MOLECULAR DYNAMICS STUDY OF THE CONFORMATIONAL DYNAMICS OF HIV-1 PROTEASE SUBTYPES A, B, C, AND F

By

T. DWIGHT MCGEE JR.

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2010

© 2010 T. Dwight McGee Jr.

To Danielle Haywood and my supportive family

ACKNOWLEDGEMENTS

It has been through the love and sacrifice of others that I have been able to obtain so much in my life. I would like to commend Dr. Roitberg for his continuous efforts to make sure that I was equipped with the necessary tools to succeed. I appreciate most of all the patience and understanding that Dr. Roitberg has shown me, which has greatly helped to nurture my development.

I would like to acknowledge my committee who has not only been instrumental in my development as a scientist, but as a person as well. I appreciate your patience and thoughtful advice. I am grateful to Dr. Dunn for his guidance and the opportunity to work on a project of such magnitude. I appreciate Dr. Fanucci for your advice and positive reassurance of my work and progress. I would also like to extend my thanks to Dr. Merz for challenging me in and out of the classroom and for your willingness to see me succeed. Dr. Horenstein, I appreciate how your classes strengthen my knowledge of biochemistry, which has had a profound impact on my research. I am indebted to Dr. Edwards for his invaluable advice and inspirational messages. You have been a valuable mentor.

I would like to acknowledge my parents Terry and Sylvia McGee who have been unbelievably understanding and encouraging in my pursuit of a higher education. Throughout my life, my parents have been a continuous stream of motivation and wisdom. I would like to thank all of my friends and family especially Chelsey Rodgers and Golden Marzett. Finally, I would like to acknowledge my fiancée Danielle Haywood who has been nothing but supportive throughout this entire process. I appreciate most of all your understanding on the demands, which graduate school places on my time and life.

TABLE OF CONTENTS

page

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
ARG	Arginine
ASP	Aspartate
EPR	Electron Paramagnetic Resonance
FDA	Federal Drug Administration
GB	Generalized Born
GLU	Glutamate
GLY	Glycine
gp	Glycoprotein
HIV	Human Immunodeficiency Virus
ILE	Isoleucine
LYS	Lysine
MD	Molecular Dynamics
MM	Molecular Mechanics
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
RMSD	Root Mean Squared Deviation
THR	Threonine
WHO	World Health Organization
UNAIDS	Joint United Nations Programs on HIV/AIDS

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

MOLECULAR DYNAMICS STUDY OF THE CONFORMATIONAL DYNAMICS OF HIV-1 PROTEASE SUBTYPES A, B, C, and F

By

T. Dwight McGee Jr.

December 2010

Chair: Adrian E. Roitberg
Major: Chemistry

Each year AIDS is responsible for millions of deaths worldwide. One of the major targets in anti-HIV therapeutics is protease inhibition. The HIV protease plays a critical role in the reproductive process of the virus, and its inhibition would prevent the maturation and spread of the virus to neighboring cells. Previous studies have shown that, of approved FDA protease inhibitors, HIV-1 subtype B protease is more responsive to drug therapy than that of subtype C, F or A. Using Molecular Dynamics, we explore how the different sequences of subtypes A, B, C, and F determines their preferred conformation. A better understanding of the dynamic motion of the HIV protease would allow researchers to potentially develop new compounds to fight the HIV-1 subtypes C, F, and A viruses.

CHAPTER 1
INTRODUCTION

1.1 HIV-1 Epidemic

Human Immunodeficiency Virus (HIV) is a lentivirus, which is responsible for causing Acquired Immunodeficiency Disease Syndrome (AIDS)[1]. Recent advances in antiretroviral therapy have improved the life expectancy of those infected but have had little impact on controlling the spread of the HIV virus. The World Health Organization (WHO) considers AIDS a pandemic. According to data provided by the Joint United Nations Program on HIV/AIDS (UNAIDS) and WHO, an estimated 33.4 million people are currently living with the disease as of 2008. Sub-Saharan Africa accounts for 22.4 million of the world's HIV population. In 2008, approximately 2 million deaths were due to AIDS and 1.4 million of total deaths were in Sub-Saharan Africa. This startling fact can be attributed to the inability of these countries to afford antiretroviral treatment, lack of education, social stigmatism and etc.

1.2 Genetic Variability

The high genetic variability of HIV can be attributed to the activity of the enzyme reverse transcriptase, which will be discussed in later sections. HIV can be divided into HIV-1 and HIV-2. HIV-1 can be divided into groups M "main", N "non-M/non-O", O "outlier", and the newly discovered P[2,3,4,5]. Group M is the largest and accounts for the majority of infections. This group is further segregated into subtypes A-D, F-H, J, K and circulating recombinant forms (CRF) such as CRF01_AE[2]. The subtypes' relationships and classifications were based on sequence alignments of gag, pol, and env regions[2,6,7].

Subtype A is predominant in West and Central Africa. Subtype B is dominant in North and South America, Europe and Australia. Subtype C can be found in Southern and East Africa, and throughout Eastern Asia. Subtype D is limited mainly to Central Africa. Subtype F is generally

found in South America and Central Africa. Subtype H has been seen only in Central Africa and subtype J in Central America. Lastly, subtype K has only been observed in the Congo and Cameroon. Subtypes A, B, C, and D account for the majority of infections worldwide. Subtype C accounts for the highest percentage of those with HIV[8]. However, it has been noted that those persons infected with subtype D develop AIDS at a much faster rate and have a higher mortality rate[9].

1.3 HIV-1 Genome

HIV is ~120 nm in diameter and has a spherical shape. It is comprised of two copies of single stranded RNA enclosed by the viral protein p24 (matrix). Its genome is rather small and is ~9.4 Kb in length. The gag gene encodes the structural proteins matrix (MA), capsid (CA), and nucleocapsid (NC). The two envelope proteins g41, gp120 are encoded by the *env*, which is responsible for viral tropism. The *pol* gene encodes for the three enzymes, reverse transcriptase, integrase and protease. The function and importance of these enzymes will be discussed in later sections of this chapter. The other genes *rev*, *nef*, *vpr*, *vpu*, *vif*, and, *tat* encode for the six accessory proteins. Three of the accessory proteins, *tat*, *rev*, and *vif*, play a role in the replication of the virus[10,11]. The other three accessory proteins *vpr*, *nef*, *vpu* play a role in regulating viral promoters, acting as a negative regulatory factor for viral expression and promoting the CD4 degradation and viral release[12,13].

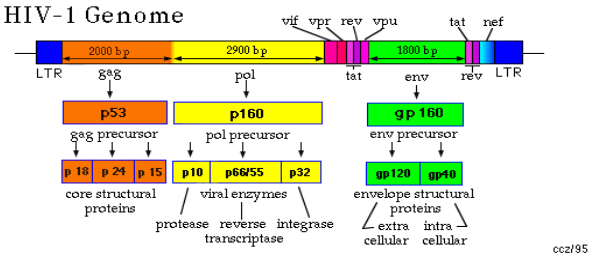

Figure 1-1. Schematic diagram of the HIV-1 Genome. Accessed and adapted on July 2010 from http://www.yale.edu/bio243/HIV/genome.html

1.4 HIV-1 Life Cycle

In general, the surface glycoproteins of enveloped viruses have an essential role in viral invasion of host cells. The Human Immunodeficiency Virus type 1 (HIV-1) envelope glycoprotein (gp) consists of two noncovalent subunits gp120 and gp41. First, gp120 is responsible for viral tropism by binding to target-cell receptors and gp41 directs fusion between cellular and viral membranes[14,15]. gp120 binds to the CD4 receptor and the CCR5 or CXCR4 co-receptors, which belong to the chemokine receptor family[14]. Macrophage (M-tropic) strains of HIV-1 use the CD4 and CCR5 receptors as a means to gain access to the cell[16]. In contrast, T-tropic strains use the CD4 and CXCR4 receptors to gain cell entry[16]. These receptors can be found present in such cells as T lymphocytes and macrophages. The interactions between gp120 and g41 are altered by a conformational change that occurs when gp120 binds to the CD4 receptor and the chemokine receptor[17]. These changes allow the virus to inject its viral genome into the host through fusion. Soon after entry, the process of reverse transcription occurs in which the enzyme reverse transcriptase transcribes a complementary, double-stranded DNA molecule from a single-stranded RNA genome. The process of reverse transcription is very error prone and results in mutations due to its inaccuracy. As a result, these mutations allow for drug resistance. The high frequency of mutation is a direct result of having an average error of $3x10^{-5}$ per nucleotide base per cycle[18]. The newly formed DNA is then transported to the nucleus of the cell where it is integrated into the host DNA by an enzyme known as integrase. Inclusion of the virus genetic material into the host cell genome results in the viral DNA becoming a provirus[19]. The provirus is then transcribed into mRNA, which is further spliced into smaller pieces. This allows it to be transported to the cytoplasm where it is translated into the Tat and Rev proteins. The Tat protein allows for increased transcription of proviral DNA[20]. The Rev protein binds to the unspliced mRNA and allows them to leave the nucleus, which allows for the translation of

the pol, env, and gag genes[19]. The last and final step of the HIV life cycle is assembly and budding. The protease enzyme is responsible for cleaving the gag and gag/pol, which is vital for viral maturation[21,22]. The structural proteins matrix (MA), capsid (CA), nucleocapsid (NC) are derived from the gag precursor. The pol produces such enzymes as reverse transcriptase, integrase and protease.

1.4 HIV-1 Protease Structure and Mechanism

A protease is an enzyme that directs proteolysis and cleaves the peptide bonds that link amino acid to form polypeptide chains. Proteases are divided into several groups: serine, theronine, cysteine, aspartyl, metallo, and glutamic[23]. HIV protease is classified as an aspartyl protease because it has two aspartic residues in the active site and substitution of the aspartic residue leads to a loss of catalytic ability[24]. HIV protease is able to cleave the peptide bonds between TYR/PHE and PRO residues and no other enzyme in humans shows the same specificity[23]. As described in earlier sections the protease plays an essential role in the life cycle of HIV and inhibition of its function would result in an immature, non-infectious viral particle.

The three-dimensional structure of HIV protease was not determined until 1989[25,26]. It is composed of two identical monomers each containing 99 amino acid residues, which give rise to a symmetrical homodimer[26]. Each monomer has one small α helix and a four-stranded β sheet, which are formed by the N and C terminal strands[23]. The active site is located at the interface of the two monomers where each monomer contributes the sequence Asp25-Thr26-Gly27[26]. These active site residues are known as the catalytic triad. The active site loops are held together by a conformation known as the “fireman’s grip”. It is a conformation in which the amide group of each Thr26 monomer donates a hydrogen bond to the opposing Thr26 carbonyl. Furthermore, each individual loop is stabilized by the hydrogen bond between Oδ1 of the aspartate and the

amide group of Gly27[26]. The active site is covered by flexible flaps, which open to allow the substrate to enter and leave. In the apo form, the flaps populate a conformation known as semi-open; computational studies have simulated flap openings as wide as 32 angstroms[27].

HIV protease cleaves its substrate through a general acid/base mechanism. The flaps open and allow the substrate to enter and close in order to secure the substrate in the active site. It has been proposed that the flaps are held in a closed conformation by the presence of a water molecule through hydrogen bonding of the amide group of Ile50[28]. Plots of the rate vs. pH, done by Meeks et al. reveal a bell-shaped curve, indicating that one Asp acts as acid and the other as a base[29,28]. The mechanism proposed by Meeks et al. is that Asp25' accepts a proton (acting as base) from water. The newly formed nucleophile is very potent and attacks the carbonyl of the substrate. The tetrahedral intermediate is stabilized by the carbonyl oxygen accepting a hydrogen from Asp25 (acting as an acid). The Asp25 accepts the hydrogen back from the oxygen as it begins to reform its double bond with the carbon. Simultaneously, the carbonyl is reformed as the N—C bond begins to elongate and eventually breaks as the nitrogen accepts a hydrogen from the Asp25'[29]. The catalyst is regenerated, the flaps open and the product leaves.

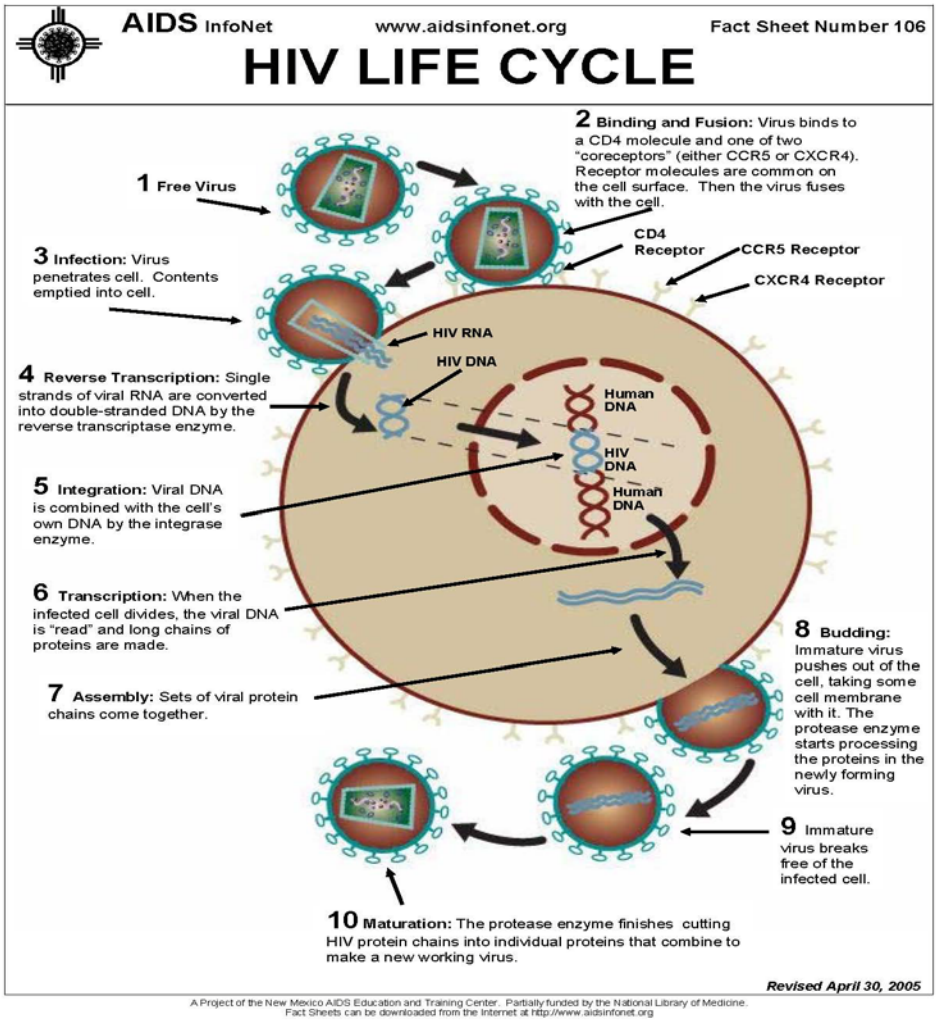

Figure 1-2. A schematic depiction of the life cycle of HIV. Accessed and adapted on July 2010 from http://www.aidsinfonet.org/

CHAPTER 2
THEORY AND METHODS

2.1 Molecular Dynamics

The first molecular dynamics (MD) simulation was published by Alder and Wainwright, in which they reported on the interactions of hard spheres[30]. The first protein simulation was performed on the bovine pancreatic trypsin inhibitor, paving the way for the application of MD to biological systems[31]. Molecular Dynamics has proven useful in validating experimental results, structure refinement, drug-design, enzymatic reactions, etc.

Molecular Dynamics is based on Newton's laws of motion, particularly his second law, which states that F=ma, where *F* is the force exerted on the particle, *m* is its mass and *a* is its acceleration. Integration of the equations of motion yields a trajectory. In order to calculate a trajectory the initial positions, velocities and accelerations must be known. For instance, the initial positions can be gathered from an x-ray crystal structure, the velocities and the acceleration are determined by the gradient of the potential energy function[32,33]. The integration of the equations of motion can be easily solved analytically for such systems as a one-dimensional harmonic oscillator. On the contrary, this is not feasible for such complex systems as proteins and these equations must be solved numerically[34]. Many numerical algorithms have been developed in order to accomplish this, for example, Verlet, leapfrog, verlocity-Verlet, and Beeman's[35,36,37,38]. All of the algorithms use a Taylor series expansion to make the approximation about the positions, velocities and acceleration. As stated above the leapfrog algorithm is commonly used:

$$\vec{q}\left(t+\Delta t\right)=\vec{q}\left(t\right)+\vec{v}\left(t+\frac{1}{2}\Delta t\right)\Delta t \tag{2-1}$$

$$\vec{v}\left(t+\frac{1}{2}\Delta t\right)=\vec{v}\left(t-\frac{1}{2}\Delta t\right)+\vec{a}(t)\Delta t \tag{2-2}$$

in which *q* and *v* correspond to the coordinate and velocity vectors of all the atoms[36]. The acceleration is determined from the forces and masses of the atoms. In the leapfrog algorithm the velocities are first calculated at the time (t - ½Δt) and the positions are determined from the velocities previously calculated. Thus, the velocities leap over the positions and the positions leap over the velocities. The advantage of this method is the velocities are explicitly calculated but the drawback of this method is that the calculation of the positions and velocities are deduced in the same step.

Langevin Dynamics are often used in conjunction with implicit solvation models. The Langevin equation of motion is a stochastic differential equation in which two force terms have been added to Newton's second law of motion in order to account for the omitted degrees of freedom. In the description of stochastic dynamics the force, which a particle it sees can be attributed to three different factors. The first force felt by a particle depends on its position relative to the other particles[32]. The second force is described as the force felt by a particle as it moves through the solvent and can be equated to the frictional drag on the particle due to the solvent[32]. The last contributing factor is the force that a particle experiences due to random fluctuations, which are created through interactions with solvent molecules[32]. The Langevin Equation can be expressed as:

$$m_i \frac{d^2x}{dt^2} = F_i\{x_i(t)\} - \gamma_i \frac{dx_i}{dt} m_i + R_i(t) \qquad (2\text{-}3)$$

where *m* is the mass, *x* is the position, *F* is an interaction force between a particle and other particles, γ represents the collision frequency, and R(t) is the force on the particle due to random fluctuations by interactions with solvent molecules[32].

2.2 Molecular Mechanics

Molecular Mechanics (MM) makes the assumption that interactions within a system can be described by such processes as bond stretching, bond angles and rotations about a bond[32]. Furthermore, the parameters calculated from small molecules can be translated to describe larger systems such as proteins[34]. An illustration of what parameters are involved in a force field:

$$V(r^N) = \sum_{Bonds} \frac{k_i}{2}(l - l_o)^2 + \sum_{Angles} \frac{k_\theta}{2}(\theta - \theta_o)^2 + \sum_{Torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \right) \quad (2\text{-}4)$$

where the first three terms describe the covalent interactions and the last two terms describe the non-covalent interactions. *V* denotes the potential energy, which is function of the positions *(r)* of *N* atoms[32]. The first two terms in Eq. (2-3) describe energy changes that occur when the bonds and angles deviate from their reference value. In addition, k_l and k_θ are the respective force constants for the bond and angles. The third term in equation 2-4 describes the torsional angles. The torsional term is expressed in a Fourier series in which V_n is the dihedral force constant, *n* is the dihedral periodicity, ω is the torsional angle and γ is a phase of the dihedral angle θ[39]. The fourth and fifth terms of Eq. (2-4) are Lennard-Jones and Coulomb potentials. The Lennard-Jones potential is used to describe the van der Waals interactions whereby ε is the depth of the potential well, σ is the distance at which the interparticle potential is zero, r^{-12} is the repulsive tern and the r^6 is the attractive term. The Coulomb potential is used to describe the electrostatic interactions where q_1 and q_2 are the charges of atoms, *r* describes the interparticle distance, and ε is the electric constant.

The force field ff99SB was used for the parameters of our system[39]. The ff99SB has better parameters for the backbone torsion terms and the Gly residues, which are crucial due to the number of Gly residues found in the flap region of the protease[27].

2.3 Generalized Born Model

Implicit solvation is modeled as a continuum solvent instead of explicitly and is used to estimate the free energy of solute-solvent interactions. Implicit solvation has several attractive advantages. First, the computational cost associated with the use of these models in MD simulations is generally cheaper than the cost of representing water explicitly. Secondly, due to the absence of viscosity associated with explicit water environment, the molecule can explore the available conformational space much faster[40]. The GB model omits the viscous effect of solvent and thus Langevin Dynamics must be incorporated in order to compensate. One disadvantage of the Generalized Born Model (GB) is that it over stabilizes salt bridges[41,42]. GB is an approximation to the Poisson-Boltzman equation[40]. The GB equation can be expressed as:

$$\Delta G_{GB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f\left(r_{ij}, a_{ij}\right)} \tag{2-5}$$

$$f\left(r_{ij}, a_{ij}\right) = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \tag{2-6}$$

$$a_{ij} = \sqrt{a_i a_j}$$
$$D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2 \tag{2-7}$$

where ε is the dielectric constant, q represents the charge, r_{ij} is the interparticle distance and a_{ij} is the Born radii.

CHAPTER 3
MOLECULAR DYNAMIC STUDY OF THE CONFORMATIONAL DYNAMICS OF HIV-1 PROTEASE SUBTYPES A,B,C,F

3.1 Introduction

The HIV protease has become an attractive target for drug design due to its role of cleaving the gag and gag-pol polyprotein precursors. Furthermore, inhibition of the protease's natural biological function would prevent the maturation of the HIV, hence preventing the infection of neighboring cells. The flap domain has been highly studied and through NMR and crystallographic studies, has been shown to exist in numerous conformations. The flap domain is the most mobile of all domains, largely attributed to the number of Gly residues found in this region[43].

Many different experimental and theoretical studies have been conducted in order to gain a better understanding of the movements of the flaps and how they correlate to drug resistance. Galiano et al. first proposed the used of site-directed-spin-labels (SDSL) and EPR studies as a means to gain insight on how the mutations impacted the flap dynamics[44,45]. NMR studies done by Ishima et al. suggest a working model of the flap-opening mechanism[43]. Through the use of MD, Scott and Schiffer proposed the curling of the flaps as mechanism of drug resistance[46]. Hornak et al. and Boric et al. were both able to successfully simulate the closing of the flaps by placing a ligand into a protease in which the flaps were originally open[47,48]. Hornak et al. was the first to simulate an opening and reclosing of the flaps[27].

As previously stated, due to the importance of the flaps having to open and close in order for catalytic activity to occur, it has been suggested by previous works that the development of a new class of protease inhibitors that instead target the flap domain or other essential domains might be more effective than the original idea of developing a protease inhibitor that works

through competitive inhibition[27,49,50]. The effectiveness of inhibitors that competitively bind to the active site is short-lived due to mutations, which decrease the potency of the inhibitor.

While there is a wealth of research available on subtype B, few theoretical studies have been conducted on subtypes A, C, or F[51]. Likewise of the HIV-1 protease inhibitors approved by the Federal Drug Administration (FDA) none have been specifically designed for non-B subtypes. The use of computational techniques such as MD has proven useful in elucidating and confirming the main different HIV protease conformations. In this current study, we investigate conformational dynamics of the flaps, and the size of the active site in order to correlate how the sequences of subtypes A, B, C, and F allow for different conformations of the protease. Our results offer insight and suggestions, which might prove to be useful in the development of new protease inhibitors.

3.2 Methods

The starting structures of subtype B proteases were obtained from the Protein Data Bank (PDB ID 1HHP) semi-open conformation, (PDB ID 1OHR) complexed with nelfinavir, and (PDB ID 2BPX) complexed with indinavir[52,53,54]. The three proteases of subtype C were obtained from PDB (PDB ID 2R8N) open conformation, (PDB ID 2R5P) complexed with protease inhibitor indinavir, (PDB ID code 2R5Q) complexed with the protease inhibitor nelfinavir[55,56]. Subtype F protease (PDB ID code 2P3C) complexed with TL-3 protease inhibitor, and subtype A (PDB ID code 3IXO) closed conformation, were gathered from the PDB[57,58]. In the case of the proteases, which were complexed with the inhibitors, the inhibitors were removed prior to simulation. Figure 3-1 shows the structures, which were used in the eight simulations of subyptes A, B, C, and F. Prior to the simulation, three mutations were made in subtype C in order to revert back to the original sequence K7Q, I63L, and I33L. One mutation was made in subtype F K7Q. These mutations were originally made to the sequence in order to increase

stability[56,57]. Only one of the catalytic Asp25 was protonated. The missing heavy and hydrogen atoms of each crystal structure were added by the *Leap* module of AMBER 11.0[59].

The minimizations and molecular dynamics were performed using the *sander* and *pmemd* modules of AMBER. The initial parameters for the protein were gathered from the force field ff99SB[39]. The solvation was implicitly modeled using Generalized Born approach (GB). The SHAKE algorithm was used in order to constrain all bonds to hydrogens and the timestep was 1 fs. In order to control temperature, Langevin Dynamics was implemented in which the collision frequency was 1 ps^{-1}. During equilibration, each system was minimized and slowly heated up from 50 to 300K with restraints first on hydrogen atoms, then all heavy, and backbone atoms. Each system was relaxed for 1ns. In the case where mutations were introduced prior to starting equilibrations process, the system was allowed to relax for 1.2 ns. Restraints were removed entirely during the last step of equilibration in order to observe the stability of the system. The production phase consisted of running the equilibrated structures at ten different initial velocities for 22ns each and the temperature remained at 300K throughout.

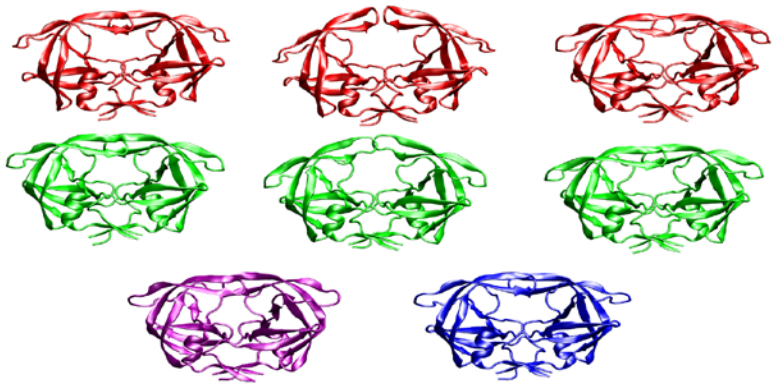

Figure 3-1. Graphic representation of the initial x-ray crystal structures used in the MD simulations. The first row in red from left to right 1OHR, 1HHP, and 2BPX, which is Subtype B. The second row from in green from left to right 2R5Q, 2R8N, and 2R5P, which is subtype C. The last row from left to right 3IXO, subtype A and 2P3C, subtype F.

3.3 Results

The next three sub-sections contain results analyzed from the eight different simulations. In the first section we compare the dynamics between a closed, semi-open, and proposed wide-open conformation. The second section is a comparison of the three different subtype B and C simulations amongst themselves. The last section will be a comparison of subtypes A, B, C and F.

3.3.1 Comparison of the Flap Dynamics

The results shown below not only compare the flap dynamics each of the ten different simulations of eight structures but their stability as well. As previously stated, each of the simulations was compared against three structures as shown in figure 3-2, where the flaps were either closed, semi-open and a proposed wide-open structure. The coordinates for the semi-open structure were derived from 1HHP. The coordinates for the wide-open structure were taken from a snapshot of the 1HHP simulation in which a wide-open conformation was observed. The closed structure used as a reference for the Subtype A simulation and Subtype F simulation was 3IXO and 2P3C. The closed structures used for the Subtype C simulations were 2R5Q and 2R5P. The Subtype B simulations used 1OHR and 2BPX as its reference for a closed structure.

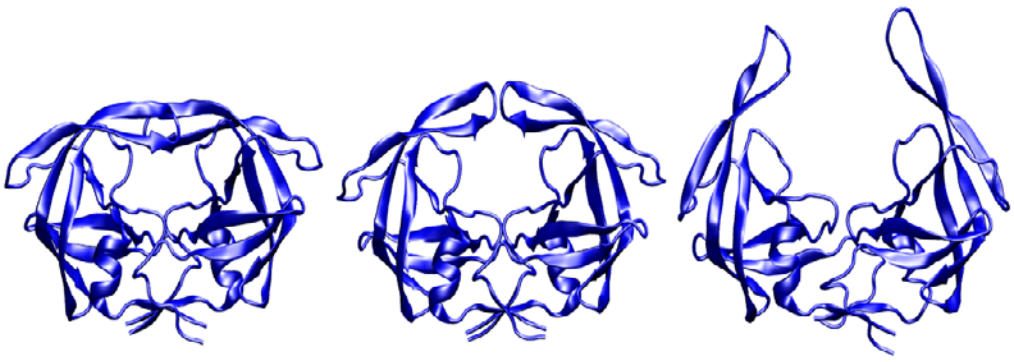

Figure 3-2. From left to right, is representation of a closed structure, semi-open conformation and wide-open conformation.

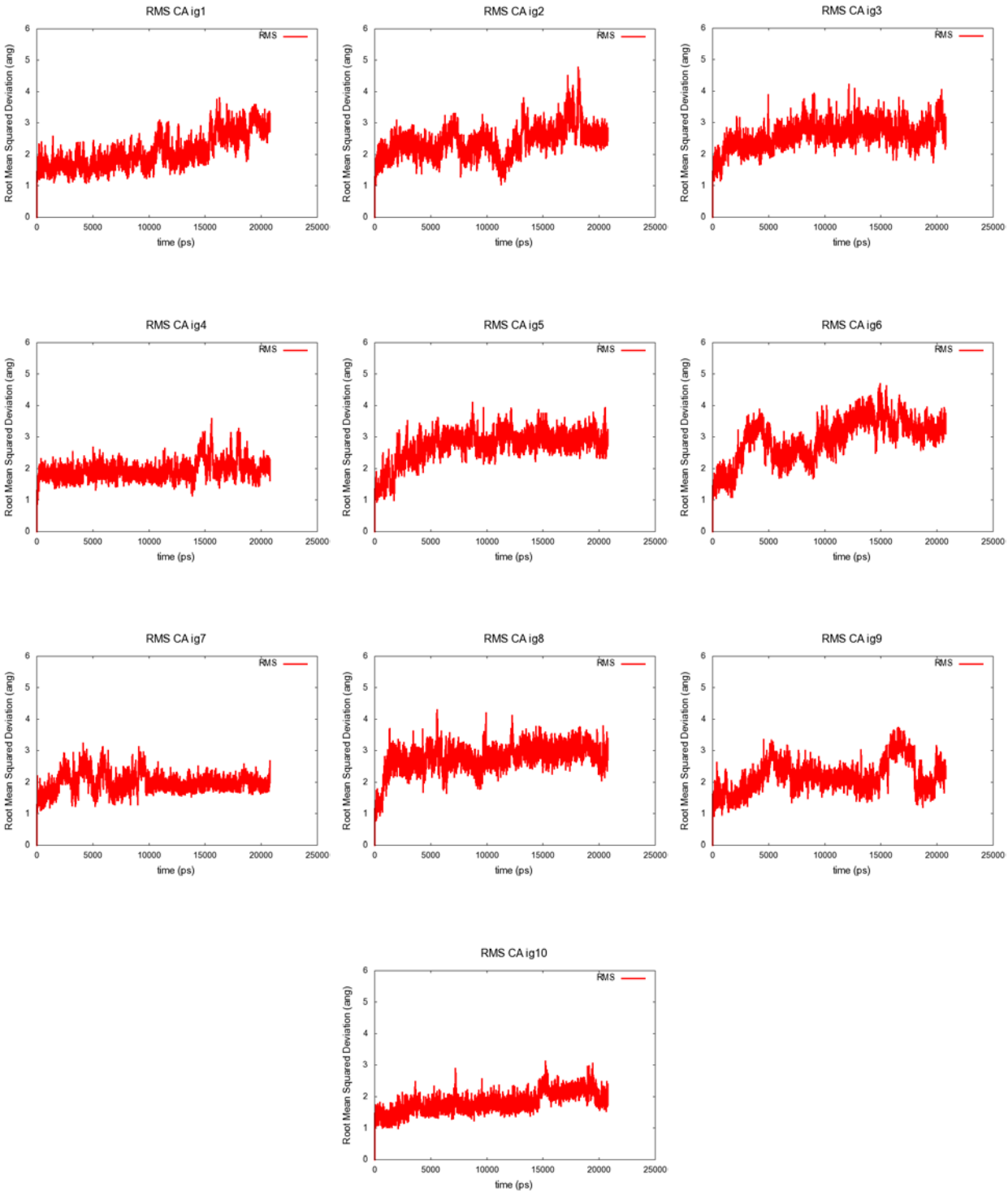

Figure 3-3. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype A (PDB ID 3IXO) closed conformation. The first frame of each simulation was used as the reference.

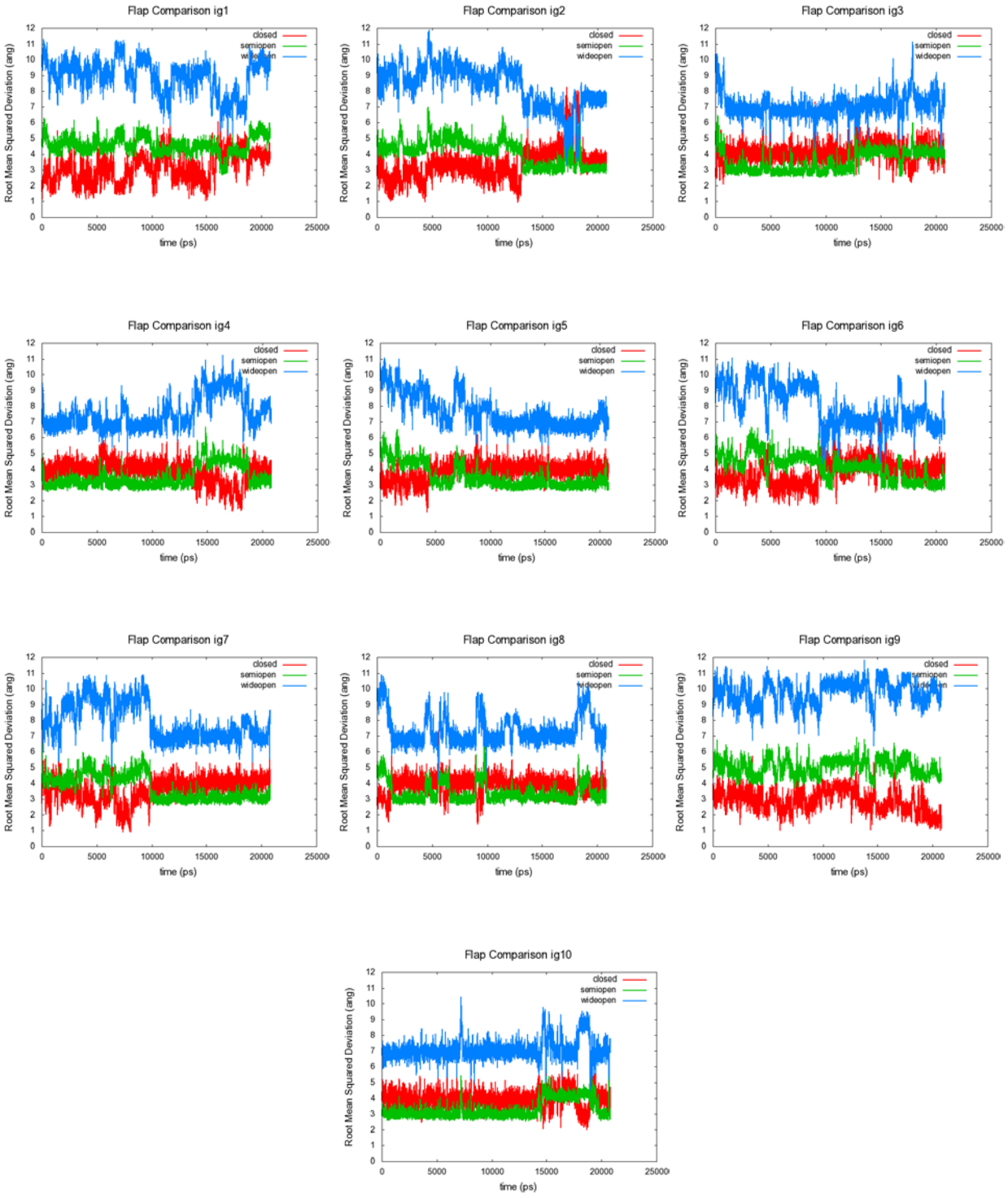

Figure 3-4. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype A. The red was versus the closed conformation. The green was versus the semi-open conformation and the crystal structure (1HHP). The blue was versus the (fully) wide-open conformation.

Figure 3-3 shows how each of the ten different simulations of subtype A differ from the first frame of their respective simulations. On average, each of the ten different simulations fluctuates between 1 and 4.5 angstroms. The large deviations in Figure 3-3 can more than likely be attributed to the movements of the flaps and not as much to the backbone. In addition, these deviations are indicative of the flaps inverting or for example going from closed conformation to an open conformation. Figure 3-4 shows how the flaps residues, which were considered to be 43-58 and 43'-58' of each monomer, deviate against the three conformations. As proposed by Hornak et al., when a ligand is not bound, the flaps are most likely to be in a semi-open conformation[27]. This hypothesis is not entirely correct because in the simulations of subtype A, the lowest fluctuation only gets within 2.5 angstroms of a semi-open conformation. However, subtype A does come within .91 angstroms of the closed conformation during several points of the simulation. There were three simulations, ig1, ig2 and ig9 in which the closed conformation is the predominant orientation of the flaps, as shown in Figure 3-4. These two simulations demonstrate a preference to the closed conformation throughout the entire simulation.

One trend that is observed in Figure 3-4 is that the flaps are more likely in a conformation that resembles closed or semi-open rather than wide-open conformation, which is rarely observed. There are points of the dynamics when neither of the three reference structures can be said to exist. The structures are believed to be either the tucked or curled structures as proposed by Kear et al[60]. In the simulations of ig2, ig6 and ig10 we do observe conformations, which get as close to 4 angstroms of the wide-open conformation. The results in Figure 3-4 do coincide with the argument made by Layten et al., although they ran MD simulations using subtype B, that the wide-open conformation is only a minor component of the structures sampled during dynamics[61].

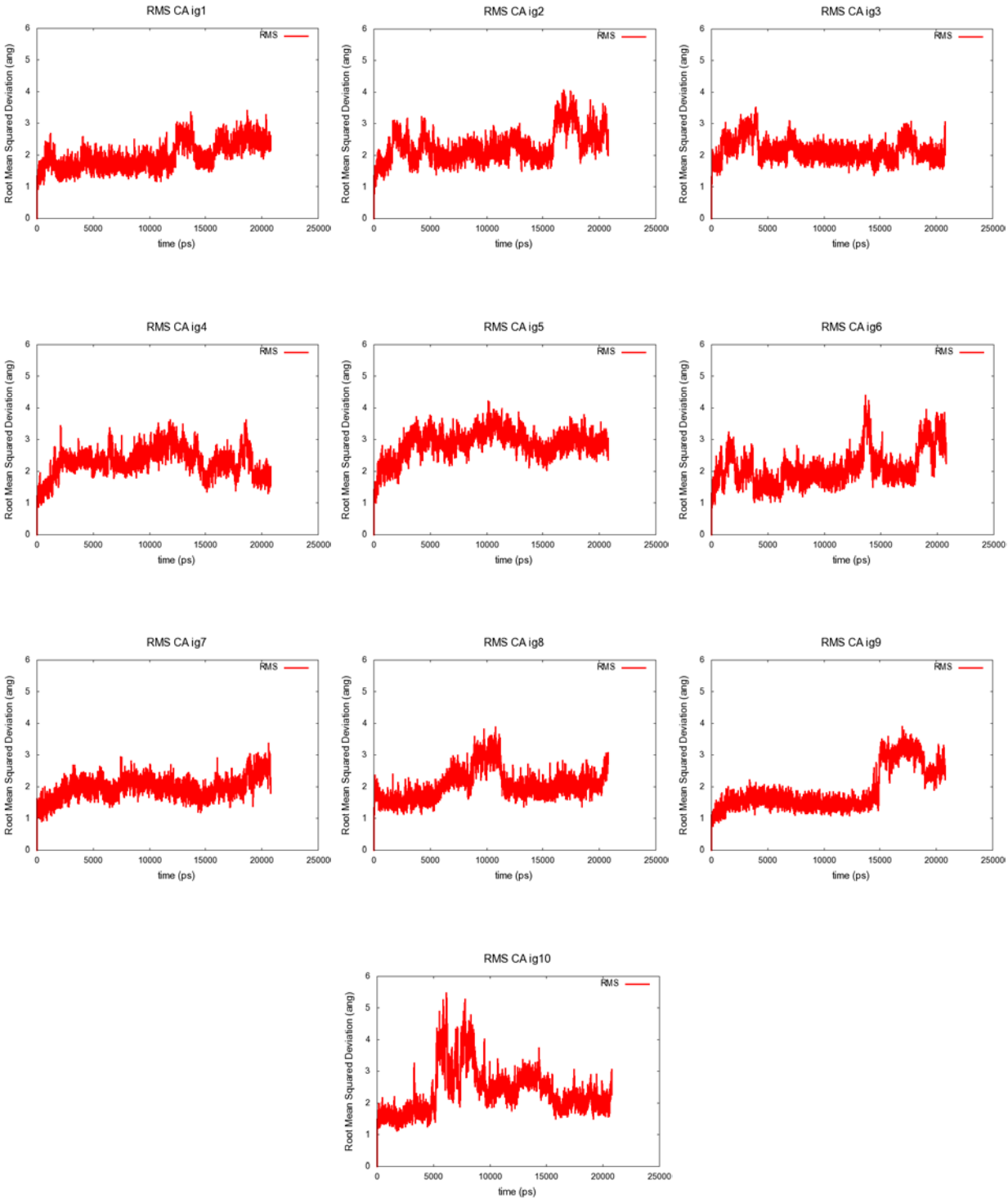

Figure 3-5. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype B (PDB ID 1HHP) semi-open conformation. The first frame of each simulation was used as the reference.

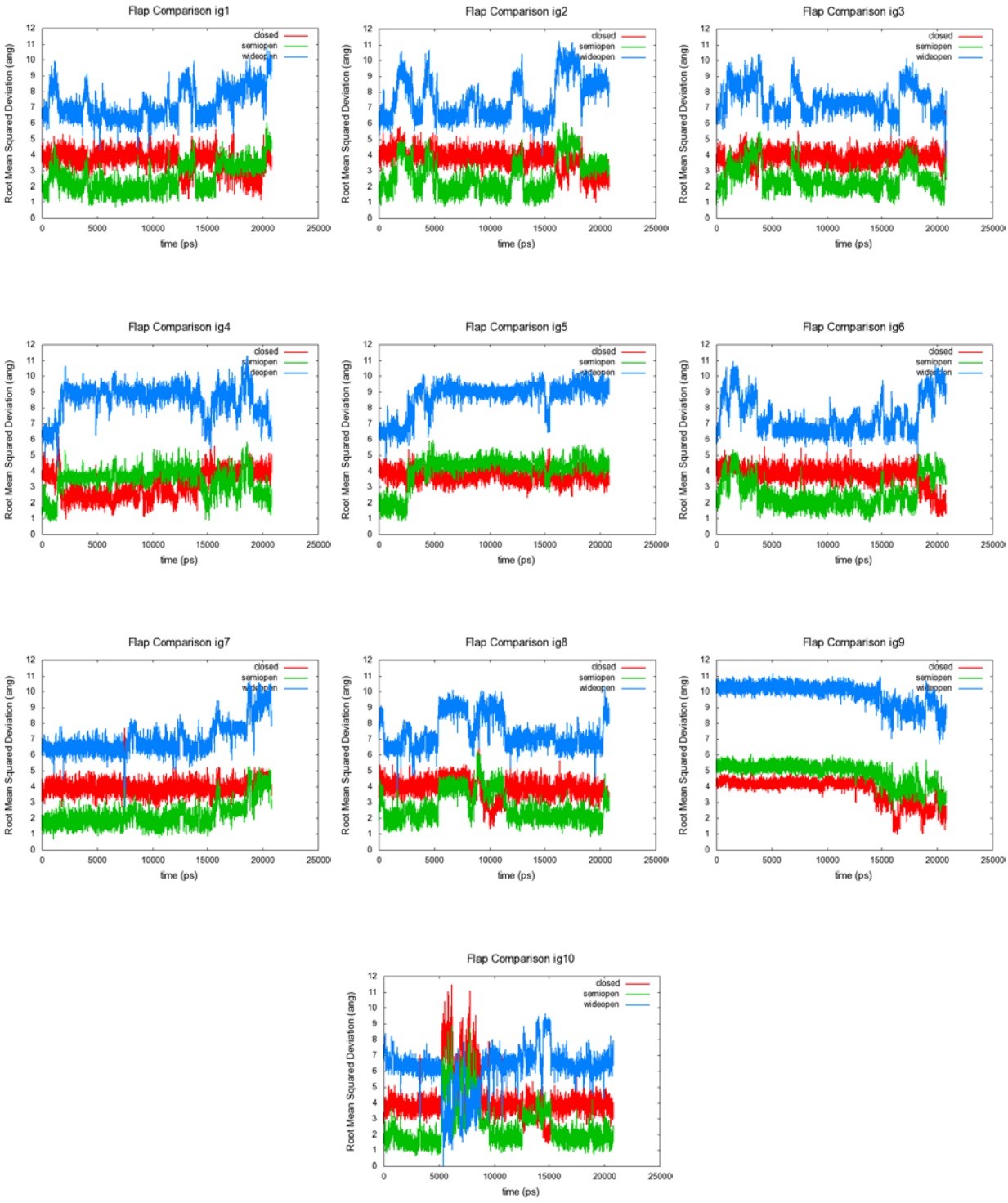

Figure 3-6. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype B (1HHP) semi-open conformation. The red was versus the closed conformation. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

Figure 3-5 shows deviations of Subtype B versus the first frame of the ten different simulations in which the starting coordinates were gathered from 1HHP. Deviations are seen on average between 1 and 5 angstroms. The ig10 simulation shows the highest fluctuations, which are observed around 5000 ps. The fluctuation becomes almost as high as 6 angstroms and then returns back to 2 angstroms, which is a clear indication that a conformational change must have occurred.

In Figure 3-6, as stated earlier we do see that the preferred conformation is semi-open. For instance, in simulations ig1, ig2, ig3 and ig7, we observe that the dynamics become as close or below 1 angstrom on several occasions to the semi-open conformation. However, there are some instances when the dynamics do become as low as 1 angstrom to the closed conformation but are rarely seen. The simulation ig5 in Figure 3-6 exhibits a unique and interesting trend. In the beginning, the ig5 simulation favors the semi-open conformation but early in the simulation we see a switch in the dynamics that favors neither the semi-open or closed conformations. The ig9 simulation in the beginning appears to favor neither the closed, semi-open or wide-open conformations. However, towards the end of the ig9 simulation we do see that dynamics favor a closed conformation. In these cases the flaps are exhibiting either a tucked or curled conformation.

The wide-open conformation is seen in the ig10 simulation around 5000 ps but only for a brief moment. This correlates strongly as to why the peak in the ig10 simulation of Figure 3-5 is present. As previously stated the wide-open conformation does not make up a significant portion of the dynamics sampled. There are more instances where the dynamics of 1HHP favors the wide-open conformation than seen in the other two subtype B simulations, which will be discussed later in this chapter.

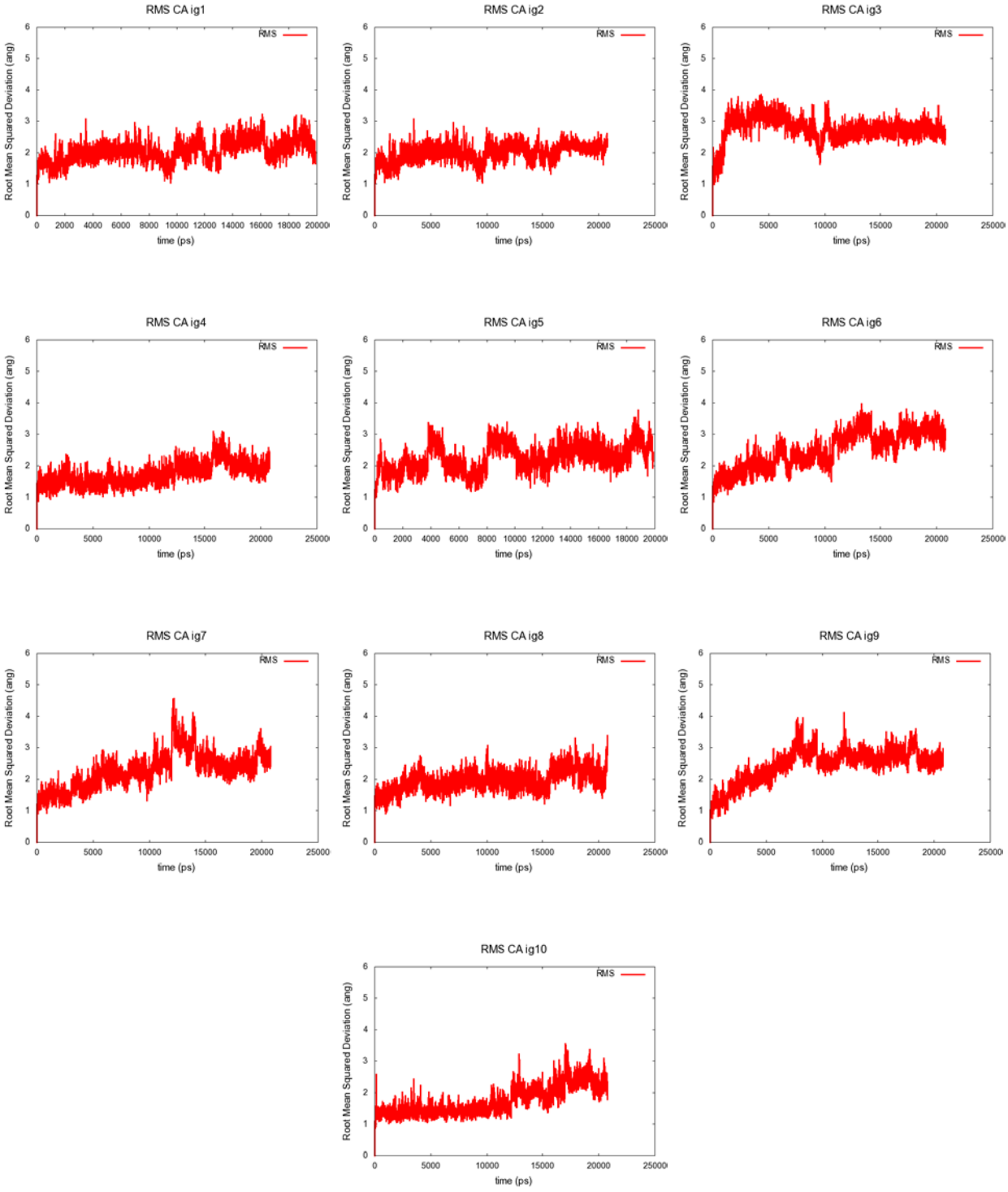

Figure 3-7. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype B (PDB ID 2BPX) closed conformation. The first frame of each simulation was used as the reference.

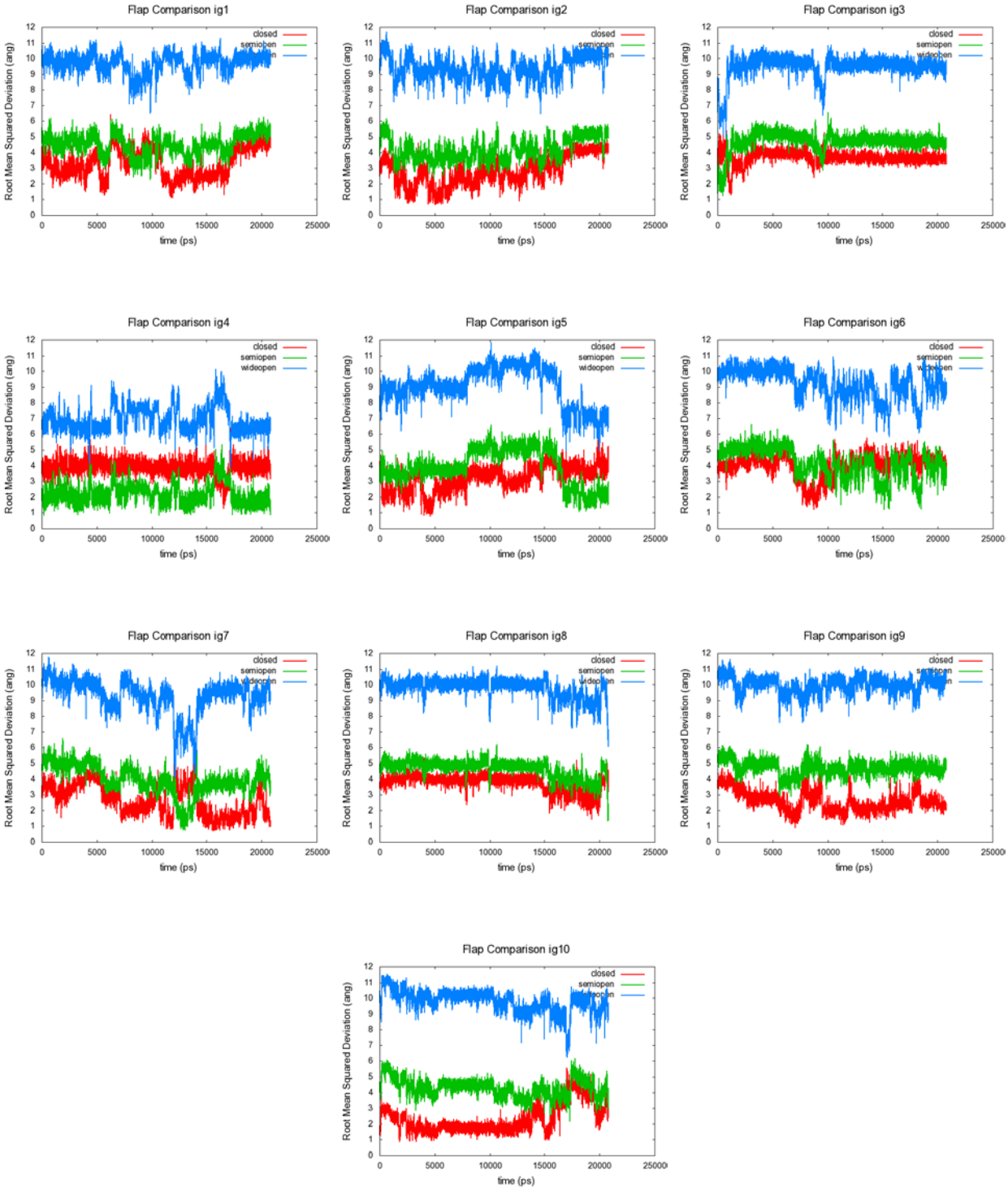

Figure 3-8. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype B (2BPX) closed conformation. The red was versus the closed conformation. The green was versus the semi-open conformation and the crystal structure. The blue was versus the (fully) wide-open conformation.

In the simulations gathered from the initial coordinates of 2BPX, deviations are observed between 1 and 3.5 angstroms in Figure 3-7. In the ig3 simulation the fluctuations quickly rise to 3.5 angstroms after the first few picoseconds and remains there throughout the entire simulation. This large fluctuation is caused by the flaps orienting themselves in a structure that looks semi-open and then quickly converting to a structure that favors more of a closed conformation. Simulations ig4 and ig10 fluctuate the least, where the highest deviation is seen around 2.5 angstroms.

In Figure 3-8 an interesting trend is observed. The semi-open conformation is not the overall dominant form of the flaps as seen in previous figures. However, in the ig4 and ig6 simulations the semi-open conformation is the dominant form. In the ig4 simulation, the dynamics fluctuate consistently between 1 and 2 angstroms from the semi-open conformation. The fluctuations in the ig10 simulation become as close as 1 angstrom in certain points of the simulation. However, for the case of the ig2, ig8 and ig9 simulations, the closed conformation appears to be the preferred form. The dynamics of ig3 simulation appear to favor the semi-open conformation but then quickly changes to favor either the tucked or curled conformation. The other simulations equally populate either the close or semi-open conformation.

In general, very few of the simulations appear to come close to resembling a structure that looked like a wide-open conformation. The ig7 simulation becomes the closest to resembling a structure that could be a wide-open conformation. The fluctuations of this particular simulation become as low as 4 angstroms during certain points of the dynamics. The lack of the dynamics favoring the more the wide-open conformation could be do to either not enough sampling or the wide-open conformation is just not prevalent structure of the flap dynamics as a whole. The large conformation of the flaps may appear on a time scale that is beyond our capabilities to simulate.

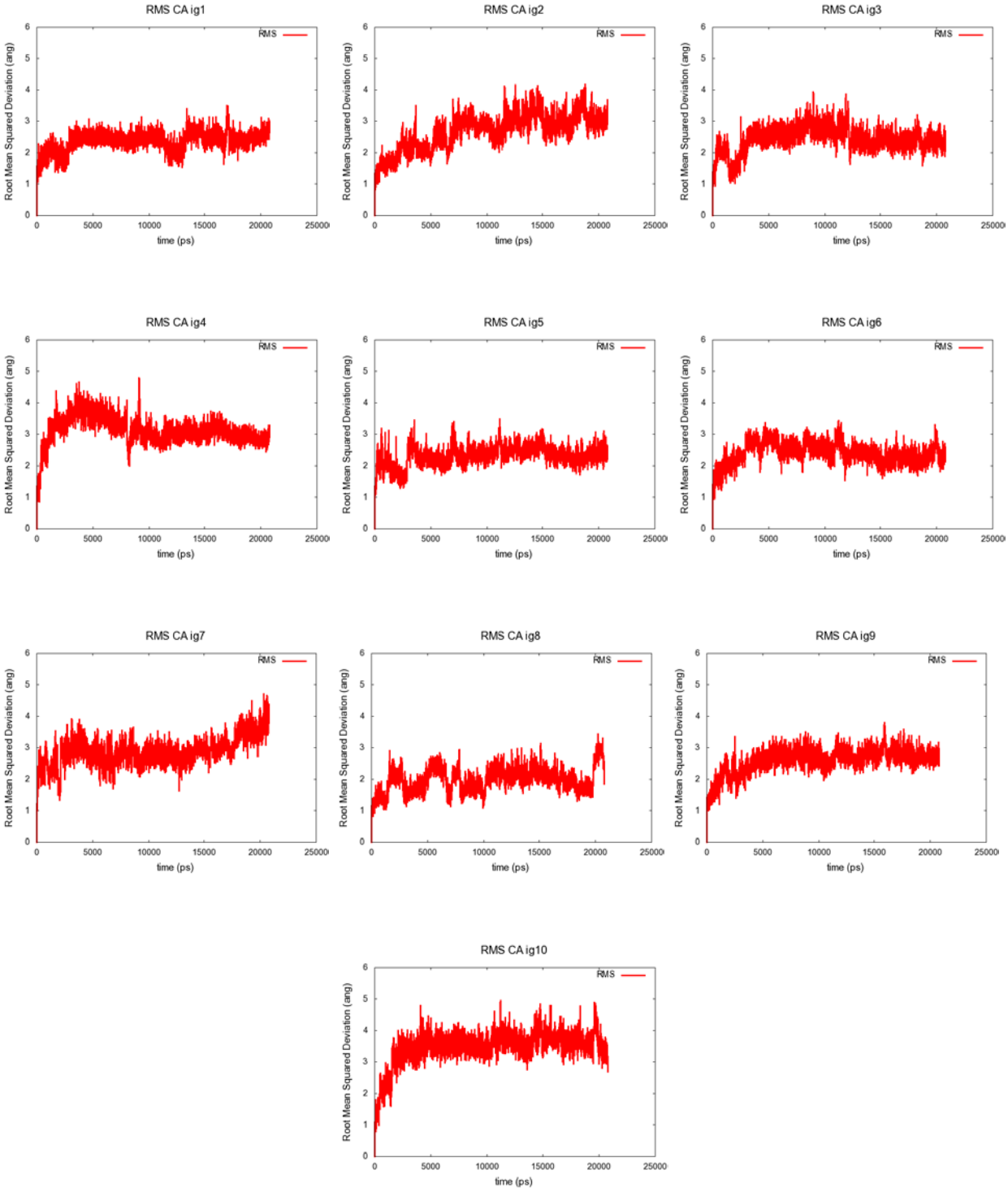

Figure 3-9. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype B (PDB ID 1OHR) closed conformation. The first frame of each simulation was used as the reference.

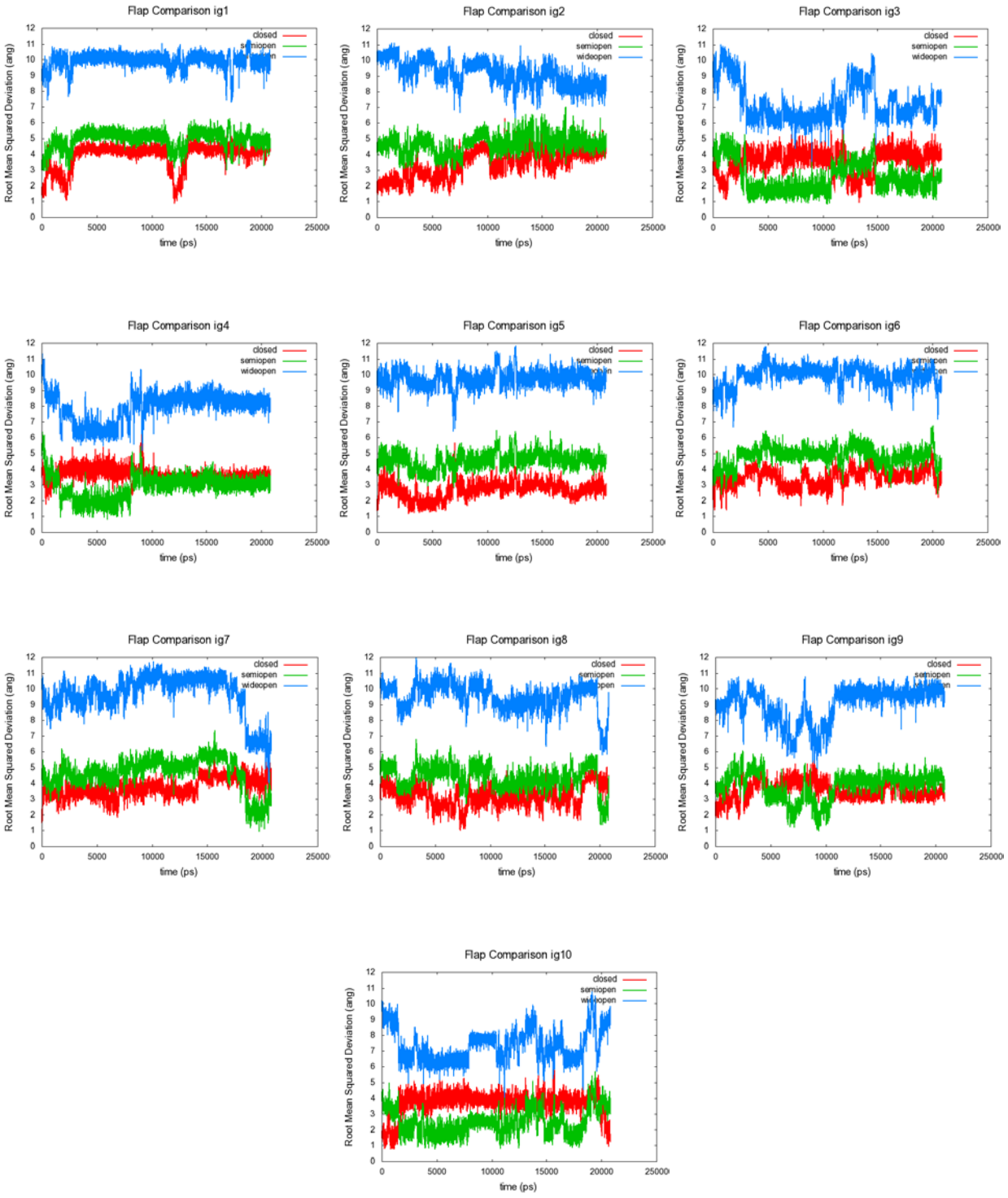

Figure 3-10. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype B (1OHR) closed conformation. The red was versus the closed conformation. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

Fluctuations are seen as low as 1 angstrom and as high as 4.5 angstroms in Figure 3-9, in which the starting coordinates were 1OHR. Simulations ig4, ig7 and ig10 show the highest fluctuations. Furthermore, the fluctuations in the ig10 simulation quickly rise to 4.5, where it remains throughout the simulation. This can be explained because in Figure 3-10 we observe that the ig10 simulation first favors a closed conformation but quickly reverts to favoring a semi-open conformation. The ig7 simulation abruptly rises to 4.5 angstroms but then decreases to 3 angstroms.

As seen in previous sections, the flaps favor a conformation other than the semi-open conformation. In contrast, in the ig5 simulation, we see that the closed conformation is preferred. The fact that the closed conformation is dominant in one simulation could be attributed to the flaps beginning the simulation from a closed structure. In the ig1 simulation, it is observed that the closed conformation is preferable at the beginning but the deviations only get as close as 2 angstroms and that is only for a brief moment. The ig1 simulation demonstrates another example of the flaps in a conformation that is neither closed or semi-open. The flaps could either be tucked or curled. The wide-open conformation is not the dominant conformation in any of the simulations. The simulations of ig3, ig7 and ig10 do however become as close as 3 angstroms to the wide-open conformations.

When comparing all the subtype B simulations similar trends can be seen in all of them. We observe that the semi-open conformation is the preferred conformation of all the simulations. On the other hand, simulations 1OHR and 2BPX do show more instances in which the dynamics favor the closed conformation. Further investigation is needed in order to explain why this is the case. Performing an energy decomposition analysis might give insight to what interactions are different between 1HHP, 2PBX and 1OHR.

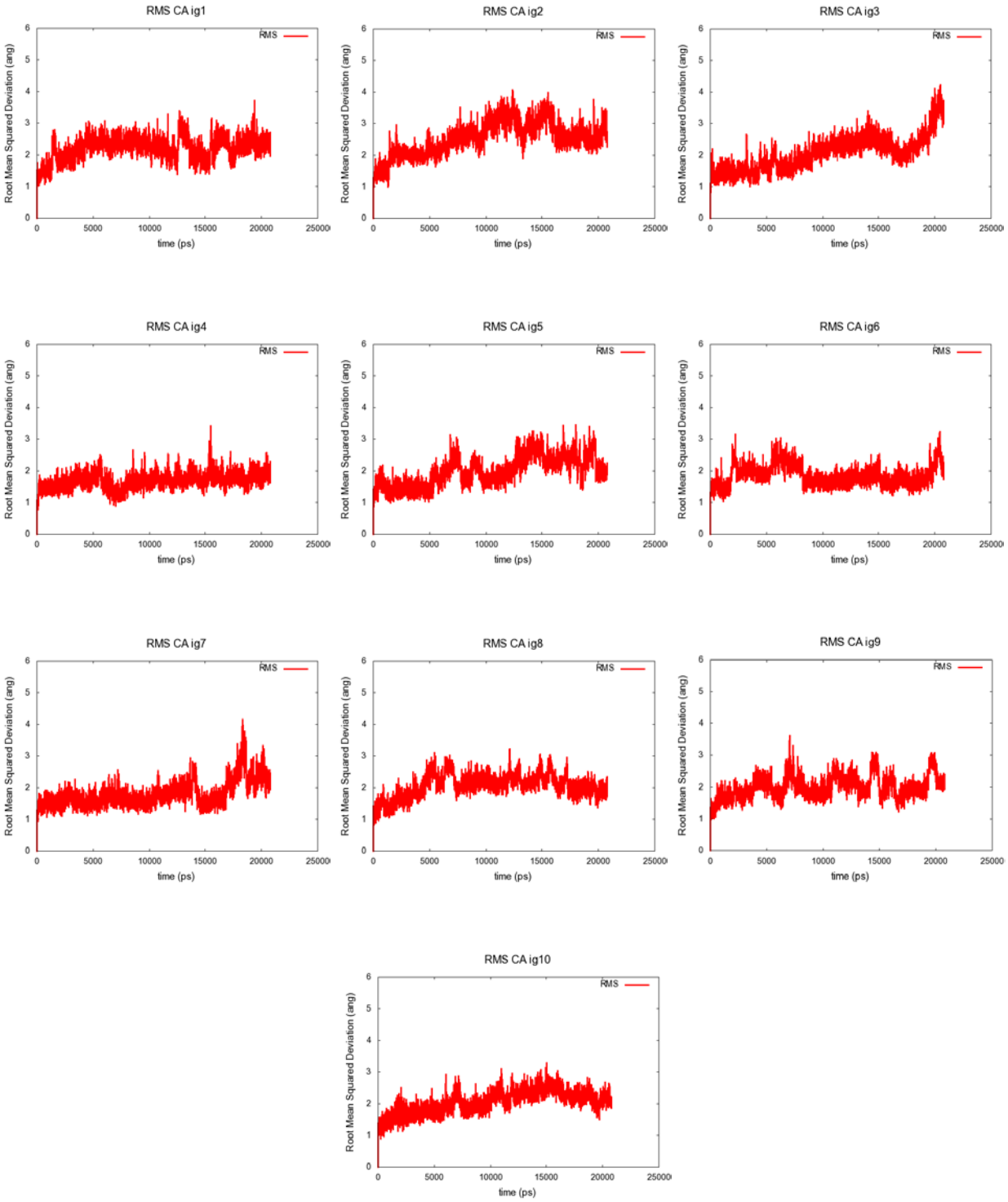

Figure 3-11. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype C (PDB ID 2R8N) wide-open conformation. The first frame of each simulation was used as the reference.

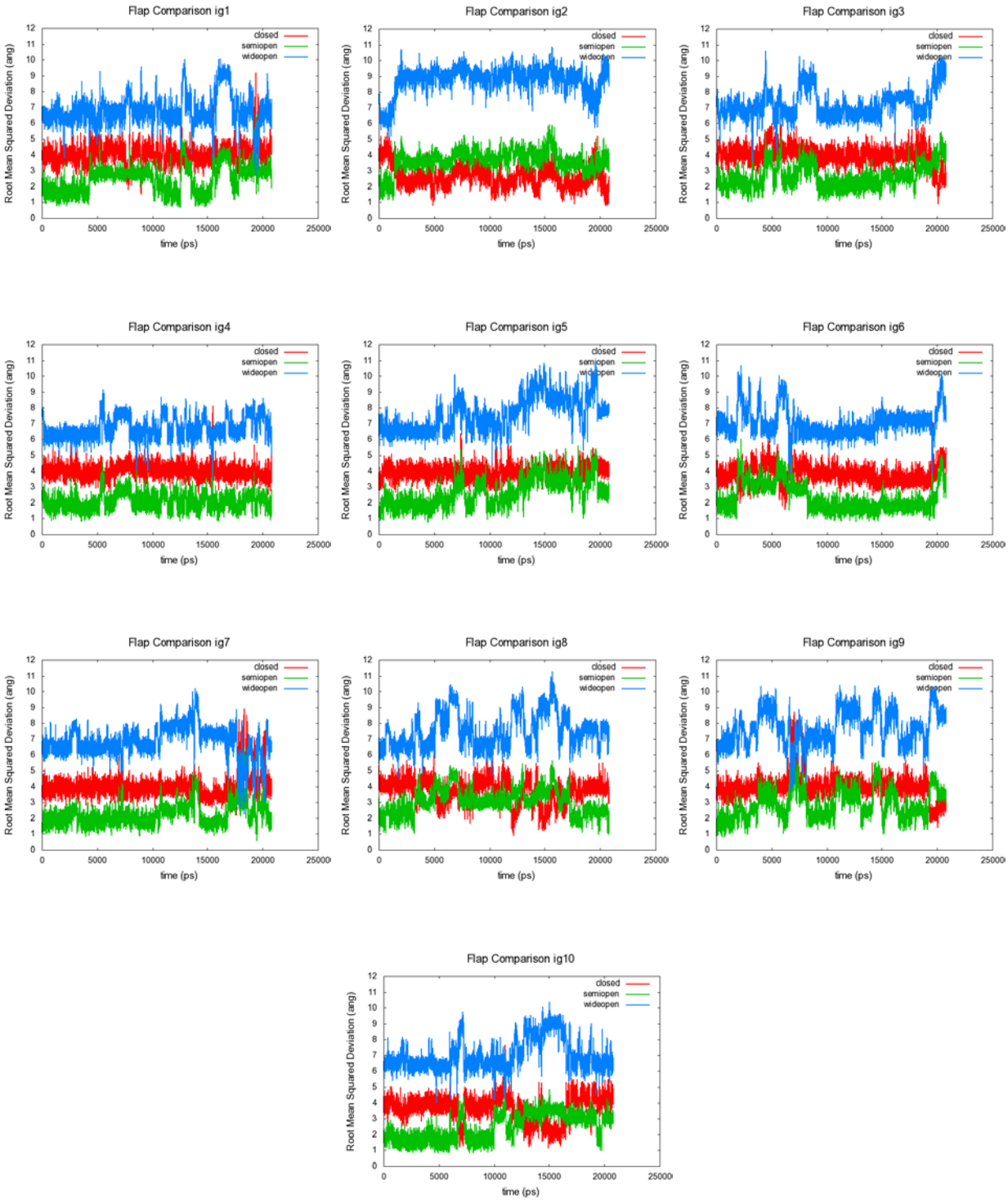

Figure 3-12. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype C (2R8N) wide-open conformation. The red was versus the closed conformation and the crystal structure. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

In Figure 3-11, where the initial coordinates were 2R8N, we observe deviations primarily between 1 and 4 angstroms. In all of the simulations there are no sudden rises or declines in fluctuations except in the simulation ig3. This sharp rise is due mainly in part that the flaps switch from semi-open to a closed conformation. From Figure 3-12, it is clear that the preferred structure is one in which the flaps favor the semi-open conformation. In many of the simulations we see that the dynamics get as close as 1 angstrom to the semi-open conformation. This strong preference for the semi-open conformation could be possibly linked to the fact that the initial coordinates of the flaps were in a conformation, which is considered to be wide-open.

Nevertheless, only the dynamics of the ig2 simulation in Figure 3-12 favor a conformation other than semi-open. In this special case the trajectory never deviates more than 2 angstroms from the closed structure except when the simulation first begins in which the deviation is at 4 angstroms. The ig9 simulation shows a very unique trend. There are points in the ig9 simulation where the fluctuations, when compared to the semi-open structure, suddenly increase. Usually when this occurs, the fluctuation when compared to the closed structure decreases; however, this trend is not seen. At that particular instance one could assume that the dynamics favor either tucked or curled conformation.

For almost all of the simulations with a few exceptions the fluctuations remain between 6-8 angstroms when compared against the wide-open structure. In particular, during the ig7 simulation we observe that the flaps come within 2 angstroms of the wide-open conformation. It could be proposed that the flaps exhibit semi-open conformation that favors more of a wide-open conformation. As seen in the subtype B simulations, the dynamics of the structure that did not begin the simulation closed favors the wide-open conformation a lot more when compared to the two simulations in which the starting structures were closed.

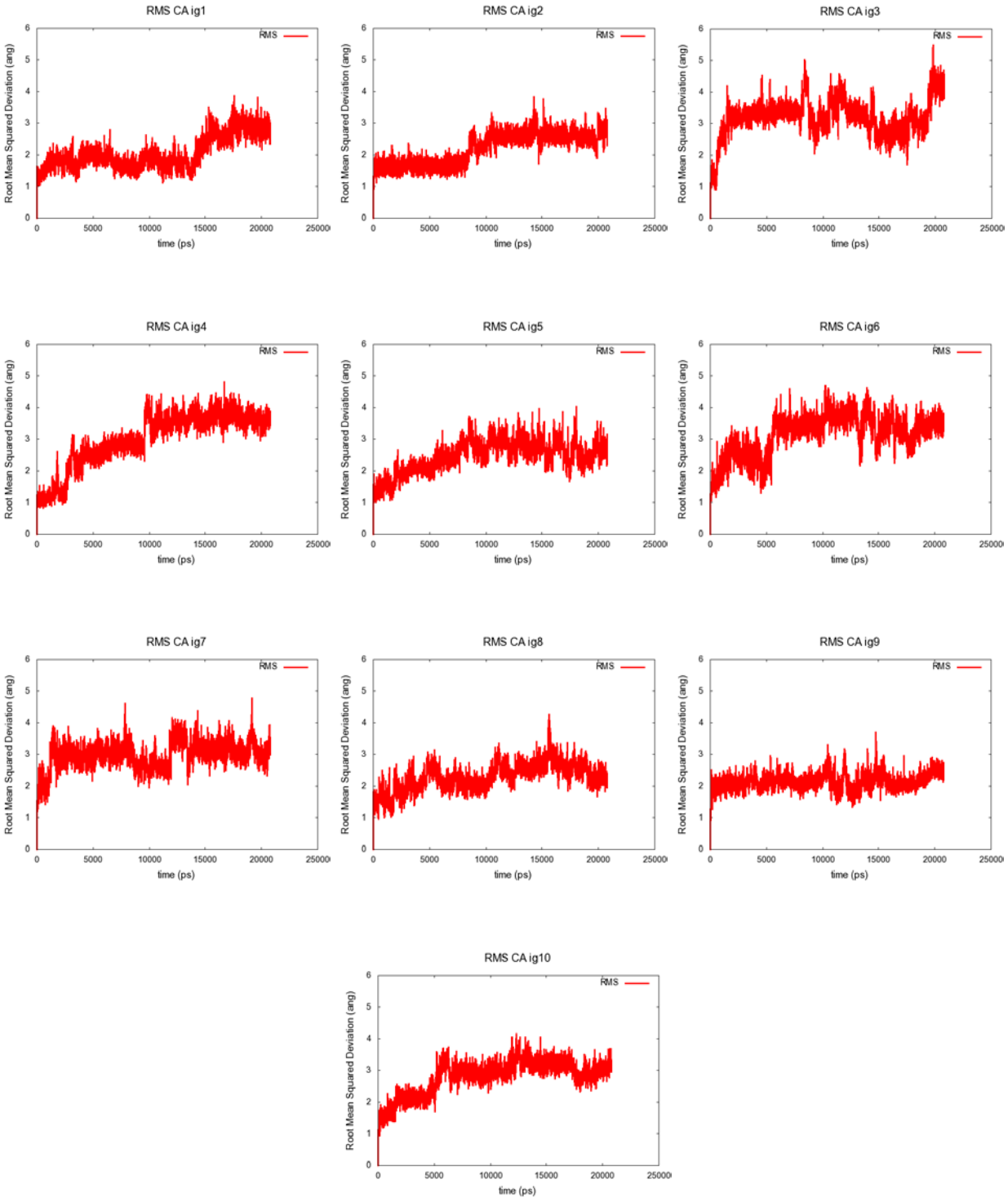

Figure 3-13. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype C (PDB ID 2R5P) closed conformation. The first frame of each simulation was used as the reference.

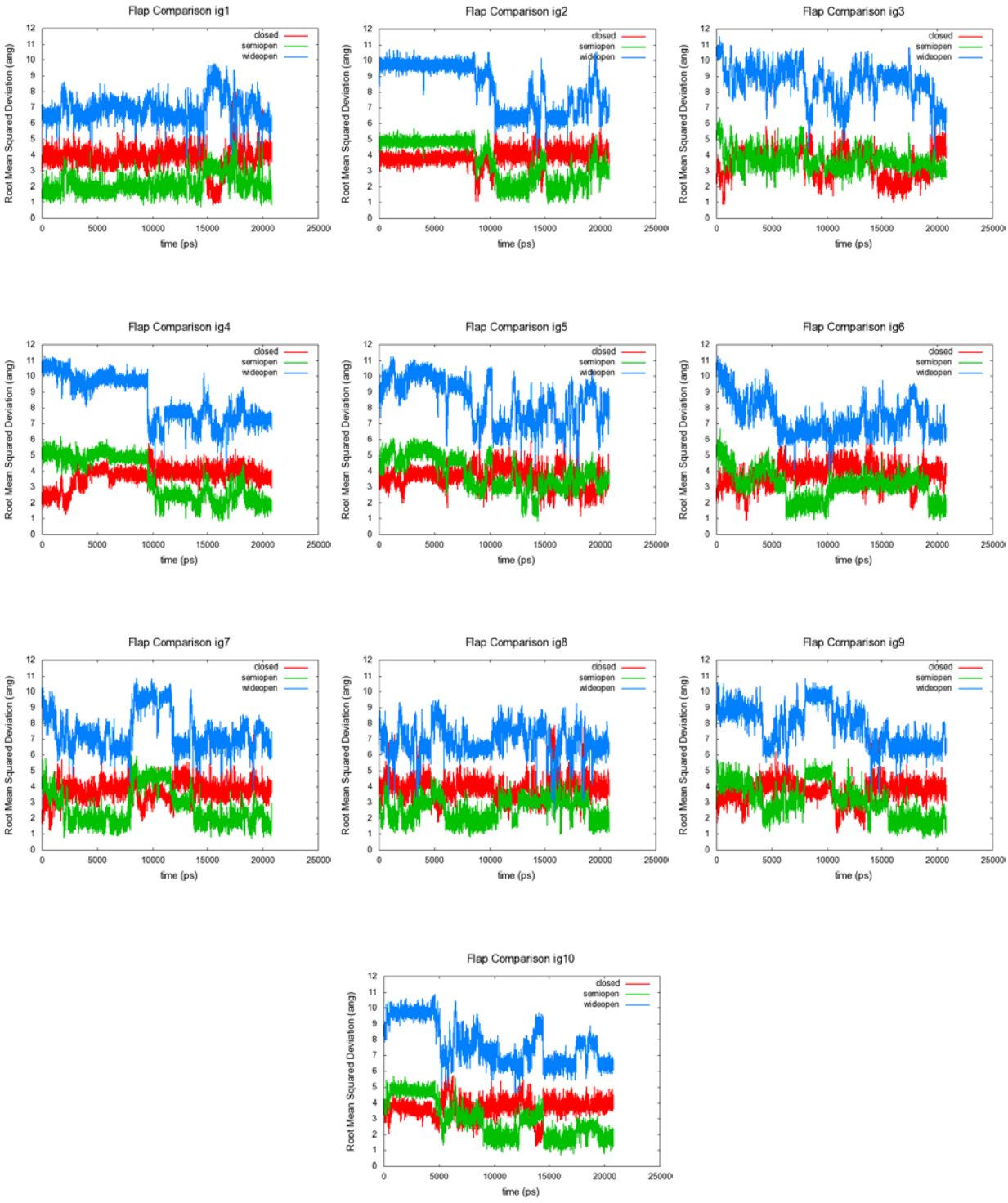

Figure 3-14. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype C (2R5P). The red was versus the closed conformation. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

The range of fluctuations on average is between 1-4 angstroms for each of the ten simulations in Figure 3-13. The initial starting coordinates were gathered from 2R5P. The highest fluctuations are seen in the ig3 simulation when deviations become as high as 5 angstroms around 20,000 ps. The rapid fluctuation can be attributed to the fact that at that same time frame we see that the flaps revert from favoring a close conformation to that of a semi-open conformation. In addition, the flaps suddenly adapt a conformation, which deviates more than 4 angstroms from the closed conformation.

It is evident from the ten simulations in Figure 3-14 that flaps favor a conformation that is semi-open. The ig2, ig4, ig9, and ig10 simulations first initially favor tucked or curled conformation but then it is seen that the semi-open conformation becomes the dominant form. The fluctuations from the semi-open structure become as low as 1 angstrom in certain simulations. Even though the dynamics began from a conformation which was closed, it still favored the semi-open conformation, which was not the case for the structures that were initially closed for subtype B. The closed, tucked and curled conformations are not the predominant form of the flaps in any of the dynamics, which is not the case in the simulations of 2R8N. In the ig2 and ig10 simulation we do see that the both the closed, semi-open, and wide-open conformation are not favored in the beginning. This suggests as stated earlier that flaps are in a tucked or curled position.

The wide-open conformation is not the dominant form in any of the simulations. We do however see points of the simulation in which the deviations become as small as 3 angstroms versus the wide-open structure as seen in the ig8 simulation in Figure 3-14. As stated earlier, the lack of the dynamics favoring the wide-open conformation could be do sampling or either the reference structure we chose to represent the wide-open conformation.

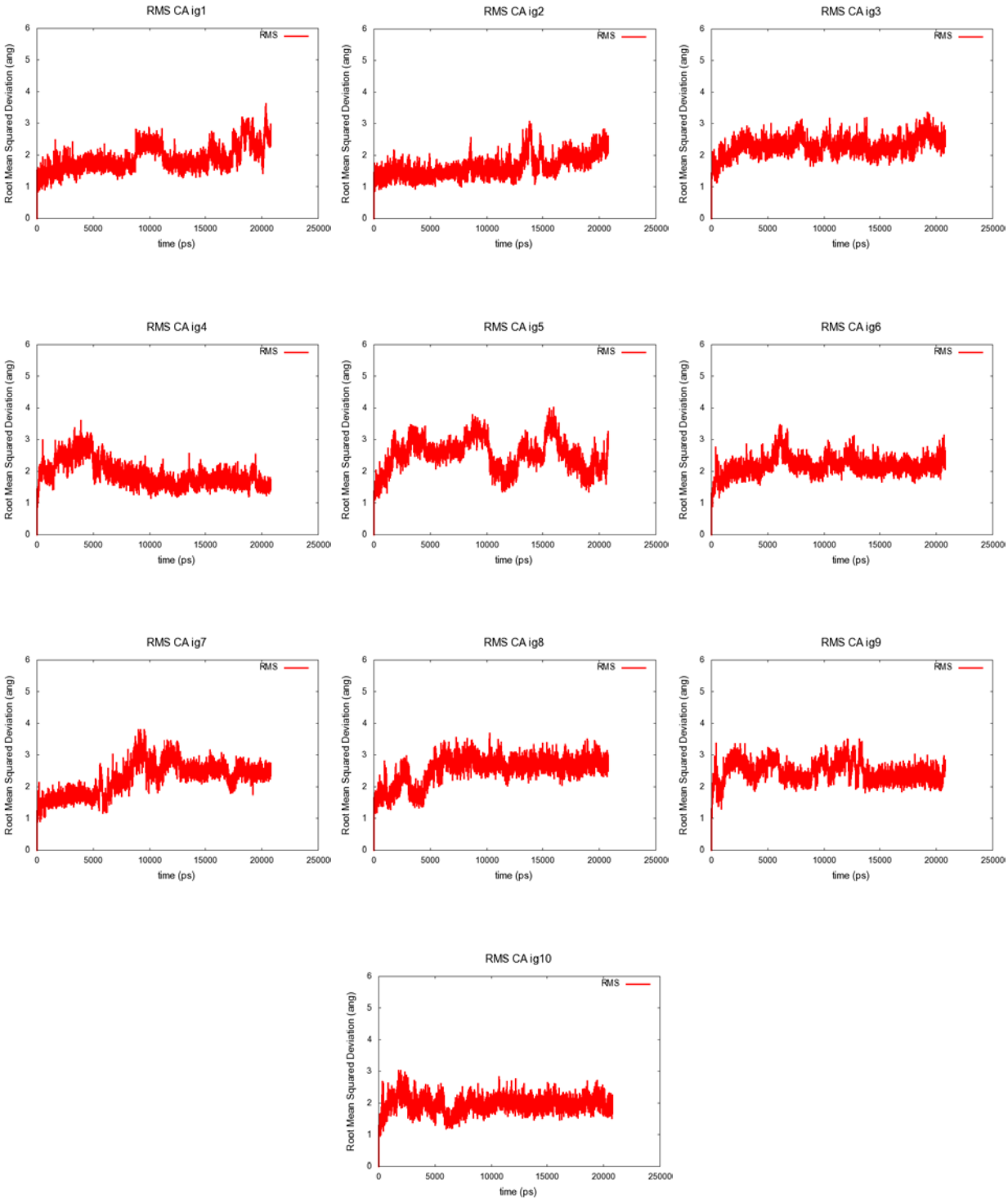

Figure 3-15. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype C (PDB ID code 2R5Q) closed conformation. The first frame of each simulation was used as the reference.

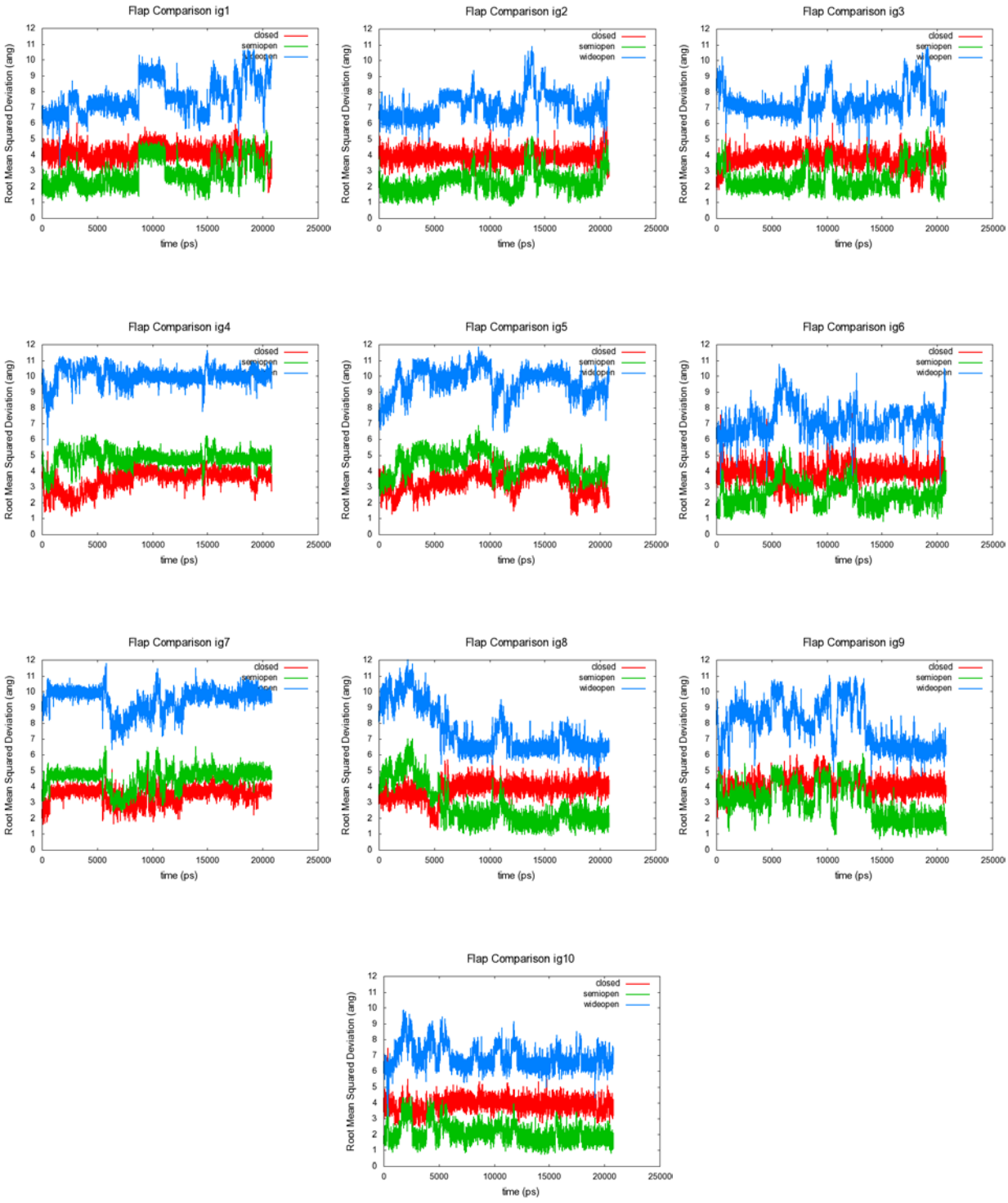

Figure 3-16. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer versus time for the ten different simulations of Subtype C (2R5Q) closed conformation. The red was versus the closed conformation. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

The initial starting coordinates were gathered from 2R5Q for the results shown in Figures 3-15 and 3-16. Only small deviations are seen in Figure 3-15 with a low of 1 angstrom and a high of 3.5 angstroms. No sudden or abrupt rises in the fluctuations are observed in any of the simulations in Figure 3-15. As seen in the other two simulations of subtype C, (2R8N, 2R5P) the main orientation of the flaps appears to be semi-open. In addition, the fluctuations become as close as 1 angstrom to the semi-open conformation as seen in the ig1 and ig3 simulations. Unlike the simulations 2R5P, there are a few simulations in which the semi-open conformation is not the preferred conformation. The ig4, ig5 and ig7 simulations are examples in which a conformation other than semi-open is dominant. In the ig4 and ig7 simulations the dominant conformation could either be curled or tucked because none of the fluctuations versus the closed, semi-open, and wide-open become lower than 3 angstroms. Further examples of the tucked or curled can be seen in the ig1 simulation. In the particular case of ig1, although the semi-open conformation is the predominate form, around 10,000 ps we see that the dynamics do not look like any of the three reference structures. The wide-open conformation as seen in the previous results is not the dominant form in any of the simulations shown in Figure 3-16. We do observe fluctuations in the flaps that become as low as 3 angstroms as seen in the ig3, ig6 and ig9 simulations. One trend that is seen throughout is that even when the dynamics favor the wide-open structure, it is only for a brief instance and further investigation is needed to explain this. One hypothesis is that it is more energetically favorable for the flaps to be in semi-open conformation rather the wide-open conformation.

All three simulations of subtypes C favor the semi-open conformation as a whole. There are fewer instances in the simulation of subtype C where the dynamics favor the closed conformation when compared to the simulations of Subtype B.

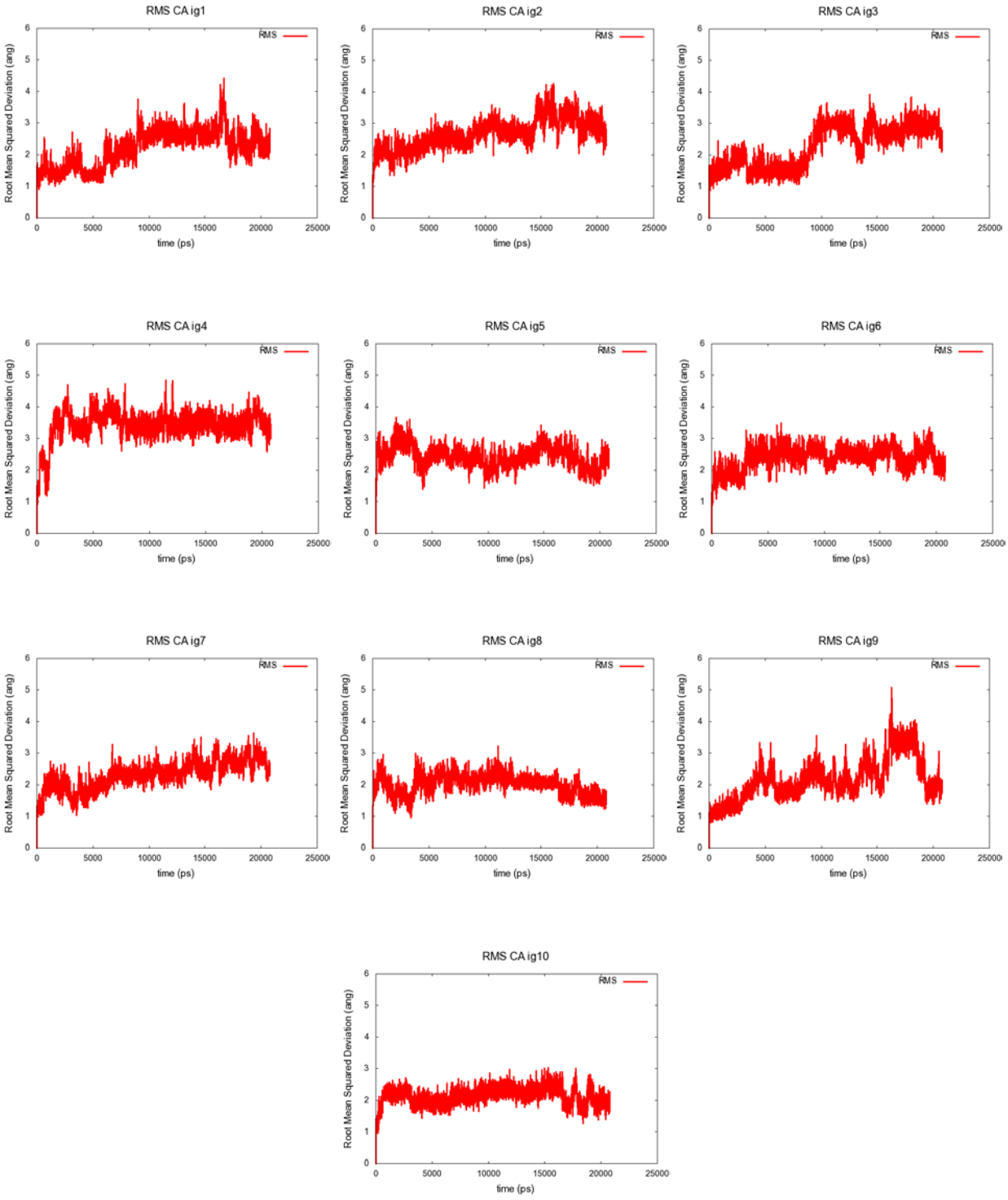

Figure 3-17. Illustrates the Root Mean Squared Deviation of the Cα versus time of ten different simulations of Subtype F (PDB ID code 2P3C) closed conformation. The first frame of each simulation was used as the reference.

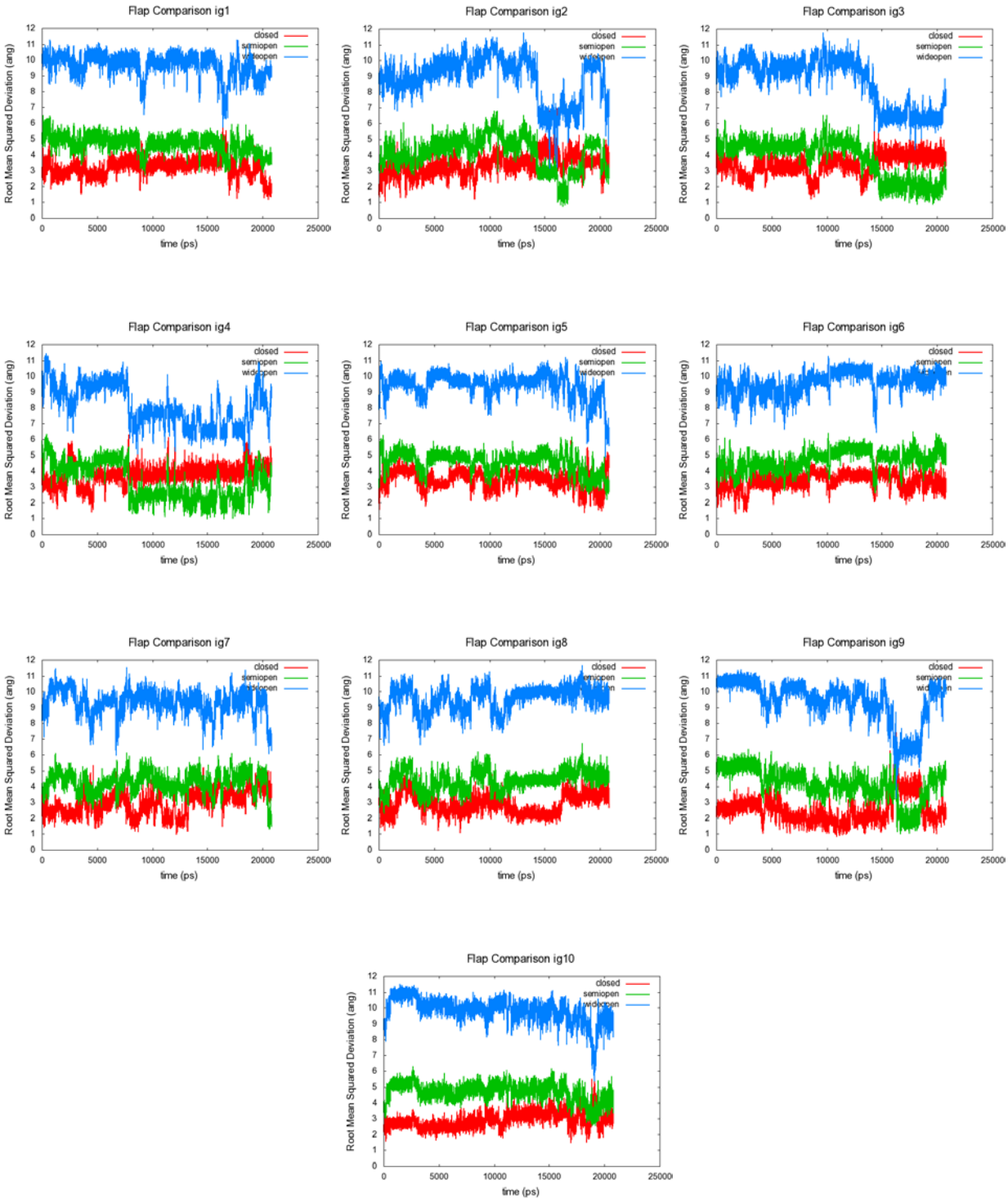

Figure 3-18. Root Mean Squared Deviation of the flap residues 43-58 Cα and 43'-58' Cα of each monomer of Subtype F 23PC closed conformation. The red was versus the closed conformation. The green was versus the semi-open conformation. The blue was versus the (fully) wide-open conformation.

Figures 3-17 and 3-18 are the results in which the initial coordinates are gathered from 23PC. Generally most of the simulations in Figure 3-17 fluctuate between 1 and 3.5 angstroms. However in the ig5 simulation, a peak is observed at around 5 angstroms, which is due to the flaps abruptly switching from a closed to a semi-open conformation. As seen with the previous simulations, the predominant form of the flaps is not the semi-open conformation. Only in the ig3 and ig4 simulations do we see that the dynamics greatly favor the semi-open conformation. It is the closed, tucked or curled structure, which is favored by most of the simulations in Figure 3-18. An example of the dynamics giving rise to either a tucked or curled structure can be seen in the ig8 simulation. Around 3,000 ps in the ig8 simulation, we see that the dynamics do not favor either a closed, semi-open, or wide-open structure. The wide-open conformation is not the dominant structure in any of the simulations. However, some of the simulations do show fluctuations as low as 4 angstroms, which is evident in simulations ig2 and ig9. The flaps of subtype F do appear to be the most rigid when compared to the other subtypes.

3.3.2 Comparison of the Three Initial Structures in Subtypes B and Subtypes C

In this next section we will compare the three different structures of subtype B and three different structures of subtype C. As mentioned earlier, the initial coordinates for two of the three structures of subtypes B and C were gathered from a closed conformation with an inhibitor bound. We hypothesize that regardless of the initial starting structure, the dynamics more or less should look the same because all three have the same amino acid sequence. Our hypothesis rests on the fact that we have sampled enough phase space as well. In order to address the issue as stated in the methods section of this chapter, each structure was simulated using ten different initial velocities. If we have not sampled enough phase space then surely the simulations, which began closed should favor each other while the simulation that began semi/wide-open should be different.

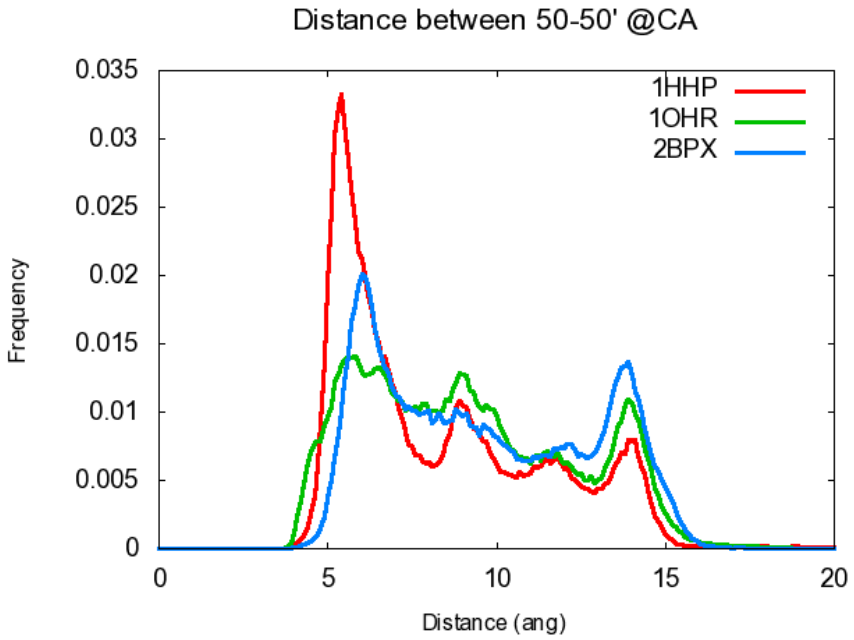

Figure 3-19. Histogram of the distance between ILE50 and ILE50'. Comparison of the Subtype B simulations

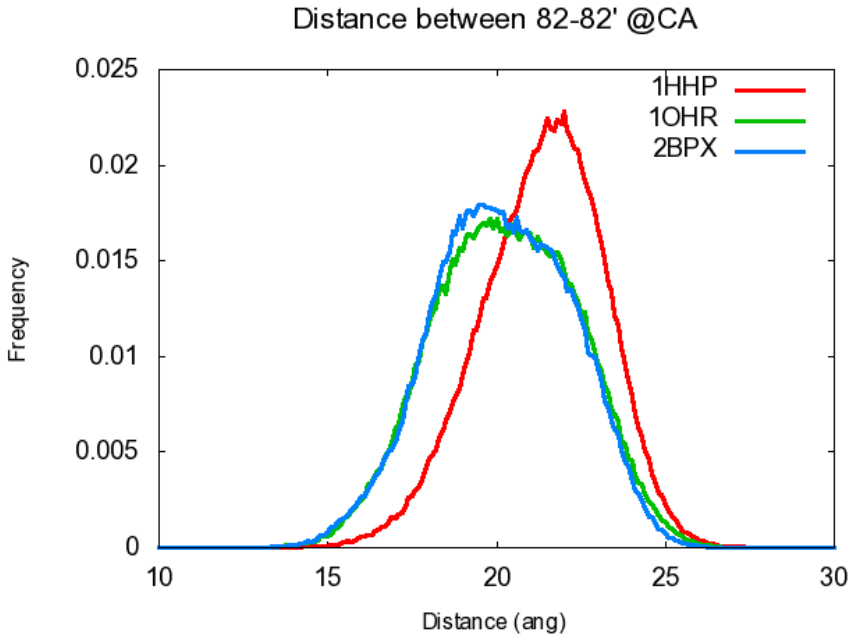

Figure 3-20. Histogram of the distance between VAL82 and VAL82'. Comparison of the Subtype B simulations.

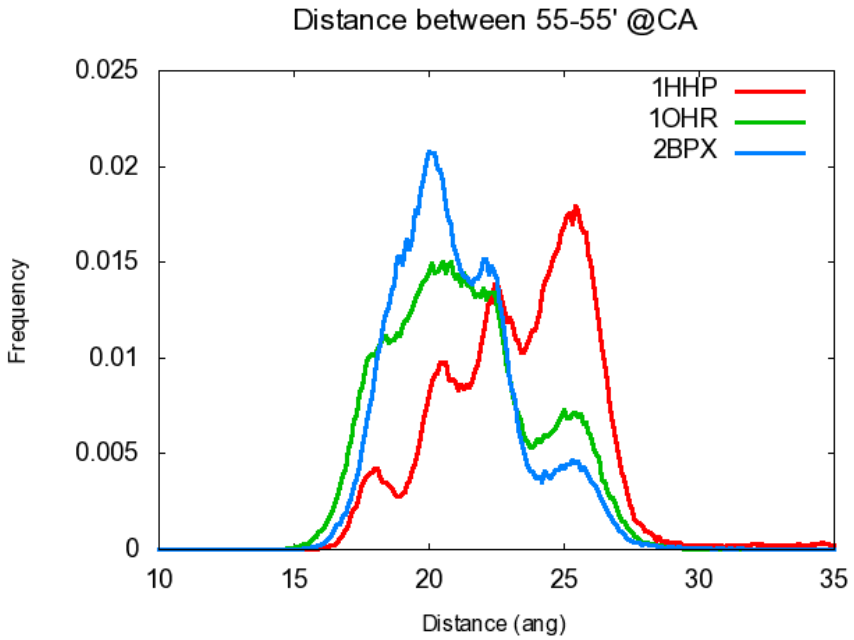

Figure 3-21. Histogram of the distance between LYS55 and LYS55'. Comparison of the three B subtype simulations.

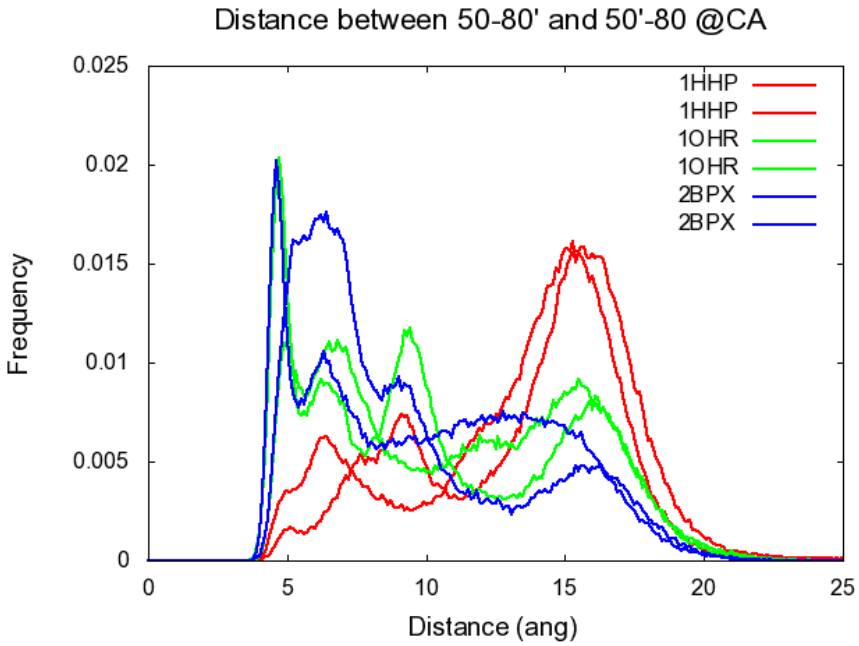

Figure 3-22. Histogram of the distance between the ILE50 and THR80' and vice versa. Comparison of the three subtype B simulations

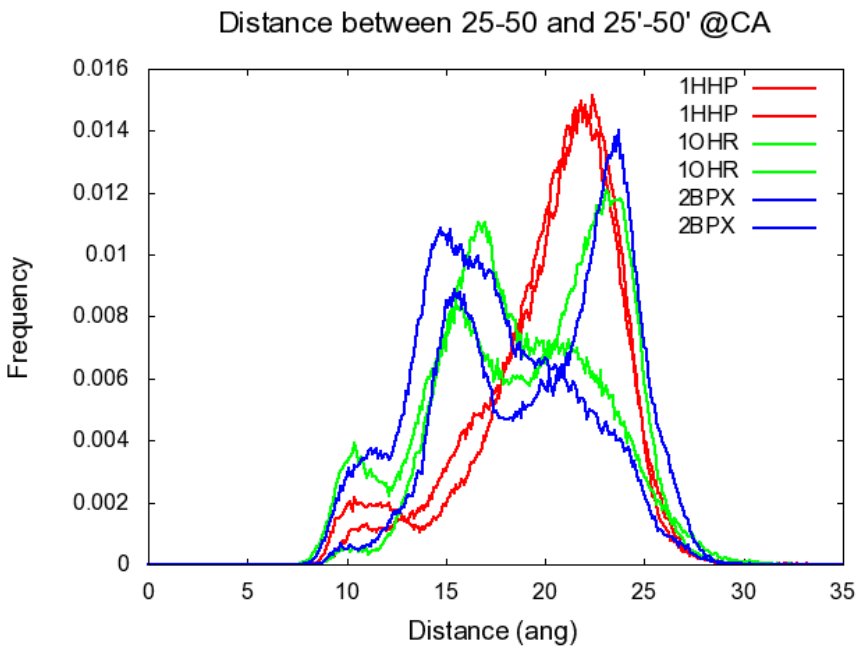

Figure 3-23. Histogram of the distance between ILE50 and Asp25 of each monomer. Comparison of the three subtype B simulations.

We will first begin with comparing the three simulations of Subtype B. The first distance taken into consideration was the distance between residues ILE50 and ILE50'. This distance gives a good indication of the flap tip separation. As seen in Figure 3-19, all three simulations have three peaks. The peaks are found at 5, 9 and 13 angstroms. Figure 3-20 is a measure of the distances between the residues VAL82 and VAL82', which are located in the active site. The 2BPX and 1OHR simulations exhibit peaks at around 20 angstroms. The 1HHP simulation, however, has a peak that is displaced from the other two simulations. The peak for the 1HHP simulation is seen at 23 angstroms. The difference between the two simulations, which started closed, and one simulation that initially started semi-open, is a lot more pronounced in Figure 3-20 than in Figure 3-19. The distance between the LYS55 and LYS55' is shown in Figure 3-21. As seen in the other figures the 2BPX and 1OHR simulations look the same. The 1HHP and the 2BPX simulation differ the most while the 1OHR simulation shares characteristics of both

simulations. Although all three simulations exhibit peaks at 26 angstroms, the peak for the 1HHP simulation is far more pronounced. Figure 3-22 is a measure between ILE50 of one monomer and the THR80' of the other monomer. When the flaps invert the distances these residues become very close. Figure 3-22 illustrates the first time in which the distribution of distances of 2BPX and the 1OHR simulations differ. Next, the distance between the ILE50 and Asp25 of each monomer was measured as shown in Figure 3-23. ILE50 is located in the flaps and Asp25 is located at the base of the active site. It is seen that the distance profiles of all three simulations look very similar, which has not been the case in previous figures. From the previous figures it is suggested that the dynamics of the 2BPX and the 1OHR look the same. The dynamics of the 1HHP simulations do not look the same.

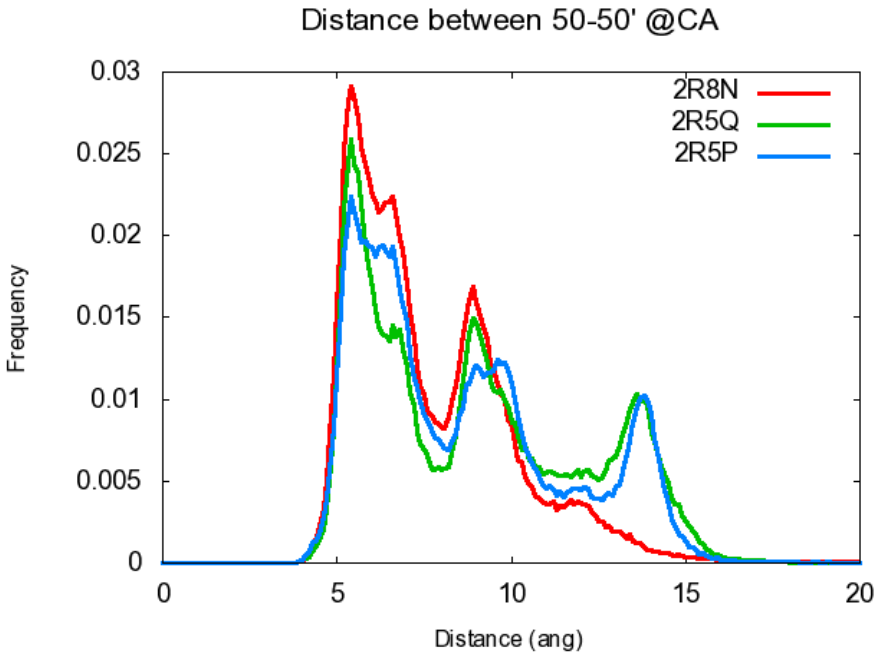

Figure 3-24. Histogram of the distance between ILE50 and ILE50'. Comparison of the three Subtype C simulations.

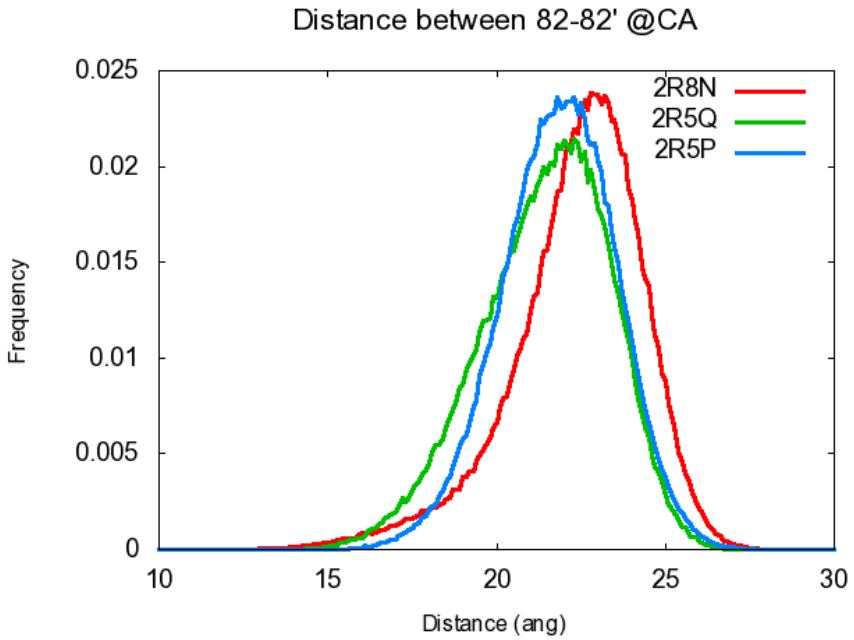

Figure 3-25. Histogram of the distance between VAL82 and VAL82'. Comparison of the three subtype C simulations.

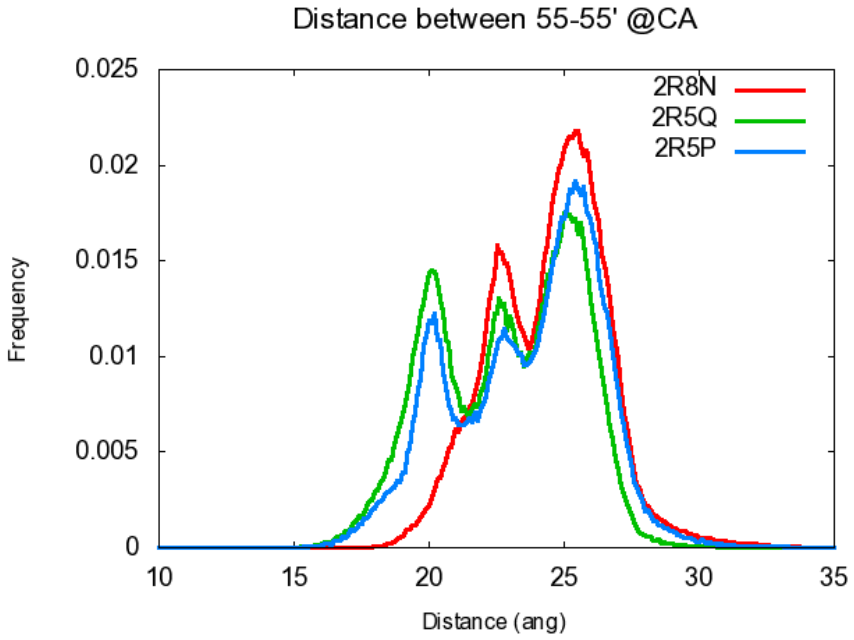

Figure 3-26. Histogram of the distance between LYS55 and LYS55'. Comparison of the three subtype C simulations.

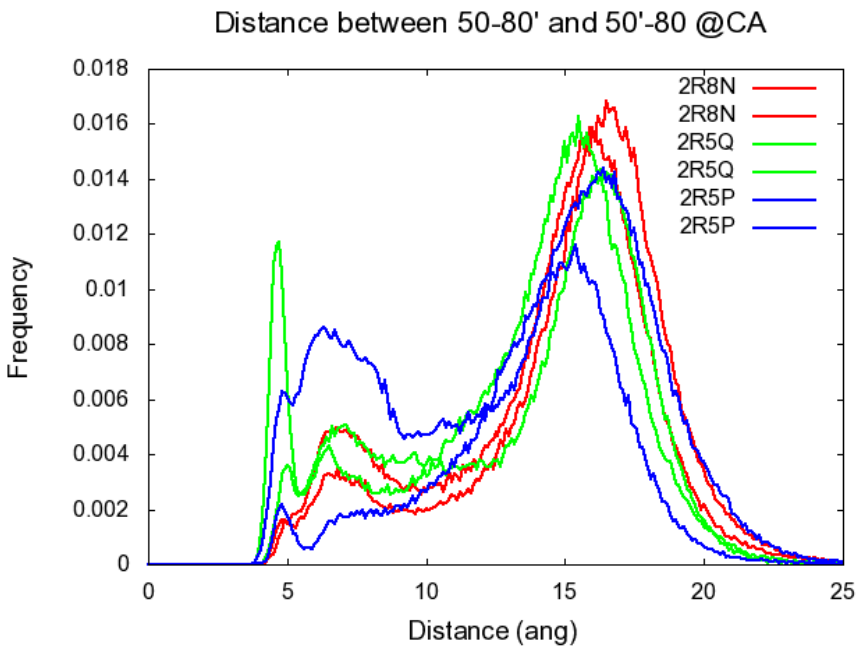

Figure 3-27. Histogram of the distance between ILE50 and THR80' and ILE50' and THR80. Comparison of the three subtype C simulations.

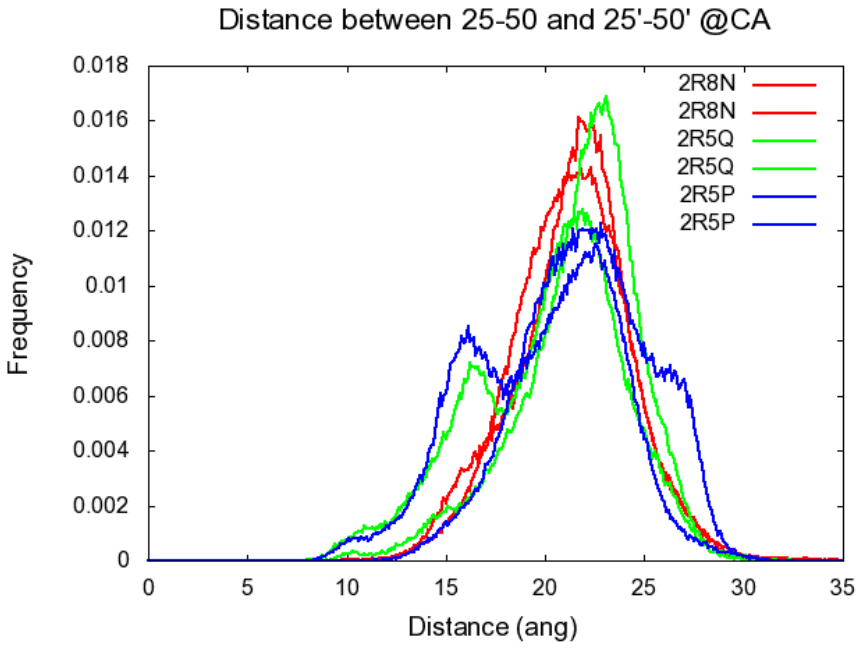

Figure 3-28. Histogram of the distance between ASP25 and ILE50' and ASP25' and ILE50. Comparison of the three subtype C simulations.

Figure 3-24 is a histogram of the distance between ILE50 and ILE50'. The simulations look very similar except that the peak at 13 angstroms is more prominent in the 2R5P and 2R5Q simulations than in the 2R8N simulations. As seen before with the subtype B simulations, the two structures, which began initially closed, look the same. In Figure 3-25, the distance between VAL82 and VAL82' was measured. The distance profiles of all three simulations favor each other. As seen when the same distance was measured for the subtype B simulations, 2R8N max peak is slightly displaced from 2R5P and 2R5Q peaks. The LYS55 and the LYS55' distances profiles exhibit very similar trends as seen in Figure 3-26. The 2R5Q and 2R5P dynamics exhibit peaks at 17 angstroms, which is not exhibited by the 2R8N dynamics. Figure 3-27 shows a histogram of the ILE50 and THR80 distances of each monomer. Other than the strong peak at 5 angstroms for 2R5Q, all three simulations look almost the same, which was not the case for the three subtype B simulations. The last plot Figure 3-28 show the distances between ILE50 and Asp25. The plots of all three look very similar with slight differences. The peaks at 13 angstroms are more prominent in the 2R5P and 2R5Q. 2R5Q does exhibit a slight peak at 26 angstroms, which is not prevalent in 2R8N and 2R5P. The dynamics of all three simulations look a lot closer than the three simulations of subtype B. The reason why there is better agreement for the subtype C simulations of our original hypothesis, that the dynamics should look the same no matter the initial coordinates, needs further investigation.

3.3.3 Comparison between the Different Subtypes

In this section, a comparison will be made between each of the four different subtypes. The nomenclature will be as follows: subtype A is 3IXO, Subtype B is 1OHR, Subtype C is 2R5Q, and Subtype F is 2P3C. The reason why 1OHR and 2R5Q were chosen as representative structures of each of their subtypes was because they begin from initially closed conformations. As mentioned in section 3.2, although subtype A was crystallized in its apo form the flaps were

found to be in a closed conformation and subtype F was crystallized with the inhibitor bound TL-3 with the flaps in a closed conformation.

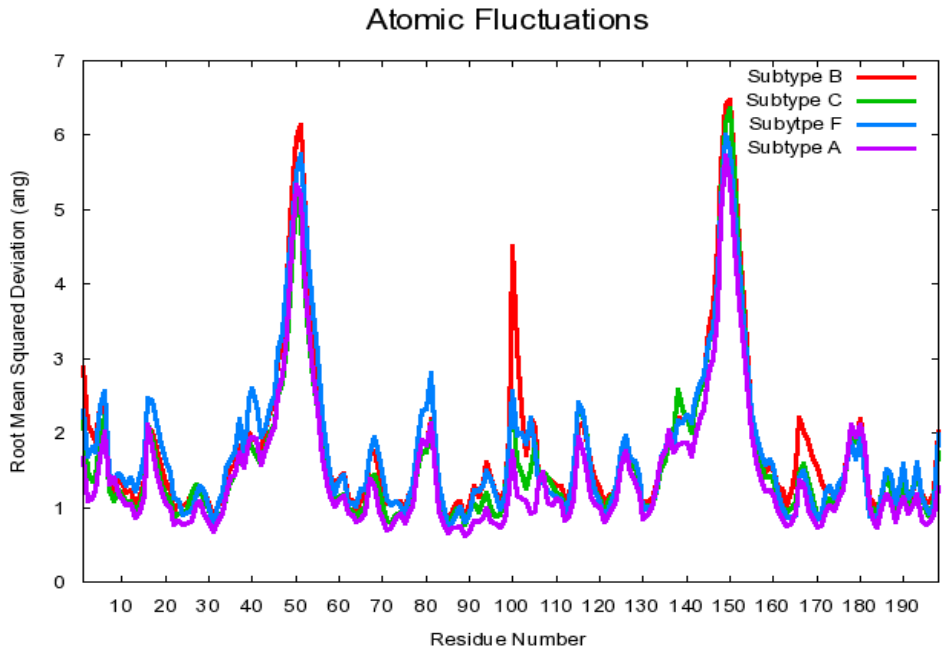

Figure 3-29. The atomic fluctuations of the Cα. Comparison of subtypes A, B, C, F

The reference structures used to produce Figure 3-29 were determined by calculating the average structure for each of the simulations using the *ptraj* module of AMBER. In Figure 3-29 all of the subtypes exhibit high fluctuations between residues 43 through 58. These particular residues are found in the flaps, so the large fluctuations are due to the fact that the flaps sample many different conformations as seen in previous figures. Of the flap residues, residues 48-52 move the most because this area is Gly rich[46,62]. In general, most of the residues fluctuate 1 or 2 angstroms from their average structure. The residue 99, which is a PRO fluctuates as high as 5 angstroms in subtype B and further investigation is needed to explain this phenomena. One explanation could be that PRO is located at the N-terrminus, which is the same for all of the other sequences, is unable to form a stable interaction with any of the surrounding residues, thus

preventing it from being so floppy. However, in subtypes A and C there is a H69K mutation and because of the longer side-chain, it can form a stable salt-bridge with the C-terminal carboxy group of PHE[58]. The stable salt-bridge adds rigidity to the PHE which prevents it from moving so much.

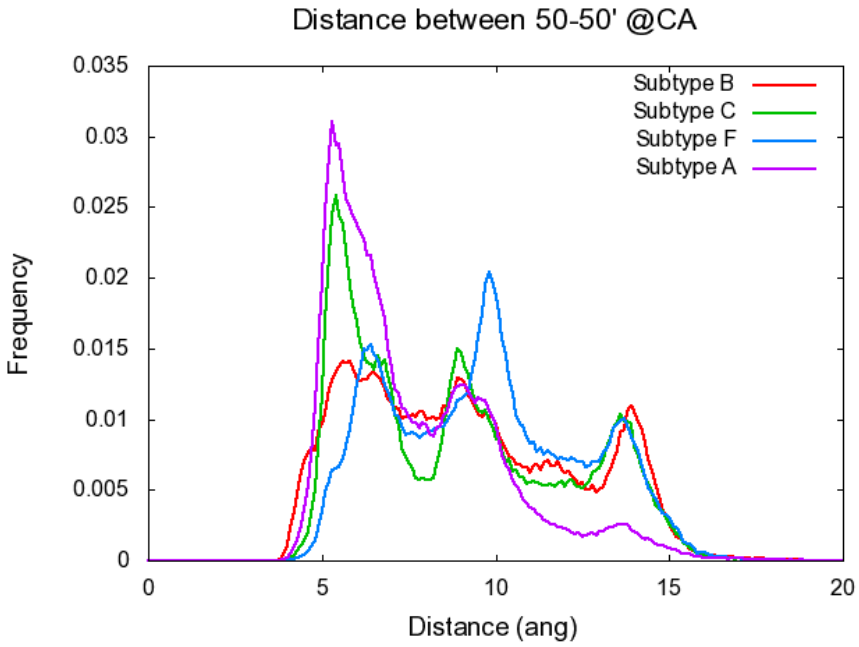

Figure 3-30. Histogram of the distance between ILE50 and ILE 50'. Comparison of the different subtypes

In Figure 3-30 we compare the flap tip separation for each of the subtypes. Three of the four subtypes exhibit three distinct peaks. Subtype A only exhibits two peaks. All the subtypes have peaks that are around 7 angstroms. Subtype A and C appear the most at this particular distance. The second peak is around 8 angstroms for all of the subtypes except for subtype F. Subtype F is commonly seen throughout the dynamics to be at 10 angstroms. All but subtype A show a significant peak at more or less 13 angstroms. It can be suggested that the flap tip separation is on average smaller for subtype A then rest of the subtypes. Figure 3-30 strengthens the argument that wide-open conformation is not a large portion of the dynamics sampled. The

distances of which we generally see the flap tip separation spans from 5 to 15 angstroms. Whether or not the ILE50 and ILE50' has any direct correlation to drug resistance is still uncertain. Although, one could propose that the distance corresponds to the ease at which a ligand could enter the active site and potentially bind.

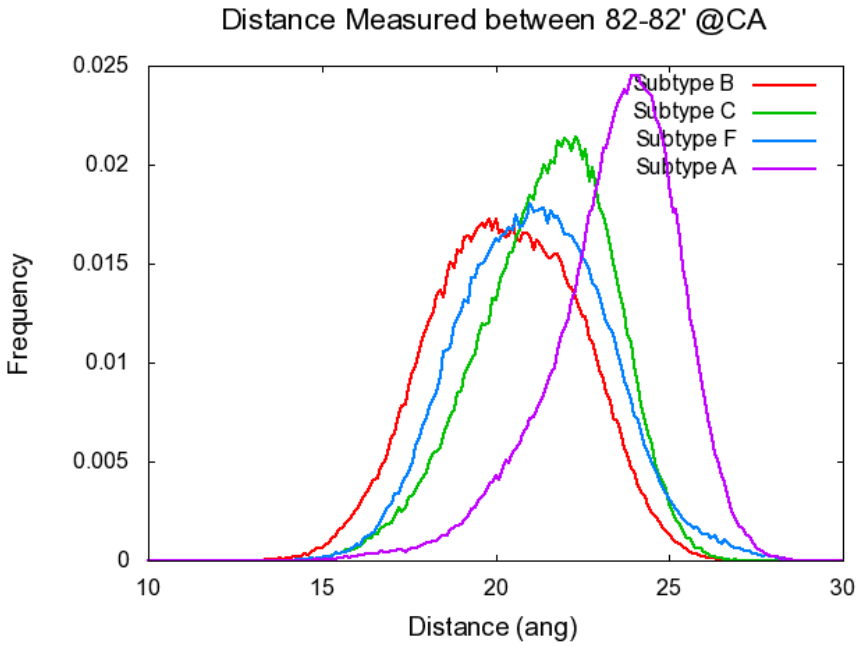

Figure 3-31. Histogram of the distance VAL82 and VAL82'. Comparison of the different subtypes.

Figure 3-31 gives some insight on the average size of the active site. Mutations are commonly seen in these active site residues, which result in the loss of hydrophobic interactions with the ligand[63,64,65]. Unlike the ILE50 ILE50' distance, this distance does correspond to drug resistance. Subtype A exhibits the largest separation between these residues with a peak at 25 angstroms. The differences between subtypes B, C, and F are less pronounced. However, the distance between the pair of residues is slightly larger for subtype C. The implications of the significance of this difference between subtype A and the other subtypes need further investigation.

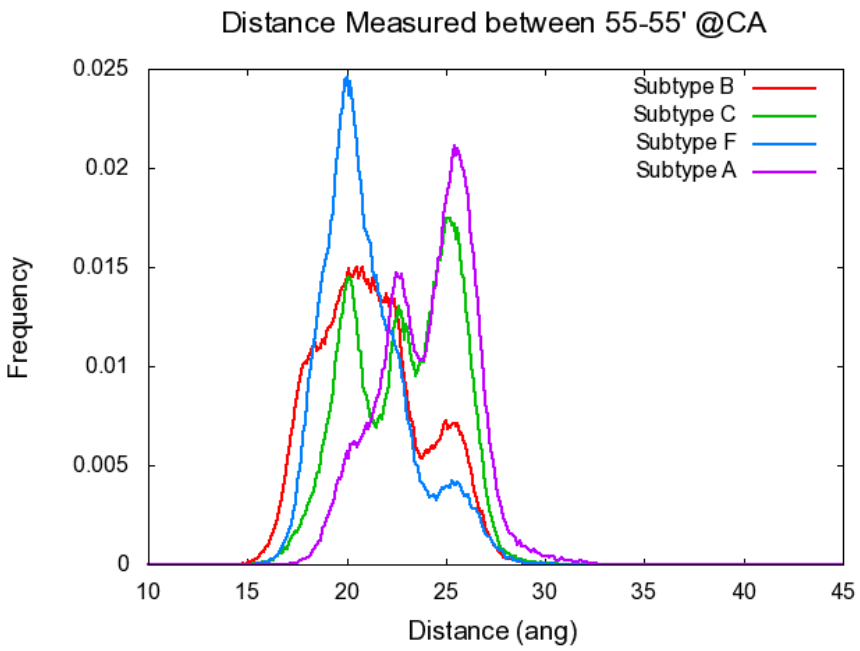

Figure 3-32. Histogram of the distance between LYS55 and LYS55'. Comparison between the different subtypes

The LYS55 LYS55' distance gives further insight into possible orientation of the flaps. The distance between the two residues seems to be the greatest for subtypes A, and C. They both exhibit strong peaks at around 27 angstroms. In the case of subtype F, the distance between the pair of residues appears to be smallest. It has the highest probability of being at 18 angstroms. Though the peaks at which we commonly see these distances do not agree with those produced by Kear et al., they do however match the trends[60]. The reason why the exact distances of the peaks do not coincide can be linked to a myriad of reasons, for example, different sequence simulated, solvent modeled implicitly, absence of spin-labels, etc. The experimental results of the EPR experiments have been able to be reproduced using computational methods[66]. The measurement of this distance has not been directly correlated to drug resistance but does however offer insight into flap orientation[66]. In figure 3-36 we show how the LYS55 LYS55' gives a better indication of the flaps when compared to ILE50 ILE50' distance.

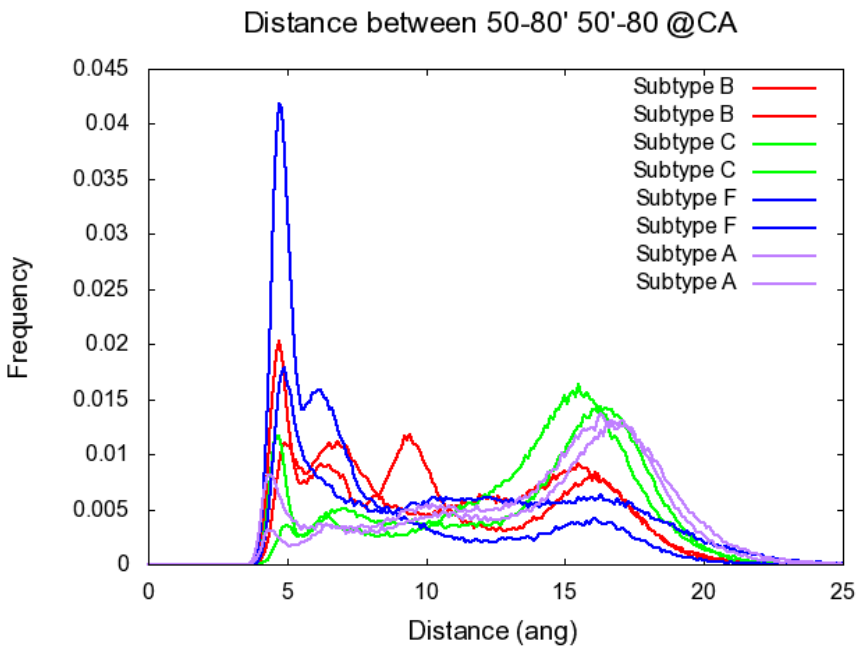

Figure 3-33. Histogram of the difference between ILE50 and THR80', ILE50' and THR80. Comparison of different subtypes.

This distance as of yet has not been correlated to drug resistance but does however suggest possible conformation of the flaps. As seen during the dynamics when the flaps invert, the distance between ILE50 – THR80' becomes small while the distance between ILE50'-THR80 becomes large, thus making inversion of the flaps possible. This is the reason why the spectrum at which we see this distance is so broad. One interesting trend observed in Figure 3-33 is that both sets of distances for each subtypes are more or less the same except for subtype F. One of ILE50-THR80' distances for subtype F is seen to be closer more frequently than the other and further investigation is needed in order to explain this. In the example of subtype F when the one flaps is interacting strongly the 80's loop of the opposite monomer that flap actually lies over the active site thus preventing access to it. The strong 50-80' interaction is stabilized by hydrogen bonding. Subtype A does have a mutation in the 80's loop that might explain why this interaction is not as prevalent in its dynamics. Subtypes B and C have the same sequence and it not clear as

to why this same strong interaction is not observed. This conformation could have some impact on ease at which such things as protease inhibitors can bind but further experiments need to be performed in order to confirm this. Another explanation could be that we modeled the solvation implicitly instead of explicitly and that this strong interaction seen between the flap tips and the 80's loop of the other monomer is due to the absence of water.

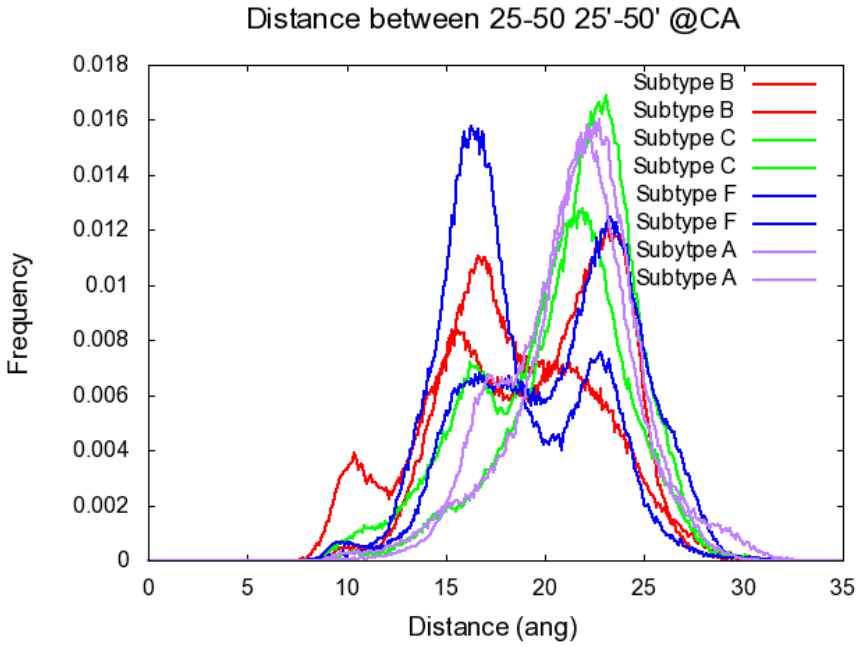

Figure 3-34. Histogram of the distance between ASP25 and ILE50 of each monomer. Comparison of the different subtypes.

Similar to the VAL82-VAL82' distance, the ASP25-ILE50 distance does provide information on the size of the active site and has been linked to drug resistance. Subtypes A and C have fairly large peaks which are seen at 25 angstroms, and it coincides with them both having a larger VAL82-VAL82' distance. There is a small peak at 10 angstroms in subtype B that is not prevalent in any of the other subtypes. Subtypes B and F exhibit fairly significant peaks at 15 angstroms. The distances of the ASP25-ILE50 and ASP25'-ILE50' of each monomer differ greatly except in the case of subtype of A. This difference is more so evident in subtypes B and

F. The reason for the distances being so different could be due to the sampling. In case of the subtype F it could be because of the strong interaction that the flap tips of one monomer makes with the opposing 80's loop. Furthermore, the ASP-25-ILE50 distance does have a large spectrum because it can become as close as 10 angstroms and on the other hand become as large as 30 angstroms from each other. One can assume that the flaps are in closed conformation when this distance is small and in a wide-open conformation when this distance becomes large. Clustering analysis needs to be performed in order to confirm this observation

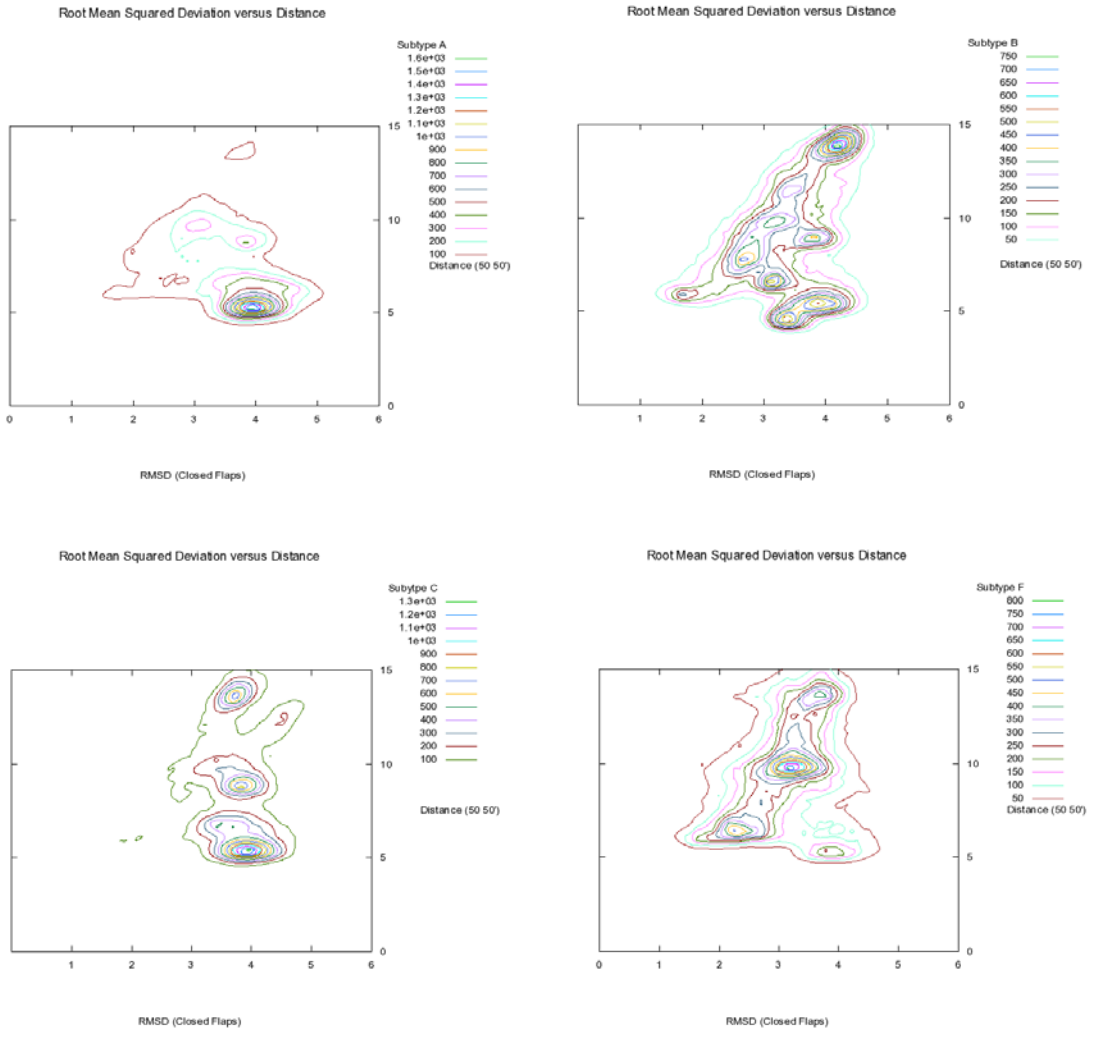

Figure 3-35. Contour plot of the Root Mean Squared Deviation of the closed conformation versus the ILE50 and ILE50' Cα. First row from left to right, Subtype A, Subtype B, and second row subtype C, subtype F.

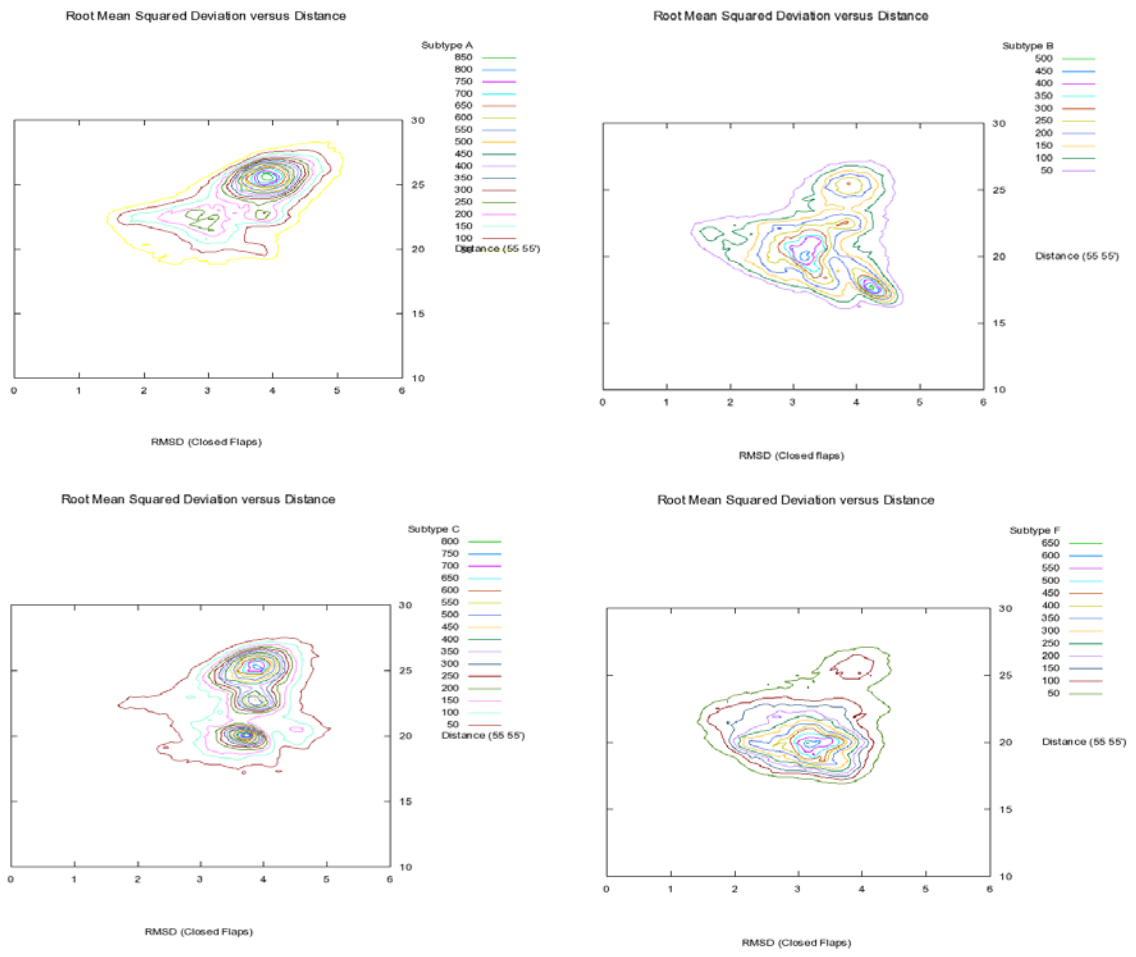

Figure 3-36. Contour plot of the Root Mean Squared Deviation of the closed conformation versus the LYS55 and LYS55' Cα. First row from left to right, Subtype A, Subtype B, and second row subtype C, subtype F.

The significance of Figures 3-35 and 3-36 is to validate whether or not one can assume a certain structure at a given distance. The reason the ILE50-ILE50' and the LYS55-LYS55' distances were chosen is because these distances indicate flap orientation, as stated earlier. It is believed that when these distances are small, the flaps are in a conformation that resembles a closed conformation and when the distances are large the flaps resemble an open conformation. In Figures 3-35 and 3-36 it is observed that at one particular distance the RMSD can fluctuate as much as 1 or 3 angstroms so caution should be excercised when using the distances to correlate flap conformation.

3.4 Conclusions

From the results above one could propose that subtypes A and C are more "open" as evident in Figures 3-29, 3-31, and 3-33. Furthermore this could explain why certain protease inhibitors are not as effective for inhibiting these proteases but further investigation is needed[56,67]. Furthermore, the fact the subtype C is more "open" might explain the decrease of the catalytic efficiency of subtype C when compared to subtype B in kinetic studies done by Coman et al[56]. On the other hand, subtype F appears to be the most rigid. This finding agrees with the hypothesis proposed by Sanches et al. which suggests that mutations in hinge region result in flap stiffening[57].

There are some salt-bridge interactions made in the flap region that could be linked or explain flap flexibility. The salt-bridge interaction of ARG41-Asp60 is formed more frequently in subtypes B and C. The LYS41-ASP60 interaction is formed but is not that frequent in subtype A, however the LYS41-GLU60 salt-bridge in subytpe F is rarely seen. Subtype F makes instead the saltbridge LYS43-GLU60 which is not seen in any of the other subtypes. Subtypes A, B and C make the the salt-bridge GLU35-Arg57 and in the case of subtype F, ASP35-LYS57. These differences in salt-bridge may coincide as to why the flaps are seen to invert more for subytpes B and C. As previously explained in Chapter 2, one must be cautious because the GB model does over-stabilize salt-bridge interactions.

Genomi et al. identified residues 11-20 and 11'-20' of each monomer as a new potential target for protease inhibition[50]. In addition, we hypothesize that residues 40-43 and 59-61 of each monomer could also be proposed as a alternative site for protease inhibition as well. The validity of this hypothesis will be tested in future work. The results presented in previous section of this chapter strengthen the hypothesis that non-active site conformations alter flap dynamics, the size

or shape of the active site[63,68]. All of these changes can have an impact on the affinity of the ligand to the protease.

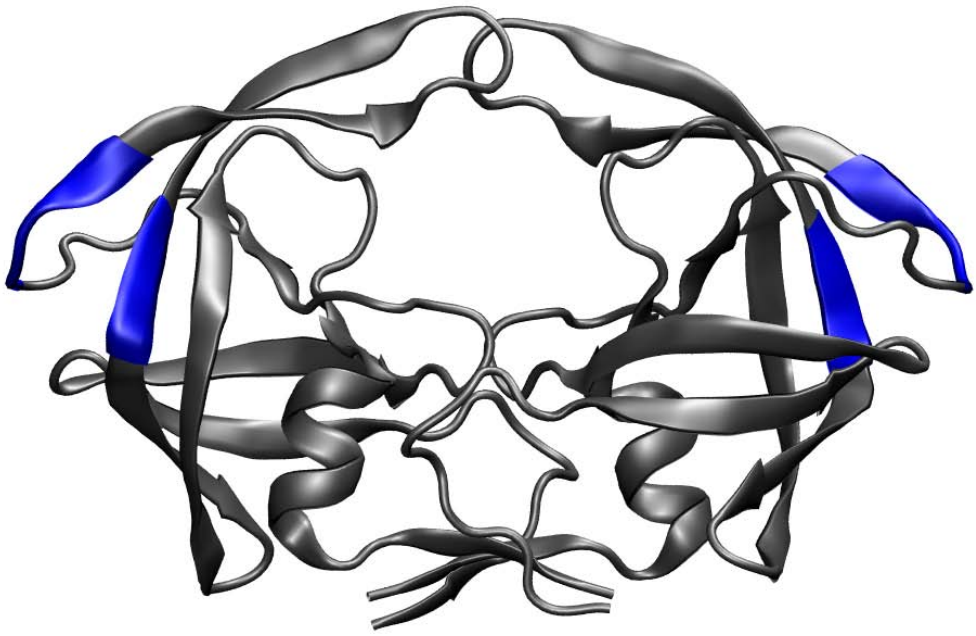

Figure 3-37. Potential new region for protease inhibition. The residues highlighted in blue are 40-43 and 59-61 of each monomer.

LIST OF REFERENCES

(1) Weiss, A.; Hollander, H.; Stobo, J. *Annu. Rev. Med* **1985**, *36*, 545-562.

(2) Robertson, D. L.; Anderson, J. P.; Bradac, J. A.; Carr, J. K.; Foley, B.; Funkhouser, R. K.; Gao, F.; Hahn, B. H.; Kalish, M. L.; Kuiken, C.; Learn, G. H.; Leitner, T.; McCutchan, F.; Osmanov, S.; Peeters, M.; Pieniazek, D.; Salminen, M.; Sharp, P. M.; Wolinsky, S.; Korber, B. *Science* **2000**, *288*, 55d.

(3) Gürtler, L. G.; Hauser, P. H.; Eberle, J.; von Brunn, A.; Knapp, S.; Zekeng, L.; Tsague, J. M.; Kaptue, L. *J. Virol* **1994**, *68*, 1581-1585.

(4) Simon, F.; Mauclère, P.; Roques, P.; Loussert-Ajaka, I.; Müller-Trutwin, M. C.; Saragosti, S.; Georges-Courbot, M. C.; Barré-Sinoussi, F.; Brun-Vézinet, F. *Nat. Med* **1998**, *4*, 1032-1037.

(5) Plantier, J.; Leoz, M.; Dickerson, J. E.; De Oliveira, F.; Cordonnier, F.; Lemée, V.; Damond, F.; Robertson, D. L.; Simon, F. *Nat. Med* **2009**, *15*, 871-872.

(6) Penny, M. A.; Thomas, S. J.; Douglas, N. W.; Ranjbar, S.; Holmes, H.; Daniels, R. S. *AIDS Res. Hum. Retroviruses* **1996**, *12*, 741-747.

(7) Archer, J.; Robertson, D. L. *AIDS* **2007**, *21*, 1693-1700.

(8) Hemelaar, J.; Gouws, E.; Ghys, P. D.; Osmanov, S. *AIDS* **2006**, *20*, W13-23.

(9) Baeten, J. M.; Chohan, B.; Lavreys, L.; Chohan, V.; McClelland, R. S.; Certain, L.; Mandaliya, K.; Jaoko, W.; Overbaugh, J. *J. Infect. Dis* **2007**, *195*, 1177-1180.

(10) Ruben, S.; Perkins, A.; Purcell, R.; Joung, K.; Sia, R.; Burghoff, R.; Haseltine, W. A.; Rosen, C. A. *J. Virol* **1989**, *63*, 1-8.

(11) Zapp, M. L.; Green, M. R. *Nature* **1989**, *342*, 714-716.

(12) Lama, J.; Mangasarian, A.; Trono, D. *Curr. Biol* **1999**, *9*, 622-631.

(13) Rogel, M. E.; Wu, L. I.; Emerman, M. *J. Virol* **1995**, *69*, 882-888.

(14) Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. *Cell* **1997**, *89*, 263-273.

(15) Pantophlet, R.; Burton, D. R. *Annu. Rev. Immunol* **2006**, *24*, 739-769.

(16) Coakley, E.; Petropoulos, C. J.; Whitcomb, J. M. *Curr. Opin. Infect. Dis* **2005**, *18*, 9-15.

(17) Chan, D. C.; Kim, P. S. *Cell* **1998**, *93*, 681-684.

(18) Robertson, D. L.; Hahn, B. H.; Sharp, P. M. *Journal of Molecular Evolution* **1995**, *40*, 249-259.

(19) Pollard, V. W.; Malim, M. H. *Annu. Rev. Microbiol.* **1998**, *52*, 491-532.

(20) Kim, J. B.; Sharp, P. A. *Journal of Biological Chemistry* **2001**, *276*, 12317-12323.

(21) Witte, O. N.; Baltimore, D. *J. Virol* **1978**, *26*, 750-761.

(22) Crawford, S.; Goff, S. P. *J. Virol* **1985**, *53*, 899-907.

(23) Whitford, D. *Proteins: structure and function*; J. Wiley & Sons, 2005.

(24) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A* **1988**, *85*, 4686-4690.

(25) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, *337*, 615-620.

(26) Wlodawer, A.; Miller, M.; Jaskólski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. *Science* **1989**, *245*, 616-621.

(27) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A* **2006**, *103*, 915-920.

(28) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J Chem Inf Model* **2007**, *47*, 1590-1598.

(29) Meek, T. D.; Rodriguez, E. J.; Angeles, T. S. *Meth. Enzymol* **1994**, *241*, 127-156.

(30) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.

(31) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585-590.

(32) Leach, A. *Molecular Modelling: Principles and Applications*; 2nd ed.; Prentice Hall, 2001.

(33) Frenkel, D.; Smit, B. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications*; 2nd ed.; Academic Press, 2001

.

(34) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; 2nd ed.; Wiley, 2004.

(35) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.

(36) Hockney, R. *Methods Comput. Phys.* **1970**, *9*, 136-211.

(37) Swope, W.; Andersen, H.; Berens, P.; Wilson, K. *The Journal of Chemical Physics* **1982**, *76*, 649, 637.

(38) Beeman, D. *Journal of Computational Physics* **1976**, *20*, 130-139.

(39) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712-725.

(40) Onufriev, A.; Case, D. A.; Bashford, D. *J Comput Chem* **2002**, *23*, 1297-1304.

(41) Zhou, R. *Proteins* **2003**, *53*, 148-161.

(42) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc* **2002**, *124*, 11258-11259.

(43) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. *Structure* **1999**, *7*, 1047-1055.

(44) Galiano, L.; Bonora, M.; Fanucci, G. E. *Journal of the American Chemical Society* **2007**, *129*, 11004-11005.

(45) Blackburn, M. E.; Galiano, L.; Veloro, A. M.; Fanucci, G. E. *Biophysical Journal* **2009**, *96*, 310a.

(46) Scott, W. R.; Schiffer, C. A. *Structure* **2000**, *8*, 1259-1265.

(47) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *J. Am. Chem. Soc* **2006**, *128*, 2812-2813.

(48) Tóth, G.; Borics, A. *Biochemistry* **2006**, *45*, 6606-6614.

(49) Bonomi, M.; Gervasio, F. L.; Tiana, G.; Provasi, D.; Broglia, R. A.; Parrinello, M. *Biophys. J* **2007**, *93*, 2813-2821.

(50) Genoni, A.; Morra, G.; Merz, K. M.; Colombo, G. *Biochemistry* **2010**, *49*, 4283-4295.

(51) Batista, P. R.; Wilter, A.; Durham, E. H. A. B.; Pascutti, P. G. *Cell Biochem. Biophys* **2006**, *44*, 395-404.

(52) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391-1396.

(53) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. *J. Med. Chem* **1997**, *40*, 3979-3985.

(54) Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. *Acta Crystallogr. D Biol. Crystallogr* **1998**, *54*, 1053-1060.

(55) Coman, R. M.; Robbins, A. H.; Goodenow, M. M.; Dunn, B. M.; McKenna, R. *Acta Crystallogr. D Biol. Crystallogr* **2008**, *D64*, 754-763.

(56) Coman, R. M.; Robbins, A. H.; Fernandez, M. A.; Gilliland, C. T.; Sochet, A. A.; Goodenow, M. M.; McKenna, R.; Dunn, B. M. *Biochemistry* **2008**, *47*, 731-743.

(57) Sanches, M.; Krauchenco, S.; Martins, N. H.; Gustchina, A.; Wlodawer, A.; Polikarpov, I. *J. Mol. Biol* **2007**, *369*, 1029-1040.

(58) Robbins, A. H.; Coman, R. M.; Bracho-Sanchez, E.; Fernandez, M. A.; Gilliland, C. T.; Li, M.; Agbandje-McKenna, M.; Wlodawer, A.; Dunn, B. M.; McKenna, R. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 233-242.

(59) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J Comput Chem* **2005**, *26*, 1668-1688.

(60) Kear, J. L.; Blackburn, M. E.; Veloro, A. M.; Dunn, B. M.; Fanucci, G. E. *Journal of the American Chemical Society* **2009**, *131*, 14650-14651.

(61) Layten, M.; Hornak, V.; Simmerling, C. *J. Am. Chem. Soc* **2006**, *128*, 13360-13361.

(62) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc* **2005**, *127*, 13778-13779.

(63) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R. S.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. *Biochemistry* **2004**, *43*, 12141-12151.

(64) Rose, R. B.; Craik, C. S.; Stroud, R. M. *Biochemistry* **1998**, *37*, 2607-2621.

(65) Logsdon, B. C.; Vickrey, J. F.; Martin, P.; Proteasa, G.; Koepke, J. I.; Terlecky, S. R.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. *J. Virol* **2004**, *78*, 3123-3132.

(66) Galiano, L.; Ding, F.; Veloro, A. M.; Blackburn, M. E.; Simmerling, C.; Fanucci, G. E. *Journal of the American Chemical Society* **2009**, *131*, 430-431.

(67) Velazquez-Campoy, A.; Vega, S.; Freire, E. *Biochemistry* **2002**, *41*, 8613-8619.

(68) Clemente, J. C.; Hemrajani, R.; Blum, L. E.; Goodenow, M. M.; Dunn, B. M. *Biochemistry* **2003**, *42*, 15029-15035.

BIOGRAPHICAL SKETCH

Terry Dwight McGee Jr. was born in Macon, Georgia. He received his bachelor's degree from the Florida Agricultural and Mechanical University located in Tallahassee, Florida in 2005. The fall of 2008 he entered the University of Florida chemistry graduate program to study physical chemistry with an emphasis in computational chemistry. Upon entering the graduate program, he joined the lab of Dr. Adrian E. Roitberg.