*, 1990) and efavirenz ((-)-6-chloro-4-cyclopropylethynyl-4-trifluoromethyl-1,4-dihydro-2h-3,1-benzoxazin-2-one) (Young *et al.*, 1995). In particular, the non-nucleoside RT inhibitors (NNRTIs) are highly effective and produce few side effects. NNRTIs bind to a common hydrophobic site, the nonnucleoside inhibitor binding pocket (NNIBP), located in the p66 palm subdomain approximately 10 Å away from the polymerase active site (Pedersen, 1999) causing a displacement of the catalytic aspartate residues and, furthermore, show lower cellular toxicity compared to NRTIs. However, the rapid mutations that make RT resistant to NNRTIs are known to arise for most of the amino acids that comprise the binding pocket as well as reducing the efficiency of the drugs (De Clercq, 1995a; Tantillo *et al.*, 1994). Therefore, it is essential to understand the detailed interactions of the inhibitors with wild type and mutant RT, to design drugs that are effective across mutations. Some structures of the NNRTIs are shown in figure 3.

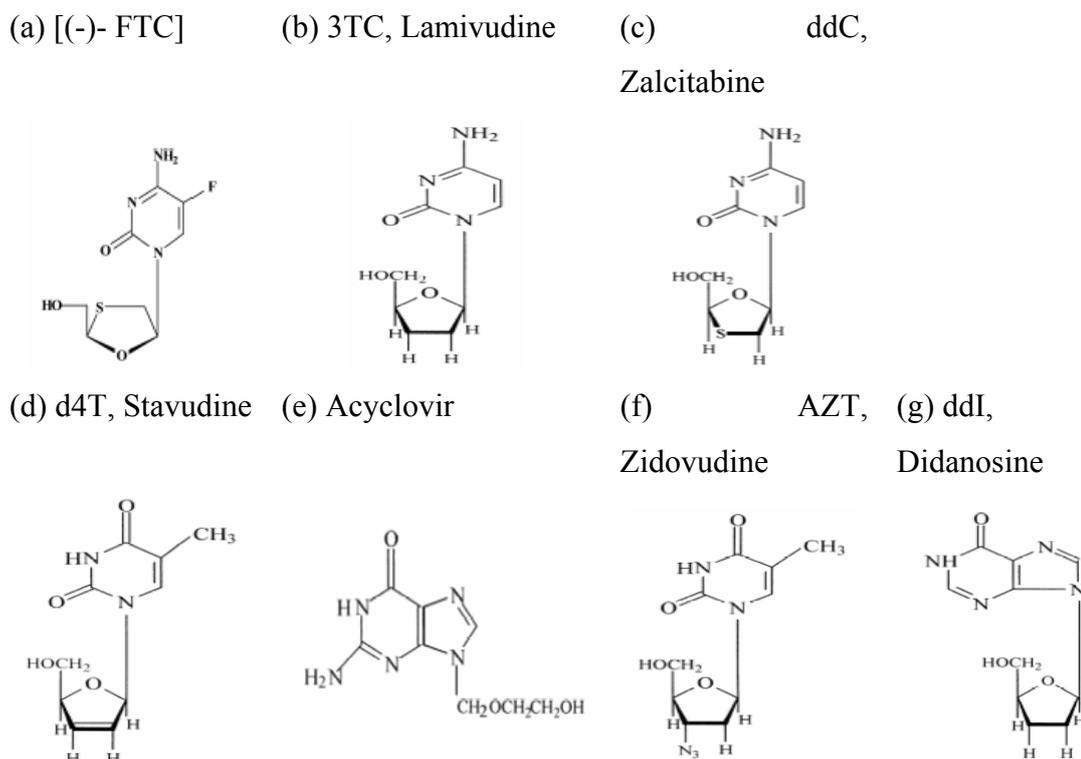
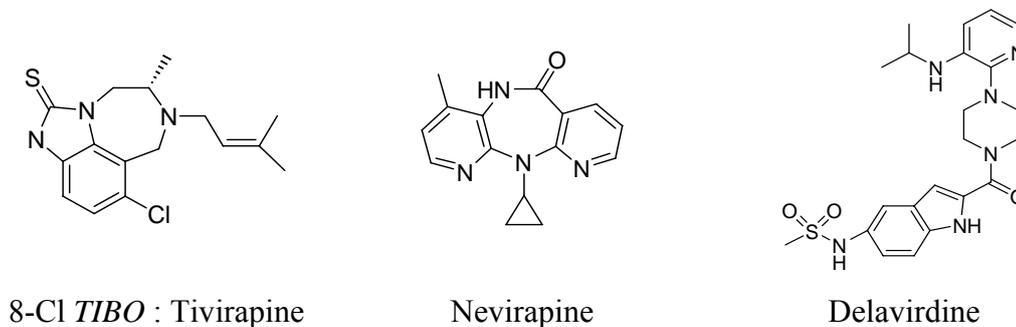


Figure 2 Structures of the anti-HIV nucleoside analogues (NRTIs).

First generation NNRTIs



Second generation NNRTIs

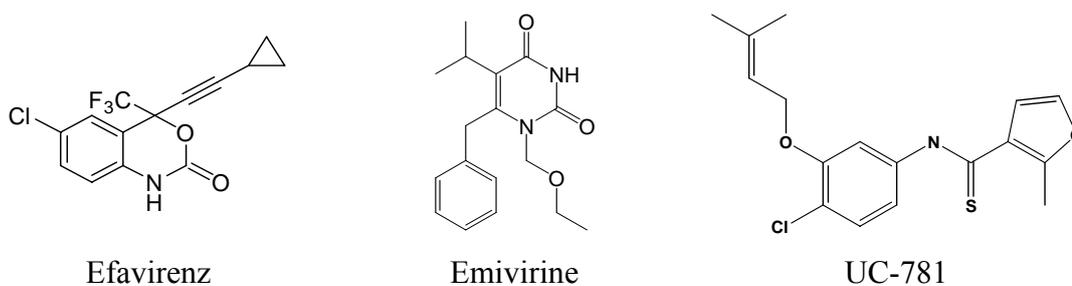


Figure 3 Structures of non-nucleoside reverse transcriptase inhibitor (NNRTIs).

Crystal structures of RT bound to several NNRTIs (Chamberlain *et al.*, 2002; Das *et al.*, 2004; Das *et al.*, 1996; Hsiou *et al.*, 1998; Ren *et al.*, 2001; Spallarossa *et al.*, 2008) show that structurally homologous inhibitors have similar binding patterns and interaction modes to RT. The interactions from various crystal structures of the HIV-1 RT with NNRTIs class had been investigated (Beyer *et al.*, 2004). Although the overall shape of the NNRTI binding site is similar among the different crystal structures of RT/NNRTI complexes, there are subtle differences among them, showing that the binding pocket is flexible and can adapt to the shape of different NNRTIs (Hsiou *et al.*, 1996; Jager *et al.*, 1994). The pocket's formation is probably induced by proximity of the inhibitor. Amino acids in the Non-Nucleoside Inhibitor Binding Pocket (NNIBP), such as Tyr 181, Tyr 188, Val 179, and Phe 227, interact with the bound NNRTI via van der Waals interactions. Hydrogen bonds occur between some inhibitors and amino acids Val 189 and Tyr 318. Water molecules also form a hydrogen bond bridge network between the inhibitor and amino acids at the mouth of the NNIBP (Zhou *et al.*, 2002).

Crystallographic studies have shown that the binding of an NNRTI induces several conformational changes (Sarafianos *et al.*, 1997). While the unliganded RT, the p66 thumb subdomain is folded into the DNA-binding cleft and lies over the palm subdomain, nearly touching the fingers subdomain, in a “thumb down” configuration almost touching the p66 fingers subdomain (Rodgers *et al.*, 1995) upon ligation of an NNRTI the p66 thumb subdomain moves to an “open” or upright position (Kohlstaedt *et al.*, 1992). As a consequence, the thumb's tip is approximately 32 Å away from the fingers. Other structural changes induced by the bound NNRTI include displacements of the base of the p66 thumb, the p66 connection, and RNase H subdomains (Hsiou *et al.*, 1996) and of the region known as the “primer grip” (Das *et al.*, 1996; Ding *et al.*, 1995b; Tantillo *et al.*, 1994). The primer grip positions the primer terminus near the polymerase active site (Jacobo-Molina *et al.*, 1993). Smaller structural changes include the reorientation of side chains to form the non-nucleoside inhibitor binding pocket (the NNIBP does not exist in the crystal structures of RT without an NNRTI) (Ding *et al.*, 1995a). There are also differences at the catalytic site between the crystallographic structures of RT/DNA and RT/NNRTI (Ding *et al.*, 1995b).

Moreover, the conformational change effected by NNRTIs binding reduces the catalytic efficiency of the enzyme.

Based on structural and biochemical information, three models have been proposed for the mechanisms of inhibition of RT by NNRTIs (Sarafianos *et al.*, 1997). The model postulates that the bound NNRTI impairs the mobility of the p66 thumb subdomain (Kohlstaedt *et al.*, 1992). The “primer grip” and the “active site” models propose that the bound NNRTI distorts the exact configuration of the primer grip (Hsiou *et al.*, 1996; Rodgers *et al.*, 1995) or of the catalytic aspartic acids (Ren *et al.*, 1995) respectively. Biochemical studies have shown that NNRTIs block the chemical step of the polymerization reaction, the formation of the phosphodiester bond (Rittinger *et al.*, 1995). The studies model could not explain the inhibition of this chemical step. However, the bound NNRTI could inhibit RT by a combination of structural and dynamical factors.

Comparison of crystallographic structures and site-directed spin labeling experiments (Kensch *et al.*, 2000) have shown that RT has regions of extreme flexibility, and it has been proposed that this flexibility is essential for the polymerization process (Hermann and Heumann, 1996). Molecular Dynamics (MD) simulations of unliganded RT, RT complexed to NNRTIs (Treesuwan, 2004) and RT complexed to double-stranded DNA have shown that the flexibility of RT depends on its ligation state, increasing upon DNA binding (Madrid *et al.*, 1999; Madrid *et al.*, 2001). The two ligation states studied (unliganded and complexed to DNA) presented different patterns of concerted motions. The motions of the amino acids that form the nonnucleoside binding pocket upon binding of the NNRTI are anticorrelated to the p66 fingers (in the RT/DNA complex) and correlated to the RNase H subdomain (in unliganded RT). Studies using the Gaussian network model have also shown that RT has the potential to undergo large concerted motions (Bahar *et al.*, 1999; Temiz and Bahar, 2002). Alternatively, the targeted MD simulations, with a water shell around the NNIBP, have been used to study the conformational changes during the association/dissociation of ligand from a K103N mutant. A series of 0.5 ns MD simulations showed that the hydrogen bond formed between N103 and Y188 plays a

role on inhibitor entry (Rodríguez-Barrios *et al.*, 2005; Rodríguez-Barrios and Gago, 2004).

The rapidity of the selection of drug resistant HIV in patients was such that single-point mutations in the virus made first generation NNRTIs such as nevirapine unusable in mono-therapy. Amongst the mutations in RT that were originally described for nevirapine resistance were those at Tyr181 and Tyr188, both of which gave rise to high level resistance. Mutation at Tyr181 has since been frequently reported in resistance studies for many other NNRTIs and the change is almost always to cysteine. In the case of the codon 188 mutation, a greater variety of changes are reported, nevirapine and HEPT select the Tyr188Cys mutation, whereas Tyr188His or Tyr188Leu are selected with TIBO or α -APA (Ren *et al.*, 2001).

As previous introduction, researchers tried to investigate the properties of NNRTIs binding to HIV-1 RT by experimental and computational method (Nunriam *et al.*, 2005; Saen-oon *et al.*, 2005; Weinzinger *et al.*, 2005; Zhou *et al.*, 2005). The experimental results showed the non-nucleoside inhibitors lost their inhibitory efficiency from 20 to 1000 fold when the mutation occurred to the amino acid in HIV-1 RT binding pocket. However, the information in depth understanding through the molecular level is still not enough. Quantum mechanics (QM) had been used to calculate the interactions between inhibitor and surrounding amino acid but still based on the crystallographic structure and QM technique can include only few atoms. The random simulations like Monte Carlo simulation also had been applied to investigate the properties; interactions and binding energy of the reverse transcriptase complex to non-nucleoside inhibitors for handle the bigger system than QM technique (Smith *et al.*, 2000). The most popular and more reliable technique to simulate large system like enzyme complex is MD simulations. Many research groups claim the successfully use of MD simulations investigated the structural properties of HIV-1 reverse transcriptase complex with non-nucleoside inhibitors such as HIV-1 RT subdomain flexibility, key interaction distances, conformational change of inhibitors as agree well with NMR or X-ray results (Bahar *et al.*, 1999; Madrid *et al.*, 1999). However, in term of energetic properties are still not clear. Many investigations revealed the

binding energy between HIV-1 reverse transcriptase and non-nucleoside inhibitors from the Free Energy Perturbation (FEP) (Kim *et al.*, 2006; Kroeger Smith *et al.*, 2008; Udier-Blagovic *et al.*, 2003; 2004), scoring function (Barreiro *et al.*, 2007; Medina-Franco *et al.*, 2004; Rawal *et al.*, 2007; Zhou and Madura, 2004), or the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) method (Kuhn *et al.*, 2001; Wang *et al.*, 2005; Wang *et al.*, 2001). It was found that the most useful method is MM-PBSA method (Fogolari *et al.*, 2003; Kuhn *et al.*, 2005; Srinivasan *et al.*, 1998) because it is based on the compromise of fast and accuracy in calculations. Now a day, all of predicted binding energy from computational simulations is still far from the experimental value.

Therefore, the main objective of this work is to introduce the key protocol implemented in MM-PBSA method to find the more reliable binding energy. In this study, Molecular Dynamics simulations of the wild type and Y181C HIV-1 RT complexes with nevirapine have been applied in order to observe the wild type to mutant conversion effect to the binding energy and the proposed induced orientating of important amino acid. The obtained results will be useful for understanding in the molecular level of nevirapine/HIV-1 RT interaction in solution and helpful as a benchmark protocol to calculate others NNRTIs. This basic statement information can lead to the development of high potent NNRTIs against mutant enzyme.

Furthermore, while the third generations of the Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors (NNRTIs) have been developing the alternatively studies on DNA sequences of HIV virus will be solving the mutation problem. Therefore, this is a very promising and challenging research to combat AIDS at the genetic level in the future.

METHODS OF CALCULATIONS

1. Theoretical background of Molecular Dynamics (MD) simulations

Molecular Dynamics simulations method is based on Newton's second law or the equation of motion, $F = ma$, where F is the force exerted on the particle, m is its mass and a is its acceleration. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Newton's equation of motion is given by

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad (1)$$

Where \mathbf{F}_i is the force exerted on the particle i , m_i is the mass on the particle i , and \mathbf{a}_i is its acceleration. The force can also be expressed as the gradient of the potential energy.

$$\mathbf{F}_i = -\nabla_i V \quad (2)$$

Combining these two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (3)$$

where V is the potential energy of the system. Newton's equation of motion can be related the derivative of the potential energy to the changes in position as a function of time.

In case of property calculations of the system,

$$\mathbf{F} = \mathbf{ma} = \mathbf{m} \frac{dv}{dt} = \mathbf{m} \frac{d^2x}{dt^2} \quad (4)$$

and the acceleration is constant,

$$\mathbf{a} = \frac{dv}{dt} \quad (5)$$

Expression for the velocity after integration is obtained,

$$\mathbf{v} = \mathbf{at} + \mathbf{v}_0 \quad (6)$$

and since

$$\mathbf{v} = \frac{dx}{dt} \quad (7)$$

therefore,

$$\mathbf{x} = \mathbf{vt} + \mathbf{x}_0 \quad (8)$$

Combining this equation for the velocity, it can be obtained the following relation which gives the value of \mathbf{x} at time \mathbf{t} as a function of the acceleration, \mathbf{a} , the initial position, \mathbf{x}_0 and the initial velocity, \mathbf{v}_0 .

The acceleration is given as the derivative of the potential energy with respect to the position, \mathbf{r} ,

$$\mathbf{a} = -\frac{1}{m} \frac{dE}{dr} \quad (9)$$

Therefore, to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic, e.g., the position and the velocities at time zero determine the positions and velocities at all other times, t . The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein.

2. Theoretical background of the Molecular Mechanics and Poisson Boltzmann Surface Area (MM-PBSA) calculations

The MM_PBSA/GBSA approach was first introduced by Srinivasan J. (Srinivasan *et al.*, 1998) and it represents the postprocessing method to evaluate free energies of binding or to calculate absolute free energies of molecules in solution. The sets of structures are usually collected with molecular dynamics or Monte Carlo methods. The MM-PBSA/GBSA method combines the molecular mechanical energies with the continuum solvent approaches. The molecular mechanical energies are determined and represent the internal energy (bond, angle and dihedral), and van der Waals and electrostatic interactions. An infinite cutoff for all interactions is used. The electrostatic contribution to the solvation free energy is calculated with the Poisson-Boltzmann method. The hydrophobic contribution to the solvation free energy is determined with solvent-accessible-surface-area- dependent term (Sitkoff *et al.*, 1994). The binding free energies are evaluated according to the strategy described in references (Srinivasan *et al.*, 1998) and (Massova and Kollman, 1999). The contribution of the change to the conformational entropy (ΔS) during complex

formation can be estimated to complete the binding free energies calculations as shown in equation (10).

$$\Delta G = \Delta H - T\Delta S \quad (10)$$

The advantage of MM-PBSA method is the fast of calculations and reliable while others like the Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) yield rigorous and accurate free energy but it consumes a lot of resource and time. Similarly, the Linear Interaction Energy (LIE) can be used to calculate binding free energy, using a semiempirical method. LIE is based on linear-response like assumptions that the binding free energy is the combination of the weighted electrostatic and van der Waals interactions between the ligand and the receptor. Although the coefficient of the electrostatic interaction energy is about 0.5, the van der Waals contribution is quite varied for different systems. This suggests that LIE might have difficulty in predicting the binding affinities for significantly different compounds (Wang *et al.*, 2001). MM-PBSA applies no empirical parameters in its free energy calculations, which makes it a promising method for ranking very different compounds from database searching. In MM-PBSA, the free energy of A + B to AB is calculated using the following thermodynamic cycle (figure 4).

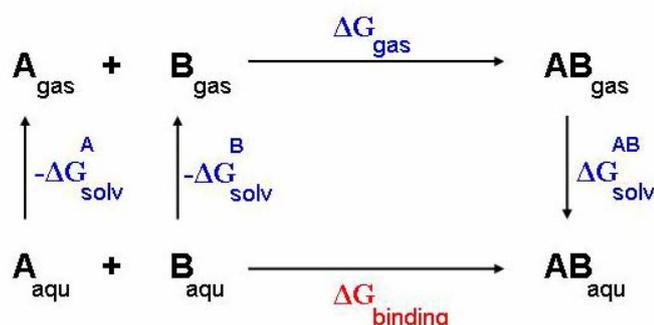


Figure 4 The schematic of thermodynamic cycle.

Where

A_{gas} , B_{gas} and AB_{gas} are A, B and complex AB in gas phase, respectively.

A_{aqu} , B_{aqu} and AB_{aqu} are A, B and complex AB in solution phase, respectively.

$-\Delta G_{\text{solv}}^A$ and $-\Delta G_{\text{solv}}^B$ are desolvation free energies for A and B.

$\Delta G_{\text{solv}}^{AB}$ is solvation free energy for complex AB.

ΔG_{gas} is the interaction energy between A and B in the gas phase.

$\Delta G_{\text{binding}}$ is the binding free energy between A and B in solution.

The solvation free energies can be estimated using a continuum approach (Poisson-Boltzmann/surface area, PBSA).

Thus

$$\Delta G_{\text{solv}}^{AB} = \Delta G_{\text{PBSA}}^{AB} = \Delta G_{\text{PB}}^{AB} + \Delta G_{\text{SA}}^{AB} \quad (11)$$

The binding free energy in solution can be derived by applying MM-PBSA approach as following:

$$\Delta G_{\text{binding}} = \Delta G_{\text{gas}} - \Delta G_{\text{solv}}^A - \Delta G_{\text{solv}}^B + \Delta G_{\text{solv}}^{AB} \quad (12)$$

$$= (\Delta H_{\text{gas}} - T \Delta S) - \Delta G_{\text{PBSA}}^A - \Delta G_{\text{PBSA}}^B + \Delta G_{\text{PBSA}}^{AB} \quad (13)$$

$$= (\Delta H_{\text{gas}} - T \Delta S) + \Delta \Delta G_{\text{PB}} + \Delta \Delta G_{\text{SA}} \quad (14)$$

Based on the assumption of

$$\Delta H_{\text{gas}} \sim \Delta E_{\text{gas}} = \Delta E_{\text{int ra}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdW}} \quad (15)$$

$$\Delta \Delta G_{\text{PB}} = \Delta G_{\text{PB}}^{AB} - (\Delta G_{\text{PB}}^A + \Delta G_{\text{PB}}^B) \quad (16)$$

$$\Delta \Delta G_{\text{SA}} = \Delta G_{\text{SA}}^{AB} - (\Delta G_{\text{SA}}^A + \Delta G_{\text{SA}}^B) \quad (17)$$

For a series of compounds with similar structures and binding modes, the entropy contribution can be omitted if one is only interested in the relative order of binding affinities.

3. Starting geometries for the models of HIV-1 RT/Nevirapine Complexes

The starting structures of both wild type and Y181C HIV-1 RT complex were taken from the previous simulation (Treesuwan, 2004). The snapshot at 1 ns was used as a starting geometry for the wild type HIV-1 RT. The snapshot at 664 ps was used as a starting geometry for the Y181C HIV-1 RT according to the conformational change in the allosteric site and nevirapine. The whole structure of both complexes was reduced its size after the residues out of 30 Å from nevirapine were removed. Amino acids within 30 Å are combined from 10 residue chains; Pro1-Val10, Lys64-Asp121, Tyr146-Val276, Glu305-Arg358, Thr369-Leu391, Pro412-Glu415, Pro579-Ile623, Trp631-Asp636, Lys686-Tyr706 and Gln954-Glu964. All solvent molecules in the range of 30 Å were also kept. Two magnesium ions were placed at the active site triad. Then new solvation was created by a water cap solvation within 20 Å from nevirapine. Finally, the counter ions, Cl⁻, were placed at the positive potential as determined by leap module for neutralized the system. The example of setting up the mutant HIV-1 RT is shown in a figure 5. The wild type HIV-1 RT model consist of 363 amino acids, 1 nevirapine, 2 magnesium ions, 11 chloride ions and 191 water molecules with the total of 6715 atoms. The Y181C HIV-1 RT model consist of 363 amino acids, 1 nevirapine, 2 magnesium ions, 11 chloride ions and 204 water molecules with the total of 6744 atoms.

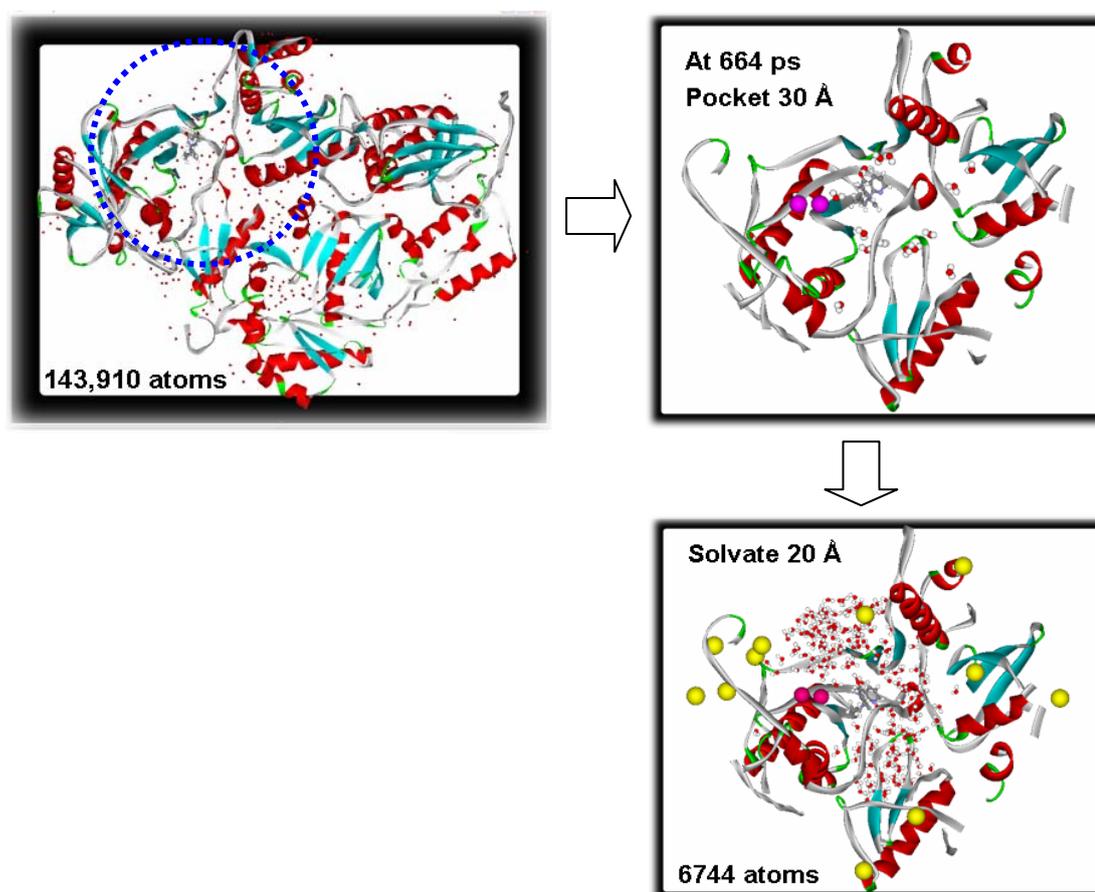


Figure 5 The diagram shows the preparing of the cap water model of the Y181C HIV-1 RT.

4. Molecular Dynamics Simulations of HIV-1 RT/Nevirapine Complexes

Molecular Dynamics simulations were performed on AMBER9 program (Case *et al.*, 2006) with the Duan *et al.* (2003) force field (Duan *et al.*, 2003; Lee and Duan, 2004) as represent a latest improvement force field for protein. Firstly, steric overlap inside the model was removed by performing through series of energy minimization steps as shown in table 1.

Table 1 Series of minimization steps use for both wild type and Y181C HIV-1 RT complex with nevirapine. Restraint weight applied on each part (kcal/mol).

	Pocket	Nevirapine	CapWAT	Ions
1. Min. Solvent & Ions	2.0	2.0	1.0	Free
2. Min. Nevirapine	2.0	Free	1.0	1.0
3. Min. Pocket	Free (Restraining Terminal)	2.0	1.0	Free
4. Min. Pocket & Nevirapine	Free (Restraining Terminal)	Free	0.5	1.0
5. Min. All	Free (Restraining Terminal)	Free	Free	Free

Solvent and chloride ions were firstly minimized by Steepest Descent (SD) 500 steps followed by Conjugated Gradient (CG) 500 steps while the enzyme model and nevirapine were kept restraint with restraint force of 2.0 kcal/mol. Next, only nevirapine was optimized while the enzyme, solvent and ions were kept restrain. Then nineteen terminal residues and nevirapine were kept restrain with restraint force of 2.0 kcal/mol while chloride ions were optimized. Then the enzyme and nevirapine were optimized. Finally, the whole system was minimized with SD 500 steps followed by CG 500 steps while only terminal residues were kept restrain with restraint force of 2.0 kcal/mol. After gradually removing bad contact inside the system, an equilibration protocol was applied step wisely to the system. Table 2 shows an example of stepwise equilibration protocol for the Y181C HIV-1 RT.

Table 2 Series of simulation steps use for Y181C HIV-1 RT complex with nevirapine. Restraint weight applied on each part (kcal/mol).

	Pocket	Nevirapine	CapWAT	Ions
1. Equilibrate, 20 ns, 1 fs	1.0 (Fix & Restrain)	1.0	0.5	Fix
2. Min. All	Free (Restrain Terminal)	Free	Free	Free
3. MDrun1, 300 K, 290ps, 2fs	Fix (Free Pocket 10 Å)	Free	Fix (Free WAT384)	Fix
4. MDrun2, 100 K, 167ps, 2fs	Fix (Free Pocket 10 Å)	Free	Fix (Free WAT384)	Fix
5. Min. All	Free (Restrain Terminal)	Free	Free	Free
6. MDrun3, 100 K, 0-1ns, 2fs	Fix (Free Pocket 10 Å)	Free	Fix (Free WAT384)	Fix
7. MDrun4, 100 - 300 K, 1-2ns, 2 fs	Fix (Free Pocket 10 Å)	Free	Fix (Free within 10 Å)	Fix
8. MDrun5, 300 K, 2-3ns, 2fs	Fix (Free Pocket 10 Å)	Free	Fix (Free within 10 Å)	Fix

The equilibration steps for the Y181C HIV-1 RT started from relaxing 10 Å of binding pocket under controlled temperature. The enzyme was kept frozen outside 10 Å radius from nevirapine while inside was kept restrain with restraint force of 1.0 kcal/mol included nevirapine. Chloride ions were kept frozen. Cap solvent was kept restrain with restraint force of 0.5 kcal/mol. Non-bonded cutoff is 20 Å. The equilibration was performed for 20 ns with time step of 1 fs. The temperature was increased from 0 to 300 K controlled by Langevin dynamics with the collision frequency of 1.0 ps. Next, the whole system was optimized with the same criteria as minimization step. Next, the simulation was performed to achieve 1 ns but the bombing occurred at 290 ps while temperature was kept constant at 300 K. The last configuration was used as the starting for next simulations. The simulation continued by lowering the temperature to 100 K and kept constant. The collision frequency was

reduced to 0.001 ps. Non-bonded cutoff was 10 Å and time step was 2 fs. However the system was still crash at 167 ps. Again, the whole system was optimized with the same criteria as minimization step to remove some bad contact from simulations. The new simulations can go smoothly after the whole system was minimized. The simulation was run for 1 ns at a constant temperature of 100 K while the enzyme pocket within 10 Å and nevirapine were freely moved. Next simulation was about to gradually increasing temperature from 100 to 300 K for 1 ns while every residues within 10 Å included nevirapine and solvent were freely moved. Finally, the system was continued performed an MD simulation for 1 ns at a constant temperature of 300 K with the same criteria as the last MD steps. The simulation was carried out under constant volume periodic boundary conditions (NVT ensemble). Cutoff distance for the Lennard-Jones interactions is 10.0 Å. The integration time step was 2 fs which SHAKE was applied to constrain bonds involving hydrogen atoms. The set of coordinate was saved every 1 ps from the last 3 ns simulations. Simulations were performed on the 2.4 GHz, 514 Mbytes, Suse10.1 PC at the Department of Chemistry, Faculty of Science, Kasetsart University.

5. Computational Analysis of HIV-1 RT/Nevirapine Complexes

The investigation is firstly focused on the energetic properties from MD simulations of the wild type and Y181C HIV-1 RT complex with nevirapine. The energetic properties can be monitored in terms of the binding free energies ($\Delta G_{\text{binding}}$) and the Decomposition energies. The snapshot structures were sampled from the last 1.5 ns of each trajectory. Thus the total of 500 snapshots sampled from both trajectories of the enzyme wild type and mutant type were used to calculate the energetic properties.

5.1. Calculations of the Binding Free Energy for HIV-1 RT/ Nevirapine Complexes

The binding free energies were calculated based on MM-PBSA approach as derived in figure 4. The entire process in calculations is carried out by MM-PBSA

script file used in AMBER. However the MM-PBSA script file needed to recall sander program. An individual solute such as the enzyme complex, free enzyme and isolated nevirapine was created 500 structures each. The calculation started by optimizing each structure then the MM energy and solvation energy were calculated later. The loop of minimization and solvation energy calculation were repeated until all of 1500 structures achieved.

The new sets of binding energies were calculated after the bridge water molecule (WAT1075) found to be important involved in binding. The MM-PBSA energies of individual solute included WAT1075 were calculated. Only hydrogen atoms on each solute were optimized then the solvation energies were acquired. Finally, each individual energies term which included energy from WAT1075 were subtract manually to obtain the binding free energies ($\Delta G_{\text{binding}}$) as shown in equation (18).

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{enzyme}} + \Delta G_{\text{Nevirapine}} + \Delta G_{\text{WAT1075}}) \quad (18)$$

The investigation was continued to obtain the most reliable technique to reproducing the experimental binding free energies. The appearance of WAT1075 can be concerned as a part of receptor to make the consistency for binding energies. Moreover, the numbers of solvent water molecules inside the binding pocket are more than one molecule. The energies of receptor would include the energies from one to five of water molecules. Thus the binding free energies can be obtained as shown in equation (19).

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{enzyme-WAT}} + \Delta G_{\text{Nevirapine}}) \quad (19)$$

Lastly, the absolute binding free energies as shown in equation (14) required the entropy term to fill in. The entropy energy can be obtained from normal mode calculations. Though one assumption in entropic energy calculations is the entropic energy would not change significantly while the structural conformation

didn't move much. Since the entropy energies calculation had known as the huge time consuming part, then only one last snapshot picked up from the simulation trajectory would be used in normal mode calculations. The calculated entropic energy also considers the hydrogen bonded-bridge water molecule as a part of solute. Details of the process to observe the entropy energies can be described as following: only residues within 10 Å from nevirapine were created as complexes of enzyme and nevirapine including WAT1075. These small systems had been optimizing by sander module under dielectric constant of 10. The minimization carried out with SD 2000 steps then followed by CG until the convergence criterion achieved. The minimization will halt when the root-mean-square (drms) of the Cartesian elements of the gradient is less than 10^{-5} kcal/mole Å. The minimization had been continued with more robust algorithm called Newton-Raphson minimization. The second derivative routines are based on expressions used in the Consistent Force Field programs (Niketic *et al.*, 1977). The Newton-Raphson minimization worked until the rms gradient converged at 0.0001 kcal/mole•Å. The memory usage could be reduced to 1/3 by setting *ismem* to 1 which means eigenvectors are not calculated. This can sometimes make calculations feasible that otherwise would not be, but only for a fairly narrow range of problem sizes. More elaborate schemes, involving sparse matrix storage, are certainly possible, nowever, this has not yet been implemented in nmode. Finally, entropy of the system was calculated by nmode module under the criteria's temperature of 298.15 K and pressure at 1.0 atm. The total entropies for translation, rotation and vibration were obtained and converted to energy (kcal/mol) or heat capacity (cal/mol K). The $\Delta S_{\text{binding}}$ could be obtained manually by subtract $\Delta S_{\text{complex}}$ with ΔS_{enzyme} , $\Delta S_{\text{Nevirapine}}$ and $\Delta S_{\text{WAT1075}}$ as same as the $\Delta G_{\text{binding}}$ in equation (18).

5.2. Calculations of Decomposition Energy

Only snapshots of the enzyme complex were used to calculate the decomposition or interaction energies. The interaction between nevirapine against both wild type and Y181C HIV-1 RT were calculated only within 10 Å radius from nevirapine. The calculation is based on MM-GBSA approach. The decomposition energies calculated as a pairwise per-residue. Then the interactions between each

residues and nevirapine were calculated. The 1-4 interactions were added to either electrostatic or vdW contributions. The newer Generalize Born model developed by A. Onufriev, D. Bashford and D.A. Case (GB^{OBC}) (Onufriev *et al.*, 2004) were implemented during the interaction energies calculations. The surface areas were computed by recursively approximating a sphere around an atom, starting from icosahedra.

RESULTS AND DISCUSSIONS

1. Equilibration of HIV-1 RT/Nevirapine Complexes

The general analysis of the MD trajectories is to observe the equilibration properties. It is the first thing to check how the enzyme complex behaved during the simulations. The important thing is only the equilibrium parameters indicated the equilibrium of the system then the trajectories could be further analyzed. The main equilibration parameters against time should be reported such as total energy (ETOT), potential energy (EPTOT), kinetic energy (EKTOT), temperature and rmsd fluctuations. The later calculations would sample from the range of stable for all parameters. In this work, the equilibration properties were reported from all periods of simulations instead of presenting only the stable period. The wild type HIV-1 RT complex simulation used the total of 7 periods to produce the stable trajectory while the Y181C HIV-1 RT complex used 5 periods as shown in figure 6 to figure 4.

Figure 6 shows energies of the wild type and Y181C HIV-1 RT complex with the simulations time. The total energy (ETOT) is a summation of kinetic energy (EKTOT) and potential energy (EPTOT). It was found that the main contribution comes from potential energy. The kinetic energies found similar between both complexes while the potential energy of the Y181C HIV-1 RT showed more negatively than the wild type HIV-1 RT. All of the energetic properties were decreasing and rising depended on temperature of the system as shown in figure 7. The total energy of the wild type HIV-1 RT was well stabilized from 2.5 ns to the end of simulations, 5.1 ns, and for the Y181C HIV-1 RT was well stabilized from 1.6 to the end of simulations, 3.3 ns.

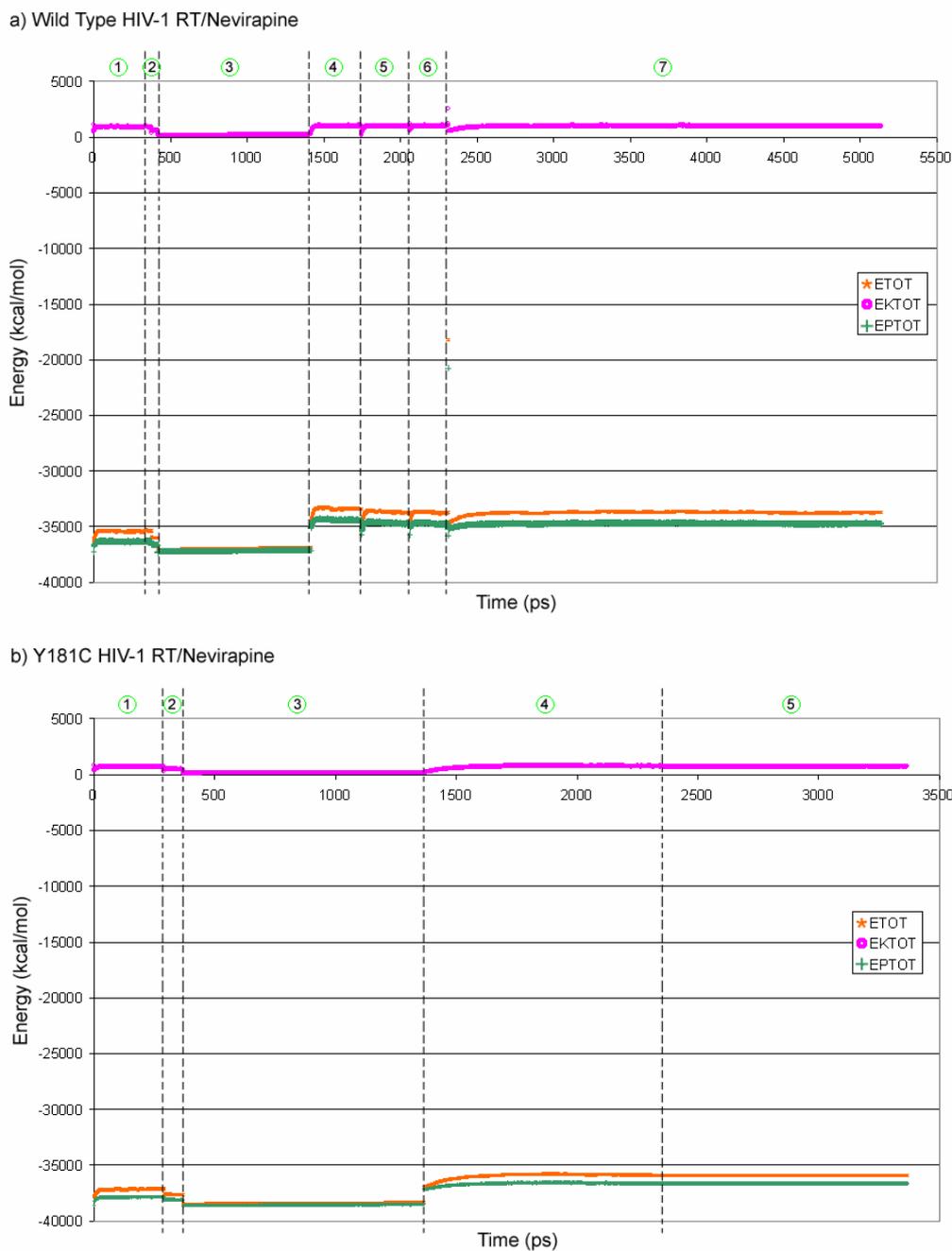


Figure 6 Total energy, potential energy and kinetic energy of the system along with simulation time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine.

Temperature of the wild type HIV-1 RT and the Y181C HIV-1 RT complexes had been observed along with the simulation time as shown in figure 7. The example case of the wild type HIV-1 RT reveals sometime the temperature was rising rapidly

at the end of period which effect to the increasing in total energy simultaneously. This indicated the unstable structure which may come from some part of the structure blow away or crash. The short minimization can solve the unstable structure. After the whole system was optimized, the next period of simulations could be continued. The target temperature in the simulation is at 300 K. Temperature in both systems was well equilibrate around 300 K controlled by Langevin dynamics with the small collision frequency of 0.001 ps. Finally, the temperature of the wild type HIV-1 RT was well stabilized from 2.6 ns to the end of simulations, 5.1 ns, and for the Y181C HIV-1 RT was well stabilized from 2.5 to the end of simulations, 3.3 ns.

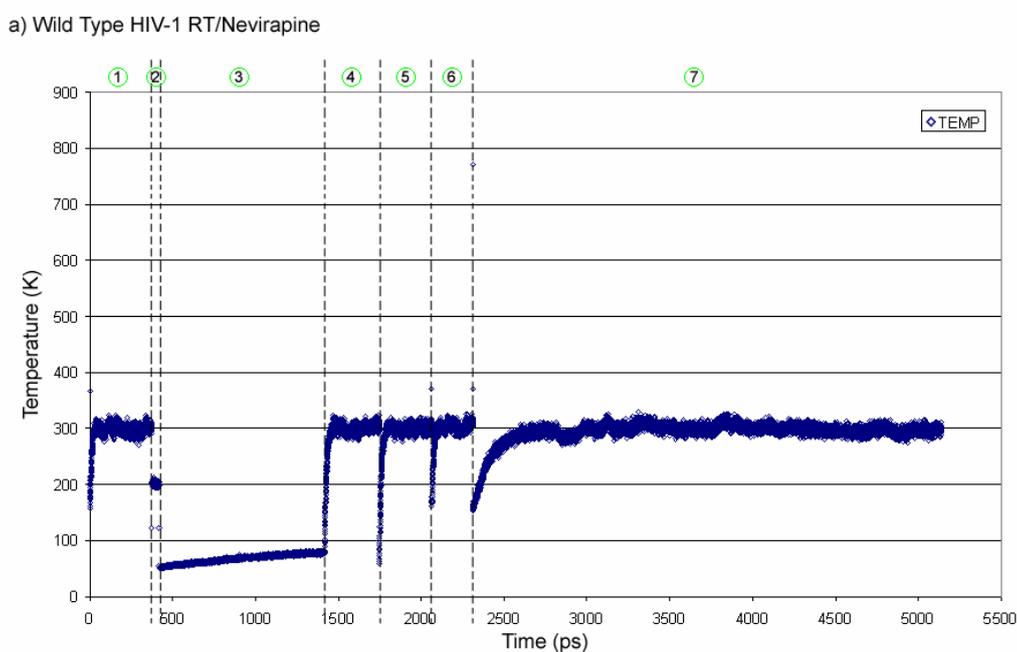


Figure 7 Temperature of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine.

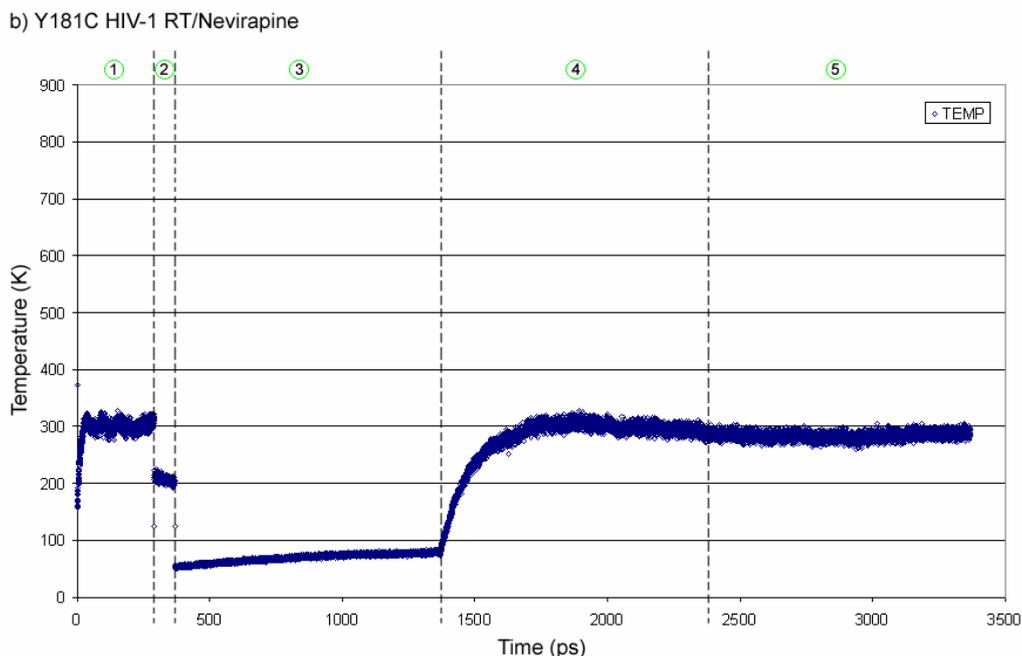


Figure 7 (Continued).

Another one important parameters is the Root-Means-Square Displacement (RMSD) which can tell how much the system is flexible. The RMSD of both wild type and Y181C HIV-1 RT were plotted against simulations time as shown in figure 8 and figure 9. The RMSD of both complexes were calculated by using the starting structure as a reference. The motion of every part in solute included translations, rotations and vibrations could be implied from the RMSD without fit function (figure 8). However the translation motions of solute was removed in the RMSD calculation by using fit function as shown in figure 9. The lower RMSD obtained after the major motion was removed by fit function. In this study, we only observed the RMSD of the interested parts which are the simulated enzyme pocket complex, the simulated enzyme pocket alone, nevirapine and WAT1075. RMSD of the pocket complex and the pocket only is nearly identical because the movement of nevirapine is very small when compared to the movement of the enzyme pocket. RMSD of the pocket complex also had very small fluctuation. It was fluctuation around 1.25 Å in the period of low temperature controlled in both complexes. Later, RMSD of the pocket complex was rising when temperature was increasing. RMSD rose to the next state and stabilized around 1.40 Å since 1.7 ns for the wild type HIV-1 RT complex.

Similarly, RMSD of the Y181C HIV-1 RT complex stabilized around 1.38 Å since 2.0 ns. The effect of temperature to the flexibility is clearer when observed the RMSD fluctuation of nevirapine and WAT1075. It showed significantly less movement due to the low temperature controlled. However, flexibility of nevirapine and WAT1075 occurred in the wide oscillation after temperature was kept constant at 300 K. But all the RMSD agrees with the criteria that the RMSD fluctuation is typically less than 1.5 – 2.0 Å (Walker, 2004).

Without fit function, nevirapine stabilized around 1.53 Å, which is higher than the RMSD of the enzyme pocket, with the fluctuation range of 1.25 to 1.99 Å, WAT1075 had fluctuation at 0.67 ± 0.26 Å at low temperature and 1.08 ± 0.81 Å in the wild type HIV-1 RT complex system. For the Y181C HIV-1 RT complex system, nevirapine stabilized around 1.08 ± 0.27 Å at low temperature and went up to 1.26 ± 0.56 Å, WAT1075 had fluctuation at 2.71 ± 0.67 Å at low temperature and down to 2.59 ± 1.25 Å.

The smaller RMSD obtained by apply fit function in the analysis. For the wild type HIV-1 RT complex, nevirapine stabilized around 1.42 ± 0.05 Å at low temperature and down to 1.38 ± 0.15 Å. The dramatic changed appear in the RMSD of WAT1075 which was nearly zero through out the simulations. For the Y181C HIV-1 RT complex, nevirapine stabilized around 0.77 ± 0.19 Å at low temperature and 0.76 ± 0.41 Å. The RMSD of WAT1075 shows similarly as in the wild type HIV-1 RT complex.

The RMSD of the enzyme pocket complex and pocket alone in both wild type and Y181C HIV-1 RT are similar. However, the RMSD of nevirapine and WAT1075 show different pattern in each complex. Nevirapine and WAT1075 in the Y181C HIV-1 RT have wider range of fluctuation than in the wild type complex. Especially, WAT1075 in the Y181C HIV-1 RT had fluctuation quite high at 2.59 ± 1.25 Å. These indicated the change from wild type HIV-1 complex as used as the starting structure to the Y181C HIV-1 RT complex. Mutation at position 181 from tyrosine to cysteine had effect directly to nevirapine and WAT1075. However, nevirapine and WAT1075

could find its new equilibrium. More details in the physical properties of nevirapine and WAT1075 motions will be further investigated.

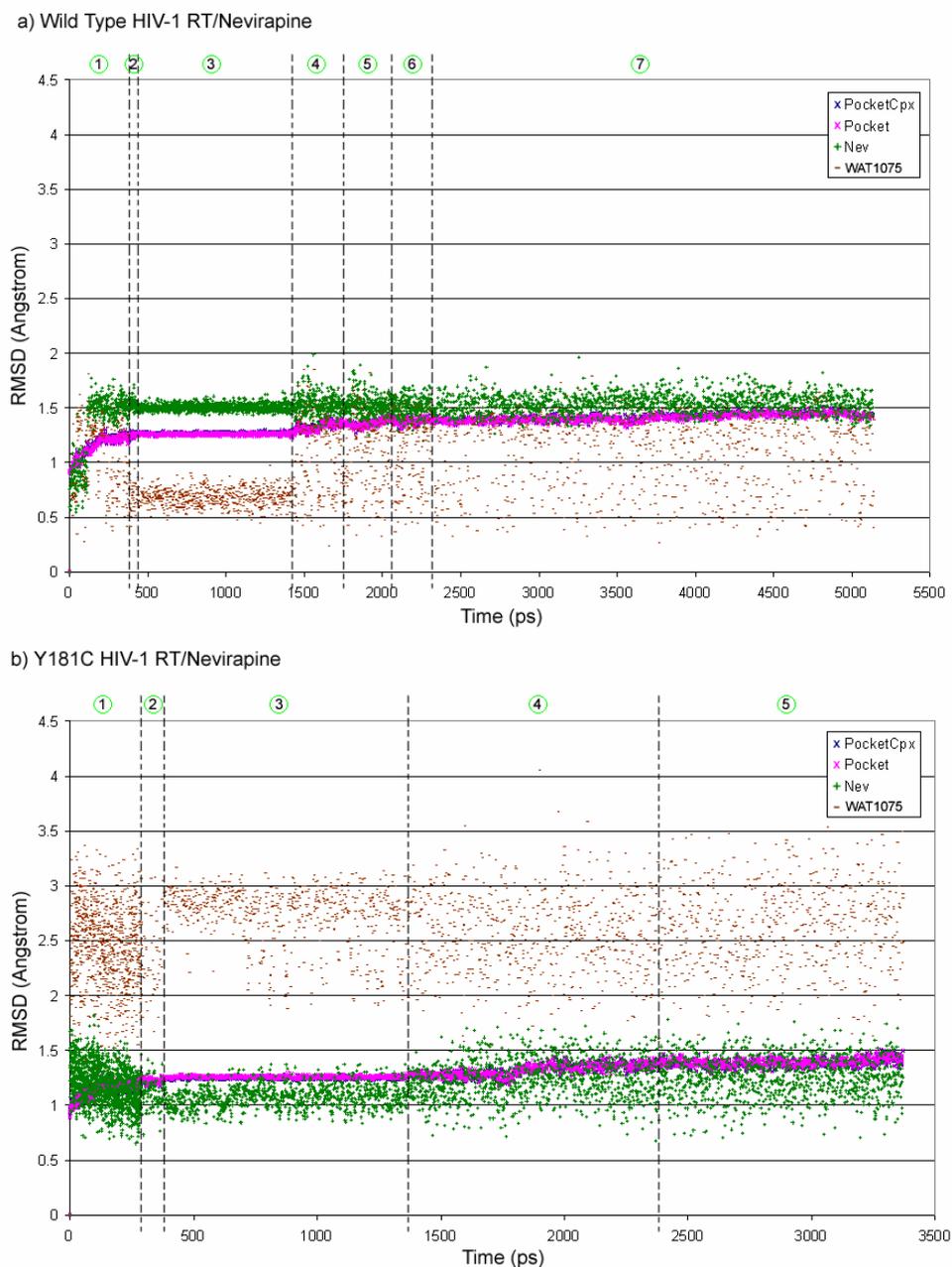


Figure 8 Root-mean-square displacement (RMSD) without *fit function* of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine. The pocket 10 Å of enzyme complex shows in blue, the pocket 10 Å of enzyme alone shows in pink, nevirapine shows in green and the WAT1075 shows in brown color.

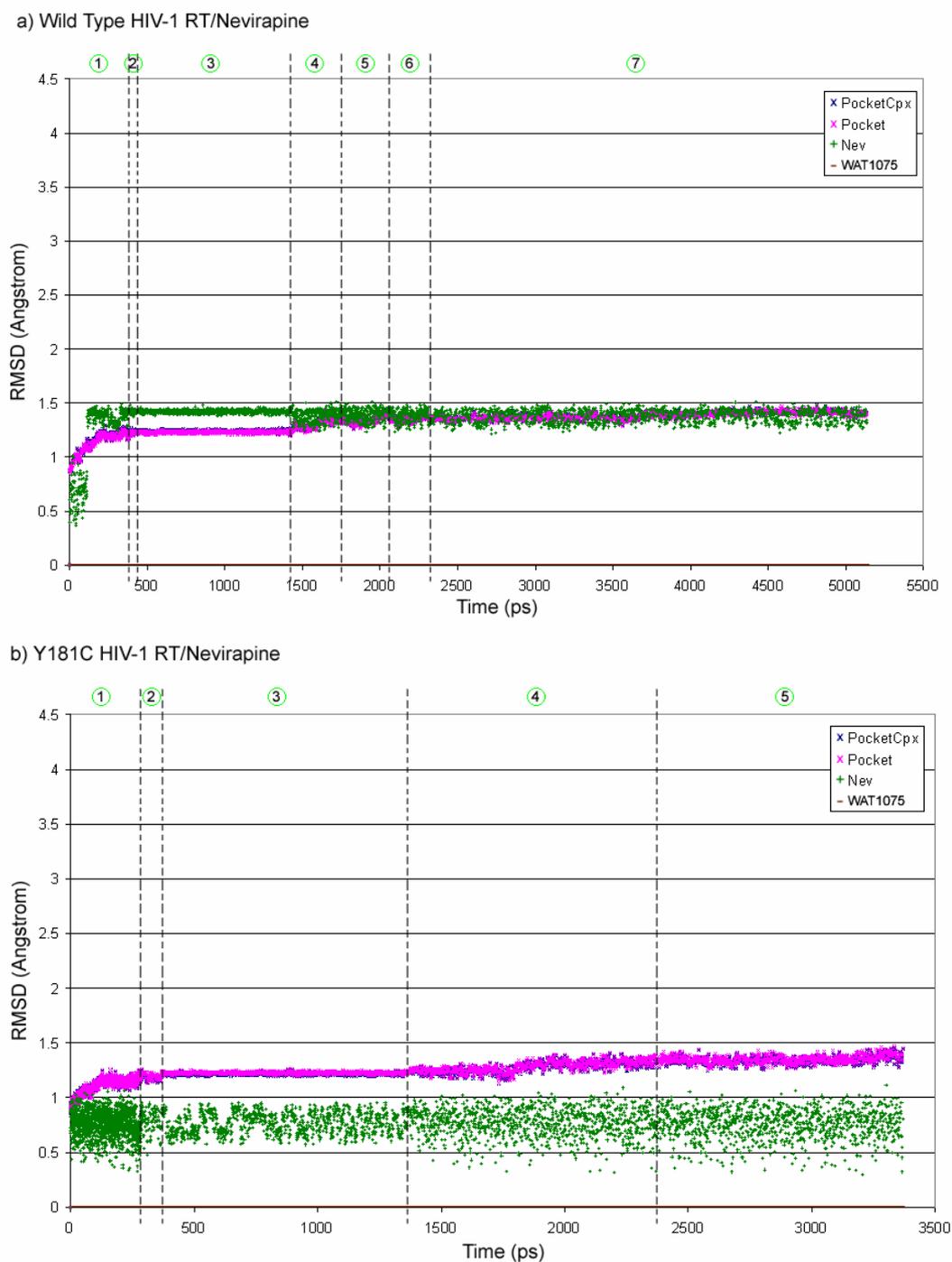


Figure 9 The root-mean-square displacement (RMSD) with *fit function* of the system along with simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine. The pocket 10 Å of enzyme complex shows in blue, the pocket 10 Å of enzyme alone shows in pink, nevirapine shows in green and the WAT1075 shows in brown color.

2. Binding Energy Calculations of HIV-1 RT/Nevirapine Complexes

The binding mechanism of the wild type and mutant type HIV-1 RT complexes with nevirapine had been studied for many years in both experimental and computational studies. One of the most important terms involved in binding mechanisms is the “binding energies”. The experimenters investigated the binding mechanism of the wild type HIV-1 RT and mutant type complexes with nevirapine through the kinetics studies. The major mutation usually occurred at the position 181 from tyrosine to cysteine when the patients were treated with nevirapine. It obviously showed that nevirapine lost its efficiency when the mutation occurred in HIV-1 RT, indicated by the higher concentration of nevirapine usage to find binding affinity for the Y181C HIV-1 RT compared to the wild type HIV-1 RT as shown in table 3.

Table 3 The experimental binding affinity between both wild type and Y181C HIV-1 RT with nevirapine (μM).

	Wild type HIV-1 RT / nevirapine	Y181C HIV-1 RT / nevirapine
1. K_d ($\text{Mg}^{2+} \cdot \text{E} \cdot \text{DNA}$) ^(Spence <i>et al.</i>, 1996)	0.025 ± 0.010	11.700 ± 4.30
2. K_d ($\text{E} \cdot \text{DNA}$) ^(Spence <i>et al.</i>, 1996)	0.019 ± 0.004	2.500 ± 1.30
3. IC_{50} ^(Hargrave <i>et al.</i>, 1991)	0.084	-
4. IC_{50} ^(Saparpakorn <i>et al.</i>, 2006)	0.060	3.200
5. IC_{50} ^(Ludovici <i>et al.</i>, 2001)	0.032	10.000
6. EC_{50} ^(Rao <i>et al.</i>, 2004)	0.044 ± 0.010	3.040 ± 1.42

The binding affinities can be observed by enzymatic kinetics studies through the parameters of K_d (the equilibrium dissociation constants), IC_{50} (50% inhibitory concentration of nevirapine against HIV-1 RT) or EC_{50} (50% effective concentration, or concentration required to protect cells against the HIV cytopathogenicity by 50%). From table 1, the binding affinities vary in a small range for the wild type HIV-1 RT complex with nevirapine and a big range for the Y181C HIV-1 RT complex with nevirapine depending on each laboratory. Nevirapine obviously has higher binding affinity with the Y181C HIV-1 RT than wild type HIV-1 RT. These observations

indicated that nvirapine lost its efficiency in binding when the mutation at 181 from tyrosine to cycteine occurred. The comparable result between experimental and computational studies is the binding energy. Thus the binding affinity can be converted to the binding energy between nevirapine and HIV-1 RT by using equation (20).

$$\Delta G = RT \ln [\text{Activity}] \quad (20)$$

Where ΔG represents the binding free energy (kcal/mol), R represents the gas constant (1.988 cal/mol•K), T represents the temperature (K) and Activity can be K_d , IC_{50} or EC_{50} .

The experimental binding free energy ($\Delta G^{\text{expt.}}$) is shown in table 3. However finding the calculated binding free energy is not straightforward using MM-PBSA methodology. After the equilibration phase of the cap simulations amino acid surrounding nevirapine inhibitor complex within 20 Å had been done, the production phase simulation was performed for 3 ns. For the test group, the samplings of 10 snapshot structures were used as the input structure to calculate binding energy by MM-PBSA method ($\Delta G^{\text{MM-PBSA}}$). The classical way of MM-PBSA calculation normally removes solvent water molecules and ions in the input structures and keeps only the enzyme and inhibitor, as the MM-PBSA method will generate the continuum solvent model during the calculation. Unfortunately, the bridge water can form strong hydrogen bonds as reported in the quantum study of HIV-1 protease-bridge water interaction (Duan *et al.*, 2007). Thus, we have a hypothesis of keeping the bridge water should significantly stabilized the complex in the MM-PBSA calculations.

The x-ray crystallographic structure, 1VRT, and the simulation trajectories reveal some water molecules inside the binding pocket. Numbers of water molecule with in 7 Å found 5 molecules in the wild type HIV-1 RT complex as WAT384 (WAT1075 in crystal structure), WAT407, WAT418, WAT477 and WAT497. Therefore, in the Y181C HIV-1 RT complex found 4 molecules, WAT381, WAT384, WAT390 and WAT393 within the same radius of 7 Å. The identity of water

molecules inside the Y181C HIV-1 RT complex corresponds to WAT1048, WAT1075, WAT1182 and WAT1218 in the crystal structure. The distance between each water molecule to nevirapine was observed with the simulations time as shown in figure 10. WAT384 (or WAT1075) occupied the closest distance from nevirapine and well stabilized around 6 Å in both case while others had higher flexibility. These implied WAT1075 should have some specific interactions and it maybe important.

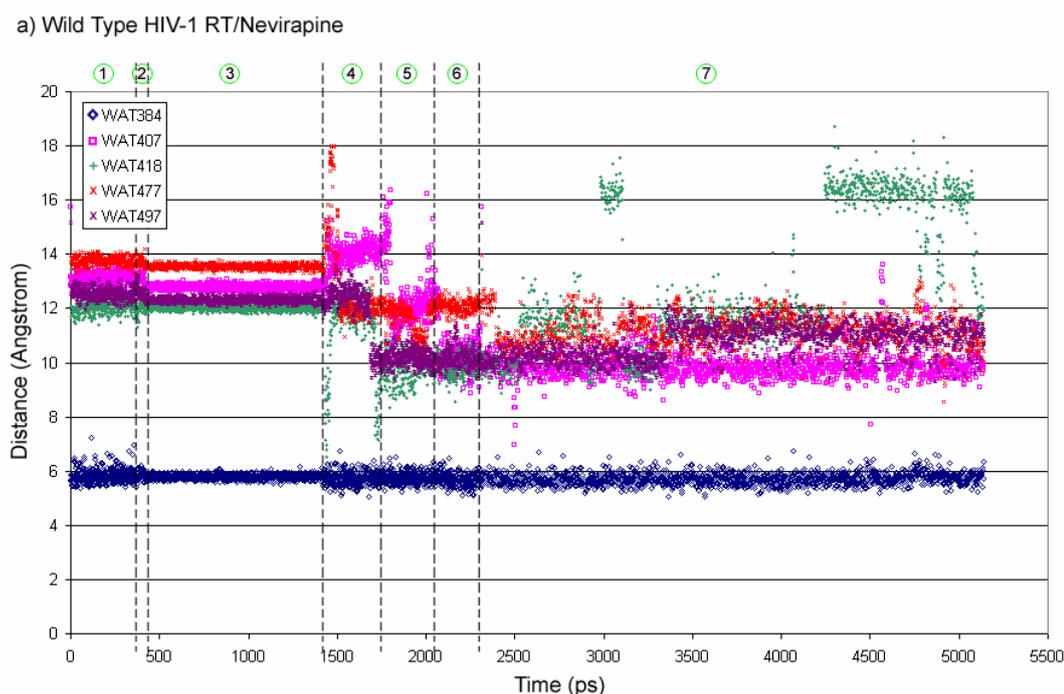


Figure 10 Distance between water molecules and the center of mass of nevirapine along with the simulations time. a) wild type HIV-1 RT/nevirapine. b) Y181C HIV-1 RT/nevirapine.

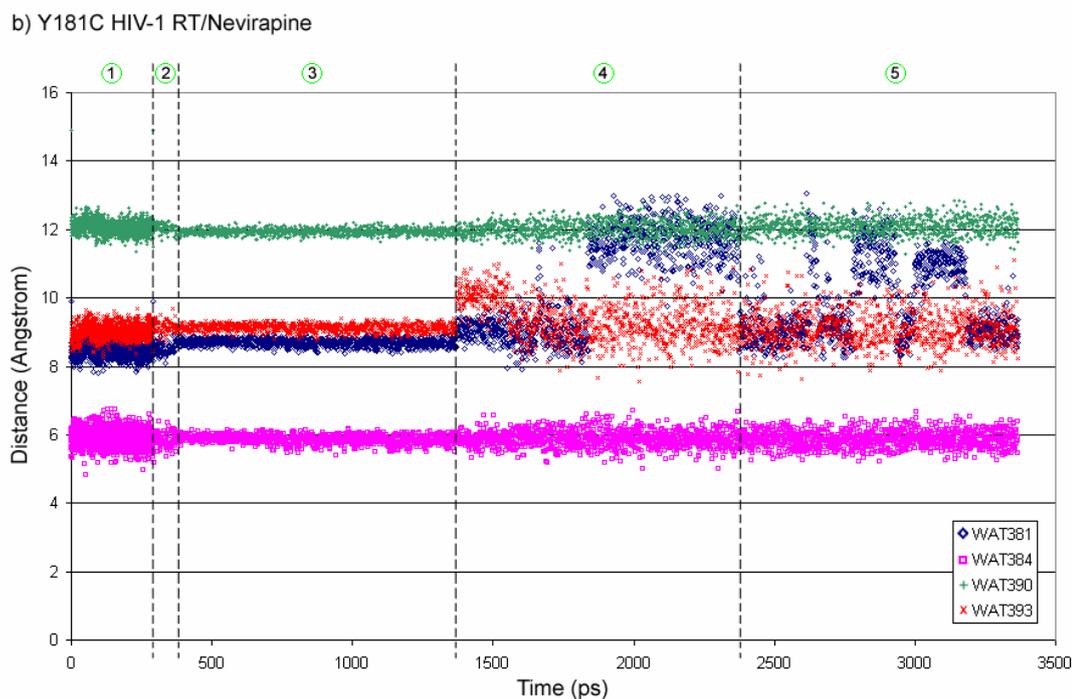
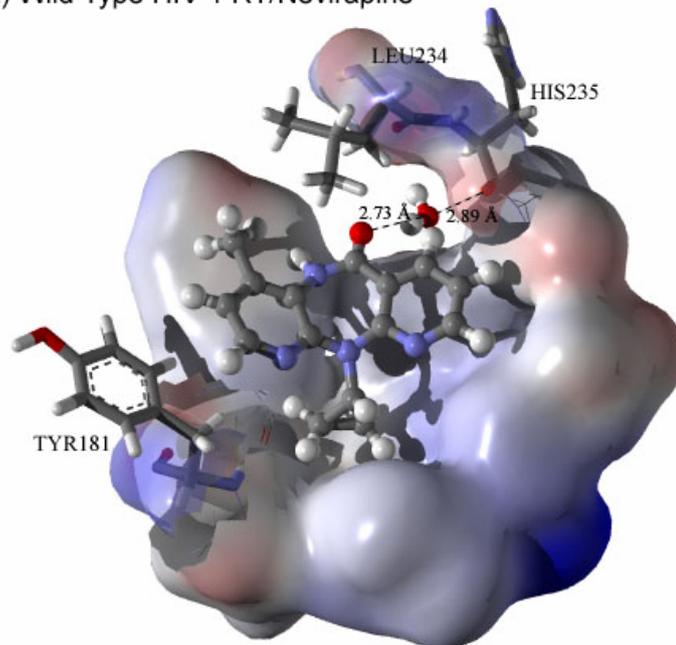


Figure 10 (Continued).

More investigations would be focused on the WAT1075. Some snapshot structures were taken from the trajectories of both wild type and mutant type complex. The physical properties such as distance were measured from WAT1075 and other residues around. From the measurement found this water molecule, WAT1075, is forming moderate/weak hydrogen-bonded bridge between nevirapine and the binding pocket of both wild type and Y181C HIV-1 RT. The hydrogen bonded-bridge water between nevirapine and the binding pocket in both of wild type and Y181C HIV-1 RT during the whole system simulation are shown in figure 11. WAT1075 acts as a hydrogen donor to O1(nevirapine), O(His235) and O(Leu234) with the distance of 2.81, 3.66 and 3.46 Å, respectively. In addition, WAT1075 is also a hydrogen acceptor from N-H(His235) with the distance of 3.61 Å. This essential hydrogen bonded-bridge was also confirmed from the Monte Carlo Simulations of HIV-1 RT complex with nevirapine or MKC-442 (Rizzo *et al.*, 2001). Thus this important one water molecule was proposed to be included in the binding mechanism of nevirapine to the HIV-1 RT.

a) Wild Type HIV-1 RT/Nevirapine



b) Y181C HIV-1 RT/Nevirapine

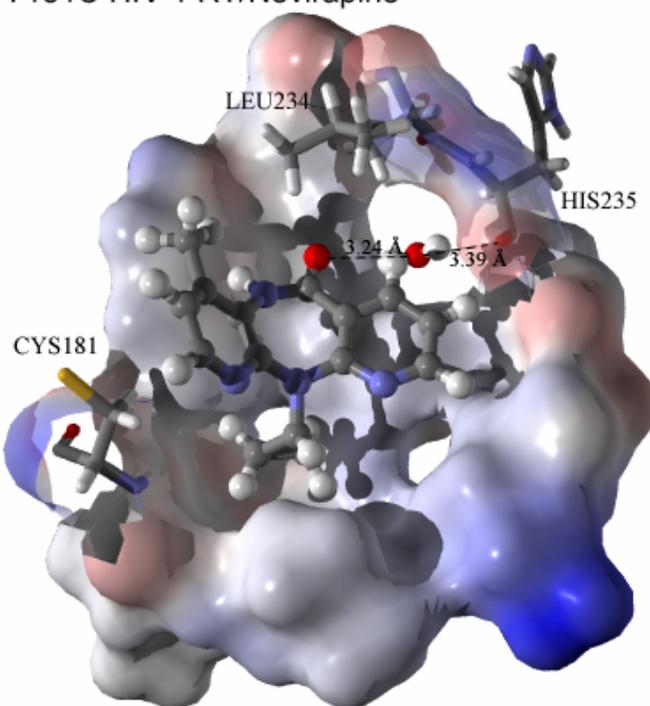


Figure 11 Distance of hydrogen-bonded bridge at 3 ns referred by heteroatomic distance. In a) wild type HIV-1 RT complex and b) Y181C HIV-1 RT complex.

After the cap simulations of 20 Å around nevirapine had done for about 3 ns, the snapshot structures were sampling from the production trajectories for 500 structures during 1.5 – 3 ns in both complex with enzyme wild type and mutant type. The MM-PBSA energy ($\Delta G^{\text{MM-PBSA}}$) of binding was calculated through the MM-PBSA calculation by sander program. The example results for the wild type and Y181C HIV-1 RT/nevirapine were shown in table 4, comparing between the calculations from included and excluded of hydrogen bonding bridge water molecule (WAT1075). The calculations of included bridge water gave lower binding energies of 6.94 and 54.70 kcal/mol for the wild type and mutant type, respectively, than the excluded bridge water model. The significant difference indicated the bound mechanism between nevirapine and the wild type RT pocket is needed a stabilized force from WAT1075.

Table 4 MM-PBSA energies (kcal/mol) of the HIV-1 RT complex with nevirapine calculated from the complex, enzyme, nevirapine and solvent water (WAT).

Models	Complex	Enzyme	Nevirapine	WAT	$\Delta G^{\text{MM-PBSA}}$
1. Enz ^{Y181C} .+Nev.	-11223.63	-11231.19	-136.90	-	144.46
2. Enz ^{Y181C} .+Nev.(min. ^a)	-11360.61	-11244.10	-138.69	-	22.18
3. Enz ^{Y181C} .+Nev.+WAT(min. ^a)	-11422.36	-11244.10	-138.69	-7.05	-32.52
4. Enz ^{RT} .+Nev.(min. ^a)	-11453.40	-11283.98	-137.22	-	-32.20
5. Enz ^{RT} .+Nev.+WAT(min. ^a)	-11467.39	-11283.98	-137.22	-7.05	-39.14

Enz^{Y181C}. = The Y181C HIV-1 RT

Enz^{RT}. = The wild type HIV-1 RT

Nev. = Nevirapine

WAT = WAT384 (WAT1075 in crystal structure)

^a = Only hydrogen atoms were minimized in solutes

While the appearance of WAT1075 proved the significantly stabilizing force in the bound mechanism between nevirapine and HIV-1 RT, especially in the mutant type complex, other water molecules were included to find its contribution as well. The rational calculation would include number of water step by step. The starting calculation was excluded water molecule. Next, only WAT1075 would be added in the model according to its closest position to nevirapine. Next, water molecule would

be adding respect to its distance close to nevirapine. The distance of water molecule within 7 Å in both of the wild type and mutant type complex was calculated as a means value from the plot in figure 10 and list in table 5 and table 6. The distance of each water molecule was ranked from the closest to the farthest to nevirapine. Some of water molecule came from the beginning as a crystal structure but some are the cap solvent added in the set up model. Water molecules in the wild type HIV-1 RT complex were ranked as WAT384 (or WAT1075 in crystal structure), WAT407, WAT497, WAT477 and WAT418 respected to the distances of 5.7, 9.9, 11.0, 11.2 and 14.0 Å, respectively, while it had been calculated during 3.5 to 5.1 ns in the trajectory file. Water molecules in the Y181C HIV-1 RT complex were ranked as WAT384, WAT393, WAT381 and WAT390 respected to the distances of 5.9, 9.1, 10.5 and 12.1 Å, respectively, while it had been calculated during 1.8 to 3.3 ns in the trajectory file.

Table 5 The closest water molecule within 7 Å from nevirapine inside the binding pocket of the wild type HIV-1 RT.

WAT ID.	Average distance (Å)
1. WAT384	5.7
2. WAT407	9.9
3. WAT497	11.0
4. WAT477	11.2
5. WAT418	14.0

WAT384 = WAT1075 (in crystal structure)

Average from 3.5 – 5.1 ns

Table 6 Closest water molecule within 7 Å from nevirapine inside the binding pocket of the Y181C HIV-1 RT.

WAT ID.	Average distance (Å)
1. WAT384	5.9
2. WAT393	9.1
3. WAT381	10.5
4. WAT390	12.1
WAT381 = WAT1048 (in crystal structure)	
WAT384 = WAT1075 (in crystal structure)	
WAT390 = WAT1182 (in crystal structure)	
WAT393 = WAT1218 (in crystal structure)	
Average from 1.8 – 3.3 ns	

The stepwise to add up each water molecule is clearly obtained from the average distance. However, the practical way to calculate the binding energies by including more than one water molecule is still have a choice. The calculated binding energies can be observed either calculated water molecule alone (as shown in equation 18) or kept it with receptor (as shown in equation 19). The comparisons of binding energies are shown in table 7 and table 8 as to evaluate the methodology. In both case of the wild type and mutant type complexes, when the binding energies calculated by concern of water molecule alone, the binding energies are still dropping after each addition of water molecule. Even addition of the first water molecule shows dramatic decrease of binding energy as 6.94 and 54.70 kcal/mol in wild type and Y181C HIV-1 RT, respectively, however, the binding energies are still dropping after the second water was added as 3.65 and 10.18 kcal/mol in wild type and Y181C HIV-1 RT, respectively. The third addition still affects to the decreasing of binding energies as 3.15 and 5.76 kcal/mol in wild type and Y181C HIV-1 RT, respectively. More reliable results were found when including water molecule into receptor type. The system was stabilized by the significantly drop of energy from the first addition, which was the important bridge water molecule. The second addition had a very small lowering of the binding energy. Lastly, after the third addition of water molecule the

energy was not changed anymore. Thus the methodology of including water molecule into receptor part would be further reference for the binding energy calculations.

Table 7 MM-PBSA energies (kcal/mol) of the wild type HIV-1 RT complex with nevirapine calculated by include different numbers of solvent molecules.

Models	Complex	Enzyme	Nevirapine	WAT	$\Delta G^{\text{MM-PBSA}}$
1. RT ^{WT} .+Nev.(min. ^a)	-11453.40	-11283.98	-137.22	-	-32.20
2. RT ^{WT} .+Nev.+1WAT(min. ^a)	-11467.39	-11283.98	-137.22	-7.05	-39.14
3. RT ^{WT} .+Nev.+2WAT(min. ^a)	-11479.19	-11283.98	-137.22	-15.20	-42.79
4. RT ^{WT} .+Nev.+3WAT(min. ^a)	-11490.39	-11283.98	-137.22	-23.25	-45.94
5. RT ^{WT} -1WAT ^{WT} .+Nev. (min. ^a)	-11467.39	-11292.52	-137.22	-	-37.65
6. RT ^{WT} -2WAT ^{WT} .+Nev. (min. ^a)	-11479.19	-11304.26	-137.22	-	-37.71
7. RT ^{WT} -3WAT ^{WT} .+Nev. (min. ^a)	-11490.39	-11315.46	-137.22	-	-37.71

RT^{WT}. = The wild type HIV-1 RT

Nev. = Nevirapine

1WAT = WAT384 (WAT1075 in crystal structure)

2WAT = WAT384 + WAT407

3WAT = WAT384 + WAT407 + WAT477

RT^{WT}-1WAT^{WT}. = wild type HIV-1 RT + WAT384

RT^{WT}-2WAT^{WT}. = wild type HIV-1 RT + WAT384 + WAT407

RT^{WT}-3WAT^{WT}. = wild type HIV-1 RT + WAT384 + WAT407 + WAT477

Table 8 MM-PBSA energies (kcal/mol) of the Y181C HIV-1 RT complex with nevirapine calculated by including different numbers of solvent molecules.

Models	Complex	Enzyme	Nevirapine	WAT	$\Delta G^{\text{MM-PBSA}}$
1. RT ^{Y181C} .+Nev.(min. ^a)	-11360.61	-11244.10	-138.69	-	22.18
2. RT ^{Y181C} .+Nev.+1WAT(min. ^a)	-11422.36	-11244.10	-138.69	-7.05	-32.52
3. RT ^{Y181C} .+Nev.+2WAT(min. ^a)	-11440.62	-11244.10	-138.69	-15.13	-42.70
4. RT ^{Y181C} .+Nev.+3WAT(min. ^a)	-11454.34	-11244.10	-138.69	-23.09	-48.46
5. RT ^{Y181C} -1WAT ^{WT} .+Nev. (min. ^a)	-11422.36	-11253.85	-138.69	-	-29.82
6. RT ^{Y181C} -2WAT ^{WT} .+Nev. (min. ^a)	-11440.62	-11271.96	-138.69	-	-29.97
7. RT ^{Y181C} -3WAT ^{WT} .+Nev. (min. ^a)	-11454.34	-11285.63	-138.69	-	-30.02

RT^{Y181C}. = The Y181C HIV-1 RT

Nev. = Nevirapine

$$1\text{WAT} = \text{WAT384 (WAT1075 in crystal structure)}$$

$$2\text{WAT} = \text{WAT384} + \text{WAT393 (WAT1218 in crystal structure)}$$

$$3\text{WAT} = \text{WAT384} + \text{WAT393} + \text{WAT381 (WAT1048 in crystal structure)}$$

$$\text{RT}^{\text{Y181C}-1\text{WAT}} = \text{Y181C HIV-1 RT} + \text{WAT384}$$

$$\text{RT}^{\text{Y181C}-2\text{WAT}} = \text{Y181C HIV-1 RT} + \text{WAT384} + \text{WAT393}$$

$$\text{RT}^{\text{Y181C}-3\text{WAT}} = \text{Y181C HIV-1 RT} + \text{WAT384} + \text{WAT393} + \text{WAT381}$$

The experimental activities can be converted to the binding free energies by using equation (20) as shown in table 9. The calculated MM-PBSA energies of the included bridge water are also reported in table 9. Unfortunately, the previous results from table 8 reveals how significant of one water molecule which forms hydrogen bonding bridge between the pocket residues and nevirapine. Then the bridge water can not be neglected in the MM-PBSA calculations due to the contribution of the stabilization energy. The later results from MM-PBSA calculations would consider of one water molecule, WAT1075, as one of the binding component in the model.

Table 9 Binding energy of wild type and Y181C HIV-1 RT with nevirapine from experimental and calculations (kcal/mol).

	Wild type HIV-1 RT / nevirapine	Y181C HIV-1 RT / nevirapine
1. K_d Expt. ^(Spence <i>et al.</i>, 1996)	-2.18	1.46
2. K_d Expt. ^(Spence <i>et al.</i>, 1996)	-2.35	0.54
3. IC_{50} Expt. ^(Hargrave <i>et al.</i>, 1991)	-9.65	-
4. IC_{50} Expt. ^(Saparpakorn <i>et al.</i>, 2006)	-1.67	0.70
5. IC_{50} Expt. ^(Ludovici <i>et al.</i>, 2001)	-2.04	1.36
6. EC_{50} Expt. ^(Rao <i>et al.</i>, 2004)	-1.85	0.66
7. Predict I ^(Zhou <i>et al.</i>, 2005)	-54.78	-48.48
8. MM-PBSA (this work)	-37.65	-29.82
9. Binding energy (this work)	-41.77	-33.52

The binding energies of both wild type and Y181C HIV-1 RT complex with nevirapine calculated from MM-PBSA calculations are shown in table 7 comparing with the binding free energy from experimental which were converted from equation (20) and the binding energy from another prediction model. The experimental binding free energies of the wild type HIV-1 RT complex show the small amount of attractive energy between wild type enzyme and ligand from the varying of -1.67 to -2.35 kcal/mol. However, the experimental binding free energies of the Y181C HIV-1 RT complex show the opposite binding mechanism by the small amount of repulsive energy between mutant type enzyme and ligand from the varying of 0.54 to 1.46 kcal/mol. The differences of binding free energy from the experiment are in the range of 0.68 and 0.92 kcal/mol for the wild type and mutant type enzyme complex, respectively. In addition, since the differences are not larger than the computed error bar from the theoretical calculations, either K_d , IC_{50} or EC_{50} can be used as the experimental references. Due to difficulty for theoretical scientist to calculate and predict the binding energy of the HIV-1 RT enzyme complex with inhibitors because of the heavy size of the enzyme, however, there was one publication predicted the binding energy of both wild type and Y181C HIV-1 RT complex with nevirapine by using the same Molecular Dynamics simulations method and used the MM-PBSA calculations to calculate the binding free energy (Zhou *et al.*, 2005). They found the binding free energy of the wild type HIV-1 RT/nevirapine was -54.78 kcal/mol and the Y181C HIV-1 RT/nevirapine was -48.48 kcal/mol by using the classical MM-PBSA calculations which removed all ions and explicit solvent molecules. The calculated binding energy from MM-PBSA calculations included complex of three components as the complex, enzyme with water and ligand as following equation.

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{receptor,water}} + \Delta G_{\text{ligand}} + \Delta G_{\text{water}}) \quad (21)$$

Therefore in this work, only the calculated MM-PBSA energy ($\Delta G^{\text{MM-PBSA}}$) of the wild type and Y181C HIV-1 RT complex with nevirapine are -37.65 and -29.82, respectively, which are one step closer to the experimental value and can improve the binding energy from the Predict I. It is clear that including the hydrogen bonded-bridge water molecule is an essential for energetic calculations. The combination of

MM-PBSA energy and entropy energy would yield the absolute binding free energy as shown in equation (22)

$$\Delta G_{\text{binding}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (22)$$

The entropic energy calculated from normal mode calculations by using only one last snapshot picked up from the simulation trajectory because of the assumption that the entropic energy would not change significantly while the structural conformation has been conserved much and also the expensive of computer resource usage. The calculated entropic energy also considers the hydrogen bonded-bridge water molecule as a part of solute which showed the entropic energies individually in table 10. The difference of the entropic energy between the complex of wild type and Y181C HIV-1 RT with nevirapine is 0.42 kcal/mol which is not significantly different, however these is also one parameter confirm the result agree with the experiment which nevirapine prefers to bind with the wild type HIV-1 RT. After the MM-PBSA energy was combined with the entropic energy, the result of binding free energy of both the wild type and Y181C HIV-1 RT complex with nevirapine are -41.77 and -33.52 kcal/mol, respectively, as shown in table 9. The entropic energy makes the absolute binding free energy value far away from the experimental results, therefore the methodology to calculate the entropic energy would be improved later.

Table 10 Binding entropy of wild type and Y181C HIV-1 RT with nevirapine from calculations (kcal/mol).

	Wild type HIV-1 RT / nevirapine	Y181C HIV-1 RT / nevirapine
Complex	7215.67	6326.98
Receptor	7012.24	6123.95
Nevirapine	180.84	180.84
Water	18.48	18.48
TΔS (this work)	4.12	3.70

3. Folded Resistance Calculations of HIV-1 RT/Nevirapine Complexes

Another way to compare the calculated result with the experiment mostly measure in the relative between enzyme wild type and mutant type because this technique is very useful which can cancel all the errors neither from human or the calculated protocol during the calculations. The relative activity of nevirapine against the wild type and Y181C HIV-1 RT from the experiment could be measured in the term of resistance folded as define in equation (23) then it can convert to the relative free energy of binding as shown in equation (24)

$$\text{Resistance Folded} = \frac{[\textit{Activity}]_{\text{against mutant type enzyme}}}{[\textit{Activity}]_{\text{against wild type enzyme}}} \quad (23)$$

And

$$\begin{aligned} \Delta\Delta G_{\text{folded resistance}} &= \Delta G_{\text{mutant}} - \Delta G_{\text{wild type}} \\ &= RT \ln (\text{Folded Resistance}) \end{aligned} \quad (24)$$

The resistance folded of nevirapine against the second most like Y181C HIV-1 RT could vary from about 50 to 500 folded which is rather wide range. After the resistance folded are converted to the relative binding free energy as shown in table 11 it would have a varying from 2.37 to 3.64 kcal/mol which are not significantly different even it come from different laboratory. The obtained results indicate that nevirapine has better performance against wild type enzyme. While the previous computational results showed the relative binding free energy of 6.30 and 3.88 kcal/mol for Predict I and Predict II, respectively, the developed method in this work can predict the relative binding free energy as 7.83 kcal/mol if consider the entropy change between each conformation is very low and can be omitted. After combination of the entropic energy with MM-PBSA energy, the relative binding free energy is increased to 8.25 kcal/mol which is not significantly different from consider the MM-PBSA energy alone. Therefore, the predicted relative binding free energy is very close to the Predict I while this work can improve a lot in the absolute binding free energy

from the previous results. Even the Predict II result is very close to the experimental but acquiring their absolute binding free energy is not clear.

Table 11 Relative folded resistance energies ($\Delta\Delta G$) in kcal/mol for Y181C HIV-1 RT mutations normalized to nevirapine.

	$\Delta\Delta G$ (kcal/mol)
1. Expt. (Spence et al., 1996) Resistance folded = 468	3.64
2. Expt. (Spence <i>et al.</i> , 1996) Resistance folded = 131.6	2.89
4. Expt. (Saparpakorn <i>et al.</i> , 2006) Resistance folded = 53.3	2.37
5. Expt. (Ludovici <i>et al.</i> , 2001) Resistance folded = 312.5	3.40
6. Expt. (Rao et al., 2004) Resistance folded = 69.1	2.51
7. Predict I (Zhou <i>et al.</i> , 2005)	6.30
8. Predict II (Rizzo <i>et al.</i> , 2000)	3.88
9. MM-PBSA (this work)	7.83
10. Binding energy (this work)	8.25

4. Decomposition Energy Calculations of HIV-1 RT/Nevirapine Complexes

Contribution of binding energies between nevirapine and HIV-1 RT is clearly observed from interaction energies. The individual interaction energies between each RT residues and nevirapine were calculated by MM-PBSA decomposition energy analysis.

Decomposition energies of the wild type and Y181C HIV-1 RT are shown in table 12 and compare with others approach. The energy plot of all approaches is shown in figure 12. Unfortunately, the main contribution comes from Tyr188 and Leu100, but not Lys103, for both wild type and mutant type. Most of the

decomposition energies for wild type complex agree well with ONIOM BBF-corr calculations. Only few interactions found an agreement between decomposition and MFCC (molecular fractionation with conjugate caps) approach (He *et al.*, 2005). These dramatic results reveal that all of the decomposition energies are attractive force. Some repulsive energy found in ONIOM BBF-corr and MFCC approach. The repulsive forces are found at Leu100, Val179, Tyr181Cys, Gly190 and Leu234. All the difference interactions between high level of calculation and molecular mechanic approach come from the starting geometry. While the ONIOM BBF-corr and MFCC approach used the starting structure based on the crystallographic structure from protein databank (PDB), the starting structure of the decomposition energy calculation comes from sampling of all possibility geometry in MD trajectory. All the difference interactions at each residue indicated the dynamics of itself during MD simulations. The patterns of decomposition energies are similar between the wild type and mutant type RT complexes. However, the difference amount of energies contribution at Tyr188, Val106 and Tyr181Cys cause the difference in binding free energies of wild type and Y181C HIV-1 RT complexes.

Table 12 Interaction energies (IE, kcal/mol) of nevirapine with individual residues.

Residues	Decomposition ^[a]		ONIOM ^[b]	MFCC approach ^[c]	
	Wild type	Y181C	Wild type	Wild type	Y181C
Pro95	-1.00	-1.29	-0.87	-	-
Leu100	-4.30	-4.24	-3.09	-0.58	2.81
Lys101	-0.71	-0.50	-1.10	-2.28	-3.08
Lys102	-0.26	-0.15	-0.37	-1.34	-1.25
Lys103	-1.92	-2.27	-1.20	-2.52	-2.38
Lys104	-0.09	-0.10	0.02	-	-
Ser105	-0.11	-0.08	-0.16	-	-
Val106	-2.99	-2.29	-0.08	-	-
Val179	-1.64	-1.89	1.47	-	-
Ile180	-0.72	-0.94	-0.50	-	-
Tyr181	-3.12	-2.44	-2.79	1.34	7.63
Tyr188	-5.74	-4.94	-5.05	-2.07	-1.35
Val189	-1.11	-0.59	-0.64	-	-
Gly190	-0.84	-0.91	1.65	-	-
Phe227	-0.67	-0.72	-1.74	-2.23	-1.46
Leu228	-0.01	-0.02	-0.20	-	-
Trp229	-2.02	-1.52	-1.19	-	-
Leu234	-3.03	-2.43	1.22	-	-
His235	-1.19	-1.15	-2.39	-1.89	-0.64
Pro236	-1.30	-1.17	-1.63	-5.28	-2.41
Tyr318	-2.09	-2.17	-1.80	-	-
Glu138 ^[d]	-0.21	-0.12	-0.09	-	-

[a] Decomposition energies on a pairwise per-residue basis.

[b] Interaction energies calculated at the MP2/6-31G(d,p) level with BSSE corrected from ONIOM3 (MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) optimization (Kuno *et al.*, 2006).

[c] Molecular Fraction with Conjugate Caps approach (He *et al.*, 2005).

[d] Glu138 taken from p51 domain of RT.

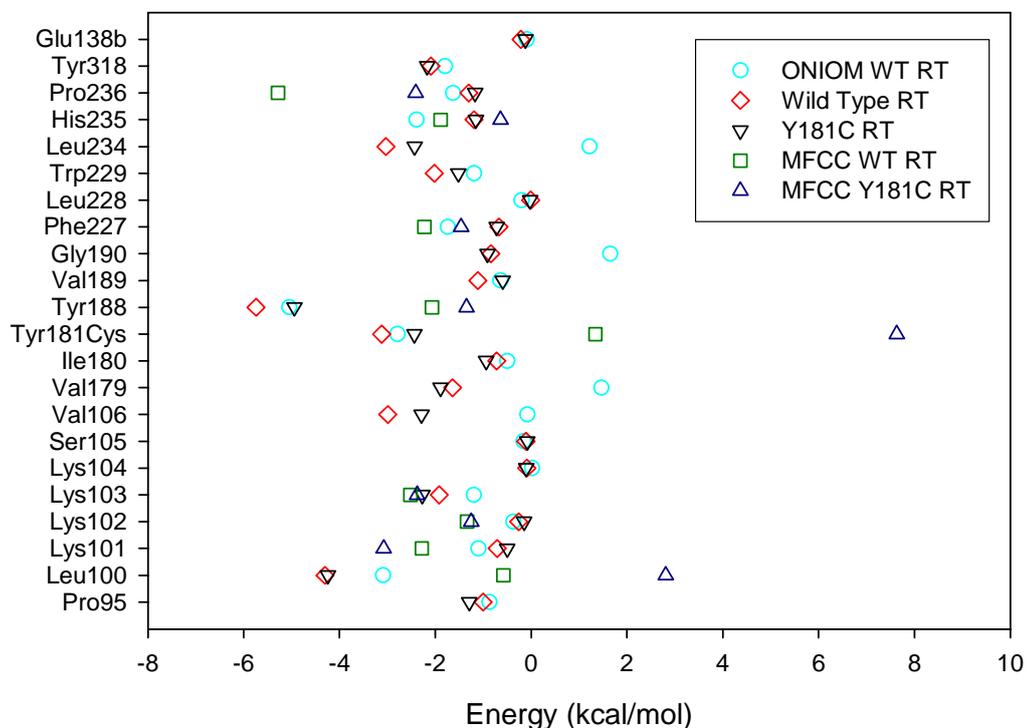


Figure 12 Plot of interaction energies (kcal/mol) of nevirapine with individual residues. The calculations from ONIOM method with the wild type RT is shown in light blue circle, decomposition method with wild type RT is shown in red diamond, decomposition method with Y181C HIV-1 RT is shown in black triangle, the MFCC method with wild type RT is shown in green square, and the MFCC method with Y181C HIV-1 RT is shown in dark blue triangle.

Moreover, the hydrogen bond distances of bridging water are clearly shown in figure 11. The contribution of hydrogen bonding interaction via bridging water leads to the calculation of decomposition energy via this bridging water molecule. The decomposition energy between RT and nevirapine via bridge-water molecule is shown in table 13. Importantly, if the decomposition energies via WAT1075 are included the attractive energies between Leu234, His235 and nevirapine would have changed dramatically. The attractive energies of nevirapine and Leu234 have changed from -3.03 to -5.99 kcal/mol for the wild type RT, and from -2.43 to -3.68 kcal/mol

for the mutant type RT. The attractive energies of nevirapine and His235 have changed from -1.19 to -5.29 kcal/mol for the wild type RT, and from -1.15 to -3.39 kcal/mol for the Y181C HIV-1 RT. If the attractive energies via bridging water molecule are included the interaction energies at Leu234 and His235 become as high as at the Tyr188. These significant interactions indicated how WAT1075 play an important role in the binding pocket.

Table 13 Decomposition energies (kcal/mol) of bridging WAT1075 with amino acid residues and nevirapine.

	Wild type	Y181C
Leu234	-0.60	-0.69
His235	-1.74	-1.68
Nevirapine	-2.36	-0.56

CONCLUSIONS

Molecular dynamics simulation is a common method used to observe structural properties and some energetic properties by combination with the MM-PBSA calculations. Including one water molecule which forms hydrogen bonded-bridge between nevirapine and the HIV-1 RT pocket of both wild type and mutant type provides better energetic properties of MM-PBSA energy, entropic energy, binding free energy and relative binding free energy than the previous one. The obtained the binding energies of -41.77 and -33.52 kcal/mol were calculated for wild type and Y181C HIV-1 RT, respectively. Although the resistance folded of 7.04 is comparable to other techniques the entropy calculation would be improved in the future. This work can introduce the new way that is suitable to calculate of both the absolute binding free energy and simultaneously the relative binding free energy which can easily compare to the experimental resistance folded. This technique can be further developed and implement to either the nevirapine analogues or the system where water play an important role of hydrogen bonding-bridge to predict the binding free energy. Therefore, it can be used as one of many tools to develop the new potential drug in the future. The decomposition energies show well agreement of interaction energies with the ONIOM BBF-corr calculations. The decomposition calculations are included the dynamic properties of the complex. Thus the interaction energies between nevirapine and individual reveal all the attractive energies. The main contribution of -5.74 and -4.94 kcal/mol comes from Tyr188 in the wild type and Y181C HIV-1 RT, respectively. The forming hydrogen-bonded bridge via WAT1075 leads to the calculation involved this water molecule. The key interaction is the attractive energies via the bridging water molecule. The contribution of interactions from WAT1075 to Leu234 and His235 are -2.96 and -4.10 kcal/mol for wild type HIV-1 RT, -1.25 and -2.24 kcal/mol for Y181C HIV-1 RT. If including the attractive interaction via WAT1075 the amino acid residue, Leu234 and His235, would have high contribution as Tyr188. The highly contribution can be observed only in the wild type RT. These significant roles of the bridging water molecule and dynamics properties are the key information to understand more on binding of the nevirapine analog in the future.

CHAPTER II: MOLECULAR DYNAMICS SIMULATIONS OF WILD TYPE, Y181C MUTANT HIV-1 RT/NEVIRAPINE AND MINOR GROOVE BINDER, THIAZOTROPSIN A, COMPLEX: THE ENERGETIC AND INTERACTIONS BY MEANS OF MM-PBSA CALCULATIONS

INTRODUCTION

Deoxyribonucleic acid (DNA) is a fundamental building block of life represented by means of the genetic code. The code is used in the development and function of all known living organisms and some viruses. In 1953, Watson and Crick determined the structure of double helical DNA (Figure 13). The double stranded DNA consists of two polynucleotide chains, binding together via hydrogen bond interactions between complementary base pairs on opposite strands (Watson and Crick, 1953). The main function of DNA molecules is the long-term saving of information and is often compared to a set of blueprints. The DNA contains the instructions needed to build other components of cells such as proteins and RNA molecules. Genes are DNA segments that carry this genetic information; some other DNA sequences have structural roles, or are involved in regulating the use of this genetic information.

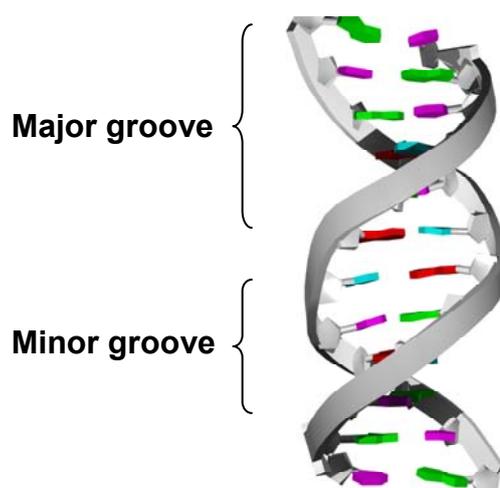


Figure 13 The molecular structure of DNA.

The DNA backbone strand is a connectivity of phosphate and ribose sugar ring as shown in figure 14. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The ribose rings are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. Thus the DNA strand has direction as a result of these asymmetric connections; in the DNA double helix, the direction of the nucleotides in one strand is opposite to their direction in the other strand and turned antiparallel. The asymmetric ends of DNA strands are referred to as the 5' (*five prime*) and 3' (*three prime*) ends.

The DNA double helix is stabilized by hydrogen bonding attraction between the bases attached to each strand. There are four bases found in DNA, which are adenine (A), cytosine (C), guanine (G) and thymine (T). These bases are attached to the sugar/phosphate to form the complete nucleotide. Adenine, cytosine, guanine and thymine bases are classified into two types; purines and pyrimidines. Adenine and guanine are purines and are fused five- and six-membered heterocyclic compounds, whilst cytosine and thymine are six-membered rings of the pyrimidine type. Each type of base specifically forms hydrogen bond with just one type of base on the other strand; purines forms hydrogen bonds to pyrimidines specifically. A forms two hydrogen bond to T only, and C forms three hydrogen bond only to G as shown in figure 14. This arrangement of two nucleotides binding together across the double helix is called a base pair. The complementary base pair sequence along the chain of DNA is the code for analysis and diagnosis of diseases at the genetic level (Vercoutere and Akeson, 2002; Zhai *et al.*, 1997). The double helix is also stabilized by the hydrophobic effect and π stacking, which are not influenced by the sequence of the DNA (Ponnuswamy and Gromiha, 1994). As hydrogen bonds are not covalent and represent weak bonds, DNA can be easily broken and reformed occasionally. The two strands of DNA in a double helix can be separated apart, either by mechanical force or high temperature (Clausen-Schaumann *et al.*, 2000). This reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.

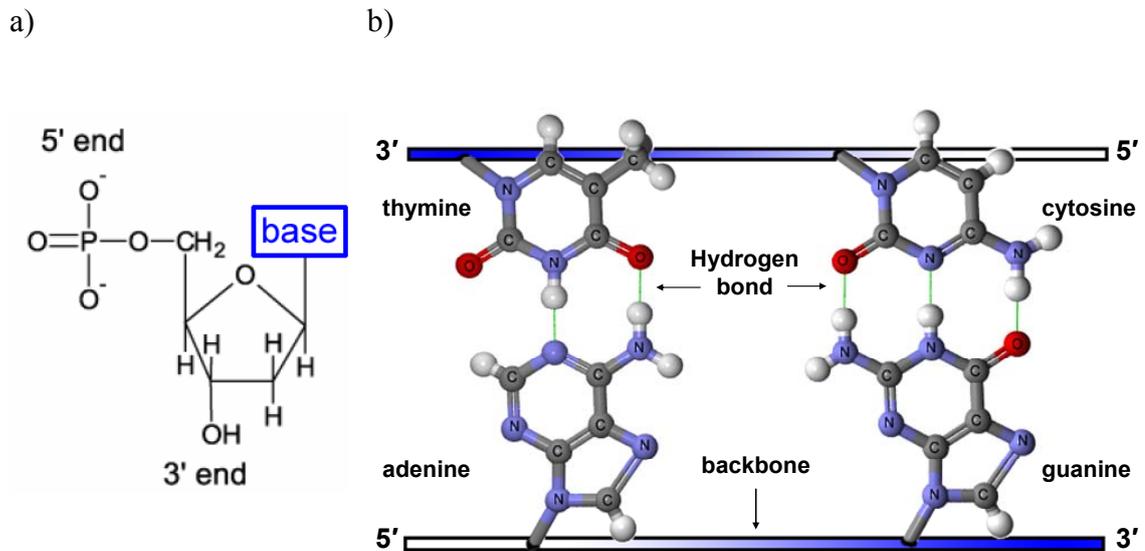


Figure 14 The back bond DNA strand components; a) DNA nucleotide, b) the four bases of DNA showing their complementary binding properties.

Double helix DNA is a right-handed spiral. As the DNA strands twist around each other, they leave gaps between each set of phosphate backbones, revealing the sides of the bases inside. There are two types of grooves observed from the spiral surface which are the major groove and minor groove. The major groove is 22 Å wide and the minor groove is 12 Å wide (Wing *et al.*, 1980). The depth and width of the groove enable access to the bases, and allows proteins like transcription factors to bind to specific sequences in double-stranded DNA by making contacts to the sides of the bases exposed in the major groove (Pabo and Sauer, 1984).

DNA damage can occur from many different sorts of mutagens, which are agents that change the DNA sequence. Mutagens include oxidizing agents, alkylating agents as well as high-energy electromagnetic radiation such as ultraviolet (UV) light and X-rays. UV light mostly damages DNA by producing thymine dimers, which are cross-linked between adjacent pyrimidine bases in a DNA strand (Douki *et al.*, 2003). On the other hand, oxidizing agents like free radicals or hydrogen peroxide cause multiple damage, including base modifications, particularly of guanine, as well as double-strand breaks (Cadet *et al.*, 1999). It has been estimated that in each human

cell, about 500 bases suffer oxidative damage per day (Cathcart *et al.*, 1984; Shigenaga *et al.*, 1989). However, the most dangerous are double-strand breaks, which are difficult to repair and can produce point mutations, insertions and deletions from the DNA sequence, as well as chromosomal translocations (Valerie and Povirk, 2003). Some ligands that have an ability to bind to DNA such as intercalators and minor groove binder can cause mutations.

Many mutagens intercalate into the space between two adjacent base pairs. Intercalators are mostly aromatic and planar molecules such as ethidium, daunomycin, doxorubicin and thalidomide. The bases must be separated in order to fit an intercalator between the base pairs. This distorts the DNA strands by unwinding the double helix as shown in figure 15. These structural changes inhibit both transcription and DNA replication, causing toxicity and mutations. The well-known DNA intercalators are often carcinogens like benzopyrene diol epoxide, acridines, aflatoxin and ethidium bromide (Ferguson and Denny, 1991; Jeffrey, 1985; Stephens *et al.*, 2000). Nevertheless, due to their properties of inhibiting DNA transcription and replication, they can also be used to inhibit rapidly-growing cancer cells in chemotherapy (Brana *et al.*, 2001).

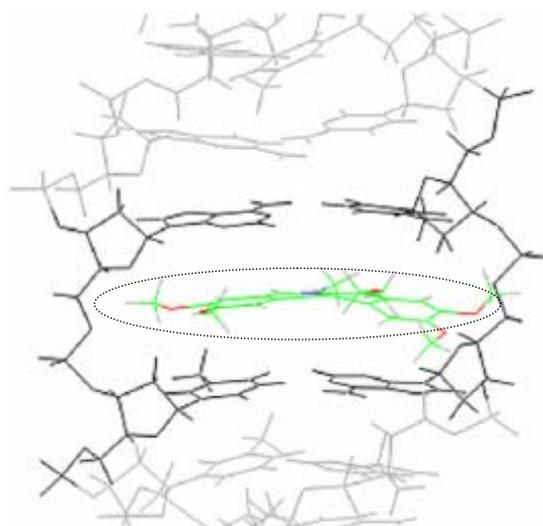


Figure 15 A intercalator bound and orientated inside the DNA sequence (Clark *et al.*, 2007).

Mutagens that can bind in the DNA minor groove called minor groove binders. Over the past two decades, there has been a great deal of research in the design of compounds that recognize the minor groove of DNA in a sequence specific manner. The evolution in the design of minor groove binders is reviewed below. Minor groove binders which are being used in biological applications are described as well as some speculation about future possibilities to be used clinically.

Amongst the groove binders there are some molecules that bind with a very substantial preference to the DNA. For example, netropsin and distamycin, the antiviral antibiotics, were the first ligand discovered that bound selectively to the minor groove of AT-rich DNA sequences. The first characterized in detail by NMR was netropsin (Figure 16); chemical shifts suggested a direct interaction between protons on the netropsin with adenosine protons in the minor groove. Later, a crystal structure of netropsin bound to an AATT site was solved, which confirmed the NMR results as well as providing more details of the interactions that stabilized binding (Kopka *et al.*, 1985a). The crystal structure showed that binding occurred in a region of narrowed minor groove with good van der Waals contact between the top and bottom surface of the ligand and the walls of the DNA groove (Kopka *et al.*, 1985b). The hydrogen bonding was between the hydrogen bond donors at the amides linking the pyrrole rings and the terminal tails, and hydrogen bond acceptors at the edges of the base pairs. The charged cationic terminal group positioned deep in the groove where the negative electrostatic potential is large. All of these interactions were considered to stabilize the complex. The association of netropsin depended on salt concentration, presumably arising from the ionic interactions of the tails. Distamycin (Figure 16) is a close homologue of netropsin, which contains three pyrrole rings and just one tail, binding with nearly the same affinity as netropsin. The structure of distamycin bound to DNA has been solved (Coll *et al.*, 1987), with very similar features stabilizing the complex. Additional, molecules in this class include Hoescht 33258, DAPI, and SN 6999 (Figure 16), which all showed the same characteristics of curved, planar aromatic cores, with positively charged groups and hydrogen-bond donors on the concave edge. The complexes of the minor groove binders and DNA that have been solved indicate that shape and functional group complementarity

between the ligand and groove are critical features for binding. To date, a great many variations of minor groove binders ligands had been synthesized based on their distamycin–netropsin parents (Bailly and Chaires, 1998).

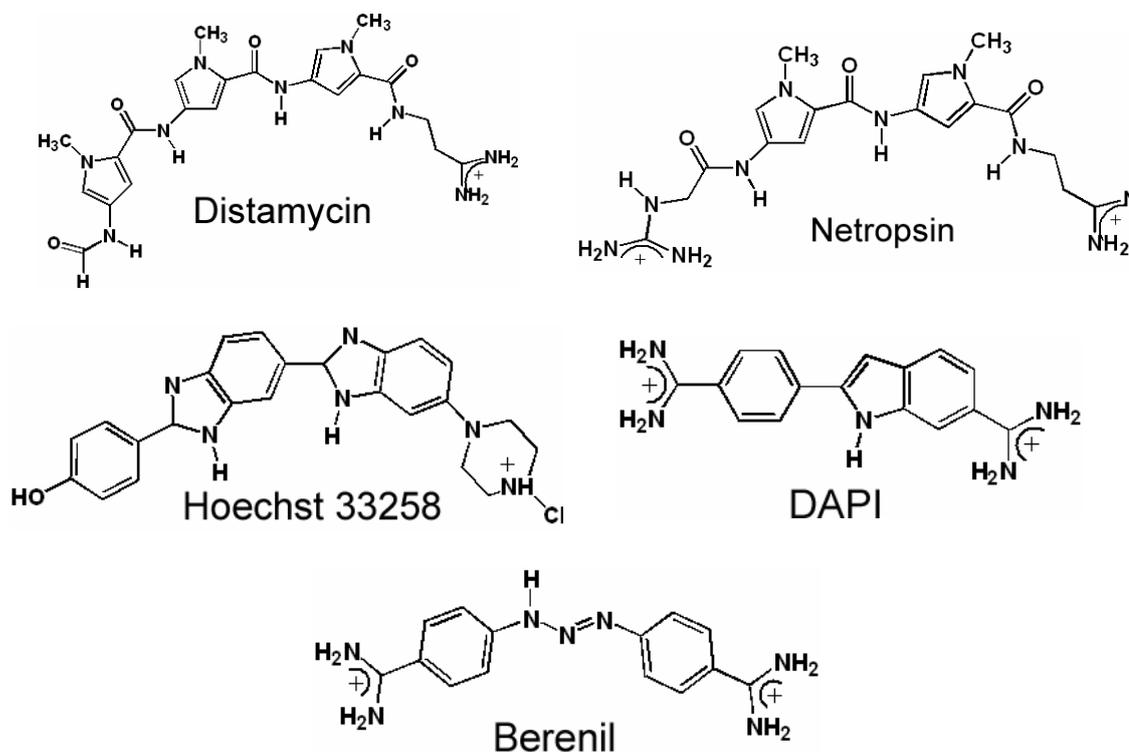


Figure 16 The structures of distamycin, netropsin, and several other minor groove ligands are shown. All have in common a planar aromatic core and positively charged end groups.

The report by Wemmer and co-workers revealed an interesting variation on the binding of distamycin through NMR studies. Distamycin binding to A,T-rich sequences, surrounding to the length of ligand, fits four consecutive AATT sequence at the center of dodecamer DNA which is sufficient for tight binding (Klevit *et al.*, 1986; Pelton and Wemmer, 1988). However, when an additional A-T base pair was added to make the target sequence of $d(-AAATT-)_2$, resonances of the complex were formed at low stoichiometries of distamycin to DNA. A new set of saturated at a 2:1 ratio of ligand to DNA duplex was indicated by the Nuclear Overhauser Effect (NOE) studies and showed that the two molecules of distamycin were bound side by side in the minor groove, running antiparallel (Pelton and Wemmer, 1989). The

crystallographic structure confirmed that distamycin bound in either orientation with respect to the asymmetric DNA sequence with nearly equal affinity. Each distamycin monomer contacted one strand of the DNA, while the the three pyrrole rings of the two ligands stacked with each as shown in Figure 17. The 2:1 complex appeared to form only when a sequence of five or more consecutive AT pairs were presented, because the positive charges on the tails were extended since each pyrrole ring contacted one base pair, and each tail did as well. As in the 1:1 complexes, the ligands are twisted slightly in order to match the curvature of the DNA. However, the thickness of the 2:1 dimer stacked ligands is much wider than a single molecule, and hence the minor groove had to widen by 3.5–4 Å compared to the 1:1 complexes (Figure 17). The relative populations of the 1:1 and 2:1 complexes have been analyzed and results indicated that the binding constant of the second molecule was about 10-fold lower than the first. This implied that the width of the DNA groove could change substantially at very low energetic cost. The flexibility of DNA came as a surprise when crystallographic studies of small DNA oligomers gave the impression that DNA was quite rigid on a local scale. The binding constants were later verified directly with isothermal titration calorimetry measurements (Rentzeperis *et al.*, 1995).

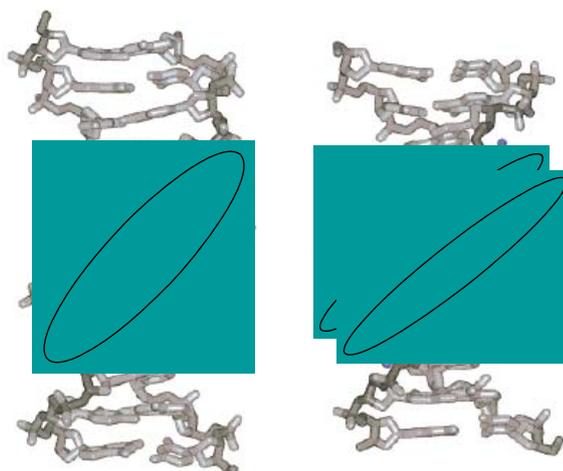


Figure 17 Structures are shown of distamycin bound in 1:1 (left) and 2:1 (right) complexes. The DNA is shown in gray, while the distamycin molecules in the minor groove are colored according to atom type. The backbone atoms are shaded darker gray to highlight the change in the minor groove width between the two forms of complex (Wemmer, 2001).

Lown and co-workers continued to develop the minor groove binders by synthesis of an impressive family of “lexitropsins”. Lexitropsins were based on netropsin and distamycin (Lown, 1988; Lown, 1994) by substituting imidazole, thiazole, triazole, pyrazole, or oxazole heterocycles in place of the N-methylpyrrole rings of distamycin. These newly designed lexitropsins had the ability to recognize and bind to sequences containing one or two GC pairs embedded in an AT sequence (Figure 18). Among the numerous lexitropsins synthesized, thiazole containing lexitropsins showed the promise of reading longer DNA sequences (Anthony *et al.*, 2004). Thiazole lexitropsins can either accept or avoid a GC base pair in their binding sites, depending on the position of the sulfur atom (Kumar *et al.*, 1991; Rao *et al.*, 1990a; Rao *et al.*, 1990b).

The lexitropsin binding mode based on a 2:1 complex has led to minor groove binders with increased recognition that include GC base pairs in the binding site. However, creating ligand binding to purely GC-specific DNA presents problems. According to computational studies, the observation that lexitropsins accommodate both AT- and GC-containing sequences came from the fact that the electrostatic potential in the minor groove of AT-rich regions is very negative (Pullman and Pullman, 1981). These electrostatic interactions between AT regions and the positively charged end groups in the lexitropsins provided the initial attraction. The presence of 1:1 or 2:1 complexes referred to mono- and di-cationic binding which had a significant effect on the electrostatic component at the AT-rich regions (Goodsell *et al.*, 1995; Marky and Breslauer, 1987; Singh *et al.*, 1994). However, distamycin in its mono-cationic form argues against the dominant role of electrostatics in sequence selectivity (Zhang *et al.*, 1993).

Complementary techniques such as quantitative footprinting methods have been used to support and provide more detail of sequence specificity coupled with main techniques such as NMR-spectroscopy and X-ray crystallography which have provided the means of structurally determining ligand/DNA complexes. A new example of lexitropsins called thiazotropsin A binds specifically to the DNA minor groove in a 2:1 ratio with extended recognition for G-DNA bases. NMR studies

showed that the new iPrTh element of the ligand enhanced both hydrophobicity and hydrogen bonding between the nitrogen of iPrTh to the floor of the minor groove via the exocyclic amine protons of G residues, and improved sequence specificity. (Anthony *et al.*, 2004). The specific recognition between thiazotropsin A and the 5'-ACTAGT-3' DNA sequence was also driven by shape and its ability to accommodate by induced-fit a second thiazotropsin A. Furthermore, the finding revealed that thiazotropsin A was able to read a six-base pair self-complementary DNA sequence d(CGACTAGTCG)₂. Further experimental support for the affinity of thiazotropsin A for ACTAGT came from capillary electrophoresis (CE) studies which also highlighted the possibility of higher order binding to DNA. Computational studies provides an informative rational explanation for the selective fit of this crescent-shaped ligand into the minor groove of DNA; in silico footprinting (ISF) data has shown the pattern of binding with high apparent affinity occurring at the ACTAGT region, although variations to flanking bases of the DNA reading frame for thiazotropsin A have been revealed (James *et al.*, 2004), but the molecular structural or energetic basis for this have not been explained. Isothermal Titration Calorimetry (ITC) can provide a means through which the energetics of the binding process in solution can be rationalized. Thus the relation of these structural investigations with binding and interaction energies between thiazotropsin A and the sequence d(CGACTAGTCG)₂, can be targeted and quantified by combining ITC experiment with molecular modeling.

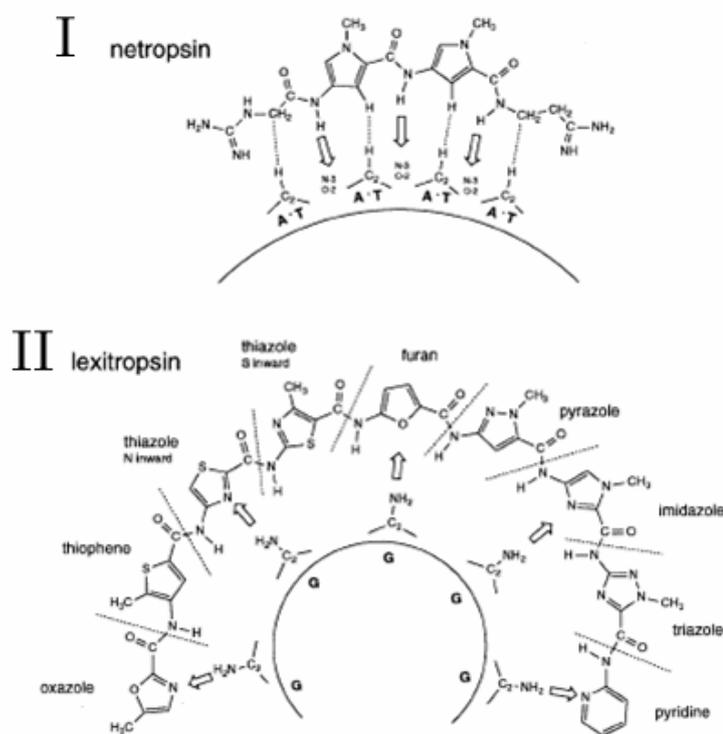


Figure 18 (I) Representation of the binding to DNA of netropsin; (II) proposed representation of a model lexitropsin molecule with guanine residues in DNA. Heavy arrows are hydrogen bonds, from donor to acceptor. Dashed lines mark close van der Waals nonbonded contacts between DNA and drug. (Bailly and Chaires, 1998)

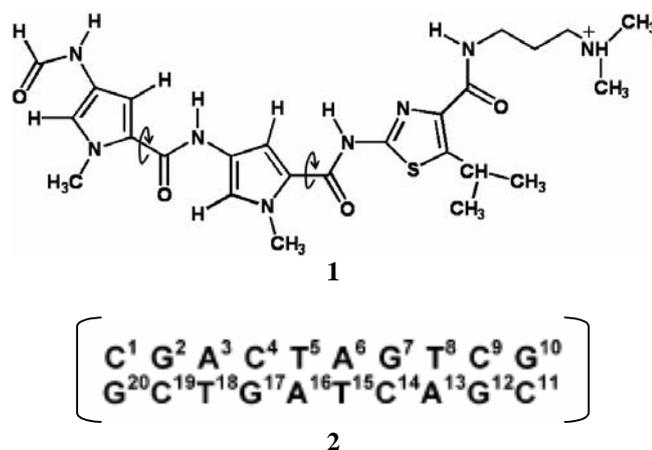


Figure 19 Compound 1, thiazole lexitropsins (thiazotropsin A), and the self-complementary DNA duplex 2 (Anthony *et al.*, 2004).

Experimental validation of modeling studies is always considered vital, especially in cases where the conformational structures are sensitive and binding modes show dramatic variations; such effects may not be evident from computing-based studies alone. However, in the longer term, as the technique of structure determination by modeling evolves, structural design alterations will be accommodated by suitable modeling approaches, and the need for experimental validation will be reduced as well as the experimental cost.

Previous studies described how MD driven insilico footprinting (ISF) is a computationally-efficient screening approach that can reliably determine sequence selectivity. Thus, the objective of this work is to conduct more fundamental thermodynamic investigations of binding that use MD simulations. The MM-PBSA methodology has provided successful free energy representations in a number of nucleic acid systems and involves conformational averaging of samples taken from extended MD simulations. Despite being a simple but approximate method based on a continuum representation of the solvent for free energy determinations, relative free energies of association can be computed that reflect experimental values as determined by ITC.

Thermodynamic studies of the 2:1 thiazotropsin A/DNA complexes will help discover how these ligands bind and recognize selective double-stranded DNA sequences and aid understanding of the binding mechanisms. To unlock the questions of quantitative structural binding energy relationships would lead to a fuller understanding of the factors governing sequence dependent binding as well as improve the minor groove binding drug design process, coupled with physicochemical, biochemical, and biological studies. The design of sequence-specific ligands has the potential to be one of the great success stories of pharmacology.

METHODS OF CALCULATIONS

1. Force Field and Parameters for Thiazotropsin A

All simulations were carried out with the AMBER8 program package. Calculations were performed on Linux 3.2 GHz and Suse9.2 Opteron 8 processors. There were 4 main combinations to construct prep files and force fields for thiazotropsin A with respect to the mapping charge method and force field parameters (bond, angle, torsion and improper torsion) which will be explained later.

1. B3LYP/6-31G**, plus ligand's force field version 1 (MD1)
2. B3LYP/6-31G**, plus ligand's force field version 2 (MD2)
3. HF/6-31G*, plus ligand's force field version 2 (MD3)
4. HF/6-31G*, plus ligand's force field version1 (MD4)

B3LYP and HF are the methods used to calculate single point energies using the Gaussian03 program with basis set 6-31G** and 6-31G*, respectively, for the thiazotropsin A monomer. Whilst B3LYP/6-31G** is a higher level of calculation, HF/6-31* is the only one used in the literature and is supposed to be compatible with the AMBER force field for DNA and proteins. Charge distributions around the thiazotropsin A molecule were calculated by the MK algorithm. The RESP mapping charge method in AMBER was then used to map the surrounding distribution back to atomic charges on thiazotropsin A.

To investigate the compatibility of the charge calculation method, another charge calculation method, NBO, was used directly from the Gaussian output with the B3LYP/6-31G** level of theory for atomic charges of thiazotropsin A.

There were two sets of force field for thiazotropsin A used in calculations. The difference between the two sets of force field was the atomic types, angles, dihedrals and improper torsion parameters as shown in the support information.

2. Force Field for DNA Sequence

Initially, AMBER force field99 was used because it has been developed for DNA simulations. The AMBER force field99 (ff99 or ff03) is a non-polarizable force field, and because DNA is a charged system, research groups have claimed that AMBER force field02 (ff02), which is a polarizable force field, is more suitable for DNA simulations than ff99 or ff03 and ff02EP (polarizable with extra charge, lone pair electrons). In this work we have used the AMBER force fields: ff03, ff02 and ff02EP for comparison.

3. Starting Models of the DNA/Thiazotropsin A Complex

The 2:1 thiazotropsin A/DNA complex starting structure was taken from the NMR refinement study as a 10 base pairs sequence: d(C₁G₂A₃C₄T₅A₆G₇T₈C₉G₁₀)₂. The InsightII program was used to add a CG terminal base pair to both ends of the DNA sequence: d(G₁C₂G₃A₄C₅T₆A₇G₈T₉C₁₀G₁₁C₁₂)₂ in order to compare the modeling and Isothermal Calorimetry (ITC) experimental results.

For the thiazotropsin A starting structure, there are two types of structure: one from the NMR refinement of the thiazotropsin A/DNA complex structure and the others from a B3LYP/6-31G** optimization. The latter was generated as follows: Optimization of thiazotropsin A monomer with the B3LYP/6-31G** method in the Gaussian03 program; convert output coordinates from Gaussian to PDB format and reorder atomic type; InsightIII program was used to superimpose thiazotropsin A monomer in PDB format to the NMR structure of thiazotropsin A; thiazotropsin A monomer was superimposed twice to fit to thiazotropsin A monomer 25 and monomer 26 from the dimer arrangement solved by NMR.

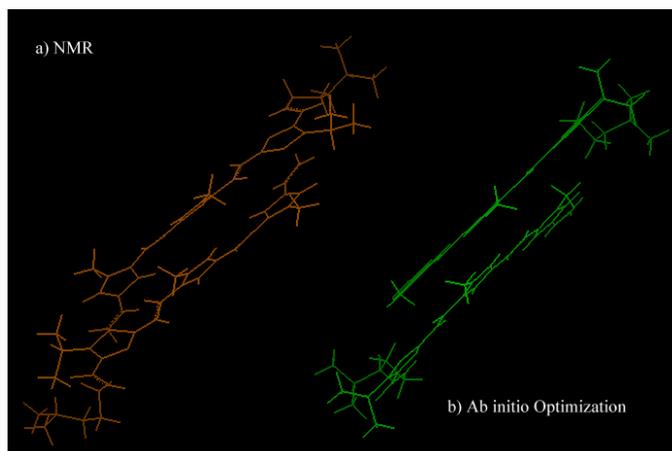


Figure 20 The starting structure of Thiazotropsin A used in the simulations. a) From NMR. b) From B3LYP/6-31G** optimization.

4. Setup the Simulations of the DNA/Thiazotropsin A Complex System

The force field for thiazotropsin A (version 1 and version 2), force field for DNA (ff03, ff02, ff02EP) and mapping for new atom types (for force field v.2: nZ, cL, cR) were loaded through the Leap module in AMBER. The pdb files for the starting complex structure (from ab initio minimization and NMR), DNA and thiazotropsin A were then loaded. The invacuo complex simulation system was saved at this stage. The solvated system was continued by adding solvent water molecule using the TIP3P water model that extended 10 Å from the solute surface; there were approximately 3900 water residues in volume about 50 x 70 x 50 Å³. Finally, 20 Na⁺ counter ions were added to neutralize the system and placed in the most negative region on electrostatic potential energy surface, as determined by Leap subprogram. If the charge fraction (such as -19.99990) appeared during add the counter ions, it would be corrected by a small fraction in the prep file of thiazotropsin A.

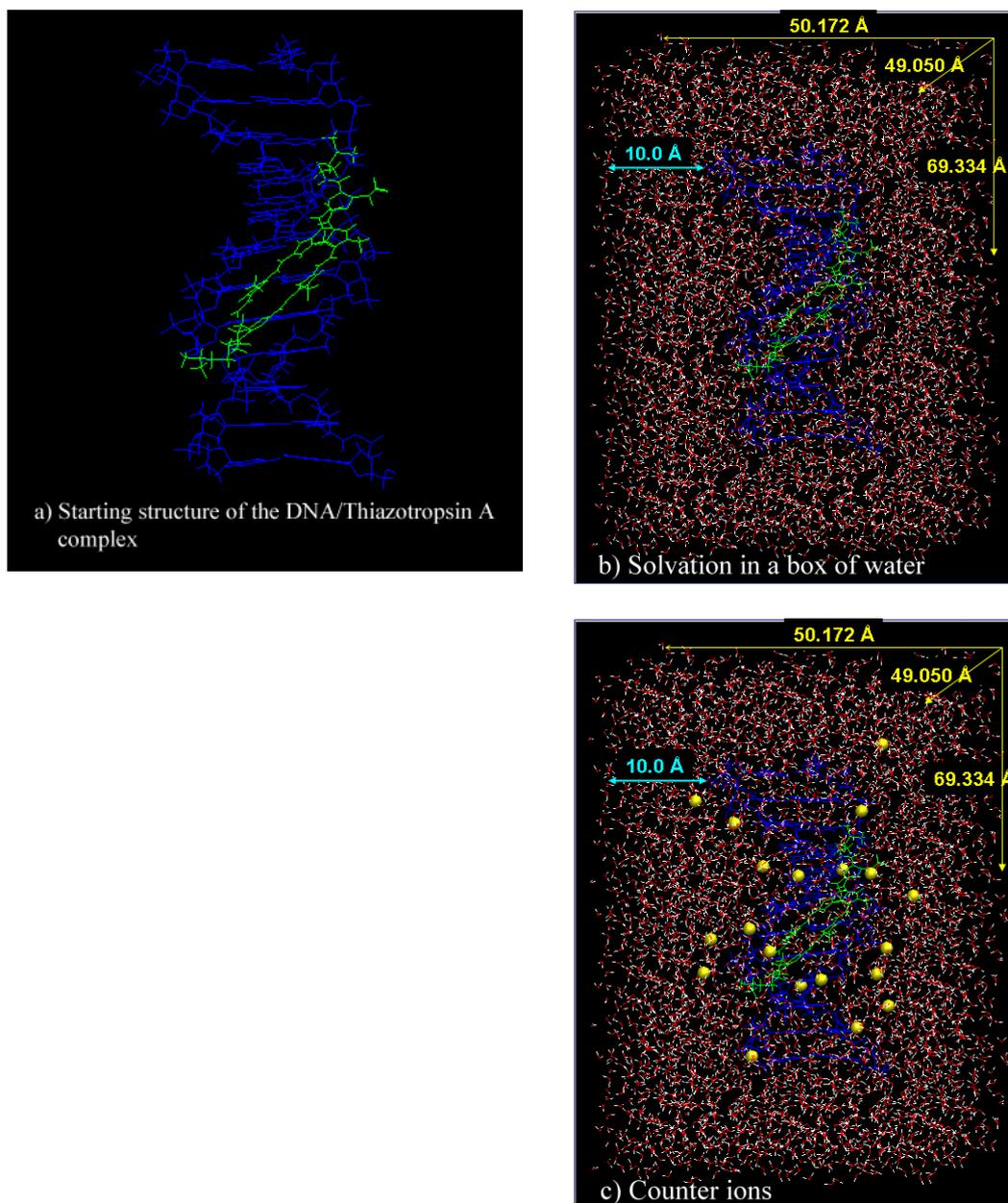


Figure 21 The setup steps of the DNA/Thiazotropsin A simulations system. A) Starting solute. B) Solvation in a box of water. C) Neutralized by adding counter ions, Cl^- .

The simulation systems were classified as follows:

1. In vacuo simulations with a restraint of 1.0 kcal on DNA
 - 1.1 In vacuo, restrain 1.0 kcal for DNA and 0.1 kcal for thiazotropsin
 - 1.2 In vacuo, restrain 10.0 kcal for DNA and thiazotropsin
2. In vacuo simulations, including counter ions, with restraint of 1.0 kcal on DNA
3. Solvation simulations, including counter ions, with restraint of 1.0 kcal on DNA
4. Solvation simulations, including counter ions, of all restraints free
 - 4.1 Protein equilibration
 - 4.2 Loop equilibration
5. Simulations using the AMBER drug/DNA tutorial equilibration

5. Computational Calculations of the DNA/Thiazotropsin A Complex

5.1 The simulations system type 1

The simulations system type 1, in vacuo simulations with restrained DNA, started from a minimization with Steepest Descents for 20000 steps, followed by Conjugated Gradients for 30000 steps, with the DNA restrained by a restraining force of 10.0 kcal/mol. An infinite non-bonded cutoff was used. The minimization was performed until the energy gradient reached the convergence criteria of 10^{-4} kcal/mol Å. Next, ten rounds of equilibration were performed by gradually decreasing the restraint force from 10.0 to 1.0 kcal/mol for DNA and 1.0 to 0.1 kcal/mol for thiazotropsin A. During the same time, the temperature was gradually increasing from 0 to 300 K. Each round of the equilibration was carried out for 10 ps with a time step of 1 fs. The non-bonded cutoff was 12 Å, and a periodic boundary of canonical ensemble with a constant temperature using the Leapfrog integrator was used with a collision frequency of 1 ps^{-1} . The time constant for heat bath coupling was 1 ps.

Finally, production simulations were performed by reading coordinates and velocities formatted from equilibration step. The simulation parameters were set as the same as in the equilibration, except for keeping the temperature constant at 300 K with a time constant for heat bath coupling of 2 ps. The system was run for 1 ns and the trajectory coordinates saved every 0.5 ps. The restraint force of 1.0 kcal/mol was applied to DNA throughout the production run.

Subsystems were also performed for complete investigation. First, a restraint force of 1.0 and 0.1 kcal/mol was applied to DNA and thiazotropsin A, respectively during the production run. Secondly, the simulation was performed only on the production run for 1 ns by using the NMR structure as a starting and the temperature kept constant at 300 K. A restraint force of 10.0 kcal/mol was applied to both DNA and thiazotropsin A.

Table 14 The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 1.

Steps	DNA	Thiazotropsin A
Min. 50000 steps (SD 20000+CG 30000)	10.0	-
Equilibration 10 ps x 10 rounds	10.0	1.0
Equilibration (round 10)	1.0	0.1
Production run 1 ns (system 1.1)	1.0	-
Production run 1 ns (system 1.2)	1.0	0.1
Production run 1 ns (system 1.3)	10.0	10.0

5.2 The simulations system type 2

The simulations system type 2, invacuo simulations including counterions with restraints on DNA were performed as referred twin simulations protocol from the simulations system type 1.

Table 15 The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropin A in the simulations system type 2.

Steps	DNA	Thiazotropin A	Ions
Min. 50000 steps (SD 20000+CG 30000)	10.0	-	-
Equilibration 10 ps x 10 rounds 	10.0	1.0	-
Equilibration (round 10)	1.0	0.1	-
Production run 1 ns	1.0	-	-

5.3 The simulations system type 3 (Protocol A)

The simulations system type 3, solvation simulations including counter ions with the restraints on DNA, started with a minimization with Steepest Descents for 20000 steps followed by the Conjugated Gradients for 30000 steps on solvent water whilst the DNA and thiazotropin A were kept restrained with a force of 10.0 kcal/mol and 1.0 kcal/mol, respectively. A non-bonded cutoff of 20 Å was used. The minimization was performed until the energy gradient reached the convergence criteria of 10^{-4} kcal/mol Å. Next, thiazotropin A was minimized using the same algorithm as the minimized solvent step, but with the restraints on thiazotropin A removed. Next, ten rounds equilibration were performed whilst gradually decreasing the restraint force from 10.0 to 1.0 kcal/mol for DNA, 1.0 to 0.1 kcal/mol for thiazotropin A and gradually increasing the temperature from 0 to 300 K. The first round of equilibration was run under constant volume dynamics for 10 ps with a time step of 1 fs. The non-bonded cutoff was 20 Å and complete bond interactions were calculated (not using SHAKE). Dynamics parameters were set to canonical ensemble with constant temperature using the Leapfrog integrator, a collision frequency 1 ps^{-1} , and a time constant for heat bath coupling of 1 ps. From the ninth to tenth step, equilibration was performed for 20 ps for each round, with an integral time step of 2 fs. The non-bonded cutoff was changed to 12 Å. Addition; SHAKE algorithm applied to constrain bonds involved H-atoms to omit bond interactions involving H-atoms. The relative geometrical tolerance for coordinate resetting in SHAKE was 0.00005 Å.

Finally, the production molecular dynamics run was performed by reading coordinates and velocities formatted from the last equilibration step. At the same time, a restraint force of 1.0 kcal/mol was applied to DNA, whilst the force on thiazotropsin A was removed. Simulation parameters were set the same as the last equilibration, except for the temperature which was kept constant at 300 K with a time constant for heat bath coupling of 2 ps. At this point, the simulation system was switched to constant pressure dynamics. Pressure regulation was added by setting molecular dynamics with isotropic position scaling, and the system was maintained at 1 atm, with a pressure relaxation time of 0.2 ps. The simulation was carried out for 2 ns and trajectory coordinates were collected every 1 ps. The simulations system type 3 will be called protocol A later in this thesis.

Table 16 The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropsin A in the simulations system type 3.

Steps	DNA	Thiazotropsin A	Ions	Solvent
Min. Solvent 50000 steps (SD 20000+CG 30000)	10.0	1.0	-	-
Min. Ligand 50000 steps (SD 20000+CG 30000)	10.0	-	-	-
Equilibration 10 ps	10.0	1.0	-	-
Equilibration 20 ps x 9 rounds	9.0	0.9	-	-
Equilibration (round 10)	1.0	0.1	-	-
Production run 2 ns	1.0	-	-	-

5.4 The simulations system type 4 (Protocol B)

The simulations system type 4, solvation simulations including counter ions without restraints, was separated into 2 subtypes according to the equilibration protocols; straight equilibration and loop equilibration.

5.4.1 The straight equilibration (Protocol B1)

The simulations started with minimization using Steepest Descents for 1000 steps followed by the Conjugated Gradient for 1000 steps of solvent water whilst DNA, thiazotropsin A and counter ions were kept restrained with a force of 10.0 kcal/mol. A non-bonded cutoff of 20 Å was used. Next, ten rounds of minimization on solvent water were performed; each round of minimization had been driven the Steepest Descents algorithm for 500 steps followed by the Conjugated Gradients algorithm for 500 steps. The restraint force was reduced by 1 kcal/mol in each round. The rate of release the restraint force was ranked from the fastest to lowest i.e. counter ions, thiazotropsin A, and DNA, respectively. Until the last round, counter ions were allowed to move. Next, thiazotropsin A was minimized using the same algorithm as in the minimized solvent step, but with the restraints on thiazotropsin A removed. Initially, equilibration was performed for 40 ps using constant volume simulations. A simulation time step of 2 fs was used whilst the SHAKE algorithm was applied to constrain bonds involving hydrogen atoms. A non-bonded cutoff of 12 Å was applied. Long range electrostatic interactions were calculated using the exact ewald summation method. The temperature of the system was controlled by rising from 0 to 300 K in a few picoseconds with a time constant for heat bath coupling of 2 ps. At the same time, a restraint force of 10 kcal/mol was applied to DNA, thiazotropsin A and counter ions. Next, seven rounds equilibrations were performed under constant pressure conditions. The system was maintained at 1 atm controlled by a pressure relaxation time of 2 ps. Equilibrations were carried out for 20 ps in each round. The temperature was kept constant at 300 K. The restraint force was reduced by 1 kcal/mol in each round. The rate to release the force was ranked from the fastest i.e. counter ions, thiazotropsin A, and DNA, respectively. The problem of a hot spot in energy arose after two rounds equilibrations. To solve this problem, the application of the SHAKE algorithm was removed and replaced by a time step of 1 fs for 20 ps for one round of equilibration. Next, five rounds equilibrations were continued using the same parameters as the previous equilibration except the direct sum Coulomb interaction for the ewald summation was switched from the exact ewald to the approximate cubic spline method. Finally, a restraint force

of 1.0 kcal/mol was kept on DNA whilst the thiazotropsin A and counter ions were free to move. The production run was performed for 2 ns and the trajectory was collected every 1 ps. The protocol was taken from the last equilibration, but with the restraint force removed from DNA allowing the entire system to move freely with a constant temperature of 300 K. The simulations system type 4 with a straight equilibration will be called protocol B1 later in this thesis.

Table 17 The stepwise to apply the restraint force (kcal/mol) on DNA and thiazotropsin A in the simulations system type 4, the straight equilibration.

Steps	DNA	Thiazotropsin A	Ions	Solvent
Min. Solvent 2000 steps (SD 1000+CG 1000)	10.0	10.0	10.0	-
Min. 1000 steps (SD 500+CG 500)	10.0	10.0	9.0	-
Min. 1000 steps (SD 500+CG 500) x 10 rounds	10.0	9.0	8.0	-
Min. 1000 steps (SD 500+CG 500)	2.0	1.0	-	-
Equilibration 40 ps, NVT, 0 to 300 K	10.0	10.0	10.0	-
Equilibration 20 ps, NPT, 300 K	10.0	10.0	10.0	-
Equilibration 20 ps, NPT, 300 K	10.0	10.0	9.0	-
Equilibration 20 ps, NPT, 300 K	5.0	4.0	3.0	-
Equilibration 20 ps, NPT, 300 K	2.0	1.0	-	-
Equilibration 20 ps, NPT, 300 K	1.0	-	-	-
Production run 2 ns	-	-	-	-

5.4.2 The loop equilibration (Protocol B2)

The simulations started with minimization using Steepest Descents for 500 steps followed by Conjugated Gradients for 500 steps of both solvent water and counter ions whilst DNA and thiazotropsin A were frozen with the ibelly option.

A non-bonded cutoff of 9 Å was used. The first 50 ps of equilibration employed the same restraints as the minimization. A simulation time step of 2 fs was used whilst the SHAKE algorithm was applied to constrain bonds involving hydrogen atoms. A periodic boundary box of constant pressure was applied and the temperature was kept constant at 300 K. Pressure was regulated by setting molecular dynamics with isotropic position scaling. The system was maintained at 1 atm controlled by a pressure relaxation time of 1 ps. A time constant of 1 ps was used for heat bath coupling, with the non-bonded list updated every 10 steps. The Particle mesh ewald (PME) method was used with the 4th order of B-spline interpolation the same as the cubic spline approximation. The direct sum tolerance was 10^{-6} .

The loop restarted with a minimization using Steepest Descents for 500 steps followed by Conjugated Gradients for 500 steps. This time, a force of 25 kcal/mol was applied to restrain only solvent water molecules whilst others were free. A non-bonded cutoff of 9 Å was used. Next, an equilibration was carried out for 20 ps with the same parameters as the last equilibration, but with a restraint force the same as the last minimization, and the temperature kept constant at 300 K.

The second loop started with five rounds of minimization using Steepest Descents for 500 steps followed by Conjugated Gradients for 500 steps. A restraint force of 20 kcal/mol was applied to DNA and thiazotropsin A in the first round of minimization. The restraint force was then reduced by 5 kcal/mol in each subsequent round. A non-bonded cutoff of 9 Å was used. Finally, the last round of minimization was performed without any restraint on solute and solvent. Next, equilibration was carried out for 20 ps with the same parameters as the previous equilibration but with all restraints on solute removed. The system temperature was controlled to rise from 100 K to 300 K in a few picoseconds. Finally, the production run was performed by a straightforward restart from the last equilibration with a constant temperature of 300 K for 2 ns. The simulation system type 4 with loop equilibration will be called protocol B2 later in this thesis.

Table 18 The stepwise to apply the restraint force on DNA and thiazotropin A in the simulations system type 4, the loop equilibration.

Steps	DNA	Thiazotropin A	Ions	Solvent
Min. Solvent 1000 steps (SD 500+CG 500)	Fixed	Fixed	-	-
Equilibration 50 ps, NPT, 300 K	Fixed	Fixed	-	-
Min. 1000 steps (SD 500+CG 500)	-	-	-	25.0
Equilibration 20 ps, NPT, 300 K	-	-	-	25.0
Min. 1000 steps (SD 500+CG 500) x 5 rounds	20.0	20.0	-	-
Min. 1000 steps (SD 500+CG 500)	0.0	0.0	-	-
Equilibration 20 ps, NPT, 100 K to 300 K	-	-	-	-
Production run 2 ns, 300 K	-	-	-	-

5.5 The simulations system type 5

The simulations system type 5 refers to the equilibration protocol used in the AMBER drug/DNA tutorial. Force field parameters for thiazotropin A were created differently from all former parameters. The structural parameters of thiazotropin A were mapped from the optimization at the B3LYP/6-31G** level using the Gaussian03 program. The charge parameters of thiazotropin A were generated using the AM1-BCC charge method in the antechamber module. The plain force field for thiazotropin A was then used later in the simulations without any modifications. The solvated system was set up using an octahedral box extended by 10 Å from the solute surface.

The simulations started with minimization using Steepest Descents for 250 steps followed by Conjugated Gradients for 750 steps on solvent water whilst a restraint force of 500.0 kcal/mol was placed on DNA and thiazotropin A. A non-bonded cutoff of 10 Å was used with a periodic boundary condition of constant

volume applied. Next, the whole system was minimized for 1500 steps (Steepest Descents 500 steps followed by Conjugated Gradients 1000 steps) using the same parameters as previously, but with all restraint forces removed. Next, the equilibration was carried out for 20 ps in a constant volume box. A non-bonded cutoff of 10 Å was set. A simulation time step of 2 fs was used whilst the SHAKE algorithm was applied to constrain bonds involving hydrogen atoms. The temperature was raised from 0 to 300 K in a few picosecond. The dynamics parameters were set to canonical ensemble with constant temperature using the Leapfrog integrator, a collision frequency of 1 ps^{-1} , and a time constant for heat bath coupling of 1 ps. DNA and thiazotropsin A were restrained with a force of 10.0 kcal/mol. Finally, the production run was performed for 1 ns by restarting from the last equilibration step. The system was then switched to a constant pressure box, maintained at 1 atm controlled by a pressure relaxation time of 2 ps. The temperature of the system was kept constant at 300 K. The rest of the parameters were the same as the equilibration step but with all restraints removed.

Table 19 The stepwise to apply the restraint force on DNA and thiazotropsin A in the simulations system type 5.

Steps	DNA	Thiazotropsin A	Ions	Solvent
Min. Solvent 1000 steps (SD 250+CG 750)	500	500	-	-
Min. All 1500 steps (SD 500+CG 1000)	-	-	-	-
Equilibration 20 ps, NVT, 0 to 300 K	10.0	10.0	-	-
Production run 1 ns, NPT, 300 K	-	-	-	-

Summary of combination systems

Table 20 Summary of combination systems in the MD1, B3LYP/6-31G** and force field version 1 for thiazotropsin A, performed in the simulations (✓).

Starting from optimized structure			
System	FF03	FF02	FF02EP
1. Invacuo, restrain 1.0 kcal for DNA	✓	-	✓
2. Invacuo and counter ions, restrain 1.0 kcal for DNA	✓	-	✓
3. Solvation and counter ions, restrain 1.0 kcal for DNA	✓	-	✓
4. Solvation and counter ions, free			
4.1 Straight equilibration	✓	-	-
4.2 Loop equilibration	-	-	-
Starting from NMR structure			
System	FF03	FF02	FF02EP
1. Invacuo, restrain 1.0 kcal for DNA	✓	-	✓
2. Invacuo and counter ions, restrain 1.0 kcal for DNA	✓	-	✓
3. Solvation and counter ions, restrain 1.0 kcal for DNA	✓	✓	-
4. Solvation and counter ions, free			
4.1 Straight equilibration	✓	-	-
4.2 Loop equilibration	-	-	-

Table 21 Summary of combination systems in the MD2, B3LYP/6-31G** and force field version 2 for thiazotropsin A, performed in the simulations (✓).

Starting from NMR structure			
System	FF03	FF02	FF02EP
1. Invacuo, restrain 1.0 kcal for DNA	✓	-	-
2. Invacuo and counter ions, restrain 1.0 kcal for DNA	✓	-	-
3. Solvation and counter ions, restrain 1.0 kcal for DNA	✓	✓	-
4. Solvation and counter ions, free			
4.1 Straight equilibration	✓	-	-
4.2 Loop equilibration	-	-	-

Table 22 Summary of combination systems in the MD3, HF/6-31G* and force field version 1 for thiazotropsin A, performed in the simulations (✓).

Starting from NMR structure			
System	FF03	FF02	FF02EP
1. Invacuo, restrain 1.0 kcal for DNA	-	-	-
2. Invacuo and counter ions, restrain 1.0 kcal for DNA	-	-	-
3. Solvation and counter ions, restrain 1.0 kcal for DNA	✓	✓	-
4. Solvation and counter ions, free			
4.1 Straight equilibration	✓	-	-
4.2 Loop equilibration	-	-	-

Table 23 Summary of combination systems in the MD4, HF/6-31G* and force field version 2 for thiazotropsin A, performed in the simulations (✓).

Starting from NMR structure			
System	FF03	FF02	FF02EP
1. Invacuo, restrain 1.0 kcal for DNA	-	-	-
2. Invacuo and counter ions, restrain 1.0 kcal for DNA		-	-
3. Solvation and counter ions, restrain 1.0 kcal for DNA	✓	✓	-
4. Solvation and counter ions, free			
4.1 Straight equilibration	-	-	-
4.2 Loop equilibration	✓	✓	-

Table 24 Summary of the performed simulations systems 5 (✓).

Starting from NMR structure			
System	FF03	FF02	FF02EP
5. AMBER drug/DNA tutorial equilibration	✓	-	-

6. Computational Analysis of the DNA/Thiazotropsin A Complex

6.1 Calculations of the Binding Energy from the MM-PBSA Approach

Coordinates of the system were collected every 1 ps and saved into a trajectory file. Snapshots were taken from the production molecular dynamics run file by sampling the last 2000 coordinates every 20 ps. Counter ions and solvent water were removed from the snapshot structures of the solvated systems. Then 100 snapshot structures were used to calculate and observe the binding energy of the system.

The new PBSA module in AMBER8 was used to calculate MM-PBSA energy through the mm_pbsa perl script. All parameters were set up as following:

$$\Delta G_{\text{binding}} = (\Delta H_{\text{gas}} + \Delta \Delta G_{\text{PB}} + \Delta \Delta G_{\text{SA}}) - T\Delta S \quad (25)$$

$$\Delta G_{\text{binding}} = \text{PBTOT} - \text{TSTOT} \quad (26)$$

Binding energy was calculated using equations (25) or (26). The from PBSA calculation obtains a PBTOT (total PBSA energy) value calculated from the molecular mechanics energy difference between the complex and reactants (DNA and thiazotropsin A) in the gas phase (ΔH_{gas}), and the solvation free energy of DNA, thiazotropsin A and complex calculated by the PBSA method ($\Delta \Delta G_{\text{PB}}$, $\Delta \Delta G_{\text{SA}}$).

To complete the exact binding energy calculations, the entropy energy distribution was calculated by normal mode analysis using the NMode subprogram in AMBER. All parameters were set up as follows; a minimization for 5000 steps to an rms convergence of 0.5; results from normal mode calculation could be obtained as the TSTOT (total entropy energy distribution) value. The total entropic energy included translational, rotational and vibrational parameters which were different between complex and reactants (DNA and thiazotropsin A) at a reference temperature

of 298.15 K. The binding free energy of the system was determined by equation (21) by combining the PBTOT and TSTOT energies.

6.2 The RMSd Fluctuation and H-Bond Analysis

The rmsd fluctuation of the complex system, DNA and thiazotropsin A were analysed with the ptraj subprogram in AMBER. Only heteroatoms were considered with the first set of geometry used as a reference. H-bond analysis was also performed with the ptraj subprogram. The hydrogen donor and receptor atoms were selected by reference to the work of Anthony et.al. (Anthony *et al.*, 2004).

6.3 Interaction Energy Analysis

A set of the last 100 snapshot structures were sampled from each trajectory. The interaction energies were calculated as the decomposition on a pairwise per-residue. The interactions between each nucleic acid and thiazotropsin monomer were calculated. The 1-4 interactions were added to either electrostatic or vdW contributions. The newer Generalized Born model developed by A. Onufriev, D. Bashford and D.A. Case (GB^{OBC}) (Onufriev *et al.*, 2004) were implemented during the interaction energies calculations. The surface areas were computed by recursively approximating a sphere around an atom, starting from an icosahedra.

RESULTS AND DISCUSSIONS

1. Comparison of the Applied Force Fields on Thiazotropsin A

The structure of thiazotropsin was minimized using a molecular mechanics method in the SANDER subprogram. The optimized structures with force field versions 1 and 2 were aligned with a structure obtained from a B3LYP/6-31G** optimization with Gaussian. Alignment was performed by superimposition using InsightII. The Root mean square deviation (RMSD) values for force field versions 1 and 2 are 0.68 and 0.35 Å, respectively.

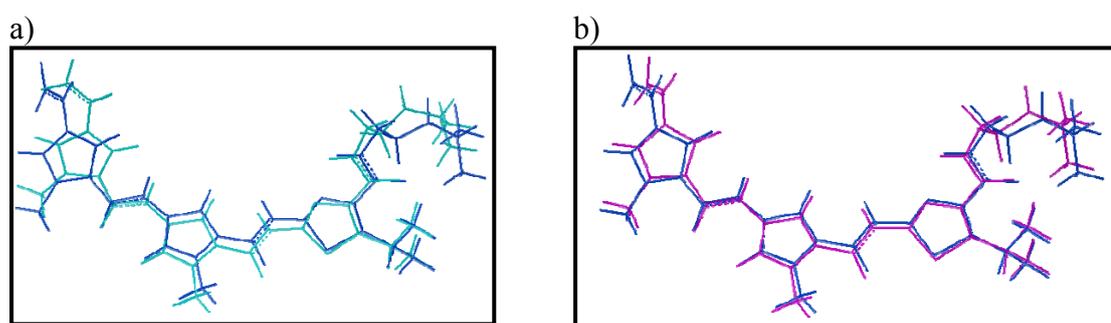


Figure 22 Superimposition of thiazotropsin a) B3LYP/6-31G** Optimized (dark blue) and minimized from force field version 1 (light blue). b) B3LYP/6-31G** Optimized (dark blue) and minimized from force field version 2 (purple).

Both molecular mechanics (MM) optimized structures are very close to the density function theory (DFT) optimized structure, showing that both force fields can be used for thiazotropsin in simulations. Although the optimized structure from force field version 2 is closer to the DFT optimized structure, a high force constant is required to maintain the first peptide torsion angle, making it less amenable for simulations.

2. The Energy Equilibration of the DNA/Thiazotropsin A Complexes

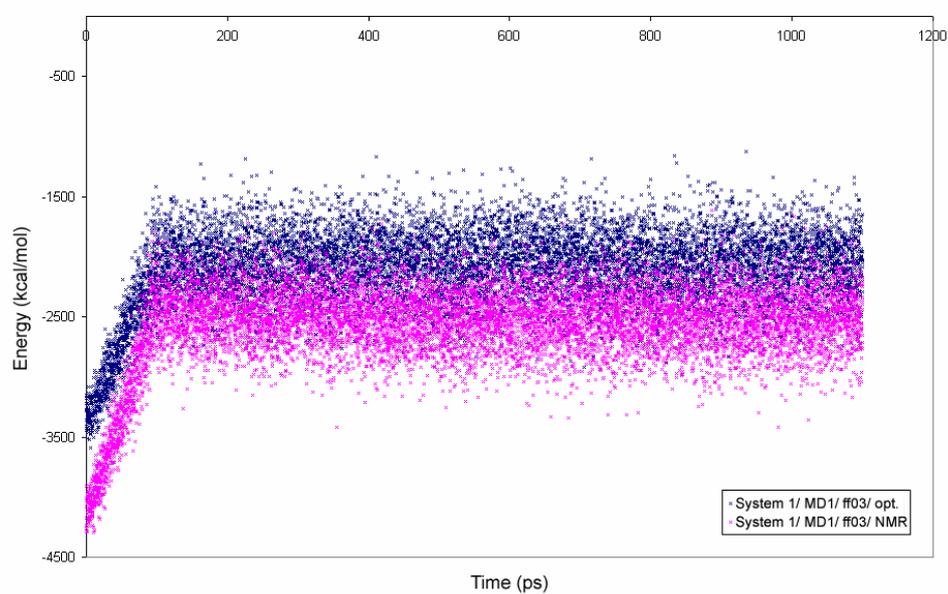
The thiazotropsinA/ d(5'-CGCACTAGTGCG-3')₂ complex system was studied under various conditions as shown in table 25. Moreover, the simulations were also performed using different sources of the starting coordinates, either from the optimized or NMR resolved structures, and with different types of force field, either non-polarizable (FF03), polarizable (FF02) or polarizable with extra-point charges (FF02EP).

In all cases, the energies of the system were well equilibrated but with different levels of energy. To show the total energy plot from all cases would be difficult to see and analyse, so only selected combinations are shown in figure 23 to figure 27. Firstly, comparison between the optimized and NMR structures in the *invacuo* simulations (system 1) reveals that both simulations have a similar pattern in the total energy plot. However, the simulations that used NMR as a starting structure equilibrated at lower energy of around -2500 kcal/mol, whereas the simulations that used the optimized structure to start equilibrated at about -2000 kcal/mol as shown in figure 23. Next, the comparison between applying the non-polarizable (ff03) and polarizable (ff02) force fields to DNA in the solvated simulations (system 3) reveals that both simulations also have a similar profile in the total energy plot. However, the simulations that used the ff02 force field equilibrated at a lower energy of around -37600 kcal/mol, whereas the simulations that used the ff03 force field equilibrated at about -37000 kcal/mol (figure 24). Moreover, it is clearly noticeable that the simulations including water molecules (such as system 3) provide a lot of stabilizing energies to the solute, lowering the energy by 14 times more than the *invacuo* simulations such as in system 1. Thus, the total energy of approximately -2500 kcal/mol for *invacuo* simulations is significantly lowered to about -37000 kcal/mol for the solvated systems. Clear comparisons of the various system types are shown in figure 25. The total energies from five simulations (system 1, 2, 3, 4(B1) and 4(B2)) can be obviously separated into two groups; the *invacuo* systems (1 and 2) and the solvated systems (3, 4(B1) and 4(B2)). Energy hot spots were observed in some solvation simulations such as system 4(B1) due to the reduction of restraint forces

during the simulations with constant pressure conditions. Nevertheless, the energy came back to the equilibration state after a few steps of MD. In the higher resolution plots shown in figure 25b and 25c, the invacuo simulations have different levels of the total energy while the patterns of energy fluctuation are similar. The total energy from the invacuo including counter ions (system 2) simulations equilibrated at a lower energy of about -5200 kcal/mol while the pure invacuo simulations (system 1) equilibrated at about -2300 kcal/mol. The total energy from all the solvation simulations (system 3, 4(B1) and 4(B2)) equilibrated at the same level; about -37100 kcal/mol. Finally, the total energies of the various topology parameters (MD1, MD2, MD3 and MD4) were compared by using an example case of the system 4 (B1) as shown in figure 26. Although all the cases equilibrated well at the same level of energy, the total energy paths before stabilization are different. The topology and parameters sets of thiazotropsin A derived from B3LYP/6-31G** level of theory (MD1 and MD2) caused two hot spot energy during the equilibration process while the usage of HF/6-31G* level of theory (MD3 and MD4) performed smoothly during the equilibration process as shown the plot overview in figure 26a. Moreover, when analyzed more closely at the higher resolution plot shown in figure 26b, the group of HF/6-31G* parameters for thiazotropsin A equilibrated at a energy around -37400 kcal/mol, which was slightly lower than the group of B3LYP/6-31G** parameters for thiazotropsin A that equilibrated around -37100 kcal/mol. Furthermore, the total energy fluctuation of system 5 which used the AMBER tutorial equilibration protocol stabilized at a lower level of energy around -48000 kcal/mol. This was due to the total number of water molecules in system 5 (5316 molecules), which was more than in the system 3 and system 4 (3879 molecules).

Table 25 The combination systems type sand model types.

	System 1	System 2	System 4		
			Protocol A	Protocol B1	Protocol B2
MD1	✓	✓	✓	✓	✓
MD2	✓	✓	✓	✓	✓
MD3	✓	✓	✓	✓	✓
MD4	✓	✓	✓	✓	✓

**Figure 23** The total energy from opt. and NMR starting.

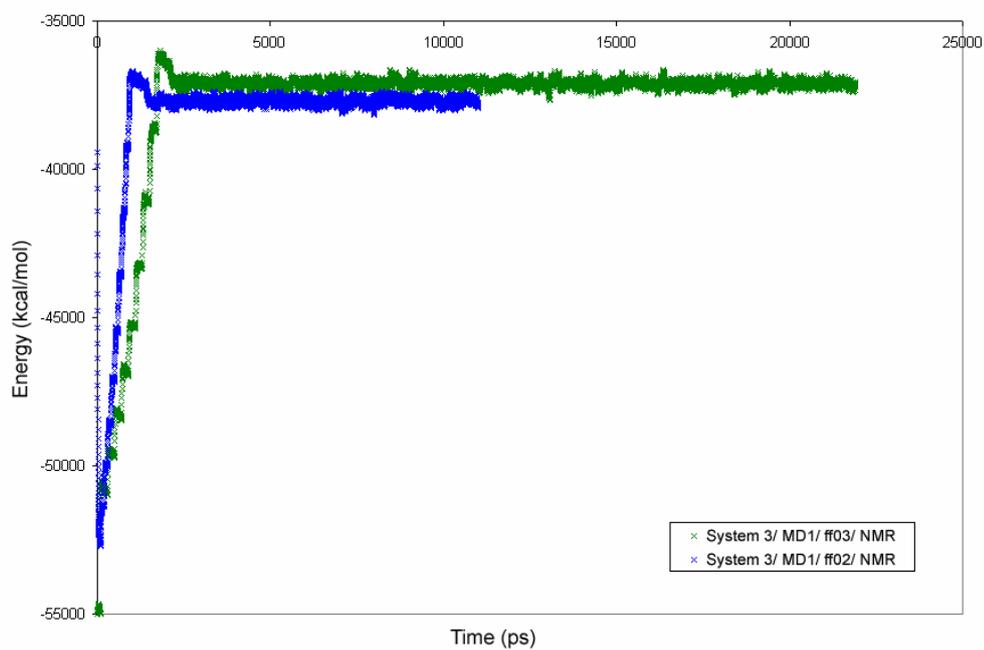


Figure 24 The total energy from ff03 and ff02 force field.

a)

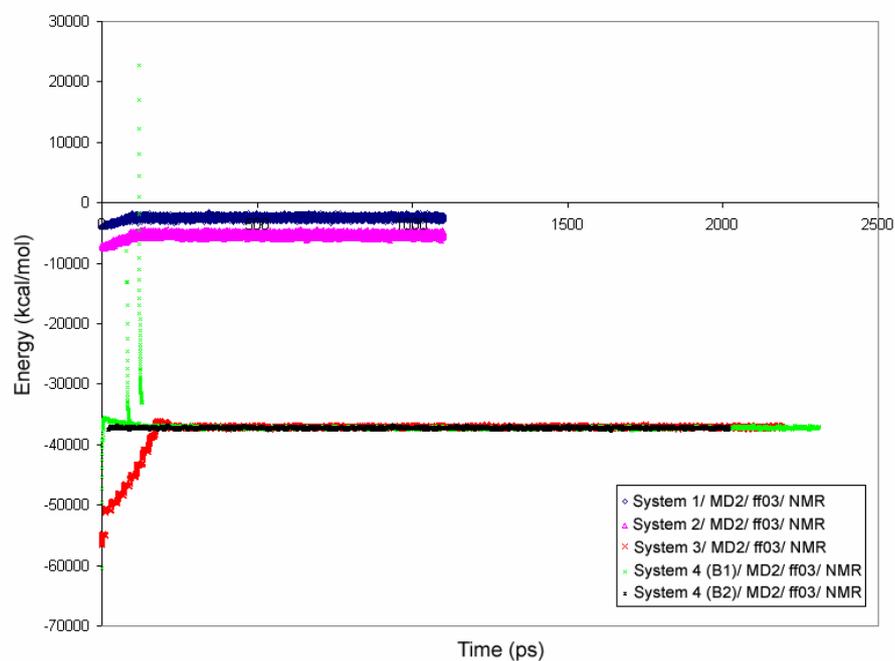
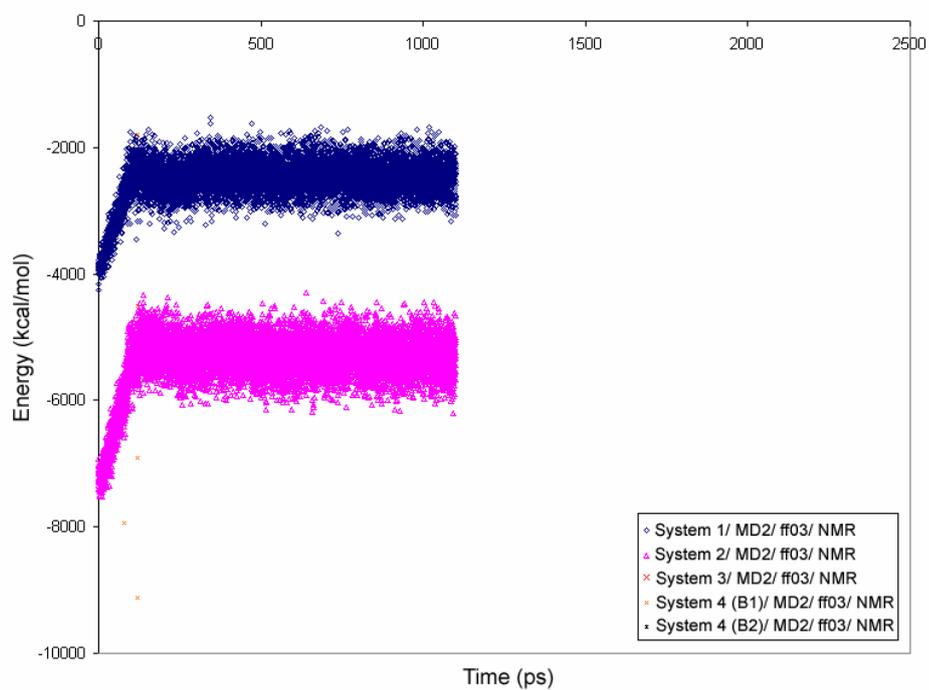


Figure 25 The total energy from various systems; a) the overview, b) focusing only in the invacuo simulations group and c) focusing only in the solvation simulations group.

b)



c)

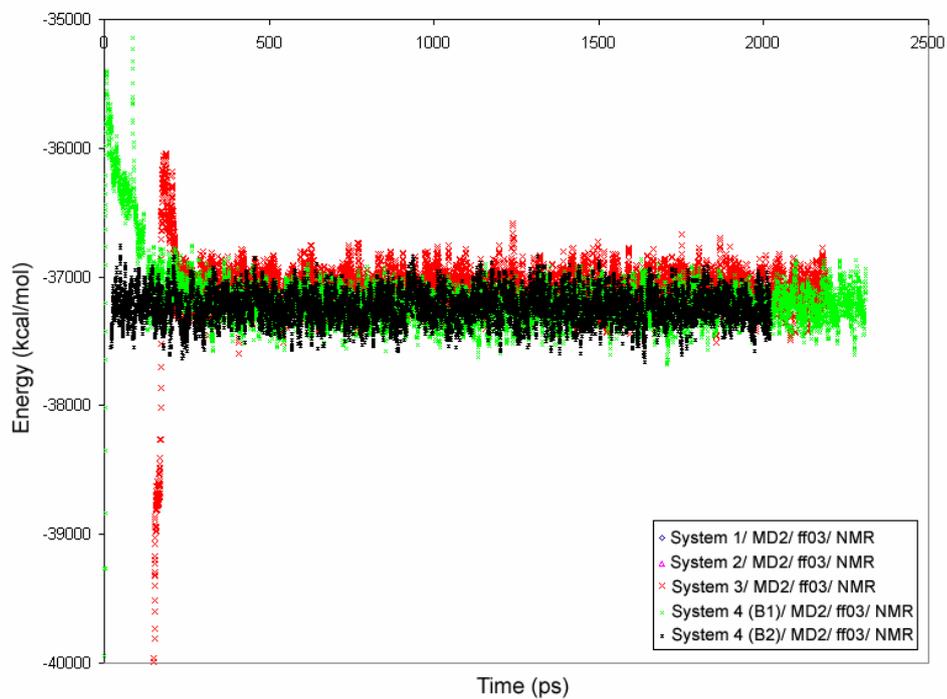
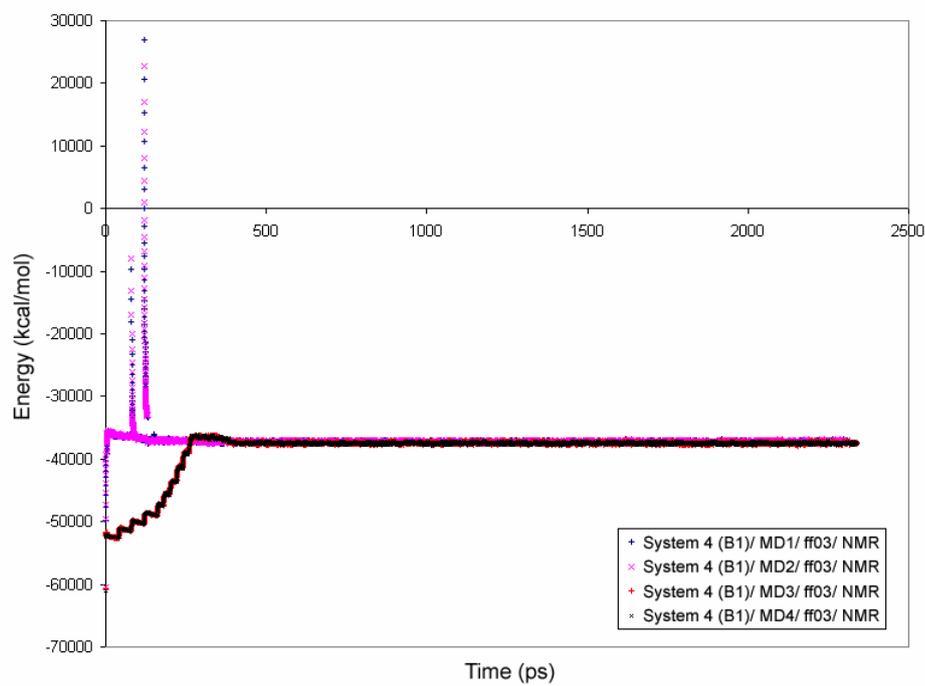


Figure 25 (Continued).

a)



b)

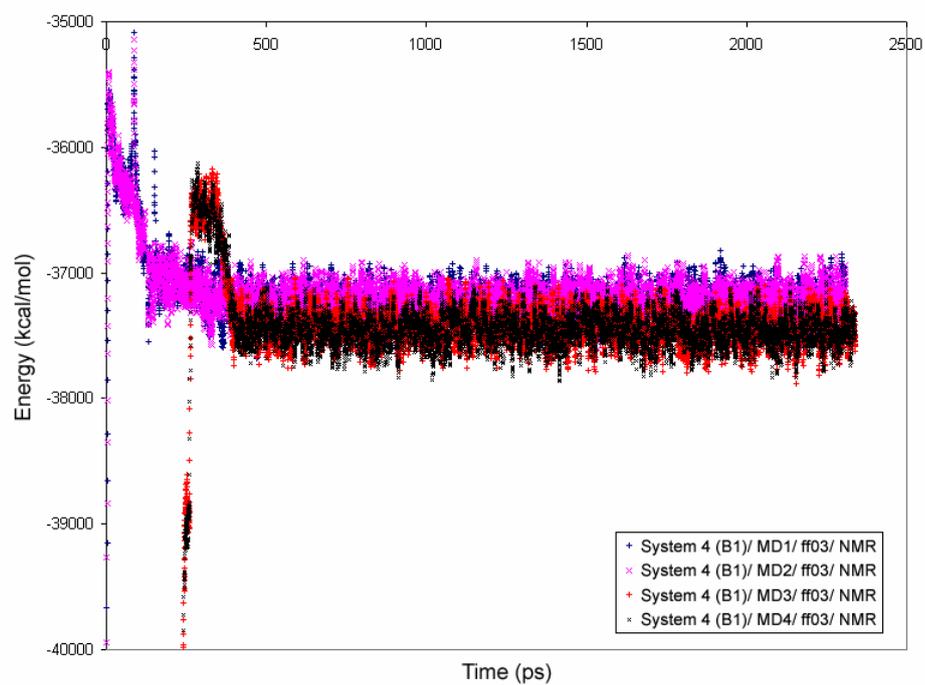


Figure 26 The total energy from various topology parameters; a) the overview and b) at higher resolutions.

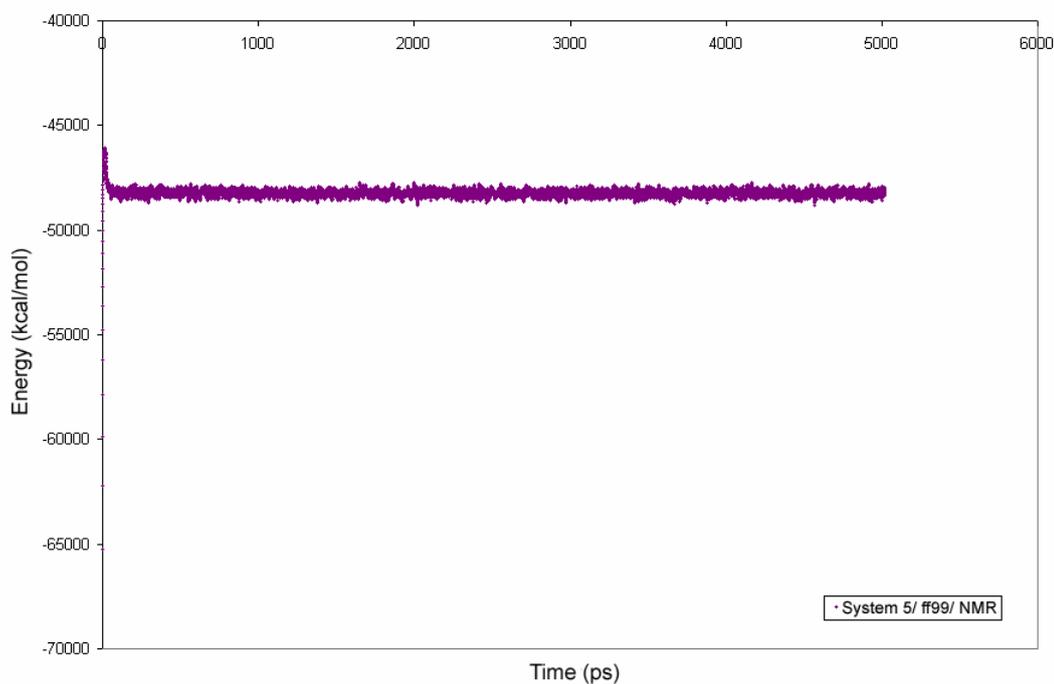


Figure 27 The total energy from simulations system 5.

3. Structural Equilibration of the DNA/Thiazotropsin A Complexes

Structural equilibration was observed by measuring RMSD positional fluctuations of the solute (DNA and thiazotropsin A). RMSD fluctuations using various combinations were analysed following the same fashion as the energy equilibration analysis with only selected combinations shown. All RMSD calculations were analyzed by excluding the terminal base pairs of DNA. Although the energy fluctuations were well equilibrated for all cases, the RMSD fluctuations were different as shown in figures 28 to 34. Systems with restraints on DNA were well equilibrated, whilst the systems using free DNA and thiazotropsin A seemed to have wider fluctuations during the equilibration and productions process.

First, the difference between using the optimized and NMR structures in the solvation simulations (system 4 (B1)) reveals that the RMSD of the complex using NMR as a starting structure was well equilibrated around 2.3 Å whilst the other shows

higher RMSD fluctuations (figure 28a). The RMSD fluctuation in DNA and thiazotropsin are shown in more detail in figures 28b and 28c. Thiazotropsin A in both cases were well equilibrated but at a high level of RMSD from the starting structure (figure 28c), which indicated that the movement of thiazotropsin A from the optimized form at the beginning caused unstability to the DNA.

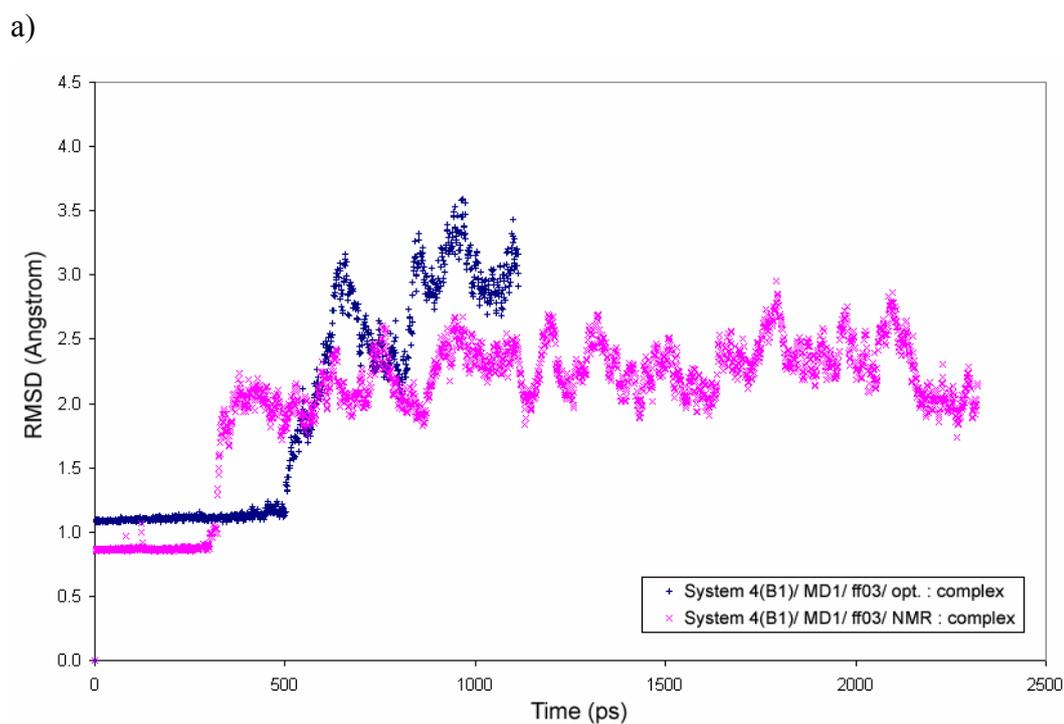


Figure 28 The RMSD fluctuation in simulations system 4 (B1) with the optimized and NMR starting structure. a) complex, b) DNA and c) thiazotropsin A.

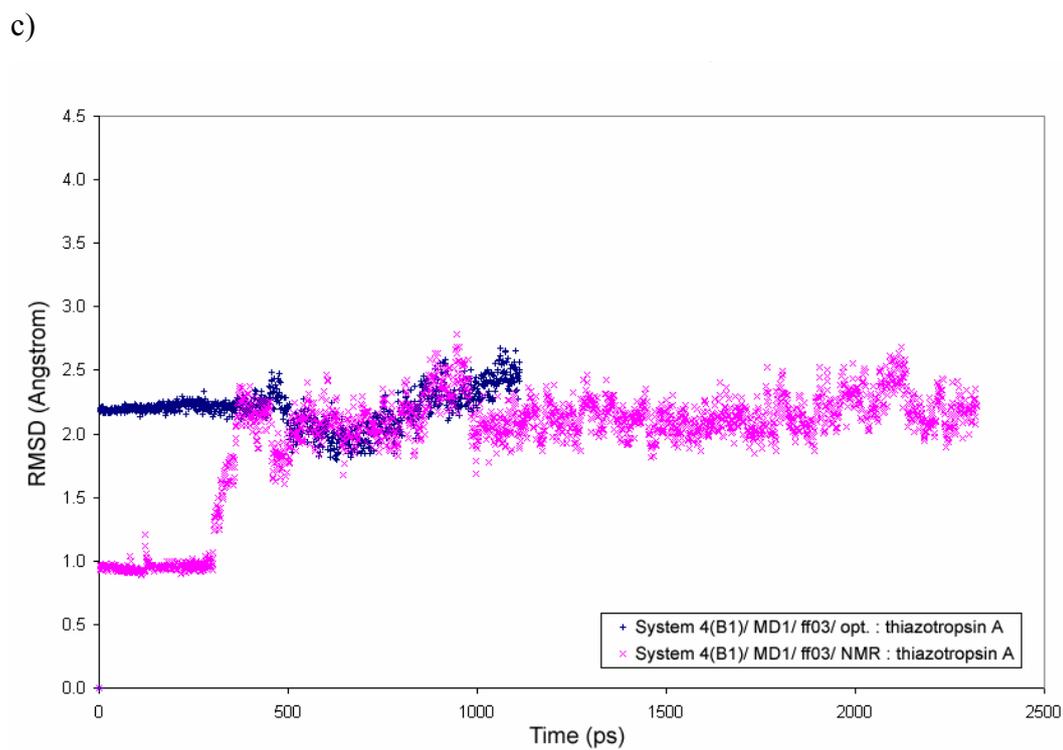
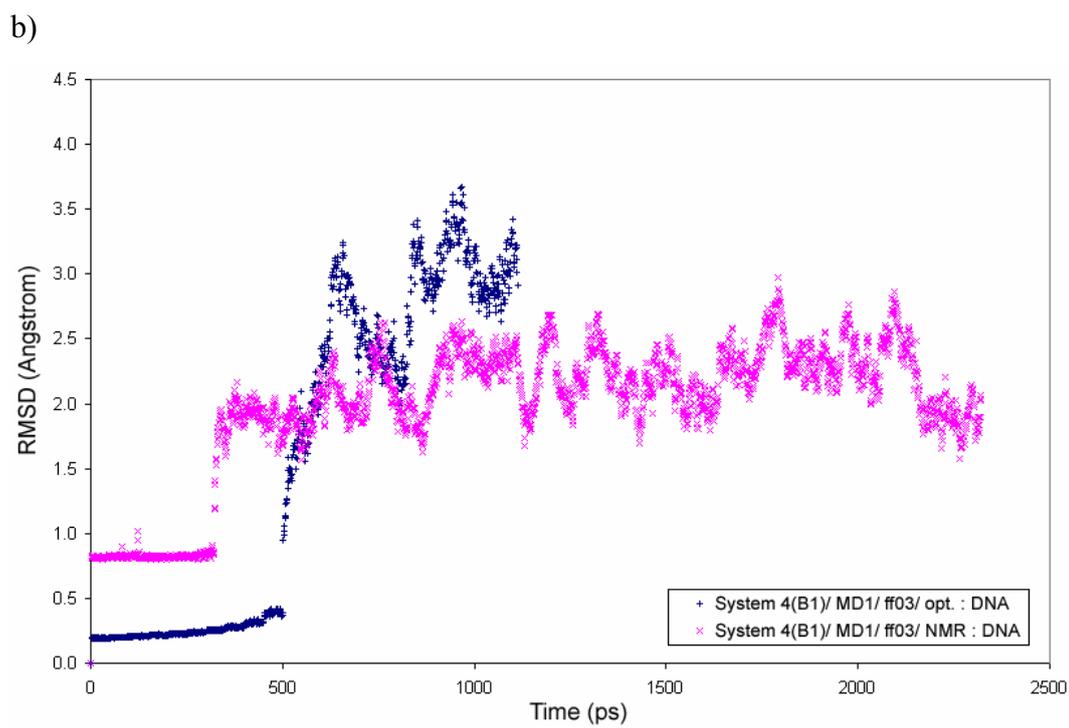


Figure 28 (Continued).

Next, comparison of the RMSD between applying the non-polarizable (ff03) and polarizable (ff02) force field to DNA in the solvation and restrained DNA simulations (system 3) reveals that the RMSD fluctuations of the complex, DNA and thiazotropsin A were well equilibrated as shown in figures 29a, 29b, and 29c. An interesting result was seen for the RMSD fluctuations of DNA (figure 29b); the RMSD fluctuation when applying the ff03 force field yielded the lower RMSD, whereas the total energy was higher than the ff02 force field in the simulations. Moreover, the restrained DNA simulation directly affected the DNA, causing the RMSD of the complex and DNA to fluctuate at a lower level around 1.0 Å, despite thiazotropsin A fluctuating at a similar level of around 2.2 Å, as shown by figures 29c and 28c.

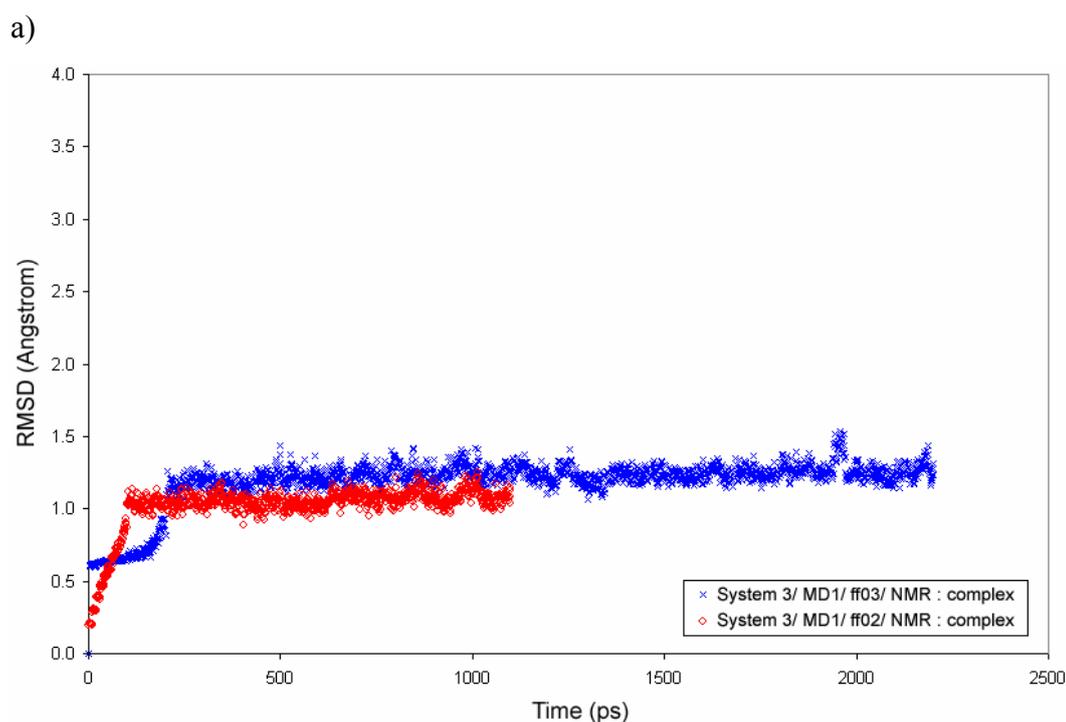
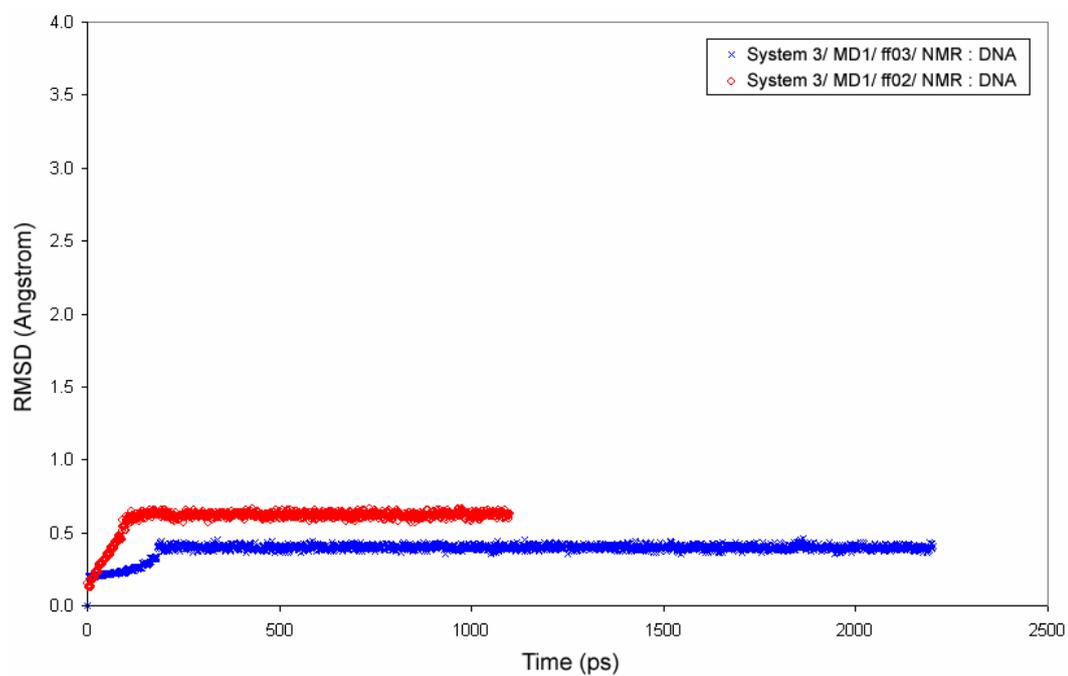
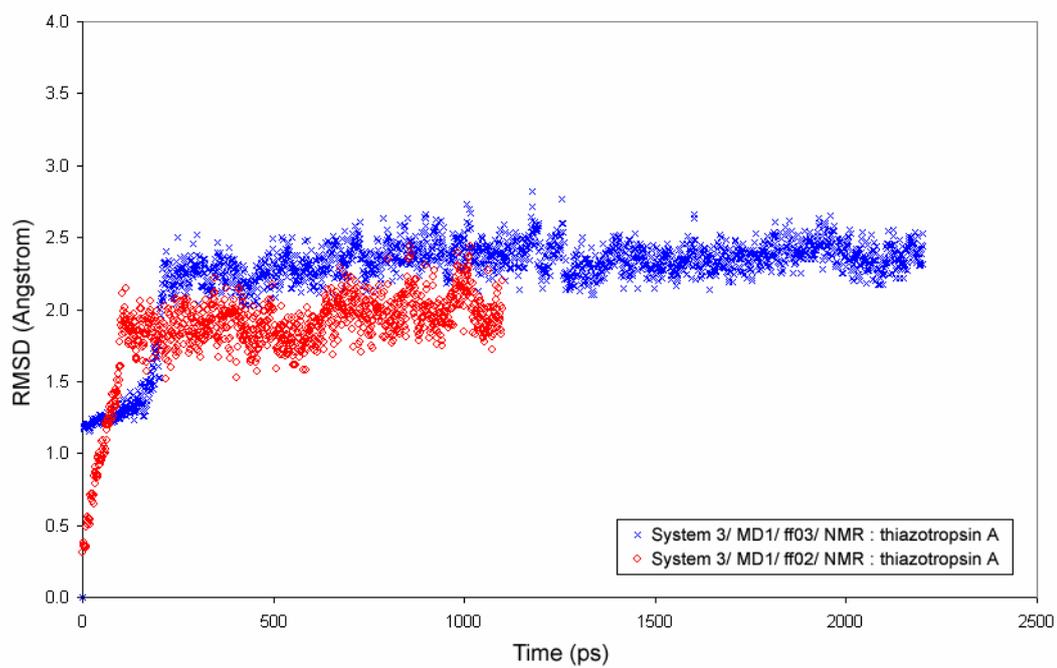


Figure 29 The RMSD fluctuation in simulations system 3 (the restraint simulations) with the ff03 and ff02 force field for DNA. a) complex, b) DNA and c) thiazotropsin A.

b)



c)

**Figure 29** (Continued).

Next, an RMSD comparison between the non-polarizable (ff03) and polarizable (ff02) force field was investigated under solvated conditions but without restraints on DNA (system 4 (B1)) as shown in figure 30. Firstly, the RMSD fluctuation pattern of the complex and DNA are the same, which is different from the previous results, yet were well stabilized for both simulations. The RMSD fluctuations of the complex and DNA from the ff03 force field had lower fluctuations at around 2.3 Å than the ff02 force field, which was fluctuating at around 2.8 Å (figures 30a and 30b). When the RMSD fluctuation of thiazotropsin A was determined (figure 30c), it also stabilized at a different level; ff03 force field was well stabilized at around 2.1 Å, whilst the ff02 was stabilized at around 2.7 Å and showed wider fluctuation. This implied that thiazotropsin A inside the DNA when using the ff02 force field had to change its conformations and adapt itself more than when using ff03 force field conditions.

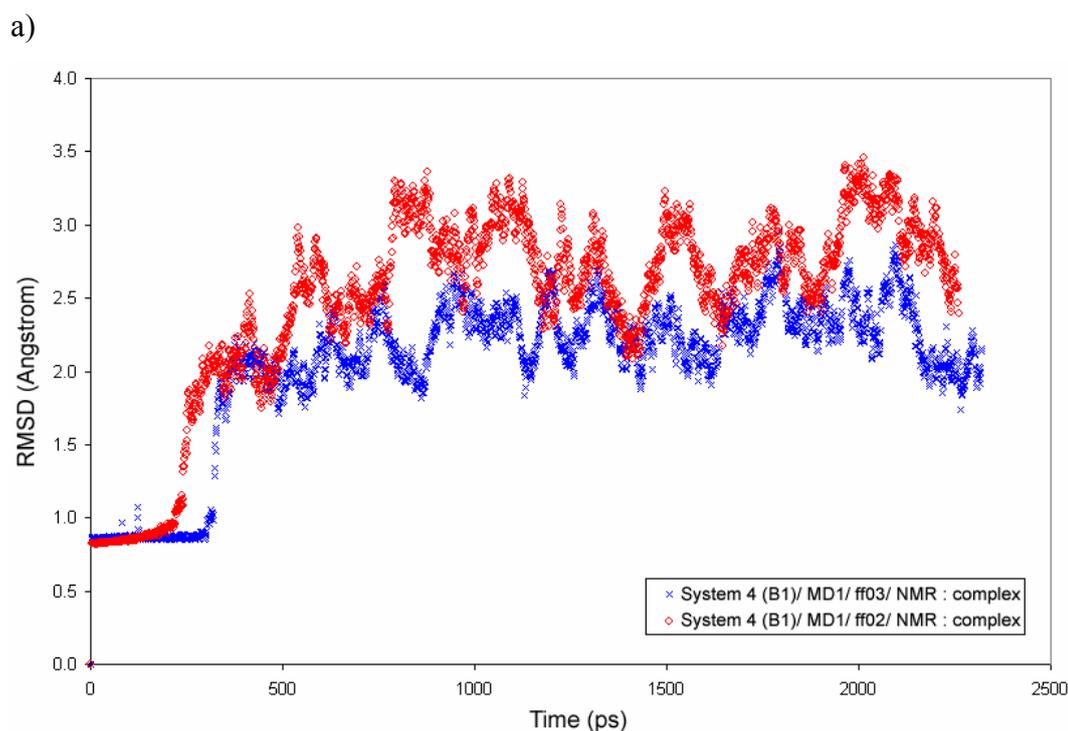
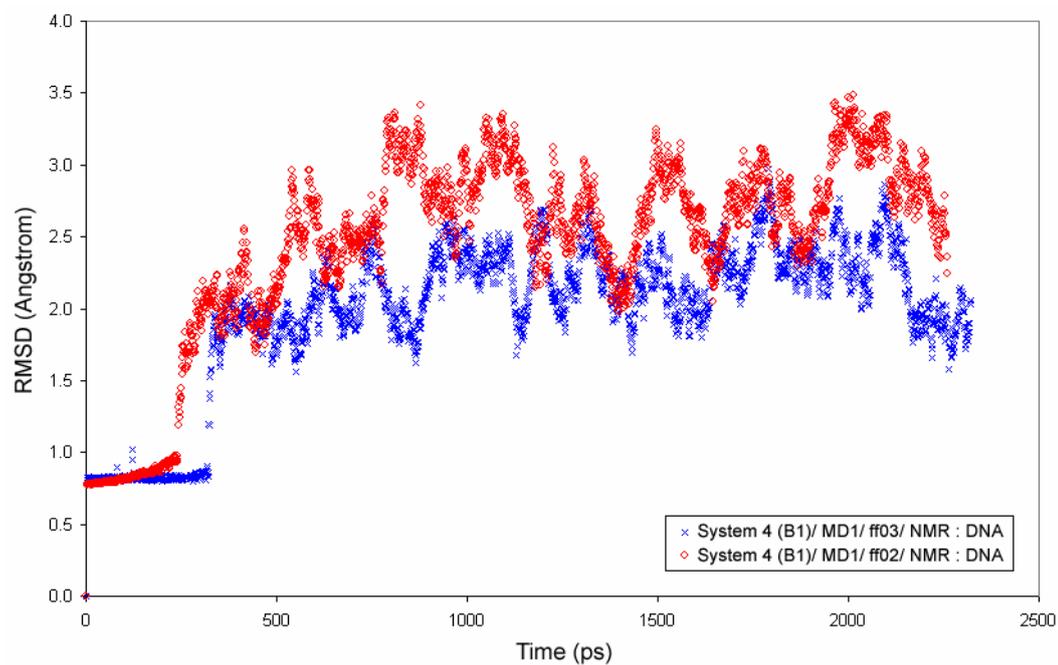


Figure 30 The RMSD fluctuation in simulations system 4 (B1) (the free simulations) with the ff03 and ff02 force field for DNA. a) complex, b) DNA and c) thiazotropsin A.

b)



c)

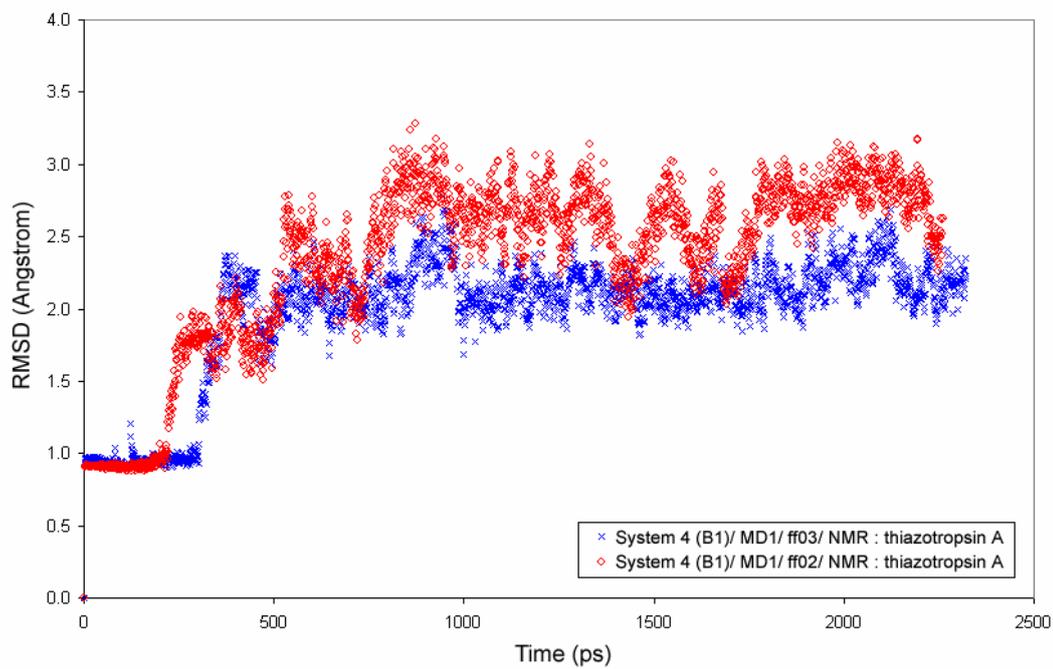


Figure 30 (Continued).

Next, we analyzed the RMSD fluctuation from various topologies and parameters of thiazotropsin A (MD1, MD2, MD3 and MD4) obtained from the simulations system 3 (restraint on DNA) with ff03 force field as shown in figure 31. The RMSD fluctuations of the complex from MD1, MD2 and MD3 stabilized in the same pattern, whereas the complex from MD4 showed a different profile; the RMSD fluctuation was stable until 920 ps, then it increased over 90 ps and stabilized again at a higher RMSD of around 1.8 Å (figure 31a). This change from one state to another indicated movement of the complex primarily from rotational motion. Moreover, this conformational change appeared to be an irreversible process. Details of the conformational change in the complex structures were investigated further by individually assessing the RMSD fluctuations of DNA and thiazotropsin A. The RMSD fluctuations of DNA in all cases were well equilibrated and stabilized at very low levels of RMSD as shown in figure 31b. What clearly evident is that the conformational change in the complex is caused by the movement of thiazotropsin A (figure 31c). The RMSD fluctuations of thiazotropsin A using MD1, MD2 and MD3 were stabilized around 2.35 Å with the same pattern of fluctuation, whilst the RMSD fluctuation of thiazotropsin A using MD4 parameters had a different pattern from the start and revealed a second conformational change at 1310 ps. However, the conformational change of thiazotropsin A driven by the MD4 parameter set did not immediately switch. There was a transition period starting from 920 to 1310 ps then it remained in two states of conformational equilibria.

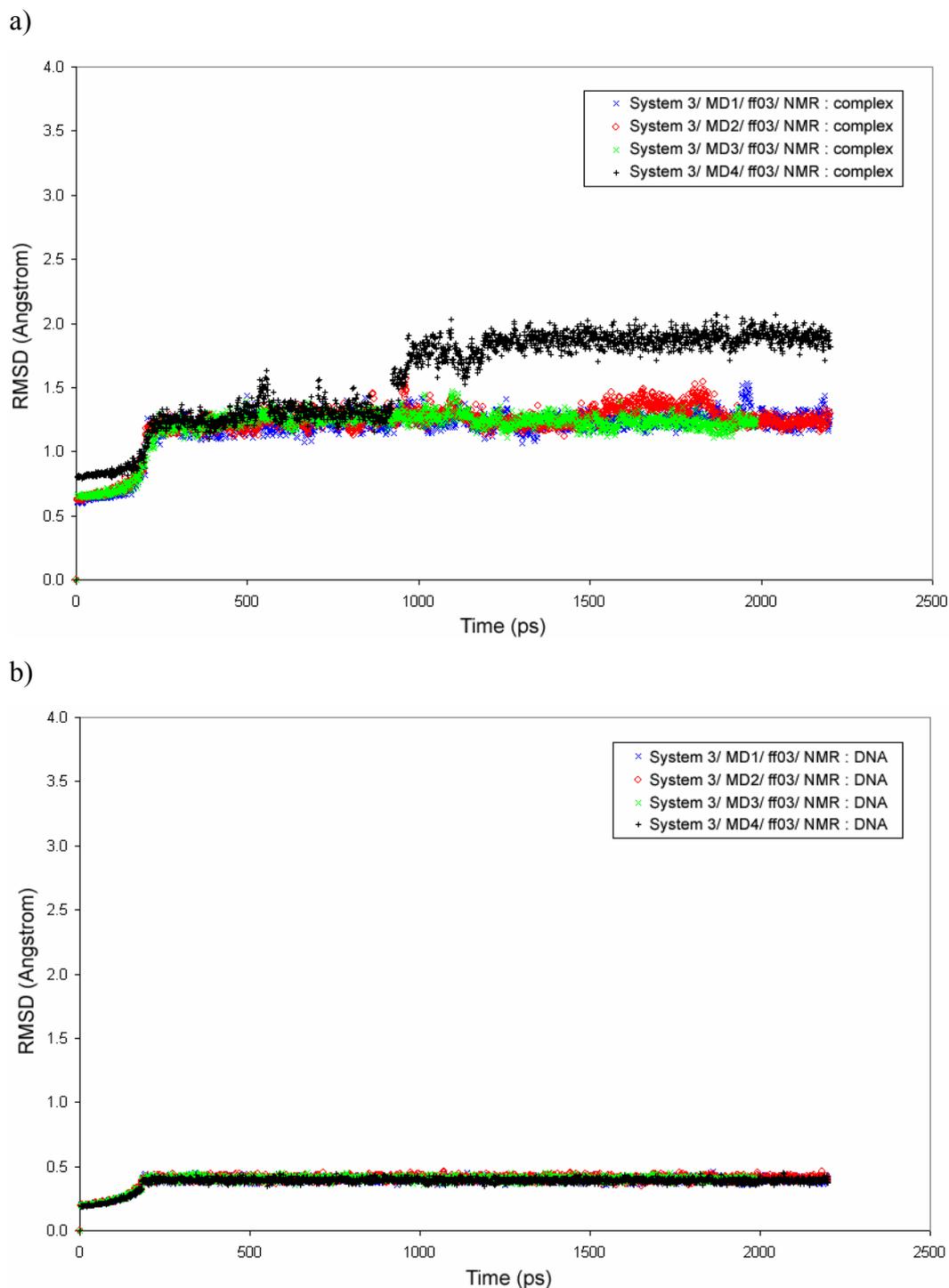


Figure 31 The RMSD fluctuation from simulations system 3 (the restraint simulations) with various topology parameters (MD1, MD2, MD3 and MD4) usage for thiazotropsin A. a) complex, b) DNA and c) thiazotropsin A.

c)

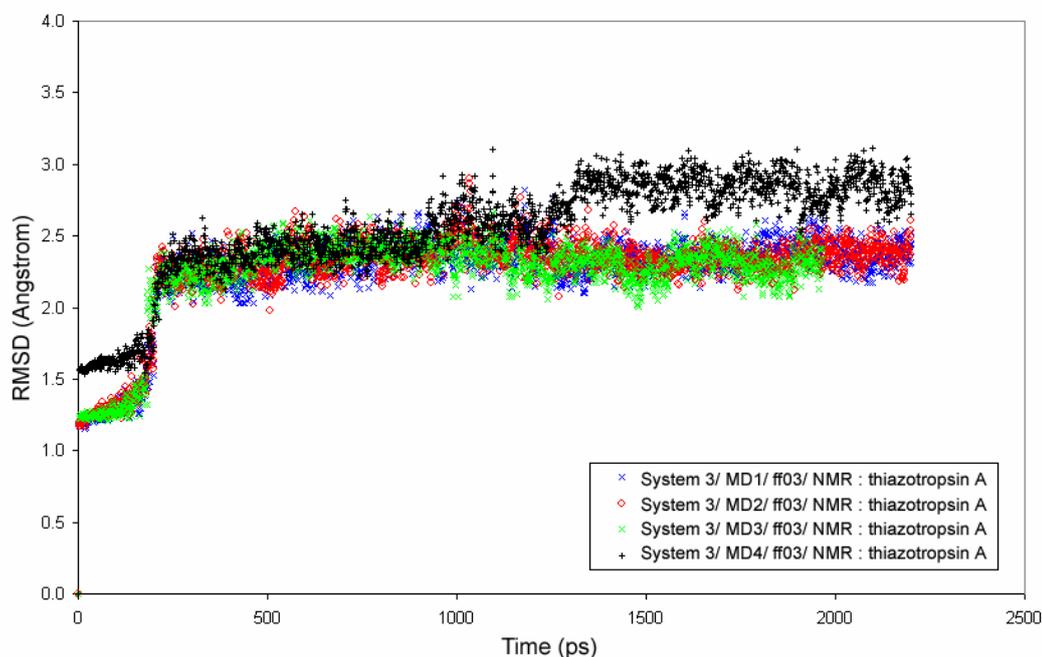


Figure 31 (Continued).

Next, the same comparison of RMSD profiles for topology parameters (MD1, MD2, MD3 and MD4) applied to thiazotropsin A was investigated, but with solvated systems in the absence restraints on DNA (system 4 (B1)) as shown in figure 32. In this investigation, the RMSD fluctuations of the complex when using MD1, MD2 and MD3 were stabilized at a similar RMSD magnitude over the last 1 ns. Contrasting results appeared for MD4, indicating that the system was not well equilibrated over the 2 ns simulation (figure 32a). Details of the structural changes in the complex structure were investigated further by individually assessing the RMSD fluctuations of the DNA and thiazotropsin A. Figure 32b reveals the same pattern for DNA as the complex. The RMSD fluctuations of thiazotropsin A in figure 32c support MD4 instability in simulations whilst the others stabilized in the same way. The results of the unrestrained DNA simulations indicate that the topology parameters show different effects to the restrained on DNA simulations and the cause of instability comes from the DNA movement although the simulations depend on the topology parameters of thiazotropsin A.

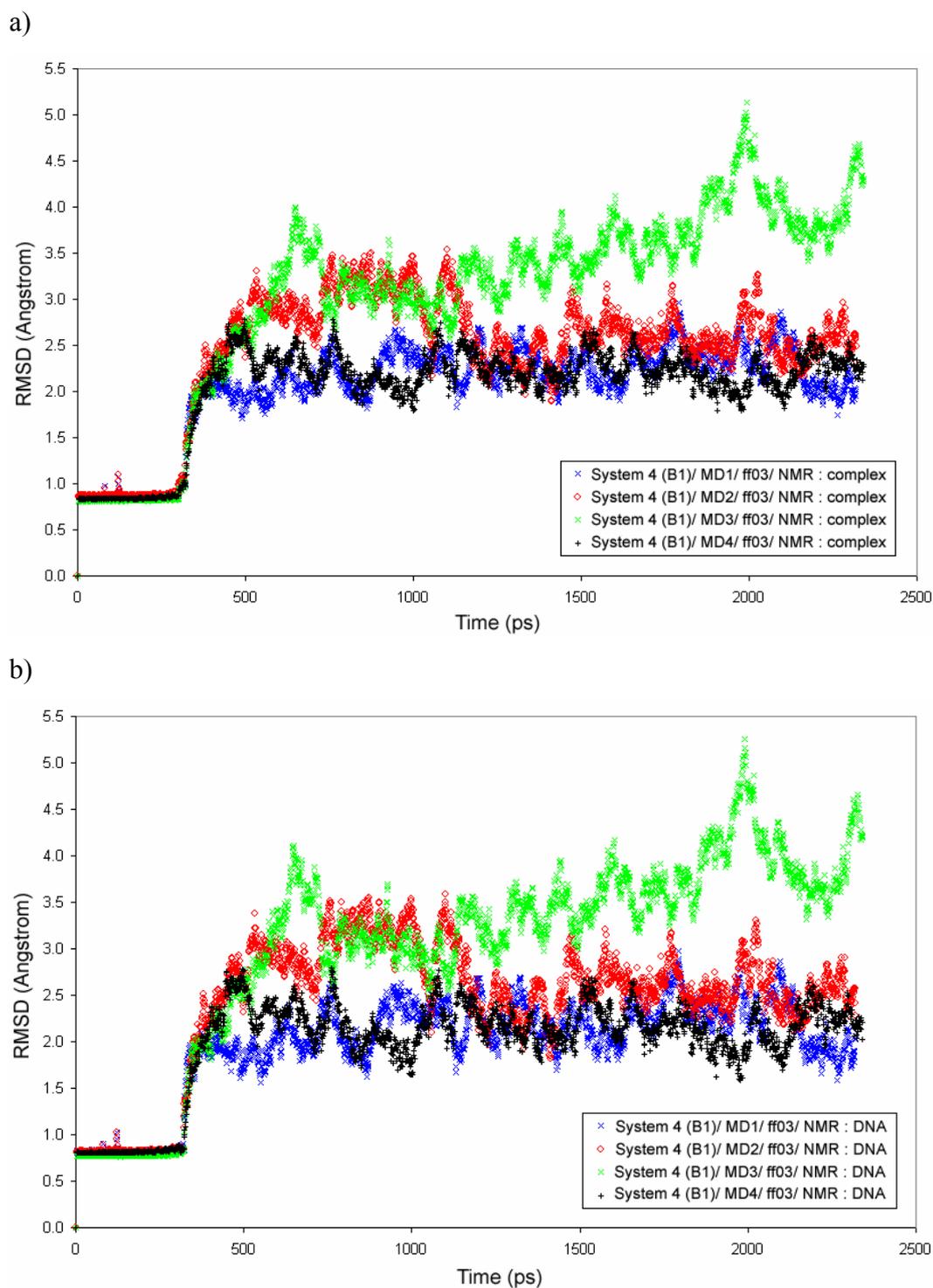


Figure 32 The RMSD fluctuation from simulations system 4 (the free simulations) with various topology parameters (MD1, MD2, MD3 and MD4) usage for thiazotropsin A. a) complex, b) DNA and c) thiazotropsin A.

c)

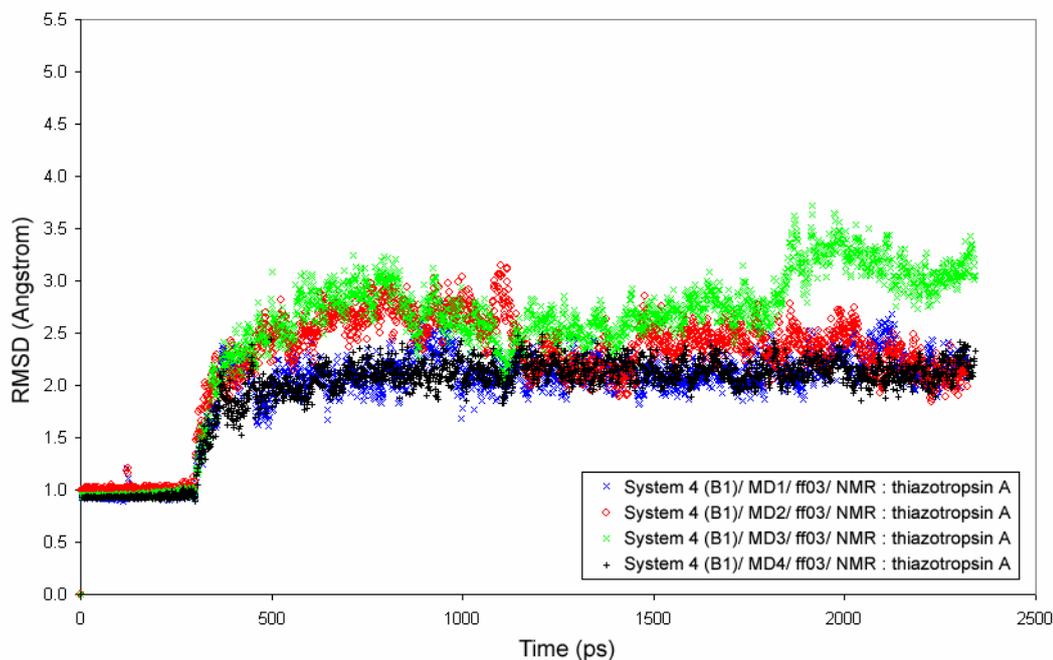


Figure 32 (Continued).

Finally, the RMSD fluctuations observed in the various simulation systems (system 1, 2, 3, 4 (B1) and 4 (B2)) with reference to the topology parameter set of MD2 with the ff03 force field as shown in figure 33. The RMSD fluctuations of the complex are clearly separated into two categories. The group with restraints on DNA (system 1, 2 and 3) were well equilibrated at a low magnitude of RMSD of around 1.2 Å whereas the group with unrestrained DNA (system 4 (B1) and 4 (B2)) showed a much higher fluctuation. The RMSD fluctuations of DNA (figure 33b), followed the same pattern as the complex, with the restrained systems (system 1, 2 and 3) were predominantly well stabilized at a low level of RMSD below 0.5 Å, which did not occur in the unrestrained DNA simulations (system 4 (B1) and 4 (B2)). When only motions of thiazotropsin A were determined (figure 33c), it stabilized and fluctuated around the same level of RMSD in all cases, with the same pattern of fluctuations found within both restrained and unrestrained DNA systems. These suggest that the solute can relax to a greater extent in a solvated environment than without solvent in the MD2 systems.

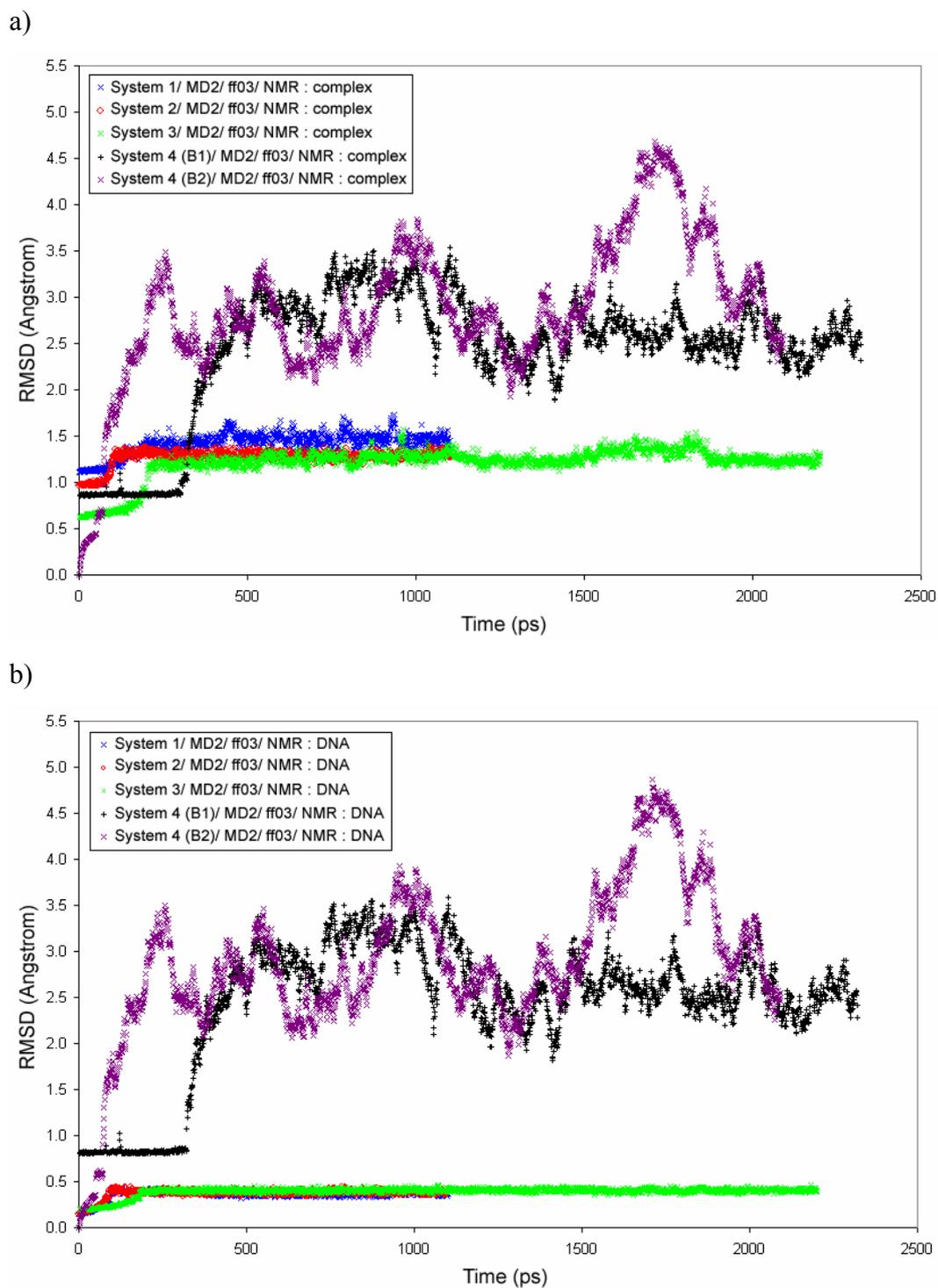


Figure 33 The RMSD fluctuation from various simulation systems (system 1, 2, 3, 4 (B1) and 4 (B2)) by using the same MD2 parameter set. a) complex, b) DNA and c) thiazotropsin A.

c)

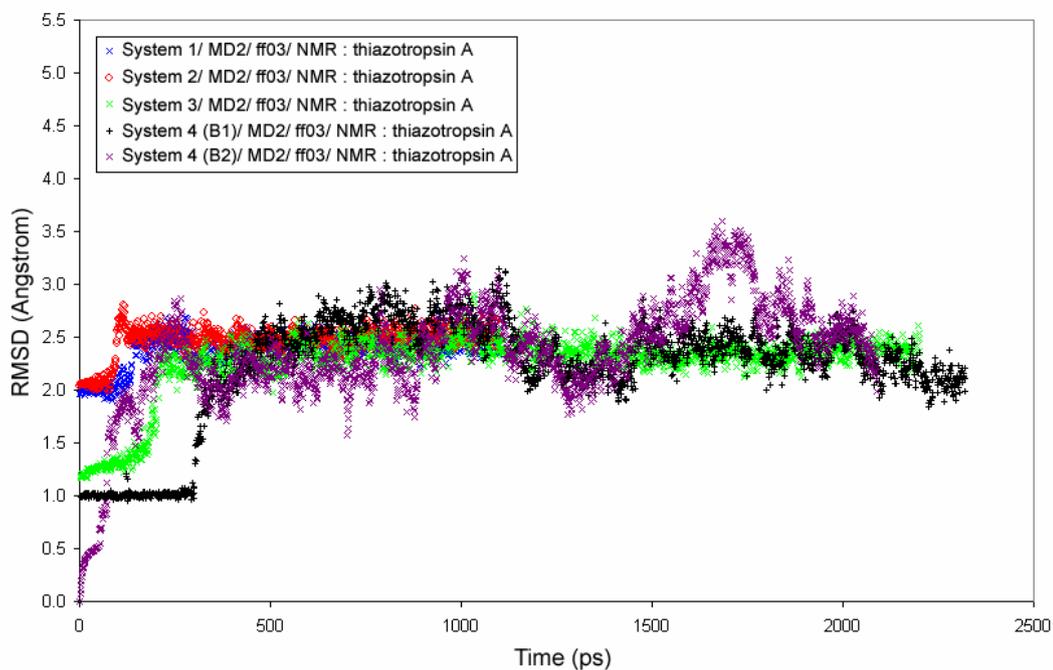


Figure 33 (Continued).

In addition, the RMSD fluctuations of the simulation system 5 using the AMBER tutorial were also investigated as shown in figure 34. The RMSD fluctuation from longer simulations that were extended to 5 ns confirmed stabilization of the complex, DNA and thiazotropsin A with previous results. The superimposition of DNA and thiazotropsin A from different simulations revealed that the backbones of DNA are similarly aligned as shown in figure 35. However, the characteristics of the base-pairs such as stack, twist, slide and roll are different. The most flexible part of thiazotropsin A appeared at its tail where the positively charged group NH⁺ is located. These were considered to be an effect of the binding mechanism.

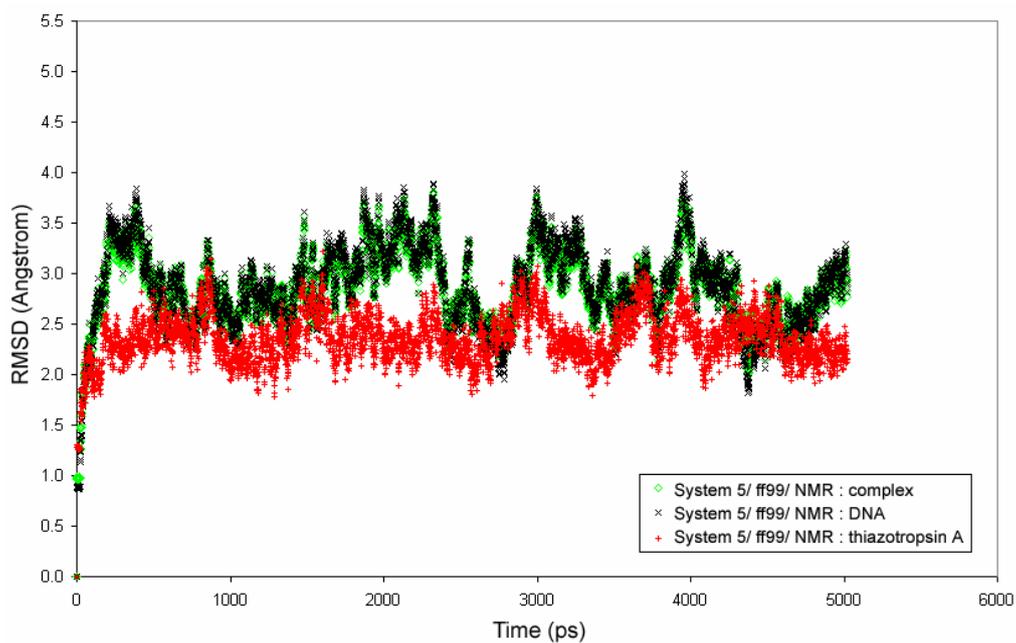


Figure 34 The RMSD fluctuation from the simulations system 5.

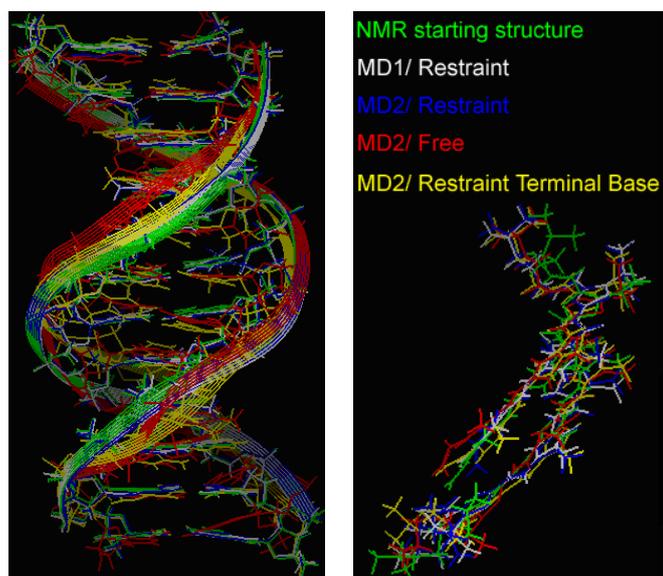


Figure 35 Superimposition of DNA and thiazotropsin A from difference simulations in solution.

4. The CPU timing in the simulations of the DNA/Thiazotropsin A Complexes

Computational resources and time usage in calculations are always considered to be a top priority when planning simulations. Therefore, the CPU timing in the different simulations of the DNA/ thiazotropsin A complexes was monitored and compared. Also the difference between computer platforms used in the studies is shown.

CPU time use in the simulations of the DNA/ thiazotropsin A are reported in table 26 and table 27. Most were performed on a computer platform with a cluster of 8 CPU nodes, although some were came out on a single CPU platform. CPU time use represents total calculation time for each simulation system. The total CPU time use of the system that applied the polarizable (ff02) force field was dramatically increased, despite one polarization step being turned off during minimization. The time consumed by the solvation simulations (system 3 and system 4) are huge compared with the invacuo simulations (system 1 and system 2), and suggests that the optimum amount of solvent should always be considered when performing simulations. Similar trends were obtained for computer platforms with the single and cluster CPU nodes. The CPU time use for system 3, MD1 and ff03 force field combination was a lot more than for the system 3, MD2 and ff03 force field combination. This was because the system 3, MD1 and ff03 force field combination was calculated on the single CPU computer platform, whilst the system 3, MD2 and ff03 force field combination was calculated in parallel on the 8 CPU node platform, and suggests that a parallel calculation on a computer platform with more CPU nodes is more preferable than using a single CPU platform. Most time consuming was system 4 (B1) with the polarizable (ff02) force field. The quantitative comparison shown in table 28 and table 29 as the time consumed per 100 steps of simulation.

Table 26 The total CPU time usage (hours) for combination systems types and model types with ff03 force field.

	System 1	System 2	System 3			System 4	
			Protocol A	Protocol B1	Protocol B2		
MD1	5.27 ^s	5.59 ^s	294.48 ^s	47.87	29.28		
MD2	5.19 ^s	5.40 ^s	52.16	48.75	46.83		
MD3	5.33 ^s	5.47 ^s	46.95	51.17	49.60		
MD4	5.14 ^s	5.53 ^s	51.21	51.31	31.31		

s : indicated the calculations from computer platform with single CPU

Table 27 The total CPU time usage (hours) for combination systems types and model types with ff02 force field.

	System 1	System 2	System 3			System 4	
			Protocol A	Protocol B1	Protocol B2		
MD1	9.17 ^s	9.51 ^s	173.13	394.17	89.29		
MD2	9.15 ^s	9.59 ^s	152.75	275.04	87.35		
MD3	9.07 ^s	9.57 ^s	68.77	179.24	87.79		
MD4	9.18 ^s	9.52 ^s	298.55	269.60	116.67		

Table 28 The CPU time usage (seconds) per 100 steps in the production MD simulations from combination systems types and model types with ff03 force field.

	System 1	System 2	System 3			System 4	
			Protocol A	Protocol B1	Protocol B2		
MD1	1.62 ^s	1.66 ^s	86.90 ^s	13.76	9.99		
MD2	1.59 ^s	1.67 ^s	14.78	13.78	16.02		
MD3	1.65 ^s	1.68 ^s	14.62	14.75	16.94		
MD4	1.59 ^s	1.70 ^s	14.83	14.79	10.43		

s : indicated the calculations from computer platform with single CPU

Table 29 The CPU time usage (seconds) per 100 steps in the production MD simulations from combination systems types and model types with ff02 force field.

	System 1	System 2	System 3			System 4	
			Protocol A	Protocol B1	Protocol B2		
MD1	2.69 ^s	2.77 ^s	55.18	60.67	14.78		
MD2	2.66 ^s	2.79 ^s	49.98	40.68	14.84		
MD3	2.63 ^s	2.78 ^s	21.94	28.39	14.95		
MD4	2.67 ^s	2.77 ^s	50.83	42.54	20.13		

5. Binding Free Energy Calculations of the DNA/Thiazotropsin A Complexes

Calculating the free energy of binding is the main purpose of this work and is based on using a combination of charge mapping, force field types for ligand and DNA, and the set up of the simulation systems. The values obtained can be benchmarked against experimental values. Binding energies were calculated from the sander, pbsa and nmode subprogram in AMBER8 based on MM-PBSA calculations (Srinivasan *et al.*, 1998) rather than from Delphi.

The simulations systems were performed using two different computer platforms; single CPU and a cluster of 8 nodes CPU. Therefore, results from selected simulations were used to check for platform error. In this example, the thiazotropsin A/DNA complexes were run on a single CPU (Vader) and Opteron cluster 8 CPU (Alien) using the same starting conformation and exactly the same protocol; the resulting energies were different after minimization. When the simulation was continued for 1 ns, it showed significantly different binding energies when using a polarizable force field with extra charge (ff02EP). The simulations system 1 combined with MD1 and ff02EP force field returned a binding energy of -9.59 kcal/mol for the single CPU platform, whereas a binding energy of -16.56 kcal/mol was found from the parallel calculation on the cluster platform as shown in table 30. A difference but less significant, was also observed in the modified simulations system 1 (plus 0.1 kcal restrain on thiazotropsin A) combined with MD2 and ff03

force field. A binding energy of -9.14 kcal/mol was calculated for single CPU platform, whereas a binding energy of -5.99 kcal/mol was found from the parallel calculation on the cluster platform. Moreover, the enthalpy and entropy components were also different. These examples suggest that the simulations are computer platform dependent. Therefore, to reduce the errors from different platform, the same computational resource should be used in the simulations and calculations.

Table 30 Binding energy (kcal/mol) results: comparison between single CPU (Vader) and Opteron cluster (Alien).

system models		ff 02 EP		
Opt. Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1 - MD1 ^s	-49.92	-40.33	-9.59	
System 1 - MD1 ^c	-50.45	-33.89	-16.56	
System models		ff 03		
NMR Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1a - MD2 ^s	-50.02	-40.88	-9.14	
System 1a - MD2 ^c	-49.00	-43.01	-5.99	

a = plus 0.1 kcal restrain on thiazotropsin A

S = single CPU platform

C = 8 nodes cluster CPU platform

The simulations system 2 combining with MD1 and the non-polarizable (ff03) force field was used as a case example to observe the effect of varying the number of snapshots used in the binding energy calculations. Within the same output trajectory, the first analysis sampled snapshots every 10 ps, whilst the second sampled every 0.5 ps. Consequently, the first sampling produced 100 sets of snapshots whereas the later sampling produced 1800 sets refer to the figures that show the energy and rmsd profiles for this protocol. The binding energies were not significantly different since the system was well stabilized in energy and position, and indicated that for the restrained DNA systems, sampling 100 snapshots were enough to obtain a binding free energy. The final values are not different, but compensation is observed between

MM-PBSA and normal mode energies. More systems need to be studied to see if it is a general trend or if the higher sampling rate should be preferred.

Table 31 Effect of the number of snapshots on the binding free energy (kcal/mol), taken from simulation system 2 combined with MD1 and ff03 force field.

Number of snapshots (sets)	ff 03		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
System 2 - 100	-56.20	-43.11	-13.09
System 2 - 1800	-54.16	-40.76	-13.40

The charge applied to the ligand force field is always an issue because varieties of charge mapping algorithms are provided in the antechamber module, and partial charge is sensitive parameter depending on the calculation method. Calculations using the natural bonded orbital (NBO) theory are well known for giving high accuracy partial charges on molecules, and the case example studied here is by applying charges from the NBO and RESP approaches to thiazotropsin A force field. The binding energies determined using different charge type was analyzed.

Comparison of the binding energies obtained from the systems using either NBO or RESP charges for the ligand suggests that RESP charges give more favorable energies, independently of the force field complementary as shown in table 32. Applying NBO charge into the ligand force field yielded positive binding energies of 8.22 and 2.87 kcal/mol for the ff03 and ff02EP DNA force fields, respectively. In contrast, for the RESP charges in the ligand force field, binding energies of -0.30 and -9.59 kcal/mol were calculated for the ff03 and ff02EP DNA force fields, respectively. This suggests that despite ligand charges being calculated from a high level of calculation such as NBO theory, they are not in agreement with either the ff03 or ff02EP DNA force fields in AMBER. The recommended charge mapping method should be from RESP approach. Moreover, although the binding energy of -9.59 kcal/mol was found to be in good agreement with the calorimetric experimental energy of -10.0 kcal/mol, this value was determined from an invacuo simulation, and more representative systems still had to be performed.

The next rational comparison was to determine which starting structure was preferable between the optimized or NMR resolved coordinates of thiazotropsin A by considering the binding energies calculated from both cases reported in table 32 and table 33 with reference to the RESP approach in table 32. The use of the starting structure from NMR yielded more attractive energies of -4.60 and -33.46 kcal/mol for ff03 and ff02EP, respectively. The more stable energies in all cases were obtained from using the starting structure from NMR, which suggests that it is better to apply the force field to experimentally determined structures from either NMR or X-Ray. Our later reports on structural parameters also confirm the use of the NMR structure as starting coordinates.

Table 32 Binding energy (kcal/mol) results for invacuo systems with restrain DNA using optimized structure of thiazotropsin A as a starting.

system models		ff 03		
Opt. Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1 - MD1 ^{NBO}	-30.74	-38.96	8.22	
System 1 - MD1 ^{RESP}	-40.33	-40.03	-0.30	
system models		ff 02EP		
Opt. Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1 - MD1 ^{NBO}	-37.09	-39.96	2.87	
System 1 - MD1 ^{RESP}	-49.92	-40.33	-9.59	

NBO = Charge from Natural Bonded Orbital calculation was mapped into thiazotropsin A force field.

RESP = Charge of thiazotropsin A was mapped from RESP algorithm into its force field

Table 33 Binding energy (kcal/mol) results for invacuo systems with restrained DNA using NMR structure of thiazotropsin A as a starting with RESP charges.

system models		ff 03		
NMR Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1 - MD1	-47.08	-42.48	-4.60	
system models		ff 02EP		
NMR Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	
System 1 - MD1	-65.62	-32.16	-33.46	

Binding energies calculated from the trajectories of the simulations system 1, 2, 3, 4(B1) and 4(B2) were reported respectively. Firstly, the binding energies from the restrained DNA invacuo simulations (system 1) were reported in table 34, as well as extra simulations based on modifications to system 1. Here, simulated structures were maintained close to their NMR structure by using restraints of 1.0 and 0.1 kcal/mol on DNA and the ligand, respectively (system 1a). The binding free energies calculated from simulation system 1a are reported in table 35. A stronger force used to maintain the simulated structure tighter to their NMR starting structure introduced by applying a restraint force of 10.0 kcal/mol on both DNA and thiazotropsin A (system 1b). The binding free energies calculated from simulation system 1b are reported in table 36.

Generally, all binding energies from the simplest simulation are shown in table 34. The combination of System 1 – MD3 gave good agreement with the experimental data (binding energy determined by ITC, -10.0 kcal/mol). However, because the starting structure was solved by NMR experiments, it leads us to further examine the testing of the average structure from NMR refinement. After restraint forces of 1.0 and 0.1 kcal/mol were kept on DNA and ligand respectively during the production phases, a range of binding free energies were generated (table 35). The binding energies from the parameter sets of MD1 and MD2 represent the closest approximation to the experiment, particularly with the ff03 force field. When a weak restraining force on DNA and thiazotropsin A was substituted by a stronger force of 10.0 kcal/mol, the binding energies are reported in table 36. Attempts were made to

constrain the DNA, but the simulations crashed. All trajectories using the stronger force produced positive binding free energies, indicating that repulsive interactions were predominant between thiazotropsin A and DNA. The tighter forces used to maintain the overall conformation of the DNA complex structure close to the average structure from NMR refinement, produced more repulsion in the binding free energies. These stepwise investigations suggest that the thiazotropsin A/DNA complex structure in solution solved by NMR refinement is unstable and approximate, and the complex structure should be relaxed in a medium such as box of water, containing high polar molecules. The binding energy calculated from the extra simulations which were performed in box of water based on the simulation system 1 with an extra 0.1 kcal restraint on thiazotropsin A (system 1a), coopted with the parameter sets of MD2 and ff03 force field is shown in table 35. The binding free energy of -9.14 kcal/mol calculated from the trajectory without solvent water, when immersed in a box of water changed to -15.27 kcal/mol. The contribution from solvent was necessary and not only observed in the total energies, but also included in the binding free energy. Although the process of calculating the binding free energies neglected all solvents and ions, the stabilizing properties from medium were important in maintaining the complex structure during the simulation. Moreover, inspection of the binding energies revealed that the ff02 force field returned more negative energies than ff03 force fields in all cases. The variation in the binding free energy came from major differences in mm-pbsa energy term or enthalpic energy, indicating that small restraints on the ligands and DNA still allow enough flexibility in the complex.

Table 34 Binding energy (kcal/mol) results for invacuo systems with restraints on DNA.

system models	ff 03			ff 02		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,						
System 1 – MD1	-47.08	-42.48	-4.60	-60.39	-40.57	-19.82
System 1 – MD2	-35.02	-38.55	3.53	-54.26	-36.22	-18.04
System 1 – MD3	-49.74	-39.24	-10.50	-49.59	-43.78	-5.81
System 1 – MD4	-35.80	-39.07	3.27	-42.76	-40.01	-2.75

* Binding energy from ITC experiment is -10.0 kcal/mol

Table 35 Binding energy (kcal/mol) results for invacuo systems with restraints on DNA and thiazotropsin A.

system models	ff 03			ff 02		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,						
System 1a – MD1	-49.67	-43.46	-7.21	-50.24	-41.21	-9.03
System 1a – MD2	-50.02	-40.88	-9.14	-56.56	-32.56	-18.00
System 1a ^s – MD2	-58.58	-43.31	-15.27	-	-	-
System 1a – MD3	-41.46	-45.06	3.60	-45.30	-43.93	-1.37
System 1a – MD4	-42.23	-43.46	1.23	-43.96	-41.60	-2.36

* Binding energy from ITC experiment is -10.0 kcal/mol

a = pluss 0.1 kcal restrain on thiazotropsin A

^s = solvation in box of water

Table 36 Binding energy (kcal/mol) results for invacuo systems with higher restraints on DNA and thiazotropsin A.

system models	ff 03			ff 02		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,						
System 1b – MD1	-17.30	-41.65	24.35	-19.63	-42.53	22.90
System 1b – MD2	-17.24	-44.79	27.55	-20.21	-44.02	23.81
System 1b – MD3	-7.95	-42.29	34.34	-10.52	-44.03	33.51
System 1b – MD4	-7.69	-42.64	34.95	-10.28	-41.32	31.04

* Binding energy from ITC experiment is -10.0 kcal/mol

b = applying a force of 10 kcal/mol to DNA and thiazotropsin A during production simulations

The binding energies from the invacuo simulations with a restraint of 1.0 kcal/mol on DNA were also investigated by including Na⁺ counter ions (simulations system 2). Most binding free energies from system 2 showed negative energies, varying from -10.97, -17.33 and -17.57 kcal/mol with parameter sets of MD1, MD2 and MD3, respectively, whilst the MD4 set produced a binding energy of 2.03 kcal/mol (table 37). The effect of counter ions on the binding energy can be observed by comparing these results with the binding energies for the invacuo systems without counter ions shown in table 34. Using exactly identical restraints on DNA, the comparison shows that the counter ions contributed to lowering the overall binding free energies. This particularly significant observation suggests that counter ions play an important role in stabilizing the simulated complexes from which the binding energies are derived. The simulations with polarizable (ff02) force field showed the same trend of binding energy as observed with ff03 force field.

Table 37 Binding energy (kcal/mol) results for invacuo systems including counter ions with restraints on DNA.

system models	ff 03			ff 02		
	$\Delta G^{\text{MM-PBSA}}$	$T\Delta S$	$\Delta G^{\text{Binding}}$	$\Delta G^{\text{MM-PBSA}}$	$T\Delta S$	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,						
System 2 – MD1	-52.33	-41.36	-10.97	-60.88	-43.71	-17.17
System 2 – MD2	-55.53	-38.20	-17.33	-66.41	-40.33	-26.08
System 2 – MD3	-55.04	-37.47	-17.57	-60.62	-37.45	-23.17
System 2 – MD4	-36.96	-38.99	2.03	-46.14	-44.91	-1.23

* Binding energy from ITC experiment is -10.0 kcal/mol

The binding energies from the invacuo simulations have been calculated in order to test the parameters and the calculation procedures to optimize the set up process, parameters and protocols. Nevertheless, to reproduce experimental binding energies by modeling simulations and to reflect actual events in biological systems, solvation of the system are required either implicitly or explicitly, with periodic boundary conditions applied to the solvated simulations.

Solvated systems were simulated with implicit and explicit solvent water as well as a restraint force of 1.0 kcal/mol applied to the DNA (system 3). The results from invacuo simulations indicated that a restrain force of 1.0 kcal/mol was not too strong and could be used to restrain the DNA whilst keeping freedom of movement for thiazotropsin A. Results using explicit solvent water found binding free energies of -13.09, -15.94, -13.07 and -12.65 kcal/mol corresponded for MD1, MD2, MD3 and MD4 parameter sets, respectively, with all energies approaching the experimental value of -10.0 kcal/mol (table 38). The example of simulations system 3 with MD2 was used to determine the level of restraint for thiazotropsin A. A binding energy of -15.27 kcal/mol obtained from simulations with a weak restraint of 0.1 kcal/mol on thiazotropsin A, was similar to the result of -15.94 kcal/mol observed from the simulation using free thiazotropsin A (simulations system 3). The comparison between weak restraints and free thiazotropsin A within system 3 revealed no significant difference between the two binding energies. This indicated that thiazotropsin A was sufficiently relaxed in the simulations, with no significant

difference observed with weak restraints when the system was immersed in a box of water, and implied that the relative position of thiazotropsin A was not significantly changed. It also suggests that within a box of water, the restraint force on thiazotropsin A is not necessary, although the key to reproducing experimental binding energies is to keep a weak restraint on DNA to prevent its deformation to far from B-DNA conformation that the experimental coordinates represent.

Table 38 Binding energy (kcal/mol) of neutralized solvated systems in implicit and explicit solvent with restraints on DNA and ff03 force field.

system models	ff 03		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,			
System 3 – MD1	-56.20	-43.11	-13.09
System 3 – MD2	-54.01	-38.07	-15.94
System 3a – MD2	-58.58	-43.31	-15.27
System 2 – MD2, $\epsilon = 4$	-32.27	-41.44	9.17
System 2 – MD2, $\epsilon = 80$		Ewald Bomb, No box for Na+	
System 2 – MD2, $\epsilon = 80$, box	-25.82	-41.37	15.15
System 3 – MD3	-49.78	-36.71	-13.07
System 3 – MD4	-45.11	-32.46	-12.65

* Binding energy from ITC experiment is -10.0 kcal/mol

a = plus 0.1 kcal restrain on thiazotropsin A

ϵ = dielectric constant (representing implicit solvent)

box = periodic boundary condition

Table 39 Binding energy (kcal/mol) of neutralized solvated systems in implicit and explicit solvent with restraints on DNA and ff02 force field.

system models	ff 02		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,			
System 3 – MD1	-57.94	-34.40	-23.54
System 3 – MD2	-69.09	-38.87	-30.22
System 3 – MD3	-61.03	-40.02	-21.01
System 3 – MD4	-49.43	-42.41	-7.02

* Binding energy from ITC experiment is -10.0 kcal/mol

The implicit solvation simulations were performed by using a dielectric constant of 4 and 80. Dielectric constants of 4 represent non-polar solvent solvation, whereas values of 80 represent the solvation with water (polar solvent). The use of dielectric constants instead of explicit solvent yielded binding free energies of 9.17 and 15.15 kcal/mol for dielectric constant of 4 and 80, respectively, and indicates that implicit solvent simulations produce repulsive of binding free energies, unlike the explicit solvent simulations. Thus it appears that implicit solvent representation is not appropriate for DNA/ thiazotropsin A complex simulations. Furthermore, implicit solvent simulations with a dielectric constant of 80 caused an ewald bombing problem that lead to simulation crashes, which was not the case in explicit solvent simulations. However, when the periodic boundary box without water was set to solve ewald bombing from Na ions flied away.

The binding free energies determined from solvated systems simulated with explicit solvent water and a restraint force of 1.0 kcal/mol applied to DNA (system 3) with the polarizable force field (ff02) for DNA are reported in table 39. Binding energies of -23.54, -30.22, -21.01 and -7.02 were observed for the MD1, MD2, MD3 and MD4 parameter sets, respectively, and most energies did not approach the experimental value of -10.0 kcal/mol in contrast to ff03 force field. The polarizable (ff02) force field lead to over estimation of binding even though a small restraint was applied to DNA.

Further investigations on the solvation simulations but without restraints (system 4) in a box of water were divided into two categories called protocol B1 and B2, according to the steps in the equilibration process. The binding free energies for non-restrained solvation simulations with protocol B1 (system 4(B1)) are shown in tables 40 and 41 using the non-polarizable (ff03) and polarizable (ff02) force fields for DNA, respectively. Binding energies were found in the range of -41.60 and -63.62 kcal/mol using the ff03 force field for DNA. Similar results in the range of -54.24 to -67.54 kcal/mol were obtained using the ff02 force field for DNA. These all represented large overestimates of binding and were only observed when the DNA was allowed to move freely during the production simulations.

Strong binding corresponded to overall conformational changes of the thiazotropsin A/DNA complex as shown by the RMSD plots, whilst simulations with a weak restraint of 1.0 kcal/mol on DNA gave binding energies value closer to the ITC experiment. Visual inspection of the trajectories using the VMD program showed bending of DNA during the course of the production run in the unrestrained systems. This conformational behavior might be a reason for the strong binding energies observed. Although the starting structures from the NMR refinement needed to be relaxed, they still retained the B-DNA form. After simulations of the complex had been performed without restraints on DNA, the structure of DNA complex started to bend and deviated from the B-DNA form as the Calladine-Drew (Calladine and Drew, 1984) projectors shown in figure 38 to figure 41 demonstrate. As a result, another type of equilibration protocol was sought in order to solve the DNA complex bending problem. Two subsequent simulations used free DNA with restraints of 1.0 kcal/mol applied to thiazotropsin and restraints on the terminal base-pairs of DNA with free thiazotropsin. The former showed that restraining thiazotropsin can improve on the binding energy by preventing conformational change of DNA whilst the latter had little effect on the simulation.

Table 40 Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B1 with ff03 force field.

system models	ff 03		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,			
System 4(B1) – MD1	-101.01	-37.39	-63.62
System 4(B1)a – MD1	-78.17	-39.22	-38.95
System 4(B1) – MD2	-95.31	-38.44	-56.87
System 4(B1)t – MD2	-99.20	-38.59	-60.61
System 4(B1) – MD3	-78.95	-37.35	-41.60
System 4(B1) – MD4	-98.00	-38.43	-59.57

* Binding energy from ITC experiment is –10.0 kcal/mol

a = plus 0.1 kcal restrain on thiazotropsin A

t = Restraint DNA only at terminal base-pairs

box = periodic boundary condition

Table 41 Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B1 with ff02 force field.

system models	ff 02		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,			
System 4(B1) – MD1	-107.16	-39.62	-67.54
System 4(B1) – MD2	-94.53	-40.29	-54.24
System 4(B1) – MD3	-100.64	-40.19	-60.45
System 4(B1) – MD4	-92.28	-37.22	-55.06

* Binding energy from ITC experiment is –10.0 kcal/mol

The simulation process was further investigated by using a loop equilibration protocol (protocol B2) which had been introduced by the group of Spacova et. al. (2001) based on 1:1 minor groove binding to DNA duplexes. Simulations of the thiazotropsin A/DNA as 2:1 binding complex with the loop protocol (system 4 (B2)) provided the binding energies as shown in table 42 and 43 complemented with non-

polarizable (ff03) and polarizable (ff02) force fields for DNA, respectively. The binding energies from system 4(B2) was in the range of -49.31 to -55.44 kcal/mol using ff03 force field for DNA, and similar results were observed in the range of -47.07 to -52.97 kcal/mol using of ff02 force field. The loop protocol (protocol B2) provided slightly better binding energies compared to those from protocol B1 by improving the MM/PBSA component for a 2 ns simulation using the non-polarizable force field (FF03). Binding energy using a polarizable force field (ff02) was expected to give a better value, but the results were nearly the same for the ff03 2 ns simulation. Again, these were all overbinding phenomena, the same as found in the system 4(B1). Furthermore, these loop protocols also did not maintain the structure of the DNA complex. The conformational change of the DNA complex still occurred as the Calladine-Drew projections show in figure 40 and figure 41. If a conformational change in DNA complex structure is demanded, the binding energies from simulations system 4(B1) and system 4(B2) were indicating, the energy consumption for the DNA deformation had to be taken into account. The binding energies needed to include deformation energies that could be obtained and calculated from separate trajectory simulations instead of single trajectory simulations. This new approach should cancel the error from the conformational change of DNA by taking into account conformations from trajectories of the DNA simulated in its unbound state, as suggested by the method of Kollman when investigating protein-RNA complexation (Reyes and Kollman, 2000).

Table 42 Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B2 with ff03 force field.

system models	ff 03		
NMR Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
System 4(B2) – MD1	-91.01	-38.67	-52.34
System 4(B2) – MD2	-90.31	-39.25	-51.06
System 4(B2) – MD3	-89.84	-40.53	-49.31
System 4(B2) – MD4	-94.69	-39.19	-55.44

* Binding energy from ITC experiment is -10.0 kcal/mol

Table 43 Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using a protocol B2 with ff02 force field.

system models	ff 02		
NMR Thiazotropsin A,	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
System 4(B2) – MD1	-85.11	-36.97	-48.14
System 4(B2) – MD2	-92.05	-39.08	-52.97
System 4(B2) – MD3	-92.59	-40.57	-52.02
System 4(B2) – MD4	-85.54	-38.47	-47.07

* Binding energy from ITC experiment is -10.0 kcal/mol

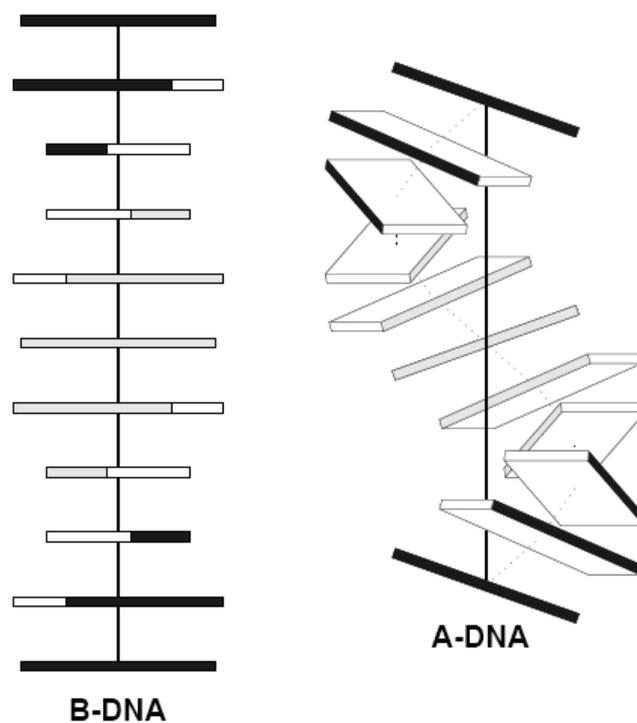


Figure 36 The 3-Dimensional drawing of nucleic acid called Calladine-Drew style, showing an ideal B-DNA and A-DNA form. (Lu, 2003)

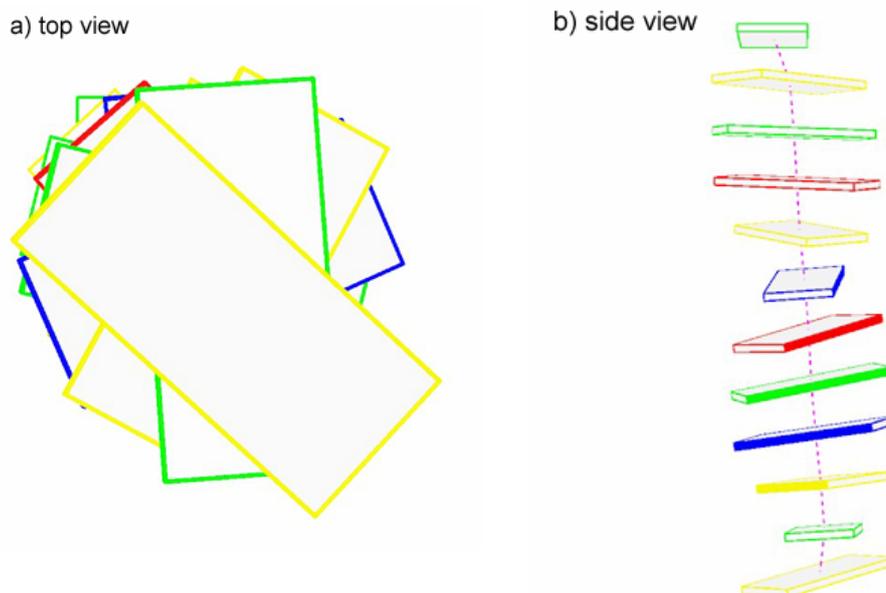


Figure 37 The Calladine-Drew projections base-centered of the starting structure Thiazotropsin A/ds(CGCACTAGTGCG)₂ complex from NMR refinement. a) top view. b) side view.

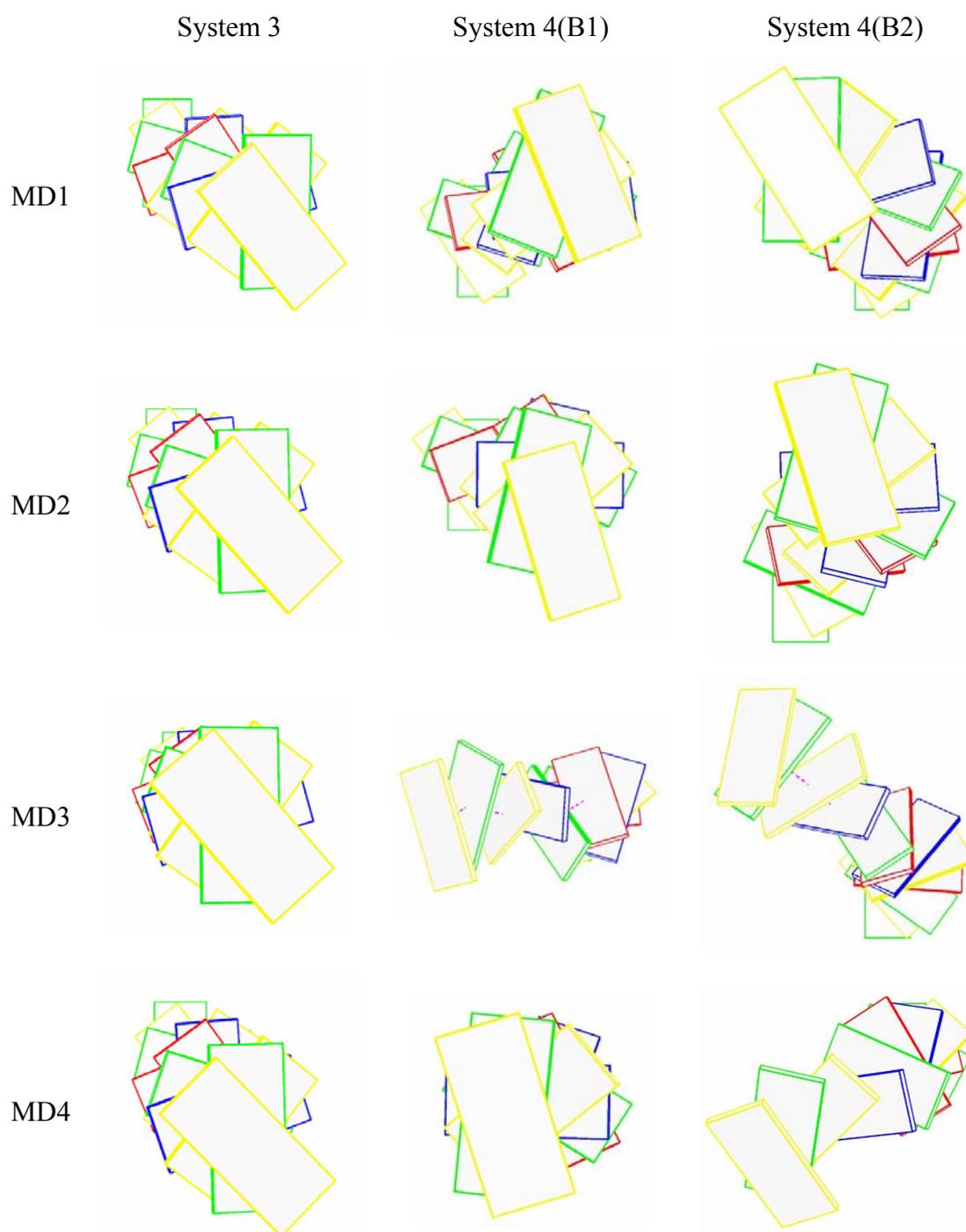


Figure 38 The Calladine-Drew projections base-centered from top view of the simulations structure Thiazotropsin A/ds(CGCACTAGTGCG)₂ complex taken the last configure from all solvation trajectories with ff03 force field.

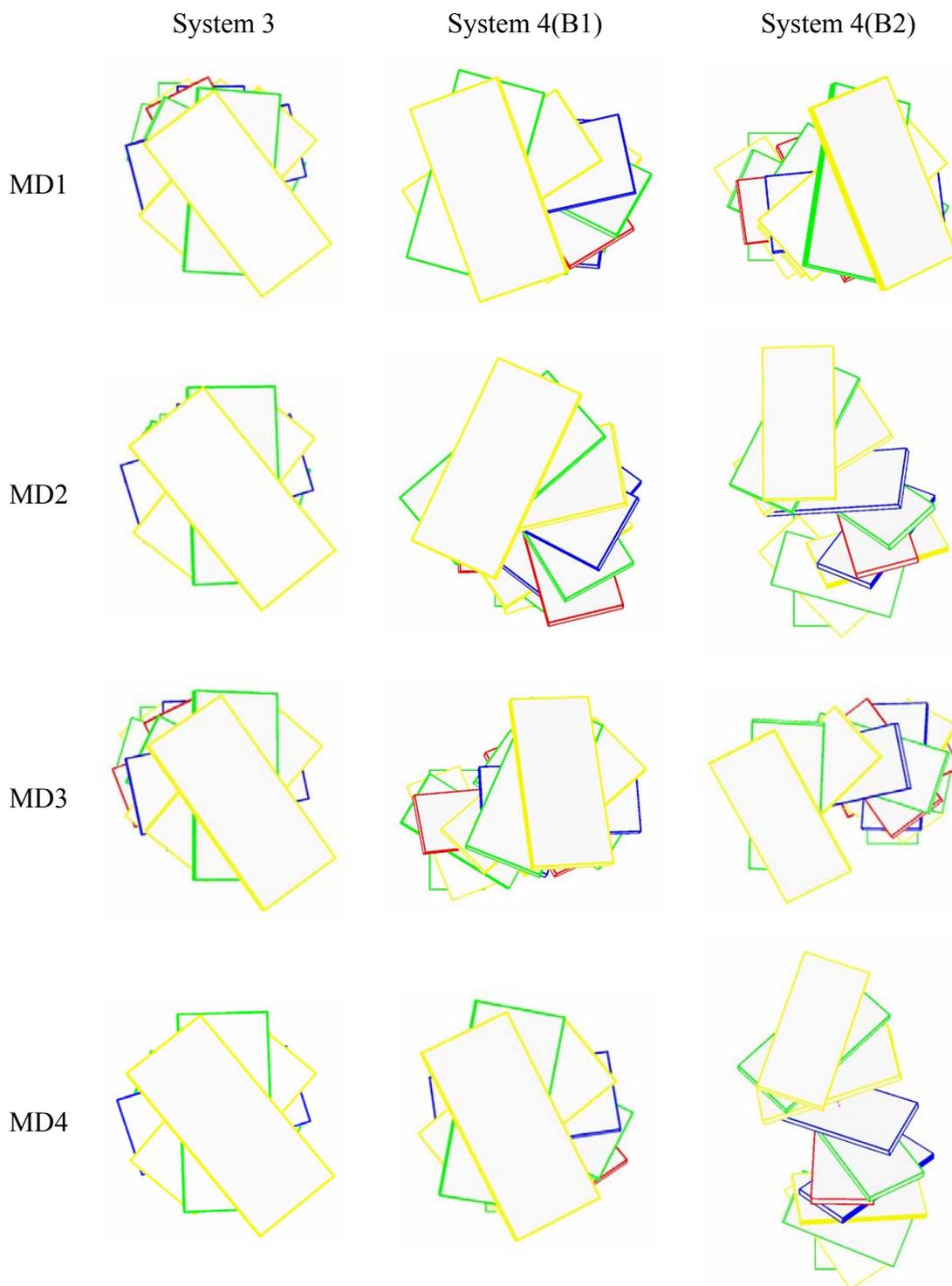


Figure 39 The Calladine-Drew projections base-centered from top view of the simulation structure Thiazotropsin A/ds(CGCACTAGTGCG)₂ complex taken the last configure from all solvation trajectories with ff02 force field.

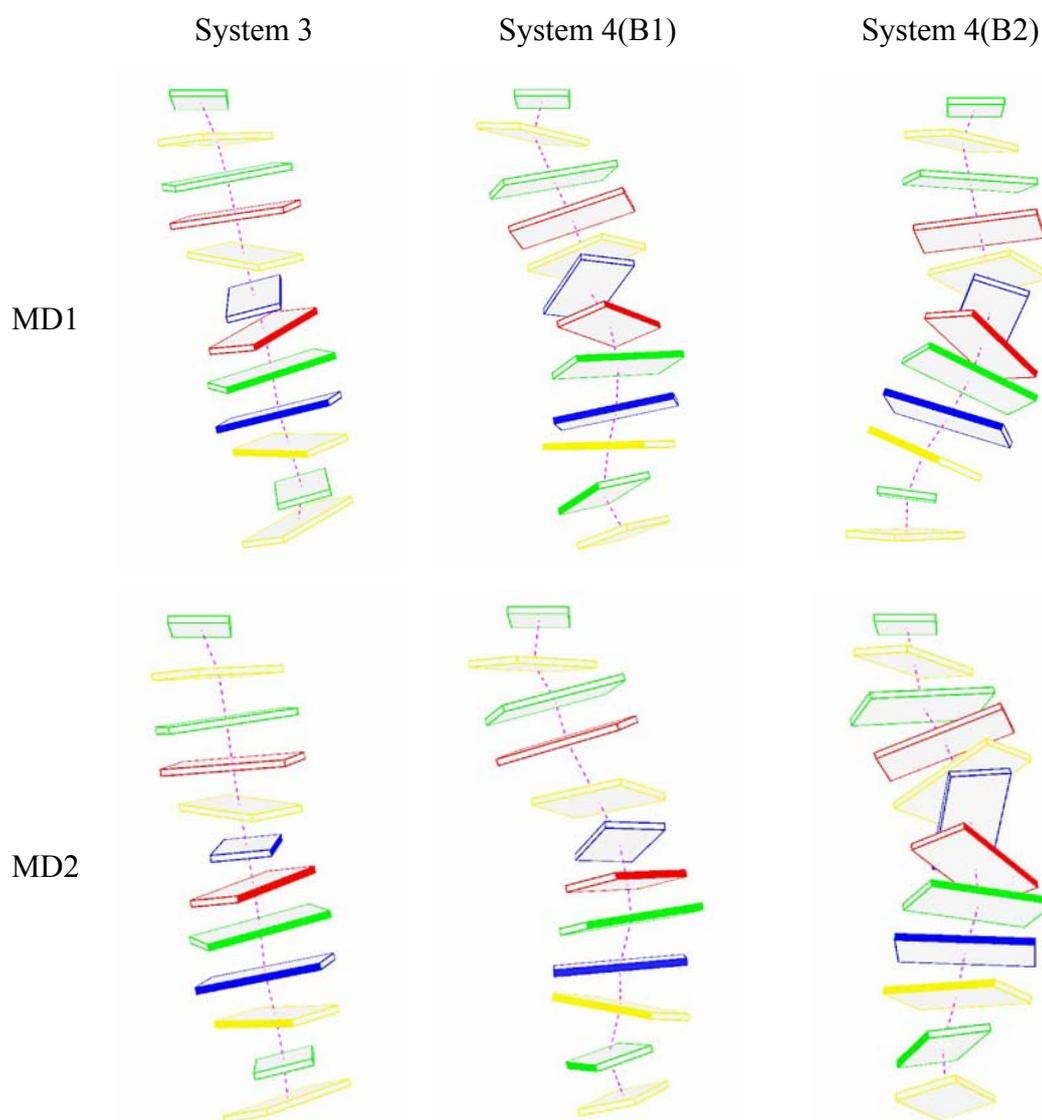


Figure 40 The Calladine-Drew projections base-centered from side view of the simulations structure Thiazotropsin A/ds(CGCACTAGTGCG)₂ complex taken the last configure from all solvation trajectories with ff03 force field.

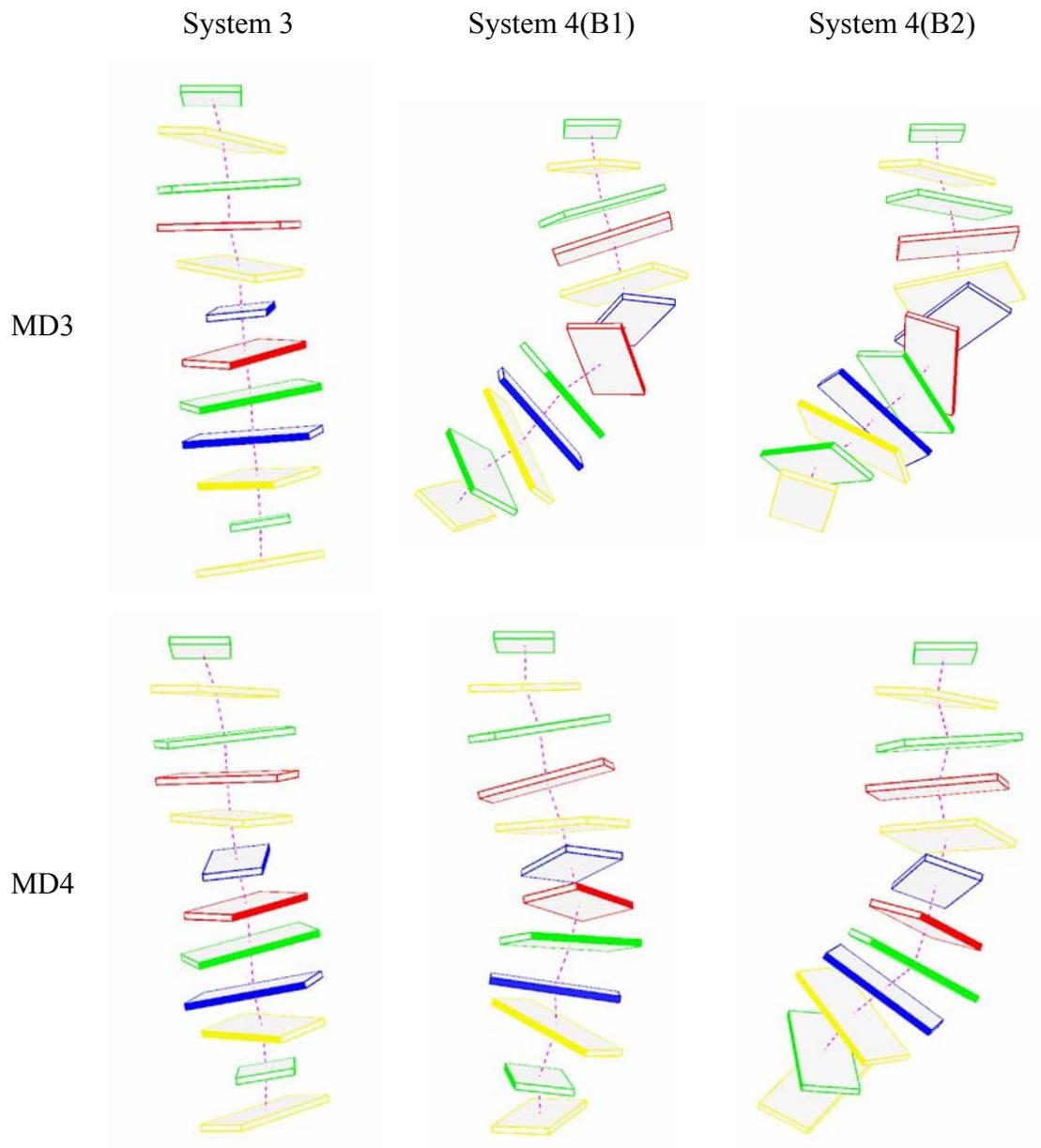


Figure 40 (Continued).

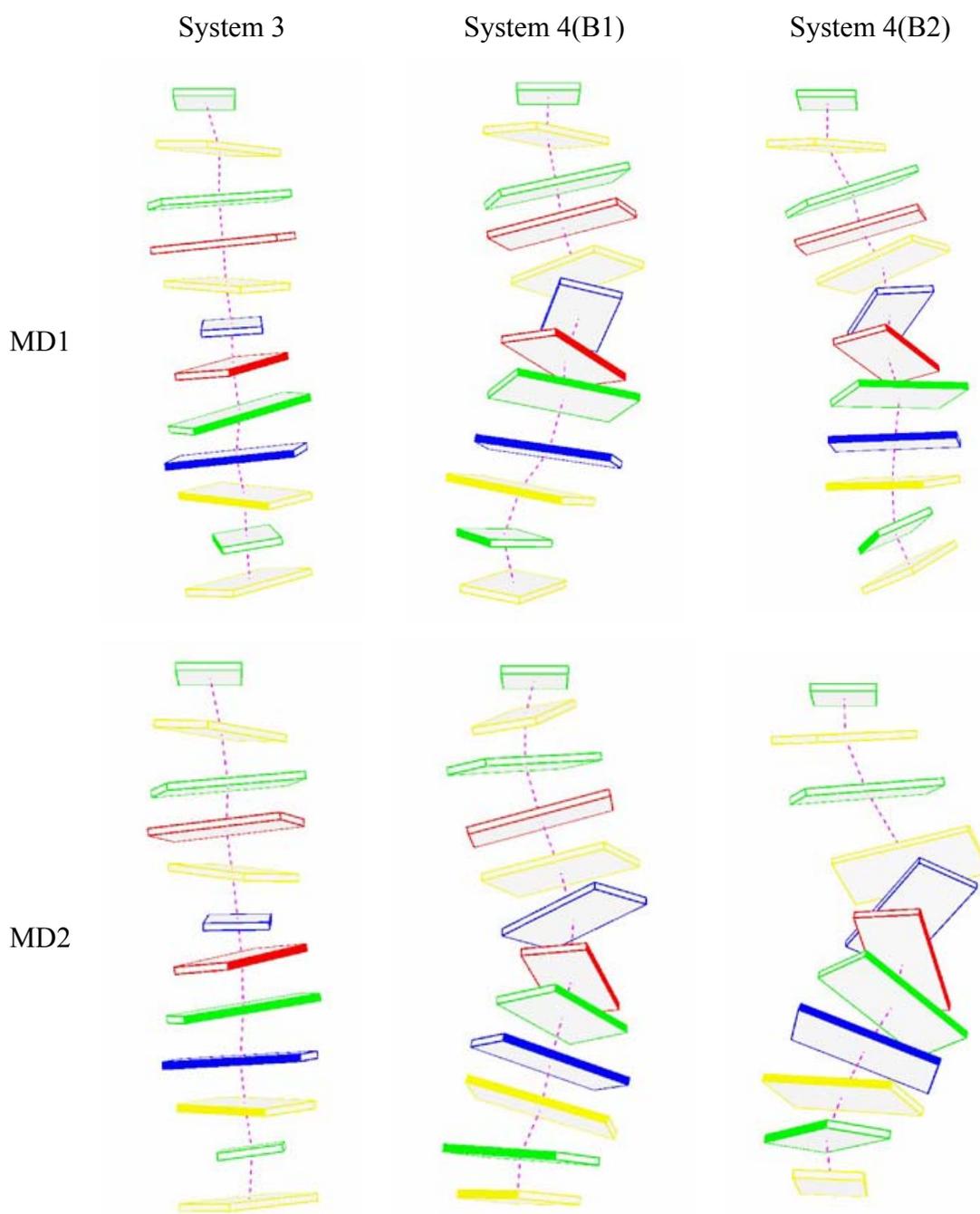


Figure 41 The Calladine-Drew projections base-centered from side view of the simulations structure Thiazotropsin A/ds(CGCACTAGTGCG)₂ complex taken the last configure from all solvation trajectories with ff02 force field.

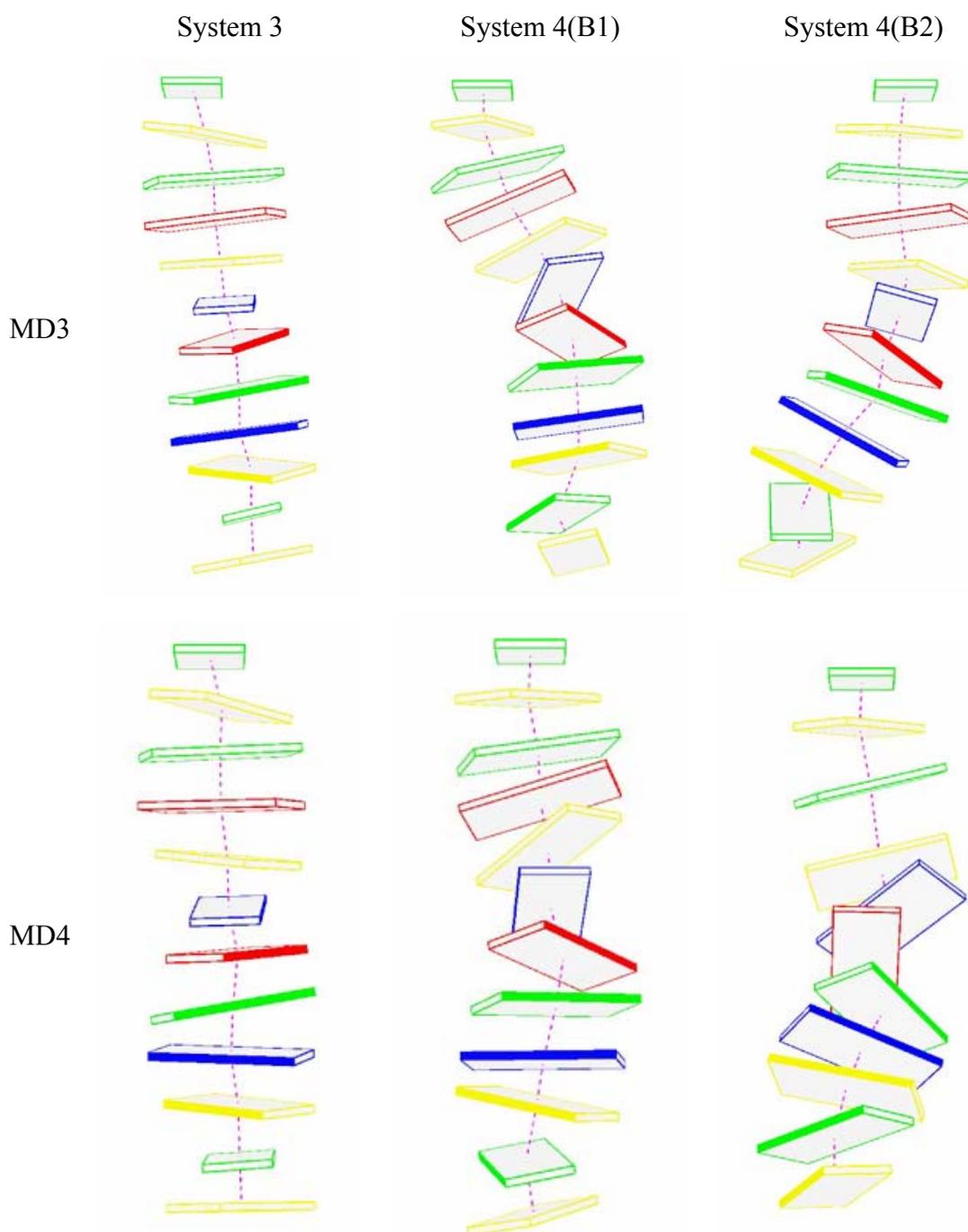


Figure 41 (Continued).

In addition, a simple simulation protocol was used to investigate the binding of the thiazotropsin A/DNA complex following the AMBER drug/DNA complex tutorial as shown in table 44. A basic force field was used for thiazotropsin A without modification of any atom types or force constants. Charge mapping was performed

using the AM1-BCC protocol directly from the Gaussian03 optimized structure. The binding energy was improved when compared with the previous protocol in both enthalpic and entropic energy components. Unfortunately, the production simulations lengthening of 5 ns showed worse binding energy than the trajectory from 2 ns simulations.

Table 44 Binding energy (kcal/mol) of neutralized solvated systems in explicit solvent with free DNA and thiazotropsin using the AMBER tutorial equilibration protocol.

system models	ff 03		
	$\Delta G^{\text{MM-PBSA}}$	T Δ S	$\Delta G^{\text{Binding}}$
NMR Thiazotropsin A,			
AMBER drug/DNA tutorial, 2 ns	-85.99	-59.30	-26.69
AMBER drug/DNA tutorial, 5 ns	-81.89	-43.57	-38.32

* Binding energy from ITC experiment is -10.0 kcal/mol

After the single trajectory calculations approach failed to reproduce the experimental binding free energy, the new calculation approach was implemented, based on the molecular dynamics simulations of DNA-DAPI complexes (Spackova *et al.*, 2003) and RNA-Protein complexes (Reyes and Kollman, 2000) using separate complex and monomer trajectories.

The binding free energies calculated from the separate trajectories are indeed an improvement over the single trajectories (table 45) when using the non-polarizable force field FF03 with both Protocols B1 and B2. The dramatic improvement came from the enthalpy energy component, whilst the entropy energy shows similar results. This indicates that the bound conformation of DNA has a high adaptation energy and conformational change from the free DNA in solution. Similar results and improved binding free energies can also be obtained from the separate trajectories when using polarizable force field FF02 (table 46). Again, the binding free energy comparison reveals that most of the group with assigned HF charges gave better energy than the group using B3LYP charges and the group using Protocol B2 gave better energies than the group using Protocol B1. Surprisingly, the MD3 model that used both

Protocol B1 and B2 gave calculated binding free energies of -10.33 and -10.06 kcal/mol, which are very close to reproducing the experimental binding free energy of -10.00 kcal/mol when using non-polarizable force field FF03, whilst only the MD4 model with Protocol B2 had a calculated binding free energy of -10.06 kcal/mol using polarizable force field FF02. Further investigations into these results are summarized in the conclusion.

Table 45 The binding free energies (kcal/mol) from *single* and *separate* trajectories calculations with non-polarizable force field FF03

System models/Protocols	<i>Single</i> trajecotries			<i>Separate</i> trajectories		
	ΔH	$T\Delta S$	ΔG	ΔH	$T\Delta S$	ΔG
MD1: Protocol B1	-101.01	-37.39	-63.62	-74.44	-38.36	-36.08
MD1: Protocol B2	-91.01	-38.67	-52.34	-54.70	-38.38	-16.32
MD2: Protocol B1	-95.31	-38.44	-56.87	-54.97	-36.38	-18.59
MD2: Protocol B2	-90.31	-39.25	-51.06	-47.46	-34.60	-12.86
MD3: Protocol B1	-78.95	-37.35	-41.60	-44.81	-34.48	-10.33
MD3: Protocol B2	-89.84	-40.53	-49.31	-45.32	-35.26	-10.06
MD4: Protocol B1	-98.00	-38.43	-59.57	-62.04	-36.03	-26.01
MD4: Protocol B2	-94.69	-39.19	-55.44	-63.71	-38.24	-25.47

Table 46 The binding free energies (kcal/mol) from *single* and *separate* trajectories calculations with polarizable force field FF02

System models/Protocols	<i>Single</i> trajecotries			<i>Separate</i> trajectories		
	ΔH	T ΔS	ΔG	ΔH	T ΔS	ΔG
MD1: Protocol B1	-107.16	-39.62	-67.54	-71.68	-38.91	-32.77
MD1: Protocol B2	-85.11	-36.97	-48.14	-52.49	-35.00	-17.49
MD2: Protocol B1	-94.53	-40.29	-54.24	-64.69	-34.91	-29.78
MD2: Protocol B2	-92.05	-39.08	-52.97	-53.66	-33.44	-20.22
MD3: Protocol B1	-100.64	-40.19	-60.45	-69.08	-40.33	-28.75
MD3: Protocol B2	-92.59	-40.57	-52.02	-55.26	-35.13	-20.13
MD4: Protocol B1	-92.28	-37.22	-55.06	-65.40	-37.29	-28.11
MD4: Protocol B2	-85.54	-38.47	-47.07	-45.10	-35.04	-10.06

The huge amount of data from the binding free energy calculations in tables 45 and 46 can be summarized into one table for easier comparison and analysis in table 47. The Protocol A simulations were performed using restraints on DNA and were not used to calculate binding free energy on the separate trajectories. Only simulations with free DNA such as Protocols B1 and B2 were treated with the separate trajectories approach where DNA conformational change energies could be observed.

It is clear that the separate trajectories approach achieves a large improvement in the calculated binding free energies as show in table 47. However, not all of the combinations between models, protocols and force field types can reproduce the experimental binding free energies. Only separate trajectories with HF mapped charges were found to reproduce the experimental binding free energies of -10.0 kcal/mol. This confirms that the conformational change energies of free DNA to bound DNA has a big influence upon ligand binding(Shaikh *et al.*, 2004) in the 2:1 side-by-side binding system.

Table 47 Summary of the binding free energies (kcal/mol) from various models, protocols and force field types.

<i>Single</i> trajectories	Non-polarizable force field FF03			Polarizable force field FF02		
	<i>ProtocolA</i>	<i>ProtocolB1</i>	<i>ProtocolB2</i>	<i>ProtocolA</i>	<i>ProtocolB1</i>	<i>ProtocolB2</i>
MD1	-13.09	-63.62	-52.34	-23.54	-67.54	-48.14
MD2	-15.94	-56.87	-51.06	-30.22	-54.24	-52.97
MD3	-13.07	-41.60	-49.31	-21.01	-60.45	-52.02
MD4	-12.65	-59.57	-55.44	-12.84	-55.06	-47.07
<i>Separate</i> trajectories	Non-polarizable force field FF03		Polarizable force field FF02			
	<i>ProtocolB1</i>	<i>ProtocolB2</i>	<i>ProtocolB1</i>	<i>ProtocolB2</i>		
MD1	-36.08	-16.32	-32.77	-17.49		
MD2	-18.59	-12.86	-29.78	-20.22		
MD3	-10.33	-10.06	-28.75	-20.13		
MD4	-26.01	-25.47	-28.11	-10.06		

All of these calculated binding free energies come from a combination of models, protocols and force field types. The benefit of performing these calculations is to observe how important model and protocol use is. We have conducted a Quantitative Protocol Energies Relationship (QPER) based on the 2:1 system of minor groove binder, thiazotropsin A/DNA complex. The correlations between models and protocols that we have investigated and plotted are show in figures 42 to figure 44. If we consider only the results from the single trajectory calculations with the same protocol, plotting values from the non-polarizable force field FF03 against the polarizable force field FF02 is show in figure 42. This indicates that the correlation between models is not observed in every protocol through the low correlation coefficient R^2 values of 0.720, 0.023 and 0.717 for Protocol A, Protocol B1 and Protocol B2, respectively. In figure 43, correlation between protocols is found in every model, as indicated by the high correlation coefficient, r^2 , values of 0.943, 0.993, 0.843 and 0.990 for MD1, MD2, MD3 and MD4, respectively. It means weather the simulation protocols usage would give not much energy different compare to the starting model usage.

This interesting point can relieve all the confusion over simulation protocol selection. The most important thing when commencing simulations are the starting model. However, this work suggests that the loop protocol, which has some repeat steps of minimization and equilibration, works better than the straight forward protocol that performed production simulation straight after minimization and equilibration. When we applied the same correlation investigation plot to the separate (figure 44), and compared them to the correlations in figure 42 by the same protocol, the separate trajectory calculations improved the correlation coefficient quite well; the correlation coefficient of Protocol B1 improved from 0.023 to 0.485 and the correlation coefficient of Protocol B2 from 0.717 to 0.961. This means that the different binding free energies from the various models can be narrowed by using the separate trajectory approach. However the group of models that used the HF/6-31G(d) mapped charge gave better binding free energies than the group that used the B3LYP/6-31G(d,p) mapped charges. A correlation between protocols with various models in the separate trajectories could not be observed due to the lack of data points.

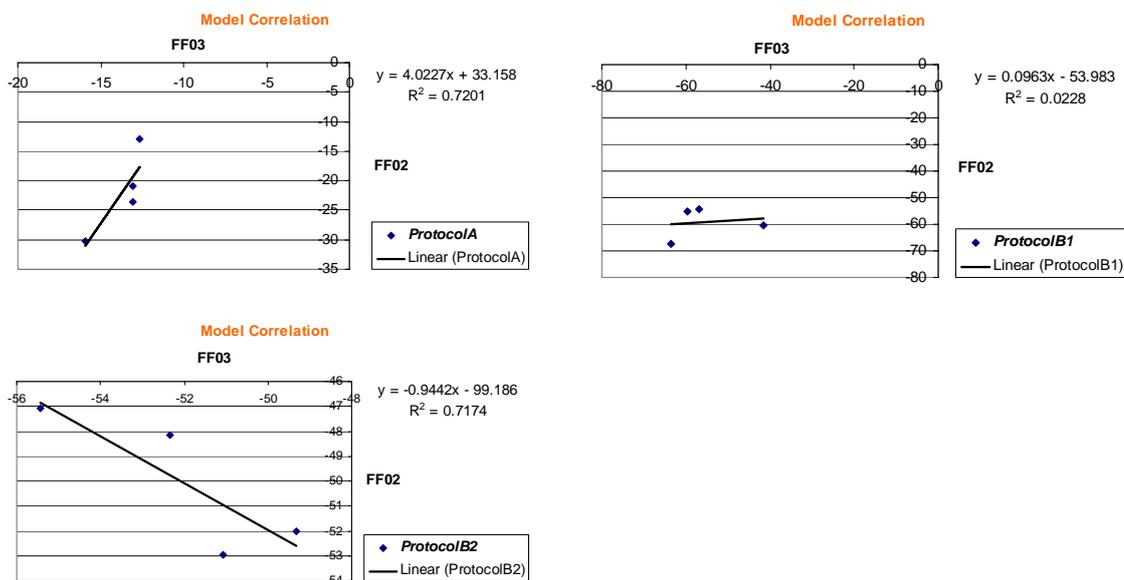


Figure 42 The correlation of model usage corresponding to various protocols for the *single* trajectory calculations.

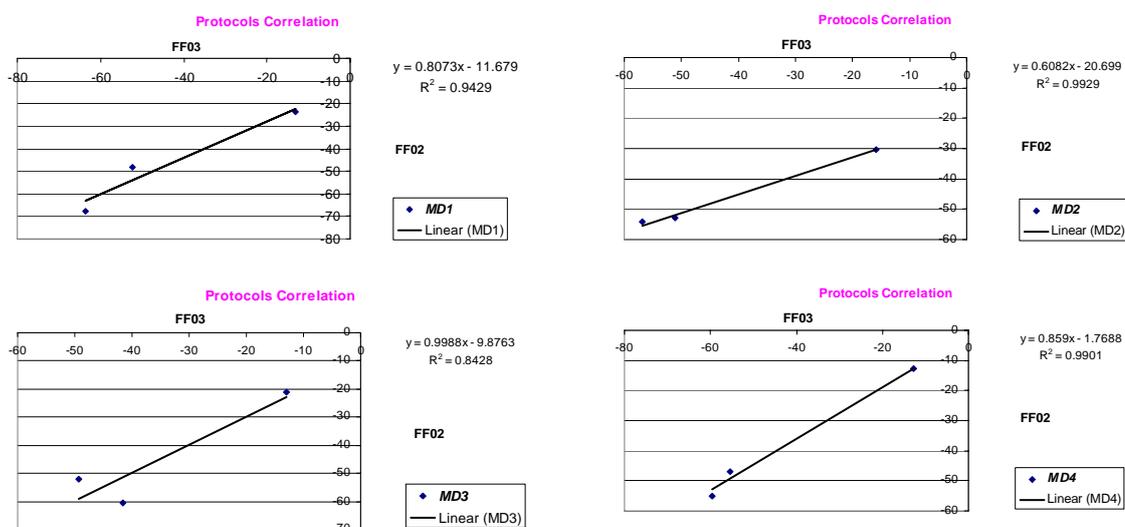


Figure 43 The correlation of protocols usage corresponding to various models for the *single* trajectory calculations.

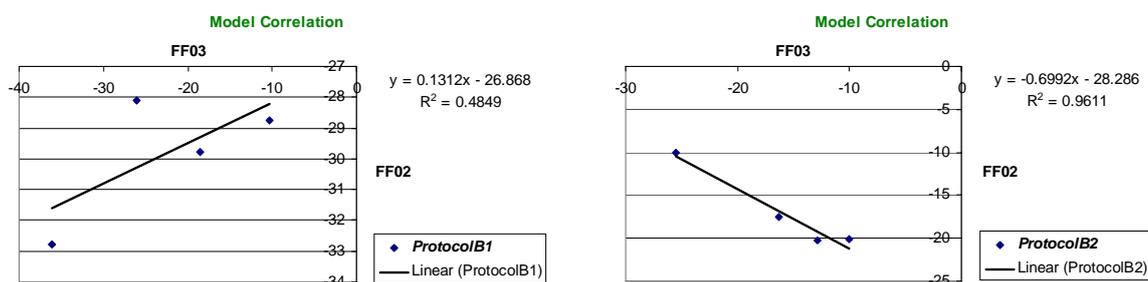


Figure 44 The correlation of protocols usage corresponding to various models for the *separate* trajectories.

6. Hydrogen Bond Interactions in the DNA/Thiazotropsin A Complexes

One of the most important interactions for thiazotropsin binding to DNA is the hydrogen-bonding interaction; thiazotropsin forms hydrogen bonds with the bases of DNA and is essential for binding selectivity (figure 45).

Table 48 Summary of the hydrogen bonds in the 2:1 complex formed between thiazotropsin A and d(CGACTAGTCG)₂ based on ¹H NMR chemical shift, labile proton exchange characteristics, and solution structure information. (Anthony *et al.*, 2004)

Distance name	Ligand atom	DNA atom
a)	H2	T ⁵ O2
b)	H9	A ⁶ N3
c)	H16	G ⁷ N3
d)	Thiazole N21	G ⁷ H22
e)	H26	T ⁸ O2

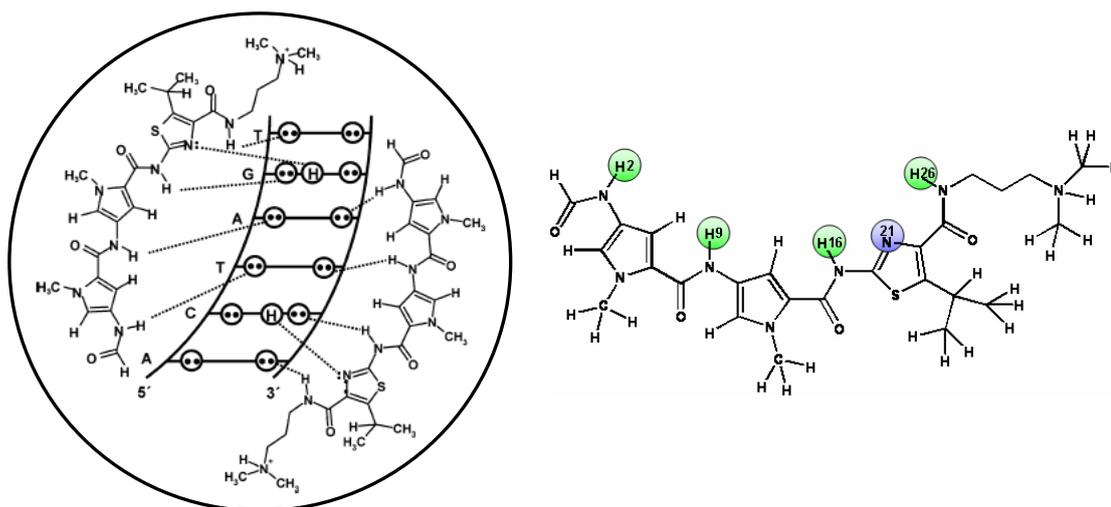


Figure 45 Interactions labeling between thiazotropsin A and DNA base-pairs correspond to NMR chemical shift assignment (Anthony *et al.*, 2004).

According to hydrogen bond formation implied from ¹H NMR chemical shifts, five hydrogen bonds distances were analysed from the solvations simulation trajectories (figures 46 and 47). Hydrogen bonds distances were measured through the equilibration and production runs starting from 0 to 2 ns between heteroatoms. The hydrogen bond distances from the NMR refinement structure were 2.33, 3.25, 3.40, 3.38 and 2.51 Å measured from distance a, b, c, d and e, respectively, and indicate that were strong hydrogen bond interactions. Surprising results were observed when

measurements were taken from all of the solvation simulations trajectories. All hydrogen bond distances within the restrained DNA simulations (system 3), increased to 6.0 and 8.0 Å from the beginning of the simulations in both ff03 and ff02 force fields. However, all distances showed stable fluctuations. This confirmed that the starting structures needed relaxation even when the NMR refinement coordinates were used, the hydrogen bond interactions inside the complex were lost and new orientations were found.

Details from non-polarizable (ff03) force field simulations (figure 46) were analyzed first. A strange pattern was found from simulations in system 3 with MD4 when the distances a and b changed in the middle of simulations at 1 ns, followed by a rise in the distances of c, d and e after 1 ns. However, the distances had stable fluctuations after increasing, indicating that interactions between thiazotropsin A and base pairs were forming again but somewhere else that was not agreement with chemical shifts found from NMR. Protocol B1 and B2 simulations that had been performed without restraints on DNA gave very stable fluctuations of hydrogen bond distances. Protocol B2's performance showed smoother fluctuations than protocol B1. The hydrogen bond distances from protocols B1 and B2 fluctuated around 3.0 Å for distances a, c, d and e, whilst distance b moved around 3.5 Å. The range of 3.0 to 3.5 Å indicates strong to medium of hydrogen bond interactions were formed which were in agreement with the chemical shift found from NMR experiment. Moreover, distance between H9 of thiazotropsin A and A⁶N3 appeared to be the most flexible interaction, occurring to the wide fluctuation around 3.0 to 3.5 Å, and indicated that this point was sensitive to conformational change.

Details from polarizable (ff02) force field simulations (figure 47) were also analyzed. Similarly, the simulations system 3 continued to show loss of hydrogen bond interactions within the restrained on DNA simulations. However, all five hydrogen bond distances observed from the simulations system 3 were smooth without changing orientation or any sign of strange patterns. Very stable fluctuations were again found in simulations systems 4 (B1) and (B2), with an average distance of around 3.0 Å for all five hydrogen bond distances. The lowest fluctuations were found

in distance a (the NH at the head group of thiazotropsin A and T⁵O₂) and indicated the strongest hydrogen bond between ligand and DNA. The highest fluctuation was found for distance b (H9 from NH peptide group of thiazotropsin A and A⁶N₃) which indicated the most flexible and sensitive interaction position between ligand and DNA. The results for distance b within the ff02 force field also corresponded to those obtained for the ff03 force field.

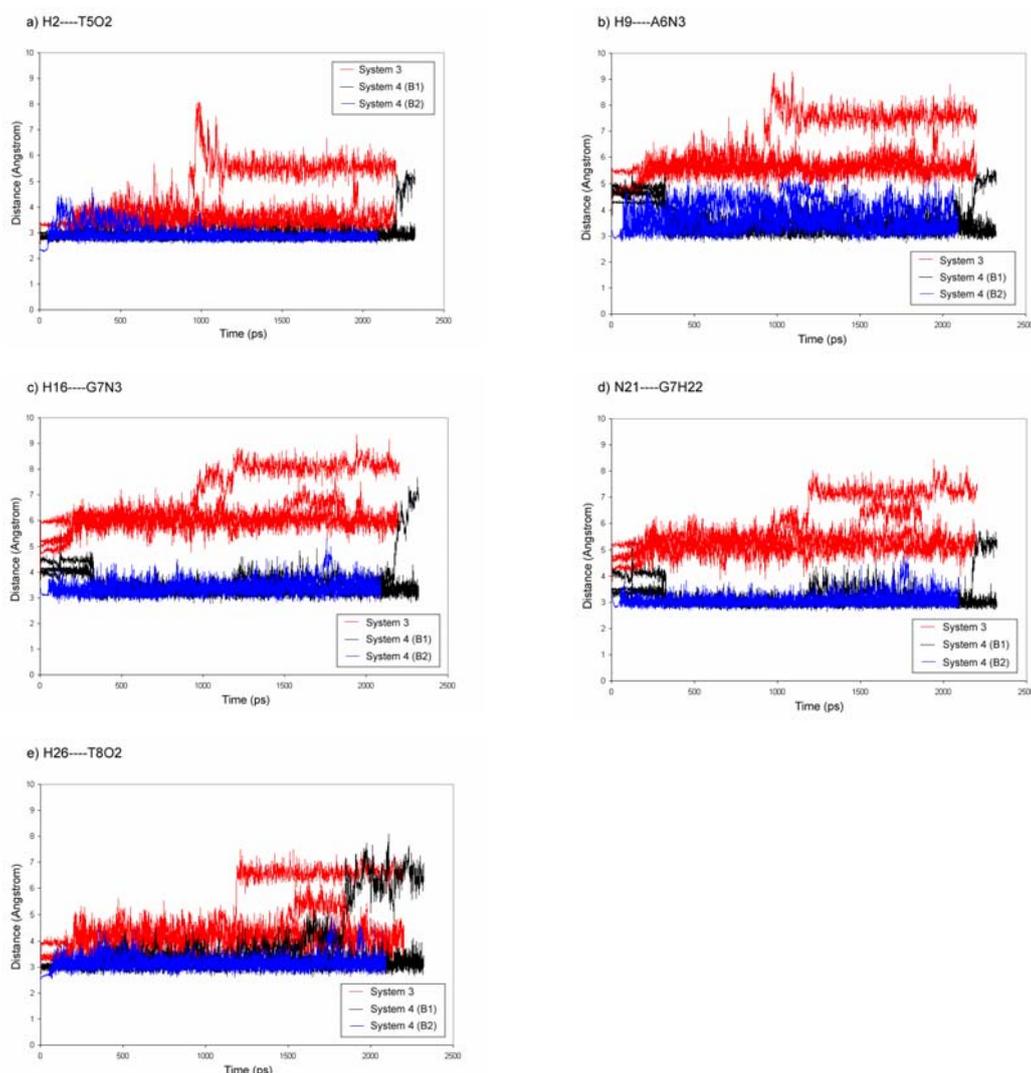


Figure 46 Five hydrogen bonds distance between heteroatom observed in the solvate simulations from MD1, MD2, MD3 and MD4 within non-polarizable (ff03) force field. The simulations system 3 labeled in red, system 4 (B1) is in black and system 4 (B2) is in blue line. a) H2--T⁵O₂, b) H9--A⁶N₃, c) H16--G⁷N₃, d) N21--G⁷H₂₂, e) H26--T⁸O₂.

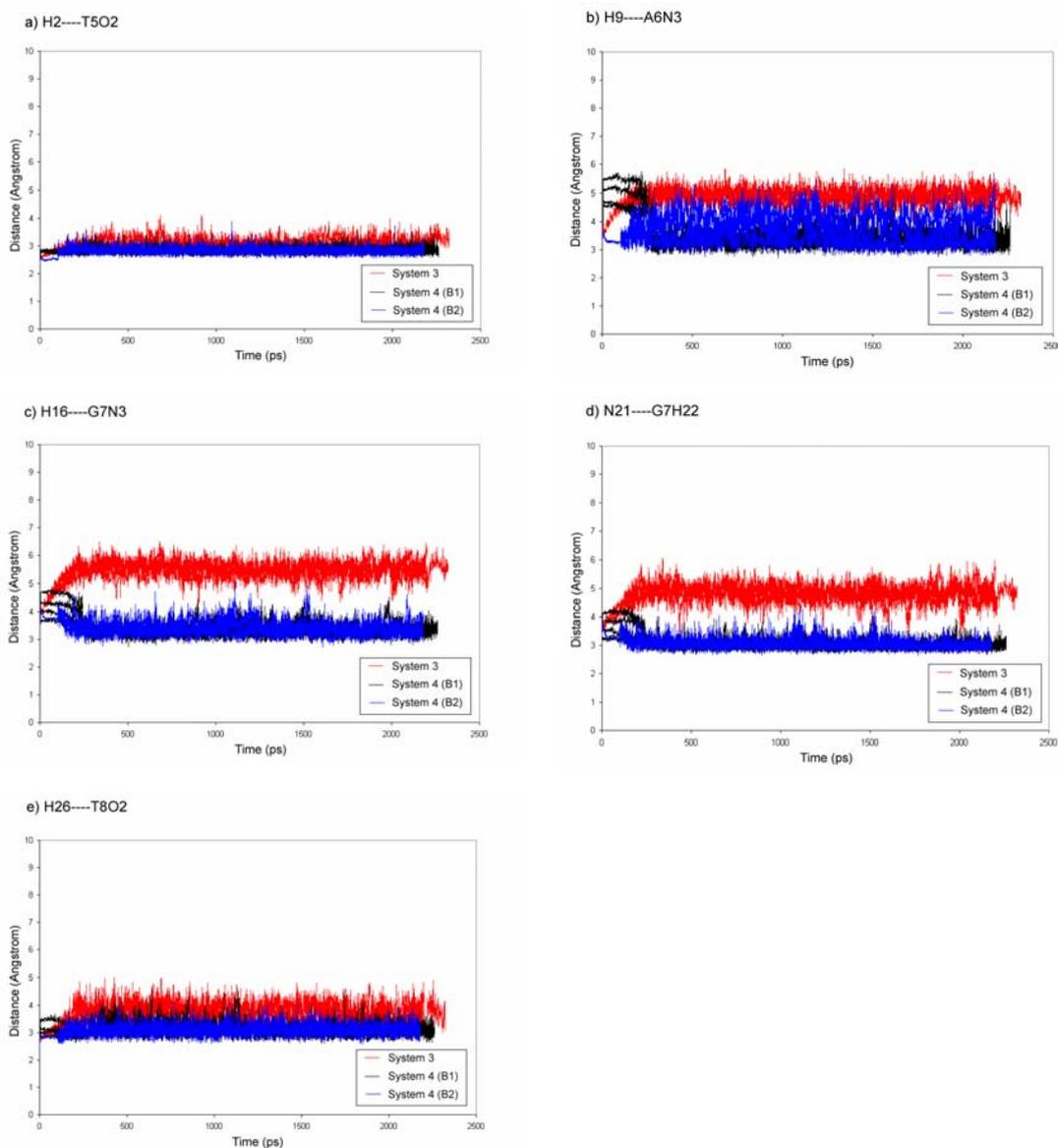


Figure 47 Five hydrogen bonds distance between heteroatom observed in the solvate simulations from MD1, MD2, MD3 and MD4 within non-polarizable (ff02) force field. The simulations system 3 labeled in red, system 4 (B1) is in black and system 4 (B2) is in blue line. a) H2--T⁵O2, b) H9--A⁶N3, c) H16--G⁷N3, d) N21--G⁷H22, e) H26--T⁸O2.

CONCLUSIONS

Molecular dynamics simulations have been used to investigate minor groove binder interactions with DNA. A new class of lexitropsin called thiazotropsin A binds to the DNA duplex minor groove, and is specific for the sequence 5'-CGACTAGTCG-3' in a 2:1 ratio, which was used as the starting structure. All simulations were performed for 2 ns using a combination of topology parameter sets MD1, MD2, MD3 and MD4 for thiazotropsin A, and simulation protocols A, B1 and B2, creating a total of simulation systems 1, 2, 3, 4 (B1) and 4 (B2). Each combination was also performed twice with the complementary non-polarizable (ff03) and polarizable (ff02) force fields for DNA. At least forty trajectories had been performing using the single trajectories analysis.

The equilibration properties were considered first. Although the total energies were well equilibrated, the numbers of water molecules in the simulations system were influential in lowering the total energies; the lowest total energy fluctuations were found in simulations system 5, which contained highest number of water of 5316 molecule, whilst other solvation systems contained only 3879 molecules. Apart from classifying the simulations system, the small variation between B3LYP and HF use represented a different pattern in the total energy fluctuations. Next, structural equilibration was observed by measuring RMSD fluctuations of the solute residues in the complex (DNA and thiazotropsin A) and of DNA and thiazotropsin A alone. RMSD position fluctuations referred to the rotation and vibration mode without translation, and most of the trajectories showed stable RMSD fluctuations over 2 ns. However, the differences could be classified; the non-polarizable force field (ff03) produced lower RMSD fluctuations than observed for the polarizable (ff02) force field. Under restrained conditions; the movement of thiazotropsin A plays a major role and is independent of the combination of topology parameter sets used, whilst the RMSD fluctuations of DNA were all flat. When restraints on DNA were removed, the RMSD fluctuations of DNA were increased dramatically and dominated the RMSD of the whole complex, which indicated an unstable starting structure taken from the NMR refinement process.

CPU times in simulations were considered to investigate the role of the speed of the computer platform in addition to the variations in the simulation systems. The 8 nodes cluster platform performed about 6 times faster than a single CPU computer platform for a simulation size of 12561 atoms. The restrained and unrestrained simulations did not show a difference in terms of CPU time. However, the use of ff03 or ff02 dramatically altered the time taken.

The MM-PBSA calculations have been used in this work with the objective of reproducing the Isothermal Titration Calorimetry (ITC) experimental binding of -10.0 kcal/mol. These major investigations involved sampling from each trajectory, a total of 100 snapshots for representation. Several combinations of force field type and simulation protocols were applied in order to obtain the closest binding free energy via molecular dynamics and MM-PBSA calculations. The calculated binding free energies results revealed the key parameters that would guide later simulations. For example, regarding the effect of force field on the result; firstly, it suggests that HF/6-31G(d) mapped charges for thiazotropsin A are better than the B3LYP/6-31G(d,p) mapped charges; secondly, simple protocols involving a straight succession of minimization, equilibration and production simulation had limitations. One of the most important improvements lay in the repetition of the minimization-equilibration phase before starting any production phase. The simulation protocols that consisted of repeat steps of minimization and equilibration before the production phase were more efficient than the straight forward simulation protocols where minimization, equilibration and production simulations occurred in one round. Finally, determining the binding free energies on the highly flexible DNA complex could be improved by using separate trajectories with the MM-PBSA methodology to accommodate the conformation change energy between free DNA and the complex. Both MD3 and MD4 models with HF/6-31 G(d) mapped charges were found to reproduce the -10.06 kcal/mol under simulations Protocol B2 with separate trajectories but with different force field type applied to DNA in each case (MD3-MD4).

The correlation between models and protocols confirmed that the simulation protocol is less important than the characteristics of the model. Therefore, future simulations of this complex should take good care to set up the models when various simulation protocols can be selected. Apart from the calculated binding free energies, we found limitations of the program package concerning the script used to calculate normal mode analysis. It cannot minimize to tight convergence criteria unless the minimization is performed manually. The methodology to calculate the entropy energies needs to be improved in future calculations with a more robust minimization algorithm. The correlation between models and protocols can be an index of technical methodology comparisons when the Quantitative Protocols Activity Relationship (QPAR) is introduced in the future.

Some structural properties were also analysed concerning the change in binding energies. The Calladine-Drew projections represent the B-DNA form of the starting structure before it changed during the simulations. The DNA structures taken from the simulations showed flexible DNA. The simulations that agreed well with the experimental energies appeared to have high structural adaptation. Unfortunately, when hydrogen bond distances were observed, most of hydrogen bonds in the restrained simulations were lost. The unrestrained simulations kept stable distances throughout the simulations and agreed well with the ^1H NMR observations; our investigation suggested that the strongest interaction is at the NH head group of thiazotropsin A and T⁵O2 base atom of DNA. The interactions between H9 from the NH peptide group of thiazotropsin A and A⁶N3 base atom from DNA appear to break and rejoin easily. The contrast obtained between energetic and structural properties suggest that these highly flexible systems need the separate simulation approach to succeed.

This work provides the basis for studying and developing the unique 2:1 ligand/DNA complex for molecular dynamics simulations. It provides informative structural and energetic determinations as well as suggesting the technical methodological approach for the new type of DNA frame-reading ligand,

thiazotropsin A. Ultimately, it can be used to predict efficient drug therapies for any DNA sequence to tackle the sources of genetic disease.

LITERATURE CITED

- Anthony, N.G., B.F. Johnston, A.I. Khalaf, S.P. Mackay, J.A. Parkinson, C.J. Suckling and R.D. Waigh. 2004. Short Lexitropsin that Recognizes the DNA Minor Groove at 5'-ACTAGT-3': Understanding the Role of Isopropylthiazole. **J. Am. Chem. Soc.** 126(36): 11338-11349.
- Baba, M., H. Tanaka, E. De Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C.F. Perno, R.T. Walker and T. Miyasaka. 1989. Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acycloauridine derivative. **Biochem. Biophys. Res. Commun.** 165(3): 1375-1381.
- Bahar, I., B. Erman, R.L. Jernigan, A.R. Atilgan and D.G. Covell. 1999. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. **J. Mol. Biol.** 285(3): 1023-1037.
- Bailly, C. and J.B. Chaires. 1998. Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. **Bioconjug. Chem.** 9(5): 513-538.
- Barreiro, G., C.R.W. Guimaraes, I. Tubert-Brohman, T.M. Lyons, J. Tirado-Rives and W.L. Jorgensen. 2007. Search for Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase Using Chemical Similarity, Molecular Docking, and MM-GB/SA Scoring. **J. Chem. Inf. Model.** 47(6): 2416-2428.
- Bavand, M.R., R. Wagner and T.J. Richmond. 1993. HIV-1 reverse transcriptase: Polymerization properties of the p51 homodimer compared to the p66/p51 heterodimer. **Biochemistry** 32(40): 10543-10552.

- Beyer, A., L. Lawtrakul, S. Hannongbua and P. Wolschann. 2004. Systematic Investigation of Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase (NNRTIs). **Monatsh. Chem.** 135(8): 1047-1059.
- Blasko, A. and T.C. Bruice. 1993. Stoichiometry and structure of complexes of DNA oligomers with microgonotropens and distamycin by ¹H NMR spectroscopy and molecular modeling. **Proc. Natl. Acad. Sci. USA** 90(21): 10018-10022.
- Brana, M.F., M. Cacho, A. Gradillas, B. De Pascual-Teresa and A. Ramos. 2001. Intercalators as anticancer drugs. **Curr. Pharm. Des.** 7(17): 1745-1780.
- Cadet, J., T. Delatour, T. Douki, D. Gasparutto, E.-P. Pouget, J.-L. Ravanat and S. Sauvaigo. 1999. Hydroxyl radicals and DNA base damage. **Mutat. Res., Fundam. Mol. Mech. Mutagen.** 424(1,2): 9-21.
- Calladine, C.R. and H.R. Drew. 1984. A base-centered explanation of the B-to-A transition in DNA. **J. Mol. Biol.** 178(3): 773-782.
- Cathcart, R., E. Schwiers, R.L. Saul and B.N. Ames. 1984. Thymine glycol and thymidine glycol in human and rat urine: a possible assay for oxidative DNA damage. **Proc. Natl. Acad. Sci. USA** 81(18): 5633-5637.
- Chamberlain, P.P., J. Ren, C.E. Nichols, L. Douglas, J. Lennerstrand, B.A. Larder, D.I. Stuart and D.K. Stammers. 2002. Crystal structures of zidovudine- or lamivudine-resistant human immunodeficiency virus type 1 reverse transcriptases containing mutations at codons 41, 184, and 215. **J. Virol.** 76(19): 10015-10019.
- Charneau, P., G. Mirambeau, P. Roux, S. Paulous, H. Buc and F. Clavel. 1994. HIV-1 reverse transcription. A termination step at the center of the genome. **J. Mol. Biol.** 241(5): 651-662.

- Clark, R.L., F.M. Deane, N.G. Anthony, B.F. Johnston, F.O. McCarthy and S.P. Mackay. 2007. Exploring DNA topoisomerase I inhibition by the benzo[c]phenanthridines fagaronine and ethoxidine using steered molecular dynamics. **Bioorg. Med. Chem.** 15(14): 4741-4752.
- Clausen-Schaumann, H., M. Rief, C. Tolksdorf and H.E. Gaub. 2000. Mechanical stability of single DNA molecules. **Biophys. J.** 78(4): 1997-2007.
- Coll, M., C.A. Frederick, A.H. Wang and A. Rich. 1987. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. **Proc. Natl. Acad. Sci. USA** 84(23): 8385-8389.
- Das, K., A.D. Clark, Jr., P.J. Lewi, J. Heeres, M.R. De Jonge, L.M.H. Koymans, H.M. Vinkers, F. Daeyaert, D.W. Ludovici, M.J. Kukla, B. De Corte, R.W. Kavash, C.Y. Ho, H. Ye, M.A. Lichtenstein, K. Andries, R. Pauwels, M.-P. De Bethune, P.L. Boyer, P. Clark, S.H. Hughes, P.A.J. Janssen and E. Arnold. 2004. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (Etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. **J. Med. Chem.** 47(10): 2550-2560.
- _____, J. Ding, Y. Hsiou, A.D. Clark, Jr., H. Moereels, L. Koymans, K. Andries, R. Pauwels, P.A.J. Janssen, P.L. Boyer, P. Clark, R.H. Smith, Jr., M.B.K. Smith, C.J. Michejda, S.H. Hughes and E. Arnold. 1996. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. **J. Mol. Biol.** 264(5): 1085-1100.
- De Clercq, E. 1995. Antiviral therapy for human immunodeficiency virus infections. **Clin. Microbiol. Rev.** 8(2): 200-239.

- _____. 1995. Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV infections. **J. Med. Chem.** 38(14): 2491-2517.
- Ding, J., K. Das, H. Moereels, L. Koymans, K. Andries, P.A.J. Janssen, S.H. Hughes and E. Arnold. 1995a. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. **Nat. Struct. Biol.** 2(5): 407-415.
- _____, C. Tantillo, W. Zhang, A.D. Clark, Jr., S. Jessen, X. Lu, Y. Ysiou, A. Jacobo-Molina and et al. 1995b. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor α -APA R 95845 at 2.8 Å resolution. **Structure** 3(4): 365-379.
- Douki, T., A. Reynaud-Angelin, J. Cadet and E. Sage. 2003. Bipyrimidine Photoproducts Rather than Oxidative Lesions Are the Main Type of DNA Damage Involved in the Genotoxic Effect of Solar UVA Radiation. **Biochemistry** 42(30): 9221-9226.
- Duan, L.L., Y. Tong, Y. Mei, Q.G. Zhang and J.Z.H. Zhang. 2007. Quantum study of HIV-1 protease-bridge water interaction. **J. Chem. Phys.** 127(14): 145101-145106.
- Duan, Y., C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. **J. Comput. Chem.** 24(16): 1999-2012.
- Ferguson, L.R. and W.A. Denny. 1991. The genetic toxicology of acridines. **Mutat. Res., Rev. Genet. Toxicol.** 258(2): 123-160.

- Fogolari, F., A. Brigo and H. Molinari. 2003. Protocol for MM/PBSA Molecular Dynamics Simulations of Proteins. **Biophys. J.** 85: 159–166.
- Gao, H. and Y. Kong. 2004. Simulation of DNA-nanotube interactions. **Annu. Rev. Mater. Res.** 34: 123-150.
- Goodsell, D.S., H.L. Ng, M.L. Kopka, J.W. Lown and R.E. Dickerson. 1995. Structure of a dicationic monoimidazole lexitropsin bound to DNA. **Biochemistry** 34(51): 16654-16661.
- Haq, I., J.E. Ladbury, B.Z. Chowdhry, T.C. Jenkins and J.B. Chaires. 1997. Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)₂ duplex: calorimetric and spectroscopic studies. **J. Mol. Biol.** 271(2): 244-257.
- Hargrave, K.D., J.R. Proudfoot, K.G. Grozinger, E. Cullen, S.R. Kapadia, U.R. Patel, V.U. Fuchs, S.C. Mauldin, J. Vitous and et al. 1991. Novel non-nucleoside inhibitors of HIV-1 reverse transcriptase. 1. Tricyclic pyridobenzo- and dipyridodiazepinones. **J. Med. Chem.** 34(7): 2231-2241.
- He, X., Y. Mei, Y. Xiang, D.W. Zhang and J.Z.H. Zhang. 2005. Quantum Computational Analysis for Drug Resistance of HIV-1 Reverse Transcriptase to Nevirapine Through Point Mutations. **PROTEINS: Struct., Funct., Bioinf.** 61: 423–432.
- Hermann, T. and H. Heumann. 1996. Strained template under the thumbs. How reverse transcriptase of human immunodeficiency virus type 1 moves along its template. **Eur. J. Biochem.** 242(1): 98-103.
- Hsiou, Y., K. Das, J. Ding, A.D. Clark, Jr., J.-P. Kleim, M. Rosner, I. Winkler, G. Riess, S.H. Hughes and E. Arnold. 1998. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside

inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. **J. Mol. Biol.** 284(2): 313-323.

_____, J. Ding, K. Das, A.D. Clark, Jr., S.H. Hughes and E. Arnold. 1996. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. **Structure** 4(7): 853-860.

Jacobo-Molina, A., J. Ding, R.G. Nanni, A.D. Clark, Jr., X. Lu, C. Tantillo, R.L. Williams, G. Kamer, A.L. Ferris and et al. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. **Proc. Natl. Acad. Sci. USA** 90(13): 6320-6324.

Jager, J., S.J. Smerdon, J. Wang, D.C. Boisvert and T.A. Steitz. 1994. Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. **Structure** 2(9): 869-876.

James, P.L., E.E. Merkina, A.I. Khalaf, C.J. Suckling, R.D. Waigh, T. Brown and K.R. Fox. 2004. DNA sequence recognition by an isopropyl substituted thiazole polyamide. **Nucleic Acids Res.** 32(11): 3410-3417.

Jeffrey, A.M. 1985. DNA modification by chemical carcinogens. **Pharmacol. Ther.** 28(2): 237-272.

Katz, R.A., P. DiCandeloro, G. Kukolj and A.M. Skalka. 2001. Role of DNA end distortion in catalysis by avian sarcoma virus integrase. **J. Biol. Chem.** 276(36): 34213-34220.

_____, and A.M. Skalka. 1994. The retroviral enzymes. **Annu. Rev. Biochem.** 63: 133-173.

- Kensch, O., T. Restle, B.M. Wohrl, R.S. Goody and H.-J. Steinhoff. 2000. Temperature-dependent Equilibrium between the Open and Closed Conformation of the p66 Subunit of HIV-1 Reverse Transcriptase Revealed by Site-directed Spin Labelling. **J. Mol. Biol.** 301(4): 1029-1039.
- Kim, J.T., A.D. Hamilton, C.M. Bailey, R.A. Domoal, L. Wang, K.S. Anderson and W.L. Jorgensen. 2006. FEP-Guided Selection of Bicyclic Heterocycles in Lead Optimization for Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase. **J. Am. Chem. Soc.** 128(48): 15372-15373.
- Klevit, R.E., D.E. Wemmer and B.R. Reid. 1986. Proton NMR studies on the interaction between distamycin A and a symmetrical DNA dodecamer. **Biochemistry** 25(11): 3296-3303.
- Kohlstaedt, L.A., J. Wang, J.M. Friedman, P.A. Rice and T.A. Steitz. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. **Science** 256(5065): 1783-1790.
- Kopka, M.L., C. Yoon, D. Goodsell, P. Pjura and R.E. Dickerson. 1985. Binding of an antitumor drug to DNA. Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. **J. Mol. Biol.** 183(4): 553-563.
- _____. 1985. The molecular origin of DNA-drug specificity in netropsin and distamycin. **Proc. Natl. Acad. Sci. USA** 82(5): 1376-1380.
- Koup, R.A., V.J. Merluzzi, K.D. Hargrave, J. Adams, K. Grozinger, R.J. Eckner and J.L. Sullivan. 1991. Inhibition of human immunodeficiency virus type 1 (HIV-1) replication by the dipyrroldiazepinone BI-RG-587. **J. Infect. Diseases** 163(5): 966-970.
- Kroeger Smith, M.B., L.H. Rader, A.M. Franklin, E.V. Taylor, K.D. Smith, R.H. Smith, Jr., J. Tirado-Rives and W.L. Jorgensen. 2008. Energetic effects for

observed and unobserved HIV-1 reverse transcriptase mutations of residues L100, V106, and Y181 in the presence of nevirapine and efavirenz. **Bioorg. Med. Chem. Lett.** 18(3): 969-972.

Kuhn, B., P. Gerber, T. Schulz-Gasch and M. Stahl. 2005. Validation and Use of the MM-PBSA Approach for Drug Discovery. **J. Med. Chem.** 48: 4040-4048.

Kumar, S., Y. Bathini, T. Joseph, R.T. Pon and J.W. Lown. 1991. Structural and dynamic aspects of non-intercalative (1:1) binding of a thiazole-lexitropsin to the decaoxyribonucleotide d-[CGCAATTGCG]₂: a proton NMR and molecular modeling study. **J. Biomol. Struct. Dyn.** 9(1): 1-21.

Kuno, M., R. Hongkrekai and S. Hannongbua. 2006. ONIOM-BSSE scheme for H. π system and applications on HIV-1 reverse transcriptase. **Chem. Phys. Lett.** 424: 172-177.

Le Grice, S.F.J., T. Naas, B. Wohlgensinger and O. Schatz. 1991. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. **EMBO J.** 10(12): 3905-3911.

Lee, M.C. and Y. Duan. 2004. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the Generalized Born solvent model. **Proteins: Struct., Funct., Bioinf.** 55(3): 620-634.

Lown, J.W. 1988. Lexitropsins: rational design of DNA sequence reading agents as novel anti-cancer agents and potential cellular probes. **Anticancer Drug Des.** 3(1): 25-40.

_____. 1994. DNA recognition by lexitropsins, minor groove binding agents. **J. Mol. Recognit.** 7(2): 79-88.

- Ludovici, D.W., M.J. Kukla, P.G. Grous, S. Krishnan, K. Andries, M.-P.d.B. thune, H. Azijn, R. Pauwels, E.D. Clercq, E. Arnold and P.A.J. Janssene. 2001. Evolution of Anti-HIV Drug Candidates. Part 1: From β -Anilinophenylacetamide (β -APA) to Imidoyl Thiourea (ITU). **Bioorg. Med. Chem. Lett.** 11: 2225–2228.
- Madrid, M., A. Jacobo-Molina, J. Ding and E. Arnold. 1999. Major subdomain rearrangement in HIV-1 reverse transcriptase simulated by molecular dynamics. **Proteins** 35(3): 332-337.
- _____, J.A. Lukin, J.D. Madura, J. Ding and E. Arnold. 2001. Molecular dynamics of HIV-1 reverse transcriptase indicates increased flexibility upon DNA binding. **Proteins: Struct., Funct., Genet.** 45(3): 176-182.
- Maga, G., M. Amacker, N. Ruel, U. Hubscher and S. Spadari. 1997. Resistance to nevirapine of HIV-1 reverse transcriptase mutants: loss of stabilizing interactions and thermodynamic or steric barriers are induced by different single amino acid substitutions. **J. Mol. Biol.** 274(5): 738-747.
- Marky, L.A. and K.J. Breslauer. 1987. Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data. **Proc. Natl. Acad. Sci. USA** 84(13):4359-4363.
- Massova, I. and P.A. Kollman. 1999. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. **J. Am. Chem. Soc.** 121(36): 8133-8143.
- Medina-Franco, J.L., S. Rodriguez-Morales, C. Juarez-Gordiano, A. Hernandez-Campos, J. Jimenez-Barbero and R. Castillo. 2004. Flexible docking of pyridinone derivatives into the non-nucleoside inhibitor binding site of HIV-1 reverse transcriptase. **Bioorg. Med. Chem.** 12(23): 6085-6095.

- Merluzzi, V.J., K.D. Hargrave, M. Labadia, K. Grozinger, M. Skoog, J.C. Wu, C.K. Shih, K. Eckner, S. Hattox and J. Adams. 1990. Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor. **Science** 250(4986): 1411-1413.
- Mitsuya, H., R. Yarchoan, S. Hayashi and S. Broder. 1990. Antiviral therapy against HIV infection. **J. Am. Acad. Dermatol** 22: 1282-1294.
- Nunriam, P., M. Kuno, S. Saen-oon and S. Hannongbua. 2005. Particular interaction between efavirenz and the HIV-1 reverse transcriptase binding site as explained by the ONIOM2 method. **Chem. Phys. Lett.** 405(1-3): 198-202.
- Onufriev, A., D. Bashford and D.A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified Generalized Born model. **Proteins: Struct., Funct., Bioinf.** 55(2): 383-394.
- Pabo, C.O. and R.T. Sauer. 1984. Protein-DNA recognition. **Annu. Rev. Biochem.** 53: 293-321.
- Pauwels, R., K. Andries, J. Desmyter, D. Schols, M.J. Kukla, H.J. Breslin, A. Raeymaeckers, J. Van Gelder, R. Woestenborghs and J. Heykants. 1990. Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives. **Nature** 343(6257): 470-474.
- Pedersen, O.S. and E.B. Pedersen. 1999. Non-nucleoside reverse transcriptase inhibitors: the NNRTI boom. **Antiviral Chem. Chemother.** 10(6): 285-314.
- Pelton, J.G. and D.E. Wemmer. 1988. Structural modeling of the distamycin A-(CGCGAATTCGCG)₂ complex using 2D NMR and molecular mechanics. **Biochemistry** 27(21): 8088-8096.

- _____. 1989. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. **Proc. Natl. Acad. Sci. USA** 86(15): 5723-5727.
- Ponnuswamy, P.K. and M.M. Gromiha. 1994. On the conformational stability of oligonucleotide duplexes and tRNA molecules. **J. Theor. Biol.** 169(4): 419-432.
- Pullman, A. and B. Pullman. 1981. Molecular electrostatic potential of the nucleic acids. **Quarterly Rev. Biophys.** 14(3): 289-380.
- Rao, A., J. Balzarini, A. Carbone, A. Chimirri, E.D. Clercq, A.M. Monforte, P. Monforte, C. Pannecouque and M. Zappalà. 2004. 2-(2,6-Dihalophenyl)-3-(pyrimidin-2-yl)-1,3-thiazolidin-4-ones as non-nucleoside HIV-1 reverse transcriptase inhibitors. **Antiviral Res.** 63: 79-84.
- Rao, K.E., Y. Bathini and J.W. Lown. 1990a. Synthesis of novel thiazole-containing DNA minor groove binding oligopeptides related to the antibiotic distamycin. **J. Org. Chem.** 55(2): 728-737.
- _____, R.G. Shea, B. Yadagiri and J.W. Lown. 1990b. Molecular recognition between oligopeptides and nucleic acids: DNA sequence specificity and binding properties of thiazole-lexitropsins incorporating the concepts of base site acceptance and avoidance. **Anti-Cancer Drug Des.** 5(1): 3-20.
- Rawal, R.K., A. Kumar, M.I. Siddiqi and S.B. Katti. 2007. Molecular docking studies on 4-thiazolidinones as HIV-1 RT inhibitors. **J. Mol. Model.** 13(1): 155-161.
- Ren, J., R. Esnouf, A. Hopkins, C. Ross, Y. Jones, D. Stammers and D. Stuart. 1995. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. **Structure** 3(9): 915-926.

- _____, C. Nichols, L. Bird, P. Chamberlain, K. Weaver, S. Short, D.I. Stuart and D.K. Stammers. 2001. Structural Mechanisms of Drug Resistance for Mutations at Codons 181 and 188 in HIV-1 Reverse Transcriptase and the Improved Resilience of Second Generation Non-nucleoside Inhibitors. **J. Mol. Biol.** 312(4): 795-805.
- Rentzeperis, D., L.A. Marky, T.J. Dwyer, B.H. Geierstanger, J.G. Pelton and D.E. Wemmer. 1995. Interaction of minor groove ligands to an AAATT/AATTT site: correlation of thermodynamic characterization and solution structure. **Biochemistry** 34(9): 2937-2945.
- Restle, T., B. Mueller and R.S. Goody. 1990. Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. **J. Biol. Chem.** 265(16): 8986-8988.
- Reyes, C.M. and P.A. Kollman. 2000. Structure and Thermodynamics of RNA-protein Binding: Using Molecular Dynamics and Free Energy Analyses to Calculate the Free Energies of Binding and Conformational Change. **J. Mol. Biol.** 297: 1145-1158.
- Richman, D., A.S. Rosenthal, M. Skoog, R.J. Eckner, T.C. Chou, J.P. Sabo and V.J. Merluzzi. 1991. BI-RG-587 is active against zidovudine-resistant human immunodeficiency virus type 1 and synergistic with zidovudine. **Antimicrob. Agents Chemother.** 35(2): 305-308.
- Rittinger, K., G. Divita and R.S. Goody. 1995. Human immunodeficiency virus reverse transcriptase substrate-induced conformational changes and the mechanism of inhibition by nonnucleoside inhibitors. **Proc. Natl. Acad. Sci. USA** 92(17): 8046-8049.

- Rizzo, R.C., J. Tirado-Rives and W.L. Jorgensen. 2001. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. **J. Med. Chem.** 44: 145-154.
- _____, D.-P. Wang, J. Tirado-Rives and W.L. Jorgensen. 2000. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Sustiva through Computation of Resistance Profiles. **J. Am. Chem. Soc.** 122(51): 12898-12900.
- Rodgers, D.W., S.J. Gamblin, B.A. Harris, J.S. Culp, B. Hellmig, D.J. Woolf, C. Debouck, S.C. Harrison and S. Ray. 1995. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. **Proc. Natl. Acad. Sci. USA** 92(4): 1222-1226.
- Rodríguez-Barrios, F.T., J. Balzarini and F. Gago. 2005. The Molecular Basis of Resilience to the Effect of the Lys103Asn Mutation in Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors Studied by Targeted Molecular Dynamics Simulations. **J. Am. Chem. Soc.** 127: 7570-7578.
- _____, and F. Gago. 2004. Understanding the Basis of Resistance in the Irksome Lys103Asn HIV-1 Reverse Transcriptase Mutant through Targeted Molecular Dynamics Simulations. **J. Am. Chem. Soc.** 126: 15386-15387.
- Saen-oon, S., M. Kuno and S. Hannongbua. 2005. Binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-Cl TIBO complex structures: quantum chemical calculations based on the ONIOM method. **Proteins** 61(4): 859-869.
- Saparpakorn, P., S. Hannongbua and D. Rognan. 2006. Design of nevirapine derivatives insensitive to the K103N and Y181C HIV-1 reverse transcriptase mutants. **SAR QSAR Environ. Res.** 17(2): 183-194.

- Sarafianos, S.G., K. Das, J. Ding, Y. Hsiou, S.H. Hughes and E. Arnold. 1997. Current perspectives on mechanisms of HIV-1 reverse transcriptase inhibition by nonnucleoside inhibitors. **Royal Soc. Chem.** 198: 328-334.
- Shaikh, S.A., S.R. Ahmed and B. Jayaram. 2004. A molecular thermodynamic view of DNA–drug interactions: a case study of 25 minor-groove binders. **Arch. Biochem. Biophys.** 429: 81–99.
- Shigenaga, M.K., C.J. Gimeno and B.N. Ames. 1989. Urinary 8-hydroxy-2'-deoxyguanosine as a biological marker of in vivo oxidative DNA damage. **Proc. Natl. Acad. Sci. USA** 86(24): 9697-9701.
- Singh, S.B., Ajay, D.E. Wemmer and P.A. Kollman. 1994. Relative binding affinities of distamycin and its analog to d(CGCAAGTTGGC).d(GCCAACTTGCG): comparison of simulation results with experiment. **Proc. Natl. Acad. Sci. USA** 91(16): 7673-7677.
- Sitkoff, D., K.A. Sharp and B. Honig. 1994. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. **J. Phys. Chem.** 98(7): 1978-1988.
- Smith, M.B.K., M.L. Lamb, J. Tirado-Rives, W.L. Jorgensen, C.J. Michejda, S.K. Rigny and R.H. Smith Jr. 2000. Monte Carlo calculations on HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor 8-Cl TIBO: contribution of the L100I and Y181C variants to protein stability and biological activity. **Protein Eng.** 13(6): 413-421.
- Spackova, N.a., T.E. Cheatham III, F. Ryja'cek, F. Lankas, L.v. Meervelt, P. Hobza and J. Sponer. 2003. Molecular Dynamics Simulations and Thermodynamics Analysis of DNA-Drug Complexes. Minor Groove Binding between 4 ϵ ,6-Diamidino-2-phenylindole and DNA Duplexes in Solution. **J. Am. Chem. Soc.** 125: 1759-1769.

- Spallarossa, A., S. Cesarini, A. Ranise, M. Ponassi, T. Unge and M. Bolognesi. 2008. Crystal structures of HIV-1 reverse transcriptase complexes with thiocarbamate non-nucleoside inhibitors. **Biochem. Biophys. Res. Commun.** 365(4): 764-770.
- Spence, R.A., K.S. Anderson and K.A. Johnson. 1996. HIV-1 Reverse Transcriptase Resistance to Nonnucleoside Inhibitors. **Biochemistry** 35: 1054-1063.
- Srinivasan, J., T.E. CheathamIII, P. Cieplak, P.A. Kollman and D.A. Case. 1998. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. **J. Am. Chem. Soc.** 120(37): 9401-9409.
- Stephens, T.D., C.J.W. Bunde and B.J. Fillmore. 2000. Mechanism of action in thalidomide teratogenesis. **Biochem. Pharmacol.** 59(12): 1489-1499.
- Tantillo, C., J. Ding, A. Jacobo-Molina, R.G. Nanni, P.L. Boyer, S.H. Hughes, R. Pauwels, K. Andries, P.A.J. Janssen and E. Arnold. 1994. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. **J. Mol. Biol.** 243(3): 369-387.
- Temiz, N.A. and I. Bahar. 2002. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. **Proteins: Struct., Funct., Genet.** 49(1): 61-70.
- Udier-Blagovic, M., J. Tirado-Rives and W.L. Jorgensen. 2003. Validation of a model for the complex of HIV-1 reverse transcriptase with nonnucleoside inhibitor TMC125. **J. Am. Chem. Soc.** 125(20): 6016-6017.

- _____. 2004. Structural and Energetic Analyses of the Effects of the K103N Mutation of HIV-1 Reverse Transcriptase on Efavirenz Analogues. **J. Med. Chem.** 47(9): 2389-2392.
- Valerie, K. and L.F. Povirk. 2003. Regulation and mechanisms of mammalian double-strand break repair. **Oncogene** 22(37): 5792-5812.
- Vercoutere, W. and M. Akeson. 2002. Biosensors for DNA sequence detection. **Curr. Opin. Chem. Biol.** 6(6): 816-822.
- Veronese, F.d.M., T.D. Copeland, A.L. DeVico, R. Rahman, S. Oroszlan, R.C. Gallo and M.G. Sarngadharan. 1986. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. **Science** 231(4743): 1289-1291.
- Wang, J., X. Kang, I.D. Kuntz and P.A. Kollman. 2005. Hierarchical Database Screenings for HIV-1 Reverse Transcriptase Using a Pharmacophore Model, Rigid Docking, Solvation Docking, and MM-PB/SA. **J. Med. Chem.** 48(7): 2432-2444.
- _____, P. Morin, W. Wang and P.A. Kollman. 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. **J. Am. Chem. Soc.** 123: 5221-5230.
- Watson, J.D. and F.H.C. Crick. 1953. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. **Nature** 171: 737-738.
- Weinzinger, P., S. Hannongbua and P. Wolschann. 2005. Molecular mechanics PBSA ligand binding energy and interaction of Efavirenz derivatives with HIV-1 reverse transcriptase. **J. Enzyme Inhib. Med. Chem.** 20(2): 129–134.

- Wemmer, D.E. 2001. Ligands recognizing the minor groove of DNA: development and applications. **Biopolymers** 52(4): 197-211.
- Whitcomb, J.M. and S.H. Hughes. 1992. Retroviral reverse transcription and integration: progress and problems. **Annu. Rev. Cell Biol.** 8: 275-306.
- Wing, R., H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson. 1980. Crystal structure analysis of a complete turn of B-DNA. **Nature** 287(5784): 755-758.
- Young, S.D., S.F. Britcher, L.O. Tran, L.S. Payne, W.C. Lumma, T.A. Lyle, J.R. Huff, P.S. Anderson, D.B. Olsen and S.S. Carroll. 1995. L-743, 726 (DMP-266): a novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. **Antimicrob. Agents Chemother.** 39(12): 2602-2605.
- Zhai, J., C. Hong and R. Yang. 1997. DNA based biosensors. **Biotechnology Advances** 15(1): 43-58.
- Zhang, Y., F.X. Chen, P. Mehta and B. Gold. 1993. Groove- and sequence-selective alkylation of DNA by sulfonate esters tethered to lexitropsins. **Biochemistry** 32(31): 7954-7965.
- Zhou, Z., M. Madrid, J.D. Evanseck and J.D. Madura. 2005. Effect of a Bound Non-Nucleoside RT Inhibitor on the Dynamics of Wild-Type and Mutant HIV-1 Reverse Transcriptase. **J. Am. Chem. Soc.** 127(49): 17253-17260.
- _____, and J.D. Madura. 2002. Docking of non-nucleoside inhibitors: Neotripterifordin and its derivatives to HIV-1 reverse transcriptase. **Proteins: Struct., Funct., Genet.** 49(4): 529-542.

Zhou, Z. and J.D. Madura. 2004. Relative Free Energy of Binding and Binding Mode Calculations of HIV-1 RT Inhibitors Based on Dock-MM-PB/GS. **Proteins: Struct., Funct., Bioinf.** 57: 493–503.

CURRICULUM VITAE

PERSONAL DATA

Name: Mr. Witcha Treesuwan
 Date of birth: January 11, 1979
 Place of birth: Bangkok, THAILAND
 Nationality: Thai

EDUCATION

: YEAR	INSTITUTION	DEGREE
1997-2000	Kasetsart university	B.Sc. (Chemistry)
2001-2003	Kasetsart university	M.Sc. (Chemistry)
2004-2008	Kasetsart university	Ph.D. (Chemistry)

SCHOLARSHIP : Higher Education Development project
 Scholarship Postgraduate Education and Research
 Program in Physical Chemistry (2001-2003).
 : Royal Golden Jubilee Scholarship (2004-2007).

THESIS TOPICS

M.Sc. (Chemistry): Investigation on flexibility of mutated Y181C HIV-1 Reverse Transcriptase enzyme by means of molecular dynamics simulations.

Ph.D. (Chemistry): Molecular dynamics simulations of mutant HIV-1 RT (K103N) and effect of enzyme flexibility to the enzymatic reaction.

PROGRAM SPECIALIST

1. Linux operation system
2. Window operation system: Microsoft office
3. AMBER program package
4. Gaussian program package
5. Sybyl program

6. Perl programming
7. X3DNA program
8. Raster3D program
9. DS Viewer program
10. HyperChem program
11. CS ChemOffice

WORKSHOP PARTICIPATIONS

AutoDock program, Gromac, DS material studio, InsightII, Catalyzt.

WORKING EXPERIENCE

1. Molecular dynamics simulation on the macro protein: HIV-1 Reverse Transcriptase.
2. Molecular dynamics simulation on the small model: the pocket of HIV-1 Reverse Transcriptase.
3. Molecular dynamics simulation on the isolate organic molecule: nevirapine and thiazotropsinA.
4. Molecular dynamics simulation on the DNA minor groove binder complex: d(GCGACTAGTCGC)₂/ThiazotropsinA.
5. Simulation protocols evaluation.
6. Energetic and structural parameters calculations.

PUBLICATIONS

Lee, V.S., K. Wittayanarakul, T. Remsungnen, V. Parasuk, P. Sompornpisut, W. Chantratita, C. Sangma, S. Vannarat, P. Srichaikul, S. Hannongbua, P. Saparpakorn, W. Treesuwan, O. Aruksakulwong, E. Pasomsub, S. Promsri, D. Chuakheaw and S. Hannongbua. 2003. **Structure and dynamics of SARS coronavirus proteinase: the primary key to the designing and screening for anti-SARS drugs.** *ScienceAsia* 29(2):181-188.

Vailikhit, V., W. Treesuwan and S. Hannongbua. 2007. **A combined MD-ONIOM2 approach for ¹H NMR chemical shift calculations including a polar solvent.** *THEOCHEM* 806(1-3):99-104.

Witcha Treesuwan and Supa Hannongbua. 2008. Binding Efficiency of Nevirapine complex to wild type and Y181C HIV-1 RT via Role of Bridge Water: MM-PBSA calculations. Submitted.

PRESENTATIONS

- | | |
|----------------------|---|
| 6- 8 Feb, 2002 | The international conference on Bioinformatics 2002 (InCOB 2002), Le Royal Meridien, Bangkok, Thailand. |
| 24-26 Mar, 2003 | The seventh Annual National Symposium on Computational Science and Engineering (ANSCSE 2003), Chulalongkorn University, Bangkok, Thailand. |
| 20-22 Oct, 2003 | 29 th Congress on Science and Technology of Thailand, Golden Jubilee Convention Hall, Khon Kaen, Thailand. |
| 1 Mar, 2004 | RGJ Seminar Series XXVII: Nano Bioscience I, Mahidol University, Thailand. |
| 12-15 May, 2004 | 1 st Asian Pacific Conference on Theoretical & Computational Chemistry (APACTC), Okazaki, Japan. |
| 21-23 July, 2004 | The eighth Annual National Symposium on Computational Science and Engineering (ANSCSE 2004), Suranaree University of Technology, Nakhon Ratchasima, Thailand. |
| 20-22 Mar, 2007 | RGJ-Ph.D. Congress VIII, Jomtein Palm Beach Resort, Pattaya, Thailand. |
| 30 Jan – 1 Feb, 2008 | Pure and Applied Chemistry International Conference (PACCON) 2008, Sofitel Centara Grand, Bangkok, Thailand |