Northumbria Research Link

MULTILEVEL MOLECULAR MODELLING OF STRUCTURE-FUNCTION RELATIONSHIPS IN ENZYMES

WARISPREET SINGH

PhD

2016

MULTILEVEL MOLECULAR MODELLING OF STRUCTURE-FUNCTION RELATIONSHIPS IN ENZYMES

WARISPREET SINGH

A thesis submitted in partial fulfilment

of the requirements of the

University of Northumbria at Newcastle for
the degree of

Doctor of Philosophy

Research undertaken in the

Faculty of Health and Life Sciences

October, 2016.

Abstract

Proteins are large flexible molecules and conformational dynamics is one of their fundamental properties which correlate the protein's structure and function [1] [2]. The crystal structure of biomolecular systems such as enzymes reveals the important atomistic details in terms of the ligand binding and possible mechanism albeit providing no information about how conformational flexibility and dynamics influences the protein structure and its key determinants. There is no information regarding the electronic structure and the chemically relevant components of the enzymes and how the protein environment affects the electronic structure. In order to provide understanding of how the conformational flexibility influences structure-function relationships of the enzymes, we applied classical molecular dynamics simulations using Gromacs [3] [4] [5] and Amber [6] [7] packages. The effect of the protein environment on the electronic structure of the active site were studied using Quantum Mechanics and Molecular Mechanics (QM/MM) [8] using ONIOM [9] [10, 11] implemented in Gaussion09 [12] [13].

Tyrosylproteinsulfotransferase (TPST): TPSTs catalyze the transfer of negatively charged sulfate group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to the hydroxyl group of a tyrosine residue of polypeptide to form a tyrosine O4-sulfate ester [14]. The binding of the substrate peptide showed more open conformation in Tyrosylproteinsulfotransferase-2 (TPST-2) enzyme in contrast to the crystal structure [15] [16]. There were identification of new hydrophobic interactions responsible for the stabilization of the enzyme dimer [16]. The binding of the substrate and cofactor to the apoenzyme contributed to the stability of the whole active complex, influenced the local interactions in the binding site and importantly, affects the pattern of the correlated motions in the entire molecule [16].

NirE an S-adenosyl-L-methionine dependent Methyltransferase: The NirE enzyme catalyses the transfer of a methyl group from the S-adenosyl-L-methionine (SAM) to uroporphyrinogen III and serves as a novel potential drug target for the pharmaceutical industry. The binding of the substrate contributes to the stabilization of the structure of the full enzyme complex [17]. The conformational changes influence the orientation of the pyrrole rings of the substrate [17]. The mutations of binding and active site residues leads to sensitive structural changes which influence binding and catalysis [17].

Matrix metalloproteinase-1 (MMP-1): The molecular dynamics studies on the Matrix metalloproteinase-1 (MMP-1) were in good agreement with the experimental observation that in the MMP-1•THP (Triple Helical Peptide) X-ray crystallographic structure MMP-1[18] is in a "closed" conformation [19]. The interactions of the THP with both the CAT and HPX domains of MMP-1 are dynamic in nature, and the linker region of MMP-1 influences the interactions and dynamics of both the CAT and HPX domains and collagen binding to MMP-1 [19]. The mutations in the MMP-1 have distinct impact on the correlated motions in the MMP-1•THP. An increased collagenase activity corresponded to the appearance of a unique anti-correlated motion and decreased correlated motions, while decreased collagenase activity corresponded both to increased and decreased anti-correlated motions.

Non-heme Fe^{2+} and 2-oxoglutrate (2OG): The non-heme Fe^{2+} and 2-oxoglutrate (2OG) dependent dioxygenases such as FTO, AlkB, PHF8 and KIA1718 perform important tasks in homeostasis through the methylation of DNA and histone proteins. The linker region

shows increase conformational flexibility and dynamics in PHF8 and KIA1718 and is important for the catalysis. The jelly roll motif structure also showed conformational stability for all demethylases and indicates its vital role in maintaining the iron geometry in the active site. The N domain of the FTO enzyme and the L1 loop region showed increased conformational flexibility and dynamics. The QM/MM optimized structure of reactant complex showed the effect of the conformational flexibility.

An important insight into the structure function relationship of different enzymes has been obtained by applying a large number of Atomistic Molecular Dynamics Simulations and Quantum Mechanics/Molecular Mechanics to different enzymes which cannot be gained experimentally. The effect of conformational dynamics, flexibility and important interactions of the active site residues can be used in chemical biology and biotechnology for structure based drug design and in the engineering of novel biocatalyst.

Acknowledgment

Above all, I praise the Lord Almighty for his grace and mercy.

First and foremost I would like to thank Dr. Christo Z Christov and Dr. Tatyana Karabencheva-Christova for their immense support, guidance and suggestions without which this research was not possible. I would like to thank Dr Christo Z Christov and Dr Tatyana Karabencheva-Christova for their time and efforts in teaching me chemistry and training me on how to conduct research effectively. They have always provided me with excellent research papers to read and then to present them in the weekly group meeting.

I would also like to thank Professor Iain Sutcliffe for giving me advise on my career and referring me to Dr. Christo Christov to pursue my PhD studies.

I would like to Mr. Billal Munir, Mujtaba Billal and Mr. Jon Ainsley for their support during my PhD studies.

I would like to thank my uncle Mr. Jagtar Singh Sidhu and his family for their immense support throughout my stay in England.

I would like to thank Mr. Imran Ahmed for his support throughout my MSc and PhD degree.

Last certainly not least I would like to thanks my respected parents for providing me with the fantastic opportunity to pursue my studies abroad and hugely appreciate the sacrifices they have made in fulfilling this pursuit. Finally my regards to my dear sister Ziart for her support and well wishes.

This thesis is dedicated to my late Grandmother Bibi Kartar Kaur Kailey

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work carried out under the supervision of Dr Christo Z Christov, Professor Iain C. Sutcliffe, and Dr Tatyana Karabencheva-Christova.

I also confirm that this work acknowledges the opinions, ideas and contributions from the work of others.

Name: Warispreet Singh

Signature:

Date:

LIST OF CONTENTS

List of Abbreviations

Abbreviation	Name
2OG	2-OxoGlutrate
3D	Three Dimensional
Å	Ångström
AMBER	Assisted Model Building with Energy Refinement
APO	Apoenzyme
CAT	Catalytic domain
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at Harvard Macromolecular Mechanics
DCCA	Domain Cross-Correlation Analysis
DFT	Density Functional Theory
ECM	Extracellular Matrix
GAFF	General AMBER force field
GROMOS	GROningen MOlecular Simulation computer program package
H	Hamiltonian
HF	Hartree Fock
HPX	Hemopexin-like domain
KDN	Lysine demethylases
MCPB	Metal Centre Parameter Builder
MD	Molecular Dynamics Simulations
MM	Molecular Mechanics
MMGBSA	Mechanics/Generalized Born Surface Area
MMP-1	Matrix Metalloproteinases-1
NirE	SAM-dependent Uro'gen III MethylTransferase
NPT	constant Number of particles (N), system Pressure (P) and Temperature (T)
NVT	constant Number (N) of particles, Volume (V), and Temperature (T)
ONIOM	Our own N-layered Integrated molecular Orbital and Molecular Mechanics
OPLS	Optimized Potentials for Liquid Simulations
PAPS	3'-phosphoadenosine 5'-phosphosulfate

PCA	Principal Component Analysis
PDB	Protein Data Bank
PME	Particle Mesh Ewald
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/ Molecular Mechanics
RESP	Restrained Electrostatic Potential
Rg	Radius of gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
SAH	S-Adenosyl-L-homocysteine
SAM	S-adenosyl-L-Methionine
SASA	Solvent Accessible Surface Area
SCF	Self-Consistent Field
SPC	Single Point Charge
TDSE	Time Dependent Schrödinger Equation
THP	Triple Helical Peptide
TPST-2	Tyrosylproteinsulfotransferase-2
VMD	Visual Molecular Dynamics
WTFC	Wild Type Full Complex

CHAPTER ONE –INTRODUCTION

Enzymes are complex biological macromolecules which catalyse various biochemical reactions in the cellular environment. The success of enzymes as biological catalysts is attributed to their ability to greatly accelerate the reaction rate by reducing the activation energy barrier and their high specificity for various ligand molecules [1] [20] [21]. Enzymology is an established field with a wealth of scientific research carried out in past decades. Understanding the mode of action of enzymes in their natural physiological environment has provided valuable knowledge for both scientific and pharmaceutical sector. The root of many diseases lies in the malfunction of enzymes, understanding these and targeting drugs towards them is now a routine procedure in the development of novel pharmaceutical compounds [22] [23] [24].

Experimental investigations in the field of Biochemistry and Biophysics have provided vital information regarding the enzyme environments influence on the kinetics, thermodynamics and mechanism of a reaction. In spite of the wealth of information on the mode of action of enzymes, there are still some open questions regarding the detailed reaction mechanisms and conformational dynamics which require further explanations. In pursuit of fostering our understanding of reaction mechanisms (through the characterisation of short lived intermediate species and even more short lived transition states) and understanding the conformational dynamics of enzymes have created the modern field of computational enzymology [25] [26] [21] [27]. Computational enzymology provides an alternative platform via simulations to obtain vital atomistic and electronic structural information on the enzyme dynamics and catalysis. Out of all the computational techniques, Molecular dynamics simulations can provide vital knowledge about the conformational flexibility and dynamics of the enzyme ligand complex which

contributes significantly to the enzymes function [1]. Molecular Mechanics (MM) relies on empirical force field data which has been parametrised to obtain potential energy functions for bimolecular systems [28]. There is no term in MM force field which describes the effect of molecular wave function or electronic density hence it is not capable of providing electronic details which are at the heart of catalysis. In order to study the electronic effect in enzyme catalysis, Quantum Mechanics (QM) is the method of choice [29]. The cluster models of the active site have provided new insights into reaction pathways in enzymes [29] [30] [31]. However due to the very high computational cost of QM methods, it is restricted to only a few hundred atoms or even less if a higher level of QM theory is used. In order to take advantage of both the QM and MM methods together in studying enzymes, QM/MM is a technique of choice [8] [32]. The combination of both MM and QM methodology in a combined algorithm is commonly known as QM/MM (Quantum Mechanics / Molecular Mechanics), where MM treats the rest of the system; QM provides details on the breaking and formation of chemical bonds, charge transfer and electron excitation in the active site of the enzymes [32] [10] [11] [33]. The QM/MM approach takes advantage of the accuracy of computationally demanding QM with the low computational cost of MM to study large complex enzyme systems, they have provided meaningful insight into catalytic process and have significantly increased our knowledge of enzyme catalysis [9] [34] [35] [36] [37] [38].

1.1 QUANTUM MECHANICS

The quantisation of energy levels proposed by Max Plank and the concept of the wave particle duality of matter laid the ground work for quantum mechanics [39]. Sub-atomic particles such as electrons behave both as particle and wave according to the De Broglie equation. The behaviour of an electron or a subatomic particle propagating as waves was described in 1925 by Erwin Schrödinger in the famous Schrödinger wave equation (equation 1.1) [39].

$$\hat{H}\Psi(\vec{r}) = E\Psi(\vec{r}) \qquad (1.1)$$

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V(x, y, z) \qquad (1.2)$$

In the Schrödinger equation; $\hat{H}$ is a Hamiltonian operator, $\vec{r}$ is the three dimensional vectors in Cartesian space, Ψ is the wavefunction and E is the total energy associated with wavefunction. The Hamiltonian operator represents the total energy (which is sum of (T) Kinetic and (V) Potential energy) of the system (equation 1.2). The Schrödinger equation is an eigenvalue equation in which the Hamiltonian operator acts on an eigenfunction wavefunction Ψ and returns a scalar eigenvalue energy component (E) and the wavefunction Ψ [39]. The wavefunction defines the complete state of a system and it is possible to obtain all the observable properties (position, energy, momentum etc.) in terms of their expectation value from the wavefunction. The wavefunction has no physical significance on its own however according to the Max Born interpretation, the product of wavefunction and it complex conjugate $|\psi(\vec{r}))\psi(\vec{r}))^*|$ represents the probability of finding a particle in $(\mathrm{d}\vec{r})$ space [40]. The expectation value of energy is obtained from the Schrödinger equation by multiplying complex conjugate of wavefunction $\psi^*(\vec{r})$ to both side of equation 1.1 and integrating to all space from $-\infty$ to $+\infty$ over the Cartesian space (equation 1.3). The denominator term in the equation 1.3 is used for the normalization of the wavefunction (equation 1.4) according to the postulates of quantum mechanics. The equation 1.5 is the Dirac representation of the expectation value of energy [41].

$$\langle E \rangle = \frac{\iiint_{-\infty}^{\infty} d\vec{r}\, \Psi(\vec{r})\, \hat{H}\, \Psi^*(\vec{r})}{\iiint_{-\infty}^{\infty} d\vec{r}\, \Psi(\vec{r})\, \Psi^*(\vec{r})} \qquad (1.3)$$

$$\iiint_{-\infty}^{\infty} d\vec{r}\, \Psi\overrightarrow{(r)}\, \Psi^*\,\overrightarrow{(r)} = 1 \qquad\qquad \textbf{(1.4)}$$

$$\langle E \rangle = \langle\, \Psi\, |\hat{H}\, |\, \Psi^* \rangle / \langle\, \Psi\, |\, \Psi^* \rangle \qquad\qquad \textbf{(1.5)}$$

The Schrödinger equation can be solved 'exactly' for an atom containing only a single electron such as a hydrogen atom by taking into account the Born–Oppenheimer approximation. This approximation is based on the fact that an electron is about 1836 times lighter than the mass of a proton and can respond instantaneously to any displacement of the proton. In solving the Schrödinger equation for hydrogen like atoms the motion of the electron is independent of the motion of the proton and the position of the nucleus is fixed in the atom. The electronic Schrödinger equation is solved for electronic energy, the fixed nuclear position is added as a parametric function to the equation giving us the concept of the potential energy surface (equation 1.6) [42].

$$\hat{H}elec\,\Psi elec = \mathrm{E}elec(\mathrm{R})\Psi elec \qquad\qquad \textbf{(1.6)}$$

The wavefunction evolves in time according to the time dependent Schrödinger equation (1.7) (TDSE). The kinetic and potential energy operators depends on the spatial component $(\Psi\overrightarrow{(r)})$ and do not change with time so the Hamiltonian operator is assumed to constant with respect to time. The TDSE is solved by applying separate of variable technique in which the wavefunction is separated into a spatial and time component $(e^{-iEt/\hbar})$. In equation 1.7, $\hbar = h/(2\pi)$ is the reduced Planck constant and i is an imaginary number [43].

$$\Psi\big(\overrightarrow{(r},t\big) = \Psi\overrightarrow{(r)}.e^{-iEt/\hbar} \qquad\qquad \textbf{(1.7)}$$

1.2 QUANTUM MECHANICAL METHODS

The various approximate methods in quantum chemistry arise due to the electron-electron repulsion term in the Schrödinger equation when we solve for Helium like atoms and larger systems [44]. The multi-dimensional wave function is normally solved by separation of variables which are not dependent on each other however in case of electron-electron repulsion term both electron one (*r1*) and two (*r2*) ($e^2/4\pi\varepsilon r1.r2$) depend on each other simultaneously which makes it not possible to solve the function by separation of variables and hence there is no analytical solution to the Hamiltonian which depends on this coulombic repulsion term and therefore we have to use approximate methods [42].

1.2.1 Variational Principal

The variational principle states that the energy of any trial approximate wavefunction would be greater than or equal to the ground state energy of the exact ground state wave function. The principle is used to determine the quality of wavefunction in terms of how well it approaches the true ground state energy for the system. The linear variational method depends on a trial wavefunction (Φ) (equation 1.8) which is represented as a linear combination of basis functions ($fn(\vec{r})$) which together form a basis set. The *Cn* coefficient acts as a variational parameter and is used to minimize the energy of a trial wavefunction [39] [45]. The minimum approximate energy in the variational methods is calculated by adjustment of variational parameters to give the best wavefunction and the energy that corresponding to it (equation 1.9) [45]. The energy in variational methods is solved by using a secular determinant which represents the Hamiltonian (H) matrix, overlap (S) matrix and the coefficient (c) in column vectors. The secular equation could be represented as a matrix analogy of the Schrödinger equation when the basis functions used in the basis set are orthogonal to each other (equation 2.0). This would result in an overlap matrix (S)

being an identity matrix which provides a matrix form of the Schrödinger equation in which the Hamiltonian matrix is represented by **H**, **c** column vector represents wavefunction and **E** represents the energy eigenvalue.

$$\Phi = \sum_{n=1}^{k} Cnfn(\vec{r})$$ (1.8)

$$\langle E \rangle = \langle \Phi | \widehat{H} | \Phi \rangle / \langle \Phi | \Phi \rangle$$ (1.9)

$$Hc = ESc \quad |Hc - ESc| = 0$$ (2.0)

1.2.2 Hartree Fock Method

In the Hartree Fock (HF) theory, we have a $\widehat{H}\,eff$ (Hamiltonian effective) Hamiltonian for each electron which is represented in equation 2.1 in atomic units. The $\widehat{V}\,eff$ (effective potential) operator (equation 2.2) is described as an effective potential energy experienced by an electron (*i*) at all possible positions in space. The electron (*i*) do not feel the columbic repulsion from electron (*j*) explicitly however it is affected in the form of electron (*j*)'s spread out charge density this is known as the mean field. Now we can solve the Schrodinger equation for individual electrons as $\widehat{H}\,eff$ acts on an electron orbital (Ψ) and give us the orbital energy (ξ) and the wavefunction (equation 2.3). However the Hamiltonian of electron (*j*) depends on the wavefunction of electron (*i*) and vice versa, making this equation a pseudo Eigenvalue problem which can be solved by a self-consistent procedure (SCF) [44] [45].

Initially a guess wavefunction (orbitals) for each electron is used to calculate the mean field potential which gives the $\widehat{H}\,eff$ Hamiltonian operator, this effective Hamiltonian is then used to obtain the new wavefunction. Using the new wavefunction we can obtain new the mean field operator (updated Hamiltonian) and vice versa. In the SCF algorithm the

question is asked in terms of whether the new generated wavefunction is the same or different, if the wavefunction changes, then the whole process is iterated until the both the mean field and the wavefunction stay the same [46].

$$\widehat{Hi}(eff) = -1/2\nabla i^2 - 2/riN + \widehat{Vi}eff(\vec{ri}) \qquad (2.1)$$

$$\widehat{Vi}eff(\vec{ri}) = \int \mathfrak{R}^3 d^3\vec{rj}\,\Psi^*(\vec{rj})\Psi(\vec{rj})\frac{1}{rij} \qquad (2.2)$$

$$\widehat{Hi}(eff)\Psi(\vec{ri}) = \xi\Psi(\vec{ri}) \qquad (2.3)$$

In HF theory the electrons are described as spin orbitals (a combination of both the spatial and spin component) and the total wavefunction is represented as a product of atomic orbitals also known as the Hartree product. The limitation of the Hartree product is that it does not follow the anti-symmetric principle; this states that all electronic wavefunctions must be anti-symmetric, any two electrons (i and j) can be exchanged as in equation 2.4.

$$\Psi(i, j) = -\Psi(j, i) \qquad (2.4)$$

Electrons are fermions with spin S=±1/2 and according to the Pauli Exclusion Principle no two electons should have the same spin and occupy the same spin orbital (i.e they must have different quantum number (n, l, m_l, m_s) otherwise the wavefucntion will be zero (equation 2.5).

$$\Psi(i, j) = \Psi(i)\Psi(j) - \Psi(i)\Psi(j) = 0 \qquad (2.5)$$

For the N number of the electrons to obey the antisymmetric principal in Hartree Fock theory, they must be represented using slater determinants (equation 2.6). The use of the Slater determinent enables the wavefucntion to be anti-symmetric in respect to the exchange of two electrons. $1/\sqrt{N!}$ is a normalization conditoin, the rows of the Slater

determinant represent different atomic orbtial (spin orbital) and the colums represent different electrons.

$$\psi(1, 2, \dots N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \cdots & \psi_N(1) \\ \psi_1(2) & \psi_2(2) & \cdots & \psi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(N) & \psi_2(N) & \cdots & \psi_N(N) \end{vmatrix} \qquad (2.7)$$

1.2.3 Hartree Fock Atomic energy for N number of atoms

 The Hamiltonian operator in atomic units (au) for an arbitary atom is showen in equation 2.8. In this, the first term is the individual kinetic energy of the ith electron and its elecrostatic attraction to the postively charged nuclei. The kinetic energy of the ith electron and the electrostatic interactions between this electron and the nuclelus are both one electron additive terms. However the second term in equation 2.8 is the two electron term which is a pairwise sum of a unique pair of electron-electron replusions. The energy of an arbitary atom is then computed by the expectation value of the Hamiltonian as described in equation 2.9. The energy in equation 2.9 is a sum of the one electron energy (i.e the kinetic energy of each electron and its electrostatic attraction to the nucleus) and the two electron energy. The two electron energy is composed of the coulomb energy and the exchange energy, the coulomb energy is the electrostatic repulsion between the two unique electron pairs and the exchange energy is due to the antisymmetric principal in slater determinant. The one electron energy operator (equation 3.0) $(\widehat{h}_i)$ and its expectation value are shown in equation 3.1. The coulomb integral and the exchange integral expectation values are shown in equation 3.2 and 3.3 respectively. The coulomb integral is the repulsion of the two charge densities of the electron i and j, the exchange integral arises due to the antisymmetric principal in quantum mechanics [39, 45].

$$\hat{H} = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 - \frac{Z}{r_iN} + \sum_{i=1}^{N}\sum_{j=i+1}^{N} \frac{1}{r_{ij}} \qquad (2.8)$$

$$E = \sum_{i=1}^{N} h_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}(J_{ij} - K_{ij}) \qquad (2.9)$$

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \frac{Z}{r_iN} \qquad (3.0)$$

$$h_i = \int \Re^3 d^3\vec{r}_i\ \Psi_i^*(\vec{r}_i)\hat{h}_i\Psi_i(\vec{r}_i) \qquad (3.1)$$

$$J_{ij} = \int \Re^3 \int \Re^3 d^3\vec{r}_i\ d^3\vec{r}_j\ \Psi_i^*(\vec{r}_i)\Psi_i(\vec{r}_i)\frac{1}{r_{ij}}\Psi_j^*(\vec{r}_j)\Psi_j(\vec{r}_j) \qquad (3.2)$$

$$K_{ij} = \int \Re^3 \int \Re^3 d^3\vec{r}_i\ d^3\vec{r}_j\ \Psi_i^*(\vec{r}_i)\Psi_j(\vec{r}_i)\frac{1}{r_{ij}}\Psi_j^*(\vec{r}_j)\Psi_i(\vec{r}_j) \qquad (3.3)$$

$$\hat{F}(\vec{r}_i) = \hat{h}(\vec{r}_i) + \sum_{i=1}^{N}(\hat{J}_i(\vec{r}_i) - \hat{K}_i(\vec{r}_i)) \qquad (3.4)$$

$$\hat{F}(\vec{r}_i)\ \Psi_i(\vec{r}_i) = \xi_i\Psi_i(\vec{r}_i) \qquad (3.5)$$

$$\xi_i = \int \Re^3 d^3\vec{r}_i\ \Psi_i^*(\vec{r}_i)\hat{F}\Psi_i(\vec{r}_i) \qquad (3.6)$$

In HF theory when electrons are in the same orbitals they don't interact with themselves because the self-coulomb and exchange integrals cancel out. The coulomb operator gives us the effective mean field potential of electron i from the charge density of electron j whilst the exchange operator exchanges the location of the two electrons i and j [39, 45, 47]. The two electron operators depend on the position of the electron i and j simultaneously. The Fock operator in HF theory for electron i is represented in equation 3.4, this operator consists of a one electron operator ($\hat{h}_i$) and over all sum of the electrons of the coulomb operator minus the exchange operator for electron i. This sum also represents the effective mean field operator ($\hat{V}\ eff$) for electron i in HF theory. The Fock operator acts on the individual atomic orbitals and gives the atomic orbital energy as these atomic orbitals are eigenfucntions of the fock operator (equation 3.5), every different

atomic orbital has a slightly different Fock operator and thus gives slightly different orbital energies. The orbital energy can be computed by taking the expectation value of the Fock operator (Figure 3.6). The application of the linear variational method to the HF equation using a basis set for atomic orbitals results in the production of the Hartree Fock Roothaan equations. Each atomic orbital is represented as a linear combination of basis functions, there are two types of basis function generally used in quantum chemistry; (i) Slater function and (ii) Gaussian function. The computation involving two electron integrals using slater function are very difficult to compute even though slater functions are a good approximations to the hydrogen atom atomic orbitals, instead Gaussian functions are used routinely in QM calculations due to their relative ease of implementation in the computational code [39, 45, 47].

A Major concern in the application of Hartree-Fock theory is its poor treatment of electron correlation, the mean field approximation and use of only a single slater determinant for the representation of spin orbitals. Post Hartree-Fock methods such as Møller-Plesset Perturbation theory [48] [49], Coupled Clusters theory [50], Configuration Interactions [51] and Density Functional Theory [52] [53] attempt to improve upon HF methods.

1.3 DENSITY FUNCTIONAL THEORY

Density Functional Theory (DFT) [52] [53] utilizes the electron density ($\rho(r)$) instead of a many electron wavefunction to obtain the ground state energy of a system. DFT is based upon the Hohenberg and Kohn theorem [52] which states that the electron density can determine a distinct Hamiltonian for a system and hence all the observable properties of the system. The electron density is an experimentally measurable quantity which is a function of three variables (r,θ,ϕ) in the spherical coordinate system [54]. The calculation of electron density in DFT reduces the computational cost and scales similarity to HF with

respect to system size, making its application in the field of computational chemistry and condensed matter physics a very attractive proposal. The Functionals of electron density provide the electronic energy in DFT based upon equation 3.7 [41].

$$E_{elc} = E^T + E^V + E^J + E^{XC} \qquad \qquad \textbf{(3.7)}$$

In equation 3.7, E^T is the kinetic energy term of the electrons, E^V is the electrostatic attraction between nuclei and electrons and the nuclear-nuclear repulsion, E^J is a coulomb electron-electron repulsion term and E^{XC} is the exchange-correlation term. The E^{XC} term accounts for both the exchange energy arising from the anti-symmetry principal of quantum mechanics and dynamic electron correlation [54]. The exchange-correlation terms requires adequate understanding before its implementation in DFT due to the unavailability of an exact functional describing E^{XC} [55]. In order to describe E^{XC} in DFT various approximate functional forms have been developed such as Local Density Approximation (LDA) which rely on the electron density, Generalised Gradient Approximation (GGA) and the meta-generalized Gradient approximations (meta-GGA) which take into account both the gradient and electron density [46] [55].

In general, hybrid functionals relies on obtaining the exchange correlation energy form HF calculations and the dynamic electron correlation energy from the DFT component. The use of hybrid functionals significantly improves the approximation of E^{XC} in DFT and one functional which has found central role in studying electronic structure in context of computational chemistry is B3LYP (Becke, 3-parameter, Lee-Yang-Parr) [55].

1.4 MOLECULAR MECHANICS

Molecular mechanics (MM) methods are based on the application of the concepts of classical physics to predict the energy of molecule as a function of its atomic nuclear positions at different conformation in Cartesian space. The Newtonian nature of the treatment of atoms in the molecule prohibits the explicit representation of electrons in molecular mechanics, instead the nuclei and electrons are combined into atom like particle [28]. This approximation makes molecular mechanics based methods unfit for studying the chemical reaction process, none the less it plays a significant role in understanding the structural properties of molecules such as their structure, dynamics and conformational flexibility. The molecular mechanics models consider atoms as hard spheres and bonds as springs which obey classical potential functions which describe bond stretching and bending. The molecular mechanics force field energy function usually consists of stretching, bending, torsion and non-bonded interaction energy terms which are additive in nature [28]. The bonded and non-bonded interactions in the force field relay on parameters obtained from experimental and quantum mechanical studies. A typical example of a force field used in describing a biomolecular system such as protein is show in equation 3.8 [28]. The common force field uses routinely in molecular modelling are AMBER [56] [6], GROMOS [57], CHARMM [58], OPLS-AA [59]. There are also specifically designed force field which deal with parametrization of ligands such as GAFF [60] and CGenFF [61].

$$V(\vec{R}) = \sum_{bond} kb(r - r_o)^2 + \sum_{angle} k\theta(\theta - \theta_o)^2 + \sum_{torsions} A[1 + cos(n\tau - \phi)] +$$

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r} \right) \qquad (3.8)$$

The MM potential energy $V(\overrightarrow{R})$ is a function of the positions vector (R) of N atoms and consists of bonded and non-bonded interactions as shown in blue and red colour respectively in equation 3.8. The first term represents the bond stretching energy between a pair of atoms and is based on Hooke's law and is modelled using a harmonic potential. It is the use of simple harmonic approximation in bonded terms which prohibit the bond breaking in molecular mechanics. The application of a Morse potential includes the effects of bond dissociation however these potential are computationally very expensive and difficult to parameterize. The *kb* parameter controls the stiffness of the bond spring, while *ro* defines its equilibrium length. The bond bending energy between the three atoms is modelled using a harmonic potential. The *Kθ* parameter controls the stiffness of the angle spring and θ_o defines its equilibrium angle. Bond torsion or dihedral angle rotation occurs between four atoms. The torsion energy is usually modelled with a periodic function in a force fields. The "*A*" parameter controls the amplitude of the curve, the *n* parameter controls its periodicity, and ϕ shifts the entire curve along the rotation angle axis τ.

The non-bonded interactions can be separated into van der Waals interactions and electrostatic interactions, non-bonded interactions occur between all atom pairs in the molecular mechanics force field. The van der Waals arise between two non-bonded atoms due to near instantaneous induced multipole-multipole interactions caused by the movement of electrons within electron cloud, they occur at short range due to London dispersion forces and rapidly die off as the interacting atoms move apart. Repulsion occurs when the distance between interacting atoms becomes less than the sum of their contact radii, the effect is generally modelled using a Lennard-Jones potential. The *εij* parameter in van der Waals represents the strength of the non-bonded interactions and *rij* is the interatomic distance between atom i and j. The second component of non-bonded interactions are electrostatic interactions which occurs between pair of non-bonded atoms (*qi* and *qj*) which are assigned partial charges based on force field parametrization. The

electrostatic contributions are modelled using a Coulombic potential. Computation of electrostatic interaction in molecular mechanics force fields can be expensive due to the relative strength in contrast to van der Waals interactions which decay very quickly. The long range electrostatic interactions are modelled with approximate methods using cutoffs schemes to reduce the complexity of the calculations.

1.5 MOLECULAR DYNAMICS SIMULATIONS

The basic idea behind molecular dynamic simulations is to solve Newton second law of motion where F is the force exerted on the particle; **m** is the mass of the particle and **a** acceleration. The force is expressed as the negative gradient (Laplacian) of the potential energy function (equation 3.9) which is obtained from the molecular mechanics force field.

$$\vec{F} = -\nabla V(\vec{r}) \tag{3.9}$$

The potential energy of the system is a function of the atomic positions ($\vec{r}$) of all the atoms in the system. The Newton second law of motion is a second order ordinary differential equation. By using the potential energy function the total force acting on i^{th} particle in the system can be obtained by vector sum of its interactions with other particles. Once the force is obtained it is possible to determine the acceleration and displacement of the particles. The potential energy function depends on N number of atoms interacting making it a many body problem whose analytical solutions are not possible [28]. Therefore in order to obtain trajectory we have to rely on numerical solutions. Some of numerical methods used for integrating the equation of are Verlet [62], Leap-frog [63] and Beeman's [64] algorithms. The main assumption in the integration algorithm is that the position, velocity and acceleration can be obtained by using Taylor series expansion. The time steps used in molecular dynamic simulations are constrained by the highest frequency motions in the system which typically include chemical bond vibration and are around 1-2 femtosecond [28].

The average value of a property of interest of a biological molecule can be calculated in Molecular dynamics simulations by following the Ergodic hypothesis which states that the time averages equals the ensemble average. This means that instead of integrating the contribution of that property (X) over the entire spatial configurations (equation 3.6), we can compute the average property of (X) over a large amount of simulated time. The denominator of equation 4.0 represents the partition function in statistical mechanics and the probability of the microstate i is proportional to the Boltzmann factor ($Pi \propto e^{-Ei}/_{kbT}$) [65].

$$\langle X \rangle = \frac{\int dr^N \, X(r^N)(e^{-E(rN)}/_{kbT})}{\int dr^N \, (e^{-E(rN)}/_{kbT}} \qquad \textbf{(4.0)}$$

1.6 QUANTUM MECHANICS / MOLECULAR MECHANICS

The QM/MM technique is considered a valuable tool in the field of computational chemistry for studying reaction mechanisms in enzymes [66]. The first application of QM/MM technique to study reaction mechanism in enzymes was performed by Warshel and Levitt in 1976 [8]. The use of QM/MM provides a valuable insight into the electronic structure of the active site and into the long range effect of protein conformation on the electronic structure of the catalytic site [9] [66]. The QM/MM technique is routinely used to correct the anomalies of cluster models in describing the overall reaction mechanism of enzymes, one such example is Isopenicillin N Synthases (IPNS) [67]. The cluster model in IPNS described the oxygen binding to non heme iron site as an endoergic process with energy cost of >10 kcal/mol. The QM/MM calculation of IPNS corrected the results by providing stabilization through van der Walls interactions from a neighboring active site residue which maintained the geometry of the iron coordination sphere, as opposed to cluster models which showed major displacement [67].

In QM/MM calculations, the total energy of the system is described by equation 4.1, where E_{QM} is the energy of QM region, E_{MM} is the energy of the MM region and $E_{QM/MM}$ is the energy of the interactions between the QM and MM region [66].

$$E_{total\,(QM/MM)} = E_{QM} + E_{MM} + E_{QM/MM} \qquad\qquad \textbf{(4.1)}$$

The QM and MM region in the QM/MM approach interact strongly with each other and requires the use of coupling terms to be taken into account before computing the total energy of the QM/MM system [9, 66]. The coupling terms describes the interactions (van der Waals terms and electrostatic terms) between QM and MM region and can be categorized into mechanical embedding and electrostatic or electronic embedding schemes. In the mechanical embedding scheme, the electrostatic interactions between the QM and MM region are treated at MM level, which prevents the polarization of the QM wavefunction by MM environment. In the electronic embedding scheme, the electrostatic interactions between the QM and MM region are treated at QM level. The electronic structure of the inner QM region is allowed to adapt to changes in the charge density of MM environment, therefore we incorporate MM point charges as one-electron terms in the QM wavefunction by performing QM calculation in the presence of an MM charge model [68] [66] [69] [70].

1.7 AIM OF THE PROJECT

Aim: (i) To provide understanding of how the conformational flexibility influences structure-function relationships and the reaction mechanism of TPST-2, NirE, MMP-1, FTO, AlkB, PHF8 and KIAA1718 enzymes; (ii) To understand the effect of the protein environment on the atomistic and electronic structural features in the chemically reactive parts of the protein.

Methodology: In order to obtain the crucial knowledge of the conformational dynamics and flexibility in enzymes, classical molecular dynamics simulations were performed using Gromacs [3] [4] [5] and Amber [6] [7] packages. To understand the electronic structure, we implemented methods such as Density Functional Theory [52] [53] (DFT) using the Gaussian09 [12] code to perform cluster Quantum Mechanics (QM) calculations on the active site of the enzymes for studying bond making and breaking events. The effect of the protein environment on the electronic structure of the active site were studied using Quantum Mechanics and Molecular Mechanics (QM/MM) [8] using ONIOM [9] [10, 11] implemented in Gaussion09 [12] [13].

CHAPTER TWO TYROSYLPROTEIN SULFOTRANSFERASE-2 AND MUTANTS

2.1 INTRODUCTION

Tyrosine sulfated proteins are the product of post translational modification (PTM) , where a sulfate group is covalently added to the hydroxyl group of tyrosine residues of the polypeptide chain [14]. The trans-Golgi networks in the cell harbor Tyrosylprotein Sulfotransferase (TPST) which is an enzyme responsible for the tyrosine sulfation reaction [14, 71]. The tyrosine sulfation is very well characterized in eukaryotes and it has been predicted that up to 1 % of tyrosine residues in the eukaryotic proteome have the potential to be sulfated [72]. Tyrosine sulfation plays a crucial role in protein-protein interactions in the extracellular environment and recently became increasingly important for biomedicine and as target for drug design [73] [74] [75] [76] [46]. For example , the interaction of HIV's gp120 protein to CCR5 in order to get entry to CD4+ T-lymphocytes requires sulfated tyrosine residues [77]. The Enterovirus71 virus responsible for neurological diseases in children depends on the tyrosine sulfation of PSGL-1 on leukocytes in order to gain entry to cell and cause infection [78]. In fact, approximately sixty immune system proteins have been shown to contain tyrosine sulfated residues [79] [80]. TPSTs catalyze the transfer of negatively charged sulfate group from the universal sulfate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to the hydroxyl group of a tyrosine residue of polypeptide to form a tyrosine O4-sulfate ester and Adenosine 3',5'-diphosphate[14]. In humans, two isoforms TPTST-1, and TPST-2, are encoded by the TPST gene have been identified [81]. The molecular weight of TPST-1 (370 residues) and TPST-2 (377 residues) isoforms are 42.2 and 41.9 kDa respectively [73]. TPST has a type II transmembrane topology i.e. a short N terminal cytoplasmic domain, 17 residue transmembrane domain, and a luminal catalytic domain [82]. The enzyme has two N-glycosylation sites, that are

four cysteine residues on the luminal oriented side of enzyme.[76] The first crystal structure of core domain of human TPST-2 (designated TPST2ΔC18 encompassing from G43 to L359), complex with PAP and the substrate peptide C4P5Y3 was solved at a resolution of 1.9 Å [83] (**Figure 2.1A**). C4PY3 contains only one tyrosine sulfate acceptor site and consists of six acidic residues, thus giving the substrate an overall charge of minus six [83]. The catalytic domain of TPTS-2 comprises a single α/β motif with a five-stranded parallel β-sheet, flanked on both sides by α helices [83]. The TPST-2 exists as a homodimer and the two subunits of the dimer are designated as protomer A and B [83]. The crystal structure of human TPST-2 reveals the important atomistic details of the enzyme and the ligand binding and possible mechanism, but shows no information about how conformational flexibility and dynamics influence protein structure, its structural determinants, key interactions with substrate and cofactor and the effects of mutations on the enzyme structure and ligand binding. Proteins are large flexible molecules and conformational dynamics is their fundamental property which correlates proteins structure and functions [1] [84]. In order to understand how the flexibility influences structure-function relationships of TPST-2, we performed 100 ns atomistic (AT) molecular dynamics (MD) simulations on the wild-type full complex TPST-2, containing the apoenzyme, the cofactor and the substrate (WT FC), and its mutant forms. The mutant forms of TPST-2 were studied experimentally [83] and contains residues involved in active site (R78, E99, K158, S285) substrate binding (P77, E99, T198, R101, R105, R122) and cofactor binding (R78, S285).
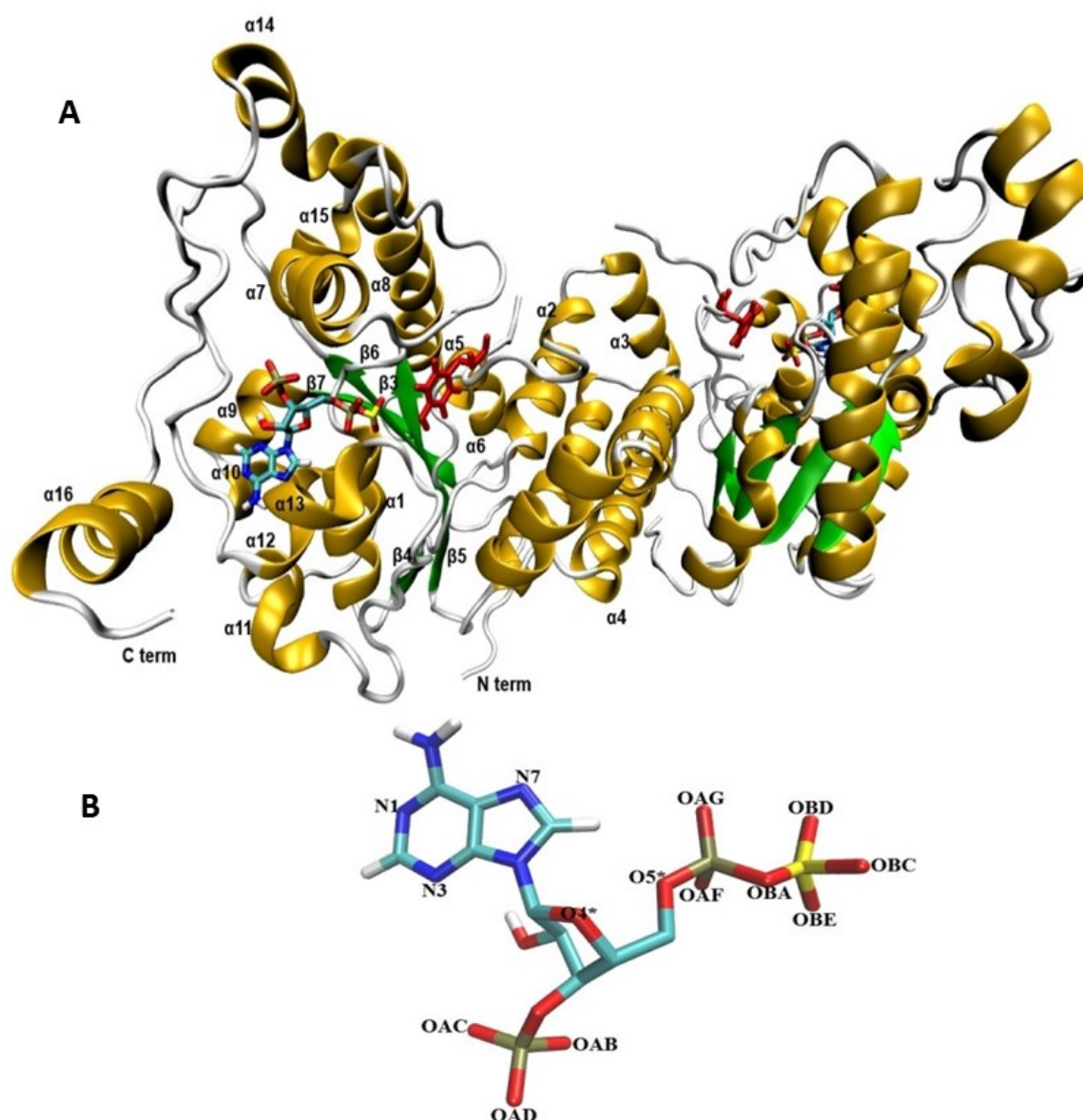
Figure 2.1: The 3D structure of human TPTS-2(PDB code: 3AP1) 15 enzyme using Visual Molecular Dynamics (VMD). (A) The homodimer TPST-2 is represented in new cartoon model with beta sheets colored in green and alpha helices in gold color. The cofactor (PAPS) and substrate (acceptor tyrosine; color red) are represented using licorice representation. (B) The cofactor PAPS is represented in licorice representation and labeled according to the nomenclature as it appeared in the md simulation. The OBA represents the bridge oxygen between the phosphate and sulfate group of the 5'-PBS region of cofactor.

2.2 METHODS

2.2.1 Initial structure preparation

The coordinates of the wild type TPST-2 were obtained from Protein Data Bank (PDB) [85] (PDB ID 3AP1) [83]. The SwissPDBViewer was used for adding missing atoms and selecting one from the alternative side chain orientations [86]. The sulfate group was added to the cofactor molecule PAP using GaussView 5.0 [87]. The resulting PAPS molecule (**Figure 2.1 B**) was used as a cofactor for the molecular dynamics simulation of TPST-2. There were eight single amino acid mutants used in this study (**Table S2.1**). These mutants were prepared by using What IF web server [88]. The single amino acid mutations were prepared in both protomer A and B of the TPST-2 enzyme. The force field parameters for cofactor were calculated using PRODRG online server [89] and were fitted to quantum chemical calculations and other published parameters of PAPS [90] .

2.2.2 Molecular dynamics simulations

In order to explore the dynamic properties of homodimer TPST-2, we performed extensive sets of Molecular Dynamics simulations for 100 ns using Gromacs 4.5.5 package [3] [4] [5] with GROMOS96 43a1 [57] force field. The hydrogen atoms were added to the protein molecule by using pdb2gmx utility in Gromacs. The protonation states of histidine residues in the protein molecule were assigned based upon the optimal hydrogen bonding conformation performed in Gromacs using pdb2gmx. In order to remove the bad contact or clashes in the protein structure which might have occurred during the crystallization, in vacuum energy minimization was performed first by using the steepest descent [91] and then by using conjugate gradient algorithm [92]. The editconf command was used to define the dimension of the cubical box and the protein molecule was placed in the box. The periodic boundary conditions were then applied to treat all the parts of the system equally both at its interior and edges. The box size was set to ensure a distance of at least 1.0 nm between the protein and the box boundaries. The energy minimized protein structure was

then solvated by using Single Point Charge (SPC) [93] water model in the cubical simulation box by using Genbox command. The system was neutralized by adding the Na^+ to various mutants and the wild-type setups (**Table S2.1**). In order to relax the solvent molecules and remove constrains from the entire system, the energy minimization of the whole system was performed using first the steepest descent followed by the conjugate gradient algorithm until the maximum force on the atoms was smaller than 100 kJ $mol^{-1} nm^{-1}$. The energy minimized structure was then subjected to position restrain dynamics for 50 ps. The simulation was performed in NVT ensemble (constant Number (N) of particles, Volume (V), and Temperature (T)) [94] at constant temperature of 300K with time step of 0.002 ps. The productive 100ns MD was carried out using NPT (constant number of particles (N), system pressure (P) and temperature (T)) ensemble at constant temperature of 300K and the initial velocities for MD simulation were drawn from Maxwell velocity distribution at 300K. The Berendsen temperature coupling and Parrinello-Rahman pressure coupling were used to keep the system at 300 K, time constant (τ_T) of 0.1 ps and 1 bar pressure, time constant (τ_P) of 0.5 ps during the simulation procedure. The MD was performed with an integration time step of 0.002 ps. The Particle Mesh Ewald (PME) [95] method was used for electrostatic interactions with Coulomb cut off of 1.0 nm, Fourier spacing of 0.135 nm tolerance of 1e-5 and an interpolation order of 4. The Lennard Jones potential was employed for the treatment of van der Waals interaction with cut off distance set to 1.4 nm (rvdw) applying switching function. The LINCS algorithm [96] was utilized to keep all the covalent bonds involving hydrogen atom rigid. The coordinates were saved after every 20 ps from multiple MD trajectories.

2.2.3 Analysis of Molecular dynamics simulations

The analyses of the trajectories obtained from the simulations were performed using tools from the Gromacs software package. The Root Mean Square Deviation (RMSD) of Cα atoms of the protein with respect to minimized crystal structure, Root Mean Square

Fluctuations (RMSF), Radius of gyration (Rg), Electrostatic interactions, hydrogen bonds, and cluster analysis were also performed using Gromacs. The visualization of MD trajectories and the structures were performed using VMD [97]software. The Free energy landscape analysis was performed by using g_sham in Gromacs. The Graph of Free energy landscape was produced by in-house scripts. The Bio3D package [98] in R was used to produce domain cross correlation analysis. Dynamic Cross correlation Analysis was performed on the wild type TPST-2 in order to understand extends of the correlated motion. The cross correlation between the ith and jth atoms are represented by Cij matrix which range from -1 to +1. The Cα atoms of the protein, substrate and all atoms of the cofactor were used to create cross correlation Cij matrix. The positive value represents the correlated motion and the negative values represent the anticorrelated motion.

2.3 RESULTS AND DISCUSSION

2.3.1 Overall Stability and Flexibility of TPST-2 Structures

The Root Mean Square Deviation (RMSDs) of WT FC and its mutants are represented in **Figure 2.2A**. The WT FC shows that system has equilibrated after 15ns. The RMSD profiles of substrate binding mutants (R101A, R105A, R122A, and T198A) equilibrated roughly after 10 ns (**Figure 2.2A**). The active site mutants (R78A, E99A, K158A and S285A) showed equilibration after 15 ns and R78A and E99A after 20 ns. The RMSD profile of the S285A and T198A mutants out of all the mutants showed the greatest structural deviation from the WT FC trajectory. The average RMSD value of all Cα atoms of WT FC is ~ 2.7 Å, whereas in mutants its ranges from 3.7 Å in S285A to 2.3 Å in R122A mutant (**Table S2.2**). The RMSDs of Cα atoms from the secondary structural elements (α-helices, β sheets) of WT FC shows that beta sheets have lowest RMSD value of 1.7 Å along the 100 ns trajectory compared to alpha helix 2.2 Å and loops 3.2 Å (**Figure S2.1**). Similar trends were seen in the mutants as well. The RMSDs of secondary structures for the mutants ranges from 3.8 Å (S285A) to 2.6 Å (R112A) for alpha helix, 2.6 Å

(T198A) to 1.8 Å (R122A) for beta sheets and 4.7 Å (S285A) to 2.8 Å (R122A) for loops (**Table S2.2**).The RMSF profiles of WT FC and the mutants are shown in **Figure 2.2B.** The detailed analysis of the Root Mean Square Fluctuations (RMSF) plot of the WT FC is shown in (**Figure S2.2**). The basal level fluctuation of the WT FC was considered to be 1 Å, 43 % of RMSFs in WT FC are below 1 Å and 57 % of residue's RMSF are above 1 Å. All the mutants have shown significant increase in the number of residues with RMSF > 1 Å apart from E99A mutant which shows slightly decrease in this percentage (**Table S2.2**). The most significant effect on the RMSFs was seen with K158A mutant. The cofactor binding region of the protein (5'PBS and 3'PBS), substrate binding region βe and many secondary core structural elements were increased in the K158A mutant relative to WT FC (**Table S2.3**). The S285A mutant also showed an increase in the 5'PBS cofactor binding site. The S285A mutant showed a significant increase in α11, α12 alpha helix and the loop connecting α12 and α13. This loop contains the S285 residue in the WT FC. The mutant E99A also showed an increase in the α2 helix. In the substrate binding residues, the most significant increase of RMSFs was seen in R101A, R105A and T198A mutants in contrast to WT FC. The results indicate that mutations affect the flexibility not only the nearby residues around the mutation site, but also have more complex structural effect on different parts of the enzyme molecule.

2.3.2 Extending the Time Scale and Multiple Simulations

In the scientific community there is missing consensus whether is better to run one longer simulation which allows exploring conformational effects at longer time scales or instead to run several shorter runs in order to clean the statistic noise. In order to explore this issue we did additional simulations. We extended therefore the trajectory of the Wild type TPST-2 to 200 ns (**Figure S2.3**). The RMSD profile remain stable and to there is not considerable difference between the averaged values to 100ns (2.7Å) and from 100 ns to 200ns (2.9Å). The mean values of three important geometric determinants discussed later

35

in the manuscript (the distance between carboxyl group of E99 and the hydroxyl group of the acceptor Y1006 from the substrate, the distance between Y1006 and the sulfur atom of PAPS and the angle of bending of substrate) were compared. In all cases the average values remain very close in the trajectory up to 100ns and from 100ns to 200ns (**Table S2.14**). Therefore for TPST2 we would be able to use 100ns trajectory for comparative atomistic analysis of the interactions in TPST2. In order to evaluate the effect of the statistical noise, two additional simulations for the WTFC of TPST-2 were run for 100ns with different initial velocities; trajectory analysis was done and compared to original simulation (**Figure S2.4** and **Table S2.14** in SI). The RMSDs and the three geometric determinants mentioned above show very similar values and trends in the three simulations, e.g., the standard deviations between the mean RMSD values are within 0.22Å. In order additionally to account for the statistical errors in mutants simulations, we run second simulation for 100ns (**Figures S2.5**). The averaged values of the RMSDs and three geometric determinants, specific for the E99A mutant (discussed in details later in the manuscript) indicate about good convergence between the two runs (**Table S2.15**) and the plausibility to use single simulations further on in this comparative MD study. The Radius of gyration from the two additional runs of WTFC showed very similar trends and points towards the equilibration of the WTFC simulation by indicating the compactness of the protein as it evolves with time (**Figure S2.6**). The free energy landscape (FEL) analysis [38] was performed on the trajectory of WTFC in order to assess the conformational space sampled by the protein during 100 ns simulation and to look for local minima which represent the metastable conformational states of the system. The Gibbs free energy was computed as a function of (i) RMSD, Rg; (ii) Rg, number of protein-protein hydrogen bonds (**Figure S2.7**). The free energy plot of Rg vs RMSD shows the presence of local minima which represents the stable state in WTFC. The total energy and potential energy of the WTFC was also stable through the simulation and (**Figure S2.8**). These results

indicating towards the structural stability and convergence of simulation which could be used to carry out further structural analysis.
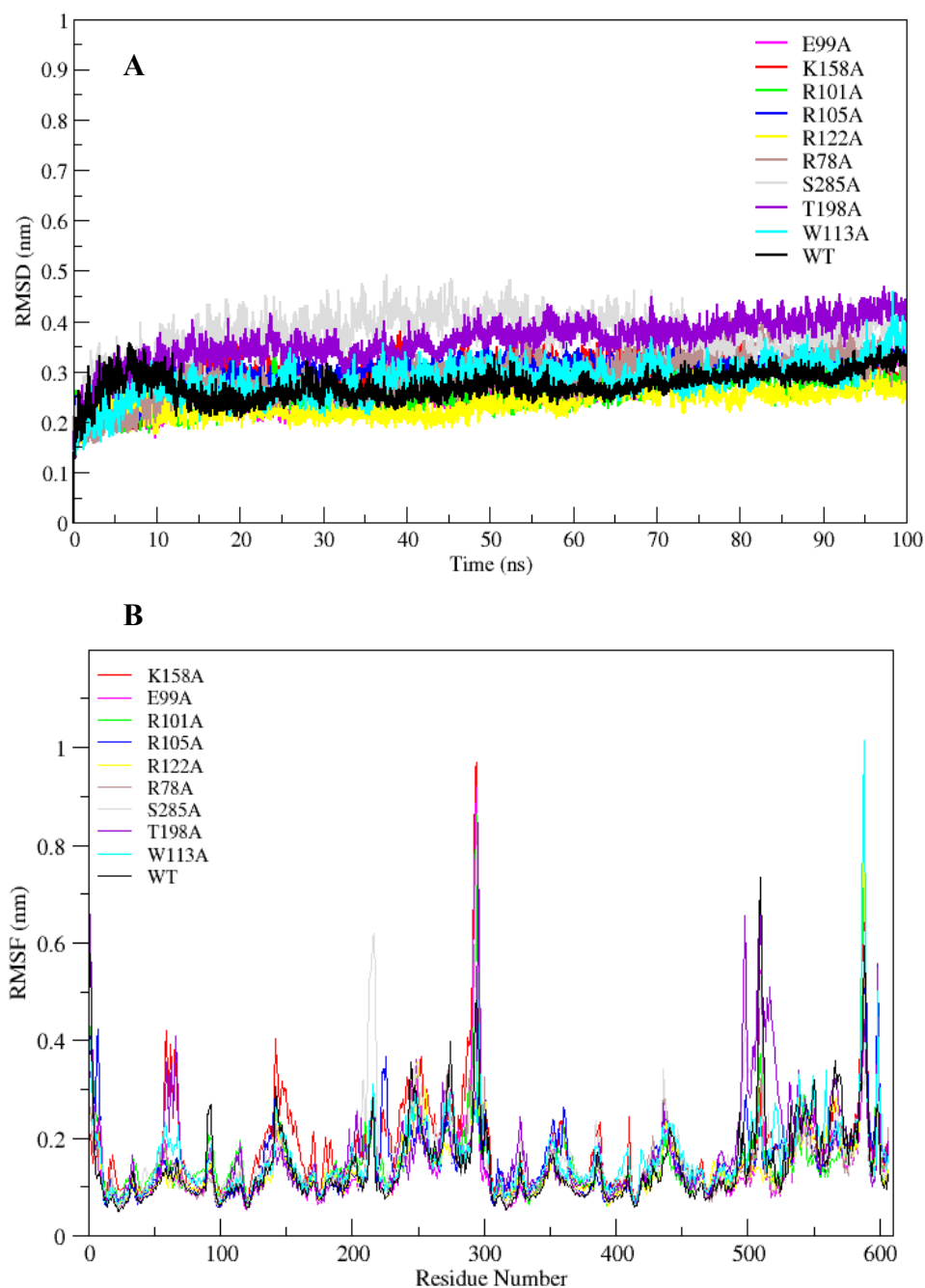
Figure 2.2 (A) The RMSD trajectory of all C α atoms in of wild type TPST-2 (WT FC) and mutants as function of simulation time in nanoseconds (ns), (B) RMSF of all C α atoms of residues of WT FC and mutants for 100ns trajectory.

2.3.3 Conformational Effects on the Cofactor Binding Site

The 5'-Phosphosulfate binding Motif (5'PBS) (residues 75-83) (**Figure 2.3B**) is located between the β3 and α1 of WT FC. The crystal structure suggests extensive interactions of the 5'-PBS with the PAPS cofactor. The average value and time evolution of RMSDs and RMSFs of the 5'PBS region show this region to be stable (**Figure S2.9**). R78, which is proposed to act as a catalytic residue in the crystal structure [83] stabilizes the 5'-phosphate group of PAPS by electrostatic interactions. The hydrogen bonding profile of the 5'PBS region of the WT FC is shown in **Table S2.4.** The sulfate group is not present in the crystal structure and was added to PAP which allows analysis of the specific interactions which stabilize and orient it for reaction with the acceptor tyrosine of the substrate. Residues S79, G80, T81 and T82 are mainly involved in hydrogen bonds with sulfate group of PAPS. The NH group of L83 makes hydrogen bonds with the OAG oxygen of the phosphate of 5'PBS region. The 3'-Phosphate binding Motif (3'PBS) (residues 180-195) (**Figure 2.3A**) spans β6 to α7. This region is next to the βe region (**Figure 2.3F**) which is involved in the substrate binding. The RMSDs and RMSFs (**Figure S2.10**) indicate that the 3'-PBS motif is also stable during the simulation. The side chain of R183 makes electrostatic interactions with the 3'phosphate group of the PAPS (**Figure S2.11**) with an average distance 3.7 Å. The side chains of S191 also make hydrogen bonds with the oxygen of the 3'-phosphate group which is stable in 67% of the 100ns trajectory. The side chain of R195 make hydrogen bonds in the crystal structure, but it is not preserved in MD studies due to the high flexibility of the R195 side chain. The K300 side chain in the crystal structure makes close contact with 3'-phosphate group of PAPS, but in the MD simulations the average distance is larger (6.03 Å). The backbone of N294 gets closer to the N7 of the adenine ring of the cofactor during the simulation and interacts by hydrogen bond

during 89% of the trajectory, which is not present in the crystal structure where the distance is 6 Å. In contrast to the crystal structure, the side chain of V293 makes hydrophobic interaction with C8 atom of the adenine ring (4.2 Å averaged distance in the simulation against 7.8 Å in the crystal structure). The Y283 in crystal structure stabilizes the nitrogen atom of the adenine ring, however in the MD trajectory the distance between these two residues is 9.6 Å. The cofactor binding motifs (5'-PBS, 3'-PBS) are highly conserved among several families of sulfotransferases such as estrogen sulfotransferase etc. [99] The simulations reveal that the sulfate group (which is not presented in the crystal structure) is stabilized by hydrogen bonds with S79, G80, T81 and T82, the vast majority of interactions in the crystal structure are stable and also that there are two new hydrophobic interactions stabilizing the adenine ring of the cofactor.

2.3.4 Substrate Binding

In the crystal structure the phenol hydroxyl group of the acceptor tyrosine Y1006 of the substrate is recognized by E99 and its aromatic ring by hydrophobic interactions with P77, whilst the backbone of Y1006 is recognized by the backbone of the short β-strand, βe (Thr198-Ala200) [83]. The bending of the substrate in the crystal structure (the L-shaped conformation) is confirmed by measurement of the angle between the Cα of Y1006, D1005 and E1004 which is 86° in the crystal structure. The same angle as a function of time shows an average value 95.9° in WT FC which indicates that there is a more open conformation of the peptide substrate and more open binding site than in the crystal structure [83]. The substrate angle for the mutant forms ranges from 90° to 103.5° (**Table S2.6**). The most significant effect on the angle was observed in T198, E99 and S285 mutants. The normalized angular distributions of angle for the mutants and the WT FC are plotted in **Figure 2.4**. The average angle for E99A mutant is 103.5° which influences the

orientation of acceptor tyrosine and can explain the experimental finding that the E99A mutant has no enzyme activity [83].The backbones of T198 and A200 interact with the backbone of the substrate tyrosine (**Figure S2.12**, **Table S2.7** and **Figure 2.3F**). The side chain of I199 makes hydrophobic interactions with the side chain of the Y1006 acceptor tyrosine during the simulation with an averaged distance of 4.1 Å. The crystal structure suggests that the P77 makes hydrophobic interactions with the aromatic ring of Y1006; however in the simulation this interaction is stable in only 41% of the trajectory. Importantly, two other residues (P160 and P161), not mentioned in the crystal structure are also involved in hydrophobic interactions with the substrate tyrosine. The P160 side chain makes hydrophobic contact with the Y1006 aromatic ring during 66.4 % of the trajectory and the F161 ring during 24 % of the simulation time. These results indicate that the aromatic ring is properly stabilized and oriented for catalysis by hydrophobic cluster containing P77, P160, and P161 residues, rather than single interaction as in the crystal structure (**Figure 2.3E** and **S2.13**). The hydroxyl group of Y1006 and E99 interact by stable hydrogen bonds as in the crystal structure (**Figure S2.14**, **Table S2.7**). The analysis of the two most populated clusters of WT FC and the E99A mutant (**Figure S2.15**, **Table S2.13**) showed that the distance between the A99 and Y1006 hydroxyl group to be 10.6 Å as compared to 2.8 Å in WT FC which explains the lack of activity of E99A. The crystal structure suggests that the backbones of D1005 and E1004 are recognized by the side chains of T198 and R101. The MD studies of WT FC show that T198 side chain makes stable hydrogen bonds with the backbone NH group of D1005 of the substrate, whilst in the T198A mutant this interaction is weaker with an average distance of 5.2 Å (**Figure S2.16**, **Table S2.9**). These results explain why T198A shows reduced enzyme activity in vitro [83] and confirms the role of T198 in the proper orientation of the substrate. The MD trajectory analysis reveals there is electrostatic interaction between the side chains of R122 and D1005 during the simulations, however, it is stronger up to 40ns (4.6Å up to 40 ns and

~ 5.8 Å after 40 ns) (**Figure S2.17**). Importantly, the simulations reveal that the side chain of R122 interacts electrostatically with the side chain of the catalytic base E99 and both interactions are lost in the trajectory of the R122A mutant (**Figure S2.18**). The side chain of R101 makes hydrogen bonding with the backbone of E1004 and D1005 from substrate for 70 ns and 90 ns respectively (**Figure S2.19**, **Table S2.7**) and the backbone of R101 interacts with side chain of D159 for entire 100 ns trajectory which does not exist in the crystal structure. D159 is located next to the catalytically important residue K158 and this interaction might be important for the proper orientation of K158 for catalysis. During the simulation the side chains of R101 and aromatic ring of F1003 interact by cation-π interactions (with average distance 4.3 Å) which most stable during the first 70 ns and were not presented in the crystal structure. The simulations of R101A mutant show the complete loss of interactions mentioned above and are in agreement with the experimental studies showing that the monomeric R101A mutant [83] exhibits nearly complete loss of enzyme activity (**Figure S2.20**, **Table S2.10**). There is stabilizing hydrogen bonding between the R105 of the enzyme and E1007 of the substrate in the crystal structure. The MD studies confirm that the both residues interact closely with averaged distance of 3.5 Å (**Figure S2.21**). In the R105A mutant the average distance between A105 and E1007 was increased to 14Å (**Figure S2.22**) which is in agreement with the finding that this mutant shows reduction in enzyme activity [83]. The MD results are in agreement with the experimental data about the effects of the mutations of key substrate binding residues on the enzyme activity, and suggest a more open conformation of the substrate than indicated in the crystal structure and indicate the appearance of a small hydrophobic cluster which stabilizes the aromatic ring of the acceptor Y1006.
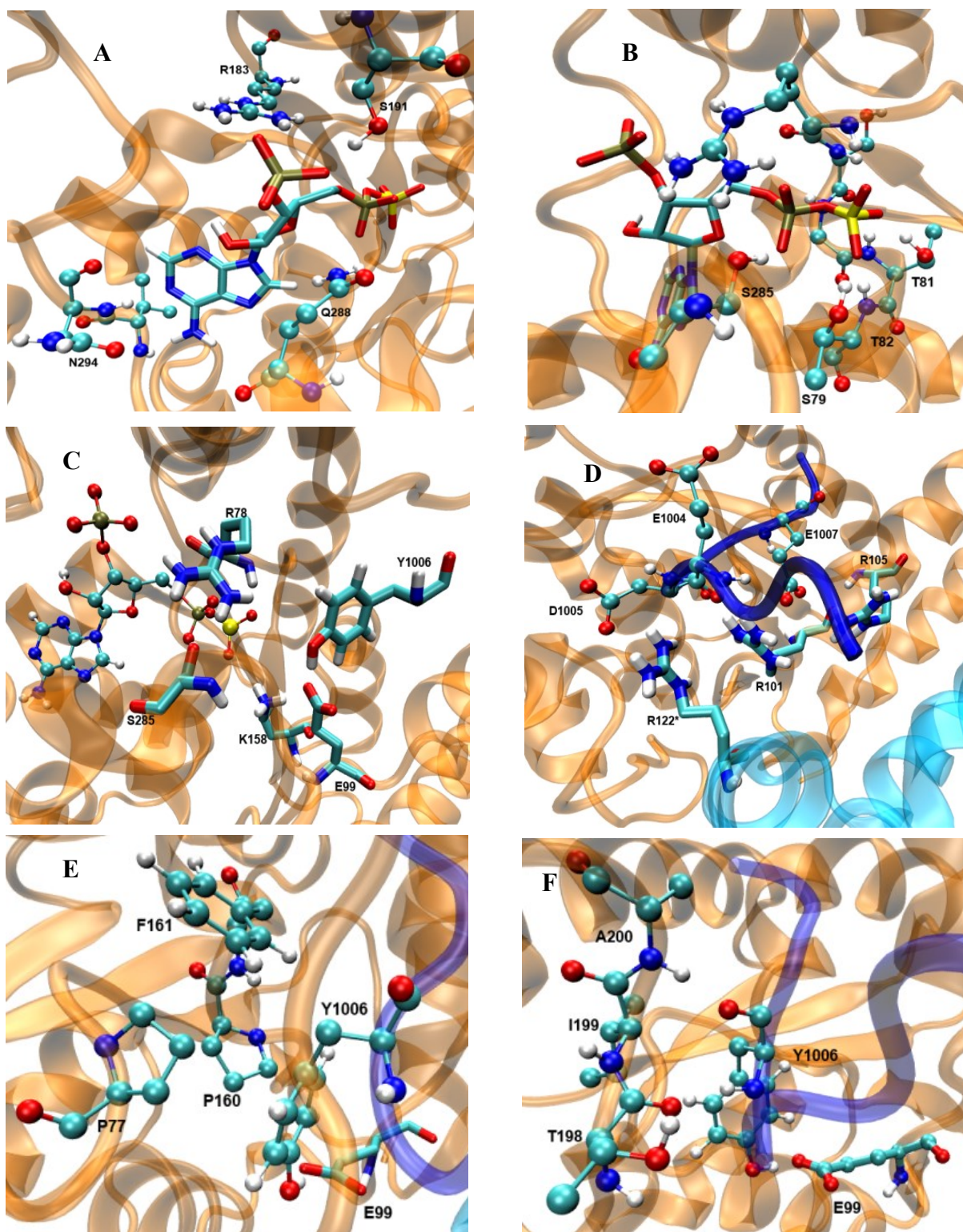
Figure 2.3 The Wild type TPST-2 enzyme and important residues obtained from the most populated cluster 100ns trajectory, (A) 3'PBS cofactor binding, (B) 5'PBS cofactor binding, (C) active site residue, (D) substrate binding site and residues, (E) hydrophobic

interactions of Y1006 acceptor tyrosine (F) backbone interactions βe (198-A200) of acceptor tyrosine residue. R122* residue of protomer B interacts with substrate peptide and residues of protomer.
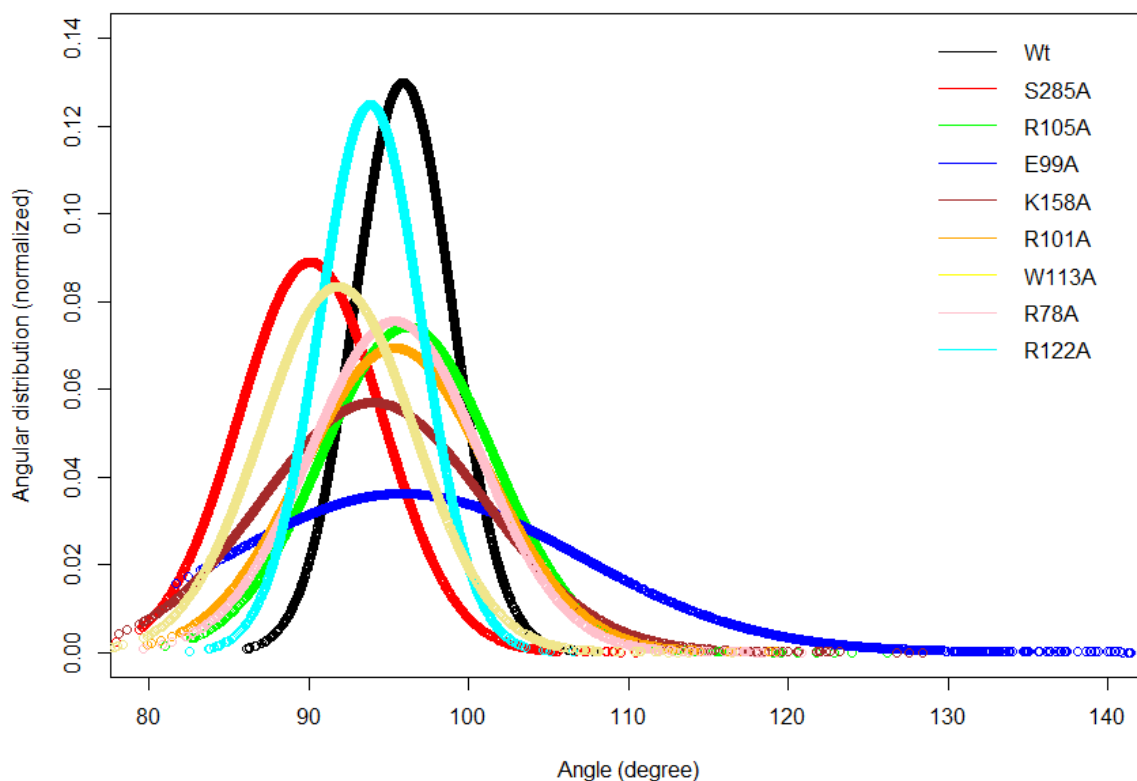
Figure 2.4 The Normalized distribution of the substrate angle of the wild type TPST-2 and the mutants. The angle was evaluated using C α atoms of Y1006, D1005 and E1004 substrate peptide residues. The plot was prepared using R by using normal distribution function available in R.

2.3.5 Catalytic Site

The catalytically important residues are R78, E99, K158 and S285A (**Table S2.5, Figure 2.3C**). R78 which is proposed to act as the catalytic acid interacts with the bridge oxygen of the leaving phosphate group as in the crystal structure (**Figure S2.23**). In the simulations the R78 side chain also interacts with the side chain of Q288 and with the hydroxyl group of S285. The backbone of R78 makes stable interaction with the hydroxyl group of the acceptor Y1006 of the substrate. In the R78A mutant (**Figure S2.24**) the interactions

between the PAPS and the A78 side chain weaken and the distance between the methyl groups of A78 and the bridge oxygen are at an average distance of 6 Å, as compared to 3 Å in the crystal structure. Importantly, in the R78A mutant E99 does not make hydrogen bond with the hydroxyl group of the acceptor tyrosine. The structural effect of R78 is in agreement with the experiments which demonstrate that R78A reduces the enzyme activity by 95 % with respect to crystal structure [83]. S285 which might be involved in the stabilization of the transition state interacts with the oxygen of leaving phosphate group in the crystal structure, an interaction which is also conserved during the simulation (**Figure S2.25**). The RMSF plot of the S285A mutant and WT FC indicates that the region spanning residues 265-280 (α11, α12 and loop between α12 and α13) shows higher fluctuations in the S285A mutant compared to WT FC. The S285 residue is located in between the loop α12 and α13. The RMSF of PAPS in the S285A mutant also showed increased fluctuations compared to WT FC equivalent, indicating the stabilizing effect of this residue on the cofactor (**Figure S2.26**). S285A mutant shows weaker contact between the OAF oxygen of the 5' phosphate group of the cofactor, than WT FC (**Figure S2.27, Table S2.11**). K158, which is also proposed to be involved in the transition state stabilization, interacts with both substrate and the cofactor in the crystal structure. The distance of side chain of K158 from the oxygen (OBE) of the sulfate group of PAPS and the hydroxyl group of acceptor tyrosine (Y1006) is 3.6 and 4.6 respectively during the simulation (**Figure S2.28**). The distances between the methyl group of A158 in the K158A mutant and hydroxyl group of tyrosine acceptor (Y1006) and to the sulfate group of cofactor shows considerable increase with respect to the WT FC (**Figure S2.29 Table S2.12**). The structural effect of this mutation is in agreement with the significantly reduced enzyme activity of WT FC. Interestingly, the RMSFs of the PAPS cofactor in the K158A mutant are higher than the WT FC, which supports its stabilizing effect for PAPS (**Figure S2.30**). E99, which is proposed to act as a catalytic base, makes stable electrostatic

interactions with the side chain of K158. In the E99A mutant the hydroxyl group of Y1006 no longer makes hydrogen bonds with the A99 side chain (**Figure S2.31, Table S2.13**) and the 5'PBS region of the cofactor. Additionally, the residues close to E99A show higher fluctuations demonstrating the stabilizing effects of E99 on the binding site and the cofactor. The sulfate group of PAPS is on average 6.0 Å away from the hydroxyl group of the acceptor tyrosine and this distance is stable during the simulation.
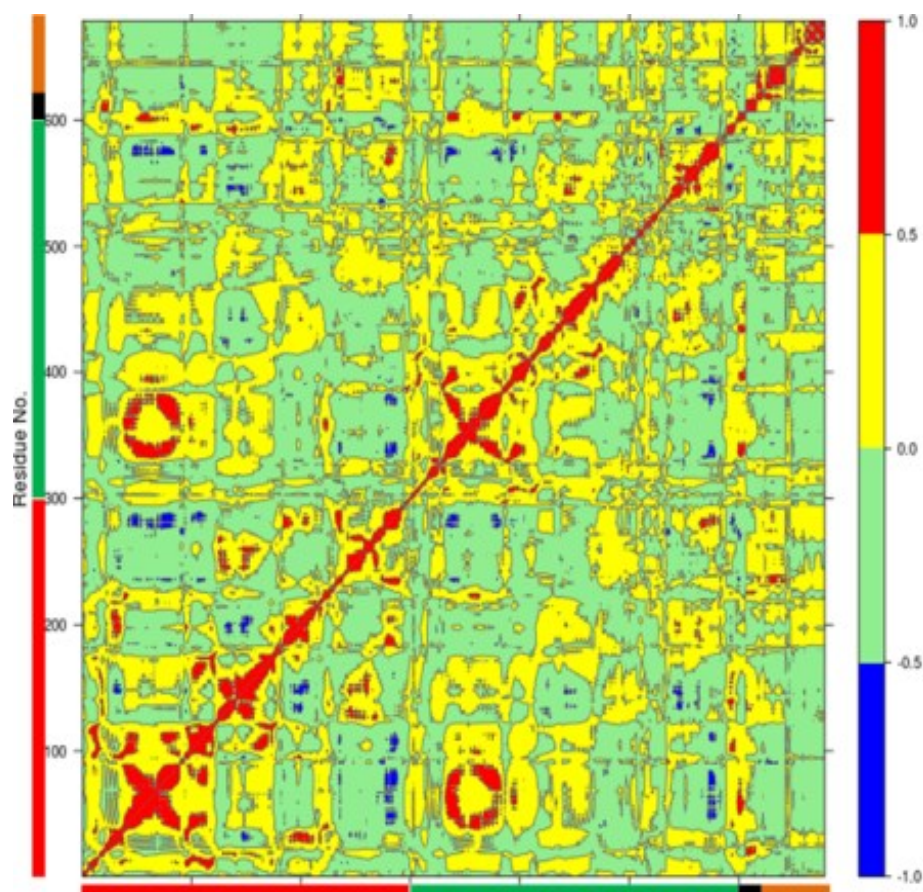
Figure 2.5 The Dynamic Cross Correlation Analysis (DCCA) of wild type TPST-2. Protomer A and B, substrate peptide and cofactor are represented in red, green, black and orange colours respectively. The positive correlated motion is shown in red colour and the negative correlated motion in blue colour. The analysis was done using R Bio3D package [98].

2.3.6 Dynamic Cross Correlation Analysis

An insight into the correlated motions between different parts of the enzyme molecule can be provided by Dynamic Cross Correlation Analysis (DCCA) [100].The DCCA of WT FC from the 100 ns trajectory is shown in **Figure 2.5**. The positive regions in the map are represented with red and yellow colour scheme is associated with correlated motion whereas anti-correlated motions (negative regions) are shown in light green and blue. Residues 78-82 (part of 5'PBS region) of WT FC show correlated motion with the sulfate group of the PAPS cofactor. The residues 70-80 (part of 5' PBS region) also showed correlation with the residues from α5-α6 and β5 (160-180, close to 3'PBS site and catalytic site). The residues 240-250, which belong to the loop between α13 and α14 (loop close to adenine ring of PAPS) move in correlation with the adenine ring of the cofactor. The key residues from the substrate binding site (R101, R105 and E99) show strong correlation with the substrate peptide residues in particular E1004, D1005 and Y1006 residues respectively. The residues from 99-110 (from substrate binding region α2) also showed correlation with the residues of α5-α6 helixes (160-170) which are close to 3'PBS. Interesting K158 residue showed strong correlation to the substrate peptide (Y1006) and weak correlation to the 5'sulfate group of the PAPS. The βe region showed correlation with the substrate residues D1005 and Y1006. There was also strong correlation between the residues encompassing α1, β4 (90-101) and α10, α11, α12 (250-270) of WT FC. The residues between α3 and α4 (115-130), involved in dimerization near the N terminal of protein show anti-correlated motion in respect to residues 324-333 which are far away at C terminal of protein. The analysis reveals that there are important correlated motions between the PBS and the cofactor, between the substrate binding site and the substrate and importantly between the PBS and the substrate binding site which might be important for formation of proper conformation for enzyme catalysis.

2.4 CONCLUSIONS

The present MD study reveals that the conformational flexibility and dynamics influences the structure-functions relationships of TPST-2. The simulations show that the sulfate group (which is not presented in the crystal structure) is stabilized by hydrogen bonds with S79, G80, T81 and T82, that the vast majority of interactions in the crystal structure are stable during the simulation and also that there are two new hydrophobic interactions stabilizing the adenine ring. The mutations influence the flexibility not only of the nearby residues around the site, but also have an effect on the different closer or distant parts of the enzyme molecule. The MD results are in agreement with the experimental data about the effects of the mutations of key residues for substrate binding on the enzyme activity, and suggest a more open conformation of the substrate and the appearance of small hydrophobic cluster which stabilize the aromatic ring of the acceptor Y1006. The important role of the conformational flexibility is of key importance for understanding of TPST-2 interactions with its ligands, understanding the effects of mutations and can be implemented in protein engineering and drug design.

2.5 SUPPORTING INFORMATION

Table S2.1: The MD setup for the wild type TPST-2 and the experimental mutants

S.No	MD-setup	Occurrence	Ions	Timescale (ns)
1	Wt-TPST-2	wild type	Na+ (10)	100
2	R78A	loop between β3-α1 (5' PSB motif)	Na+ (12)	100
3	E99A	loop between β4-α2	Na+ (8)	100
4	K158A	β5	Na+ (12)	100
5	S285A	α13	Na+ (10)	100
6	T198A	βe	Na+ (10)	100
7	R101A	loop between β4-α2	Na+ (12)	100
8	R105A	α2	Na+ (12)	100
9	R122A	α3	Na+ (12)	100

Table S2.2: Average properties of wild type TPST-2 and mutant's structures computed for RMSD and RMSF for the TPST-2

Name	RMSD C α (Å)								RMSF	
	All		Alpha helices		Beta sheets		Loops		RMSF > 1Å (%)	RMSF < 1Å (%)
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2		
WT	2.72	0.27	2.25	0.25	1.73	0.14	3.2	0.66	57	43
R78A	2.9	0.45	3.09	0.39	2.39	0.27	3.66	0.47	67	33
E99A	2.68	0.39	2.94	0.34	2.55	0.21	3.40	0.48	55	45
R101A	2.54	0.28	2.91	0.29	2.07	0.23	3.21	0.24	78	22
R105A	3.01	0.31	3.08	0.26	2.51	0.29	3.76	0.45	65	35
R122A	2.35	0.25	2.67	0.27	1.83	0.19	2.82	0.18	67	33
K158A	2.98	0.35	3.04	0.29	2.66	0.28	3.64	0.36	81	19
S285A	3.73	0.43	3.85	0.45	2.28	0.35	4.77	0.60	78	28
T198A	3.60	0.43	3.54	0.38	2.65	0.22	4.63	0.74	75	25

μ = mean, σ^2 =standard deviation

Table S2.3: Local structural RMSF of wild type and mutants of TPST-2. The region is chosen to increase or decrease if there is an average change in RMSF of > 0.3 Å in at least 50% of its residue

Name	Increased		Decreased	
	Residue span	Name	Residue span	Name
E99A	114-120 , 340-350	α3, α16	149-153, 201-210,256-261	α4-β5, βe α7-α8, α10-α11
K158A	75-85, 111-131, 184-194 ,200-230, 237-247,276-289, 340-350	5'PBS, α2α3α3-α4, 3'PBS, βeα8α8-β7, β7α9, α12-α13, α16	148-153, 328-336	α4-β5, α15-α16
R78A	105-120,174-176	α2α3, α6-β6	149-154, 200-203,	α4-β5, βe
S285A	65-70, 97-108, 132-148, 265-295	β2-β3, β4-α2α2, α4 α12α12-α13α13-α14	149-154, 200-205, 331-335	α4-β5, βe, α15-α16
R101A	102-118, 135-148, 162-176, 243-250, 265-270, 277-283	α2α3, α4α4-β5 ,α5α6, α10, α12, α12-α13	149-152, 303-308, 329-335	α4-β5, α13-α14, α15-α16
R105A	65-70 188-200, 277-290	β2-β3, α7βe, α12-α13α13	149-154, 259-262, 302-309, 330-337	α4-β5, α10-α11, α13-α14,α15-α16
R122A	170-174, 308-310	α6-β6, α13-α14	149-152, 259-262, 330-337	α4-β5, α11, α15-α16
T198A	115-130, 164-176, 228-232, 261-270	α13, α5α6, α8-β7, α11α12	149-152, 200-205, 301-308,330-336	α4-β5, βe, α13-α14, α15-α16

* Regions underlined are the loops between the alpha helix and beta sheets.

Table S2.4: Hydrogen bond profile of cofactor binding site of wild type TPST-2 for Protomer A

Donner	Acceptor	Distance(Å)	Probability %
S79 (bb)	OBC-sulfate	2.78	95.1
G80 (bb)	OBC-Sulfate	2.71	97.3
T81 (sc)	OBE-Sulfate	2.6	96.1
T82 (sc)	OBD-sulfate	2.59	98.8
T82 (bb)	OBD-sulfate	2.75	99.1
L83 (bb)	OAG- phosphate	2.83	98.8
S285 (sc)	OAF- phosphate	2.67	95.5
S191 (sc)	3'PBS-OAB	3.1	66.9
N294 (bb)	N7-ring	3.04	89.9
S79 (bb)	OBC-Sulfate	2.78	95.1
G80 (bb)	OBC-Sulfate	2.71	97.3
T81 (sc)	OBE-Sulfate	2.6	96.1
T82 (sc)	OBD-sulfate	2.59	98.8
T82 (bb)	OBD-sulfate	2.75	99.1
L83 (bb)	OAG- phosphate	2.83	98.8
S285 (sc)	OAF- phosphate	2.67	95.5

Table S2.5: Important substrate and catalytic site interactions of wild type TPST-2 averaged over 100ns

Substrate Interactions	Residue	Distance (Å)	Catalytic site Interactions	Residue	Distance (Å)
Y1006 (sc)	E99 (sc)	3.16	R78 (sc)	PAPS (OAD)	3.0
Y1006 (sc)	P77 (sc)	4.2	R78 (sc)	Q288 (sc)	4.1
Y1006 (bb)	T198 (bb)	3.5	R78 (sc)	S285 (sc)	3.7
Y1006 (bb)	A200 (bb)	3.5	R78 (bb)	Y1006 (sc)	5.8
Y1006 (sc)	I199 (sc)	4.1	S285 (sc)	PAPS (OAF)	2.6
Y1006 (sc)	P160 (sc)	3.9	S285 (sc)	PAPS (OAD)	4.8
D1005 (bb)	T198 (sc)	3.1	K158 (sc)	OBE sulfate	3.6
D1005 (bb)	R101 (sc)	3.5	K158 (sc)	Y1006 (sc)	4.6
D1005 (sc)	R101 (sc)	4.3	E99 (sc)	K158 (sc)	2.7
D1005 (sc)	R122 (sc)	5.2	Y1006 (sc)	PAPS sulfate	6.1
E1004 (bb)	R101 (sc)	3.5			
F1003 (sc)	R101(sc)	4.3			
D159 (sc)	R101(bb)	3.2			
E1007 (sc)	R105 (sc)	3.6			
R122 (sc)	E99 (sc)	4.0			

Table S2.6: The substrate angle for wild type and mutants for Cα atoms of 1004, 1005 and 1006 residues of substrate peptide

Name	Angle (mean)
WTFC	95.9°
R78A	93.5°
E99A	103.5°
K158A	94.5°
S285A	90.0°
R101A	95.3°
R105A	96.1°
R122A	91.8°
T198A	93.1°

Table S2.7: The hydrogen bond profile for substrate binding in the wild type TPST-2

Donor	Acceptor	Distance (Å)	Probability (%)
Y1006 (bb)	T198 (bb)	3.5	86.6
A200 (bb)	Y1006 (bb)	4	63.6
Y1006 (sc)	E99 (sc)	3.1	99.9
T198 (sc)	D1005 (bb)	3.1	79.1
T198 (sc)	D1005 (sc)	3.5	59.6
R101 (sc)	E1004 (bb)	3.4 (70-ns)	25.6
R101 (sc)	D1005 (bb)	0.38	0.5

Table S2.8: The cluster and minimized structure distances of mutant E99A and wild type TPST-2

	Cluster Analysis				Initial crystal structure with hydrogen
	Wild-type		E99A mutant		Wild type
Residue	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
E99 (sc) - R101(sc)	7.74	7.58	4.13	7.23	7.31
E99 (sc) - K158 (sc)	3.77	4.14	4.38	4.20	3.52
E99 (sc) –Y1006 (oh)	2.80	2.61	10.64	8.72	2.36
Y1006(oh) - sulfate	5.85	6.47	7.92	3.64	4.56
R78(bb) - Y1006(oh)	4.64	5.56	4.25	3.23	4.91
K158 (sc) - sulfate	5.87	5.65	5.34	4.77	3.80
K158 (sc) – Y1006 (oh)	4.44	4.60	9.49	6.47	4.35

Note: oh = hydroxyl group of substrate Y1006

Table S2.9: The cluster analysis and minimized structure distances of mutant T198A and wild type TPST-2

	Cluster Analysis				Initial crystal structure with hydrogen
	Wild type		Mutant T198A		Wild Type
Residue	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
T198 (bb) – Y1006(bb)	1.83	1.89	2.72	2.83	**2.09**
T198 (sc) – D1005 (sc)	2.38	2.33	3.84	4.15	**3.92**
T198(sc) – D1005 (bb)	3.05	2.75	5.15	5.35	2.89
A200(bb) - Y1006 (bb)	2.23	4.34	2.70	2.72	1.81
I199 (sc) - Y1006 (sc)	3.77	3.65	3.63	4.34	4.17

Table S2.10: The cluster analysis and minimized structure distances of mutant R101 and wild type TPST-2

	Cluster Analysis				Initial crystal structure with hydrogen
	Wild type TPST-2		K158A mutant		Wild type
Residue	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
R101 (sc)- E1004 (bb)	2.92	3.18	6.12	6.93	2.83
R101 (sc)- D1005 (bb)	3.48	2.98	4.91	5.68	2.99
R101 (sc) - D1005 (sc)	6.17	3.69	9.10	9.47	4.38

Table S2.11: The cluster analysis of S285A mutant and the wild type TPST-2

| Residue | Cluster Analysis | | | | Initial crystal structure with hydrogen |
| | Wild type TPST-2 | | S285A mutant | | Wild Type |
	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
S285 (sc) - bridge-oxygen of cofactor	3.70	4.10	5.92	6.70	3.96
S285(sc)- phosphate close to sulphate	2.52	2.83	4.77	5.55	2.55
S285(sc) - Oxygen OAF of phosphate close to sulphate	1.57	1.75	3.72	4.53	2.57
S285 (sc) – OAF of cofactor	3.35	2.88	2.29	3.24	5.64

Table S2.12: The cluster analysis of K158A mutant and the wild type TPST-2

| Residue | Cluster Analysis | | | | Initial crystal structure with hydrogen |
| | Wild type TPST-2 | | K158A mutant | | Wild type |
	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
K158 (sc) – Y1006 (oh)	4.44	4.60	10.53	11.20	4.35
K158 (sc) - sulfate group (OBE)	6.09	5.97	11.84	10.18	4.60
K158 (sc) – bridge oxygen of cofactor	7.08	6.91	12.03	11.26	4.89

Table S2.13: The Cluster Analysis of E99A and the wild type TPST-2

| Residue | Cluster Analysis | | | | Initial crystal structure with hydrogen |
| | Wild-type TPST-2 | | E99A mutant | | Wild type |
	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
E99 (sc) - R101 (sc)	7.74	7.58	4.13	7.23	7.31
E99 (sc) - K158 (sc)	3.77	4.14	4.38	4.20	3.52
E99 (sc) - Y1006 (oh)	2.80	2.61	10.64	8.72	2.36
Y1006 (oh) - sulfate	5.85	6.47	7.92	3.64	4.56
R78 (sc) -Y1006 (oh)	4.64	5.56	4.25	3.23	4.91
K158 (sc) - sulfate	5.87	5.65	5.34	4.77	3.80
K158 (sc) - Y1006 (oh)	4.44	4.60	9.49	6.47	4.35

Table S2.14 The average properties of the multiple runs of the wild type TPST-2 for 100 ns

Name	Mean (Å)	S.D (Å) within each trajectory	S.E.M (Å)	S.D (Å) In respect to the average trajectory	E99 (sc)-Y1006(OH) Å	Y1006(OH)-PAPS(sulphur) Å	Substrate angle
Run 1	2.73	0.24	0.004	0.22	3.1	6.1	95.9°
Run 2	2.81	0.15	0.002	0.21	3.4	6.5	95.5°
Run 3	2.40	0.23	0.004	0.21	3.5	6.7	94.7°
Averaged	*2.65*	*0.16*	*0.002*		*3.3*	*6.4*	*95.3°*
200ns (extension of Run 1)	2.90	0.34	0.003		3.6	6.3	96.0°

Table S2.15 The averaged properties of E99A mutant for multiple run trajectories for 100 ns run

Name	Mean (Å)	S.D (Å)	S.E.M (Å)	A99(sc) – Y1006(sc0	A99(sc) – K158 (sc)	Substrate angle
Run 1	2.42	0.25	0.006	9.5	4.3	103.5°
Run 2	2.99	0.30	0.006	8.2	4.0	98.9°
Averaged all	2.7	0.27	0.006	8.8	4.1	101.2°
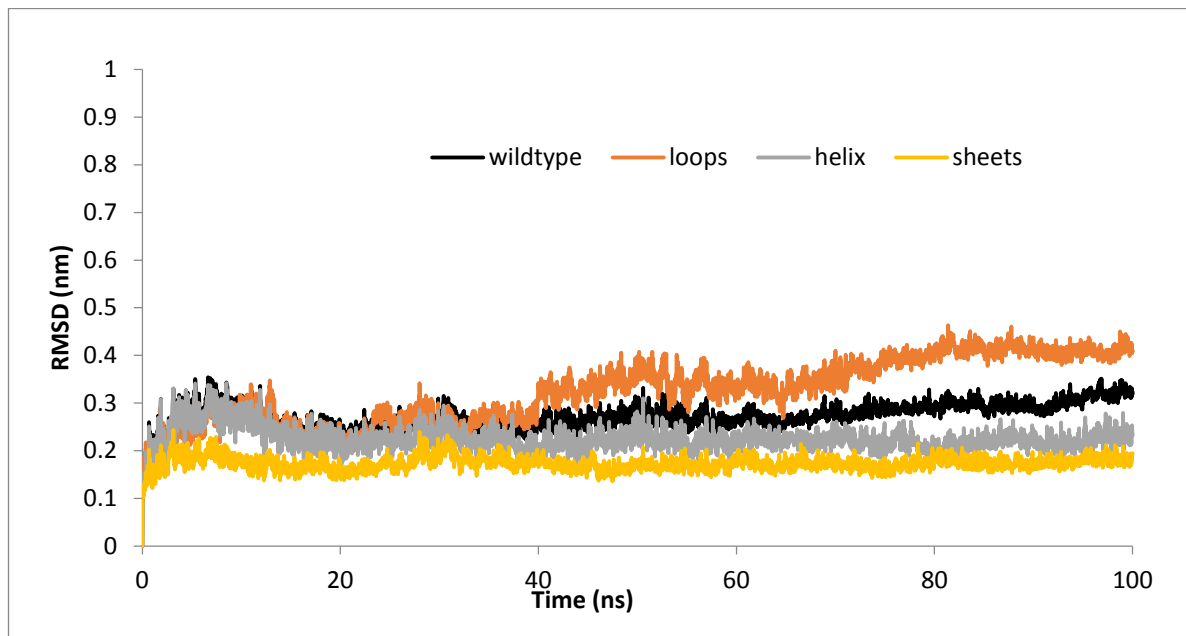
Figure S2.1 The secondary structure RMSDs of wild type TPST-2. This includes alpha helix, beta sheets and loops for 100ns.
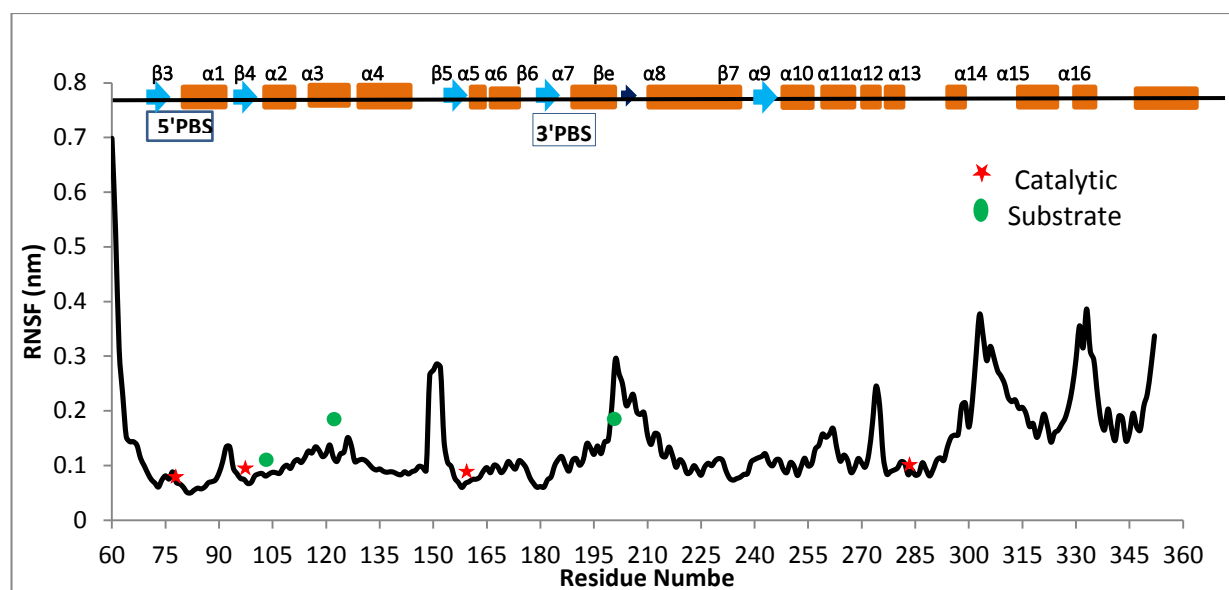
Figure S2.2 the RMSF analysis of protomer A of wild type TPST-2. The 5'PBS and 3'PBS are the cofactor binding sites and βe is the substrate binding residues in the wild type TPST-2. The catalytic residues are represented in red colour star and the substrate binding residues are in green circle representation.
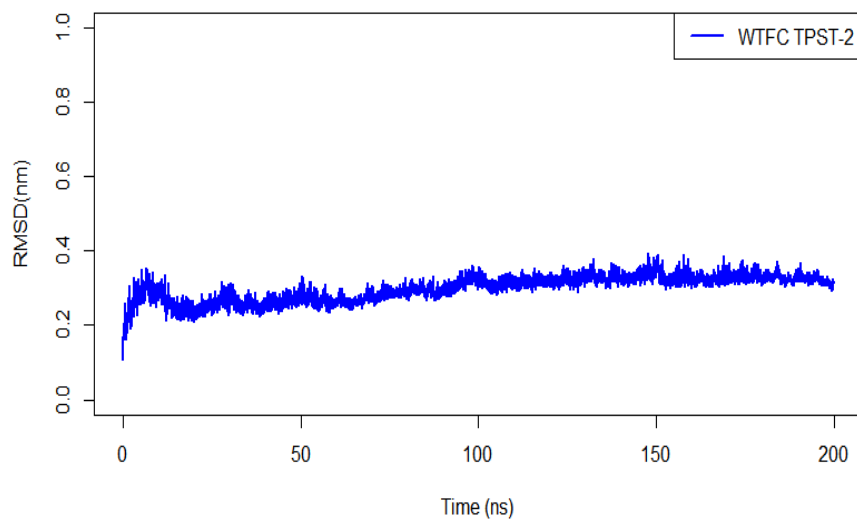
Figure S2.3 The RMSD trajectory of all C α atoms of wild type TPST-2 (WT FC) for 200ns
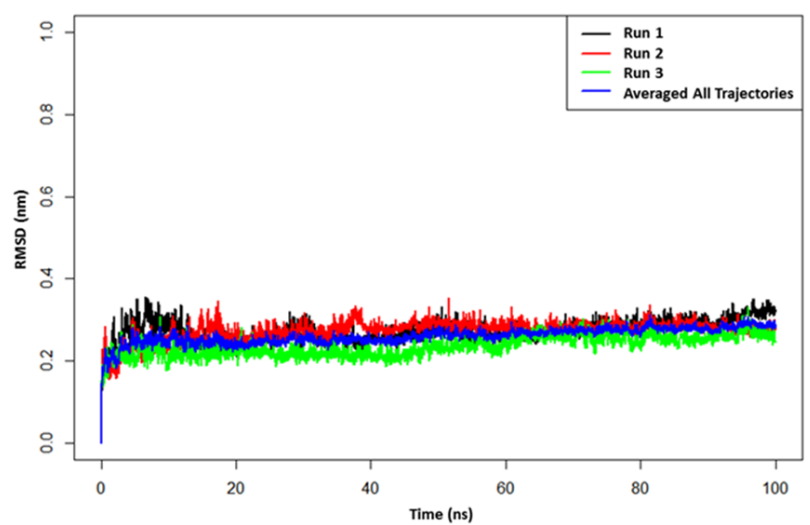
Figure S2.4 The RMSD of wild type TPST-2. The average trajectory is shown in blue colour from run 1, run 2 and run 3
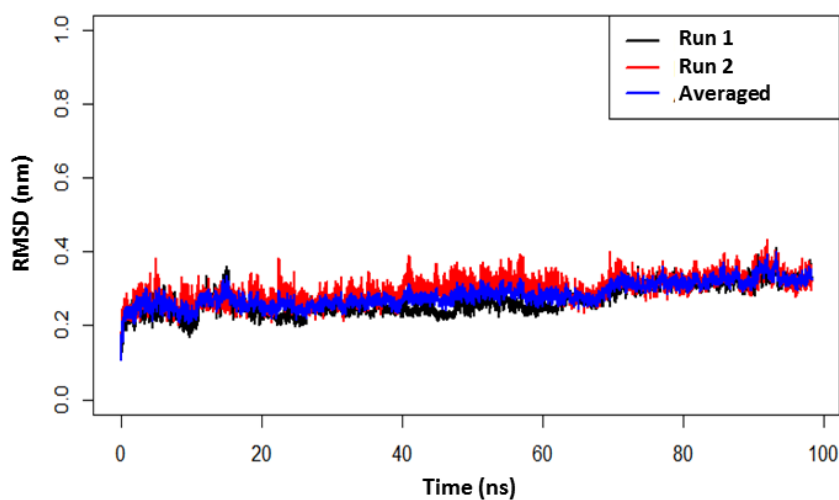
Figure S2.5 The replica runs of the E99A mutant along with the average run (100ns) for two trajectories.
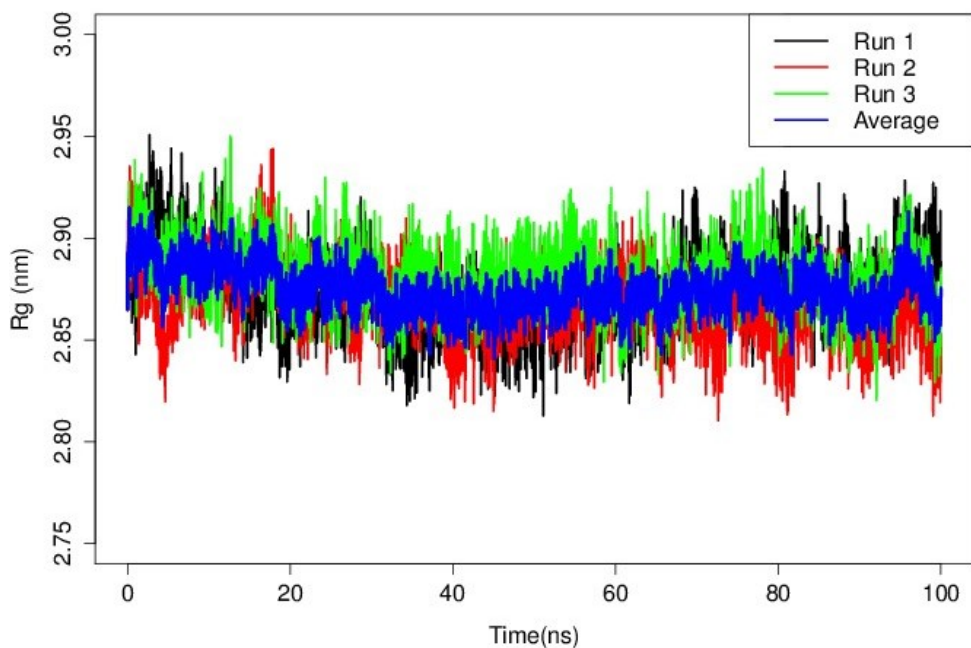
Figure S2.6 The Radius of Gyration (Rg) of WTFC for 100 ns. The average value of Rg from the independent runs has been plotted in blue colour as function of simulation time.
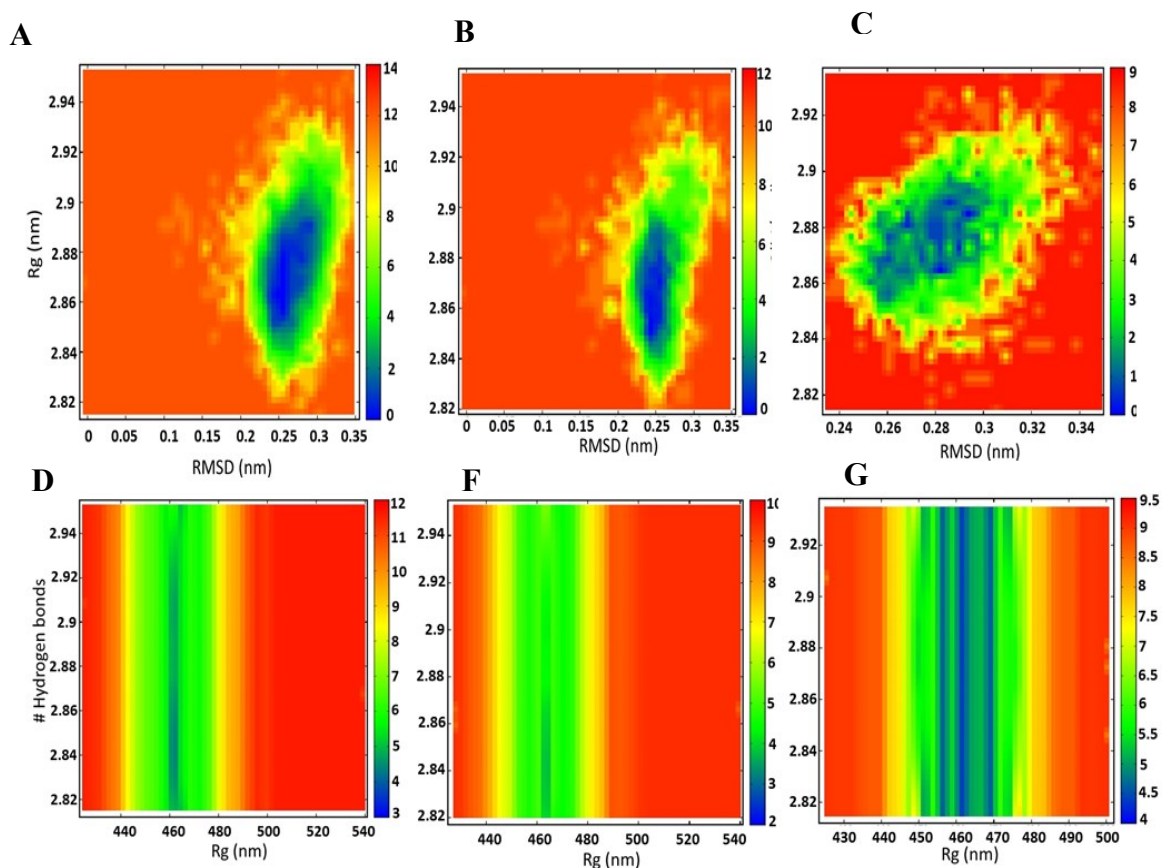
Figure S2.7 Figure S5 The free energy landscape (FEL) sampled from 100 ns trajectory for wild type TPST2 (WTFC) at 300 K. The Gibbs free energy is estimated from probability distribution of the system with respect to Radius of gyration (Rg), Root Mean Square Deviation (RMSD) and

58

protein-protein hydrogen bonds. (A) RMSD vs Rg from 0-100 ns, (B) RMSD vs Rg from 0-50 ns, (C) RMSD vs Rg from 50-100 ns, (D) Rg vs protein-protein hydrogen bonds from 0-100 ns, (E) Rg vs protein-protein hydrogen bonds from 0-50 ns, (F) Rg vs protein-protein hydrogen bonds from 50-100ns. The FEL was obtained using g_sham option in Gromacs.
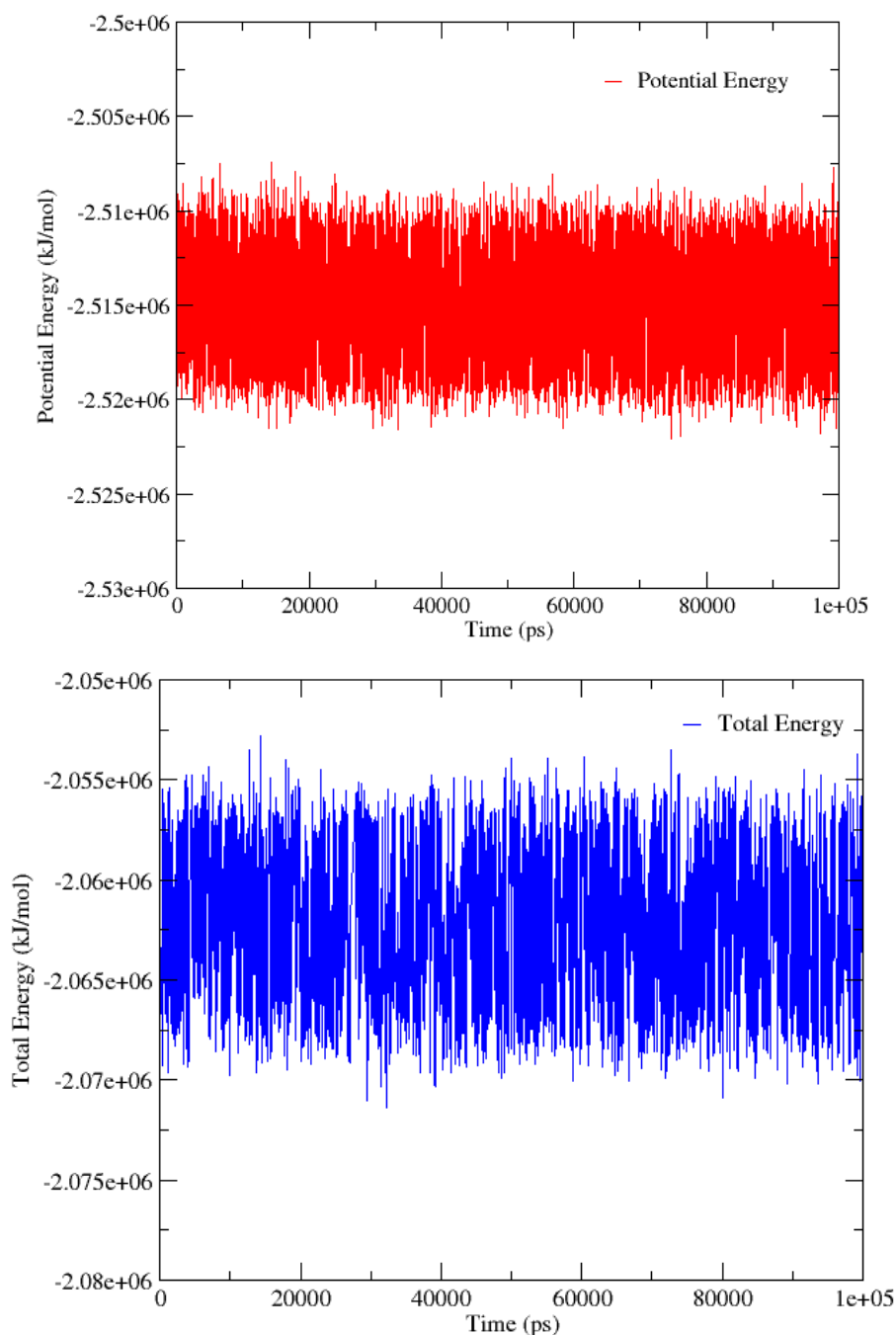
Figure S2.8 The plot of total and potential energy of WTFC vs simulation time for 100 ns.
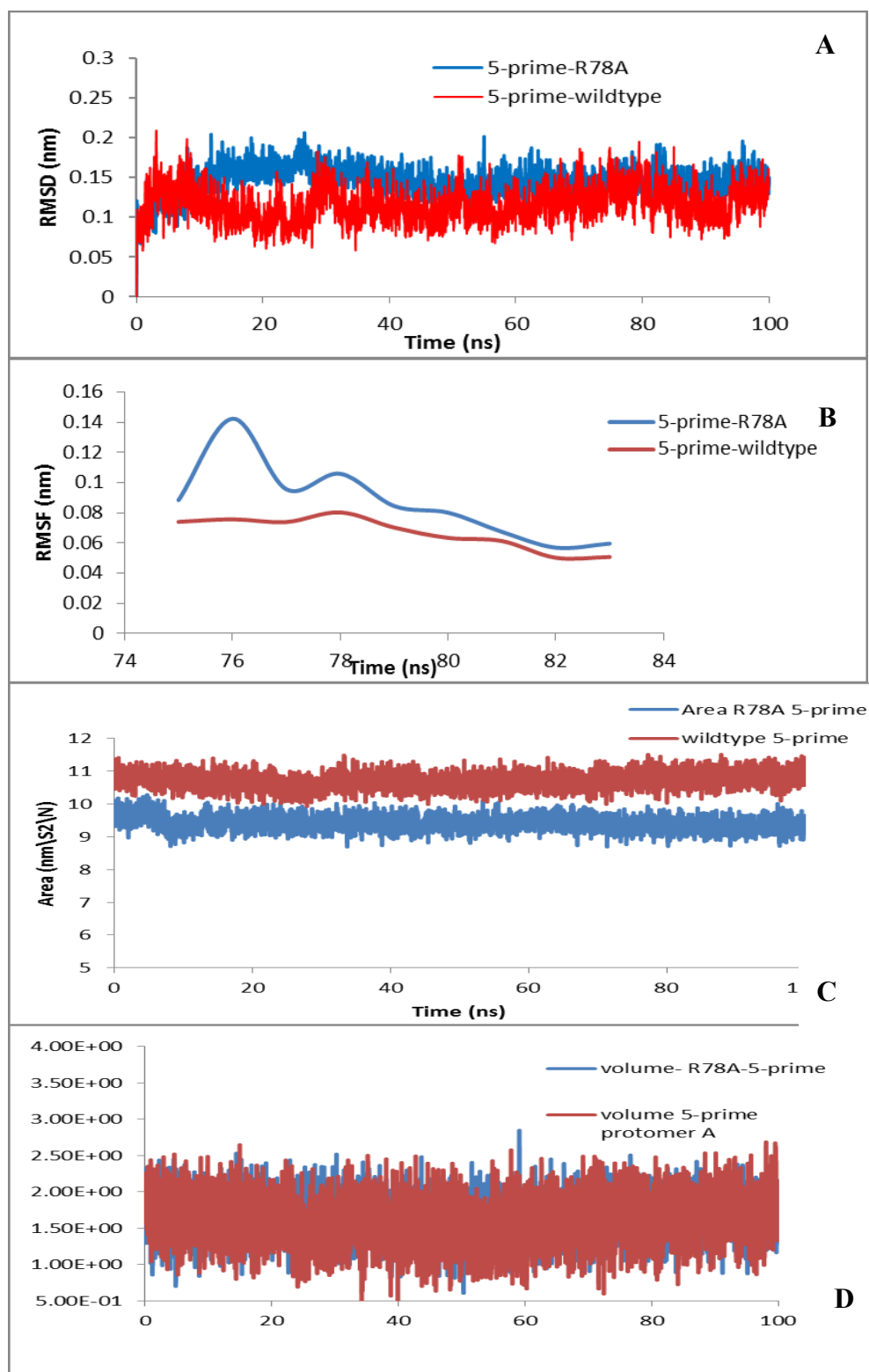
Time (ns)

Figure S2.9 The 5'PBS binding region of the wild type TPST-2 compared against R78A mutant;(A) The RMSD of 5'PBS region, (B) RMSF of 5'PBS region, (C) The SASA of 5'PBS region and (D) volume of 5'PBS region of wild type TPST-2 and the R78A mutant

60
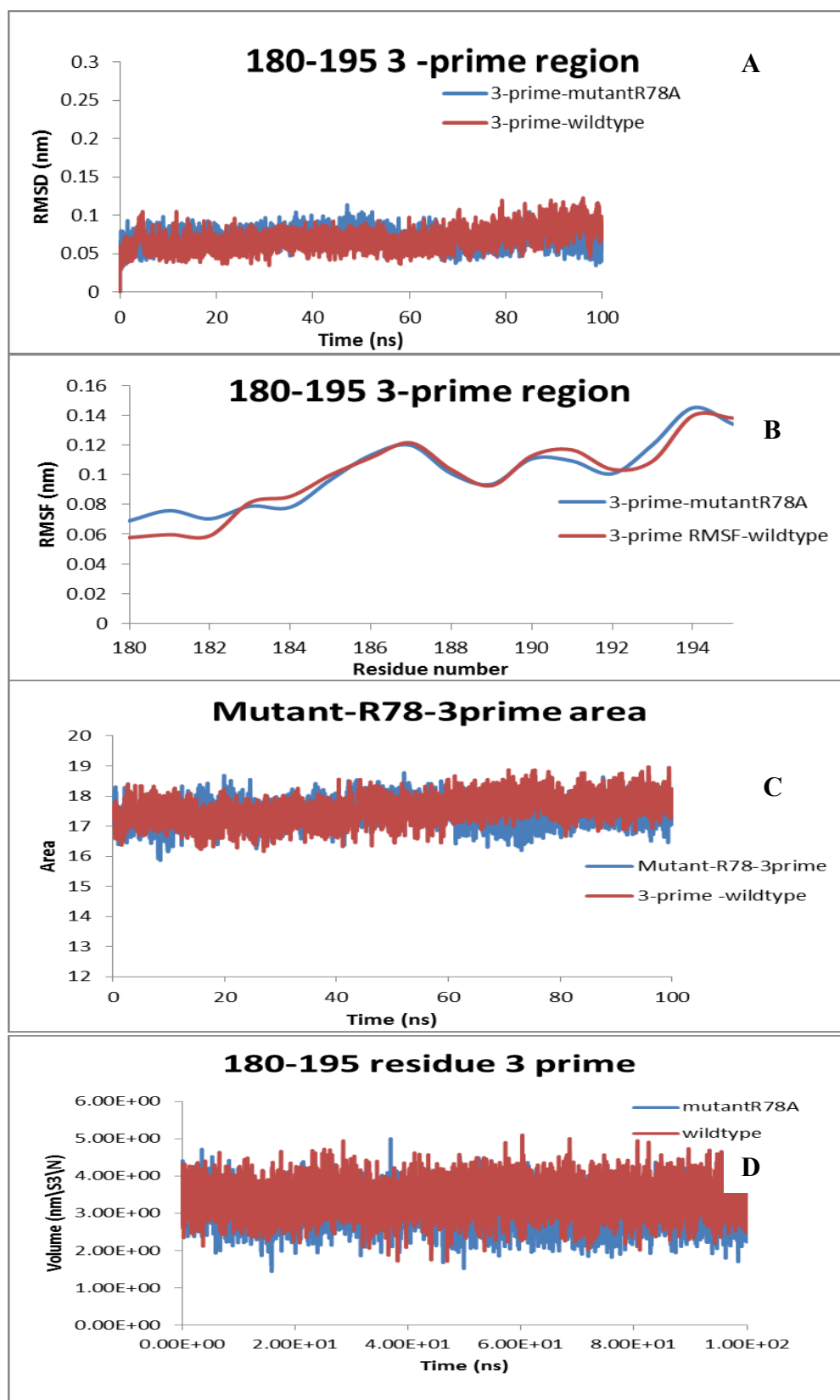
Figure S2.10 The 3'PBS binding region of the wild type TPST-2 compared against R78A mutant. (A) The RMSD of 3'PBS region, (B) RMSF of 3'PBS region, (C) The SASA of 3'PBS region and (D) volume of 3'PBS region of wild type TPST-2 and the mutant R78A.
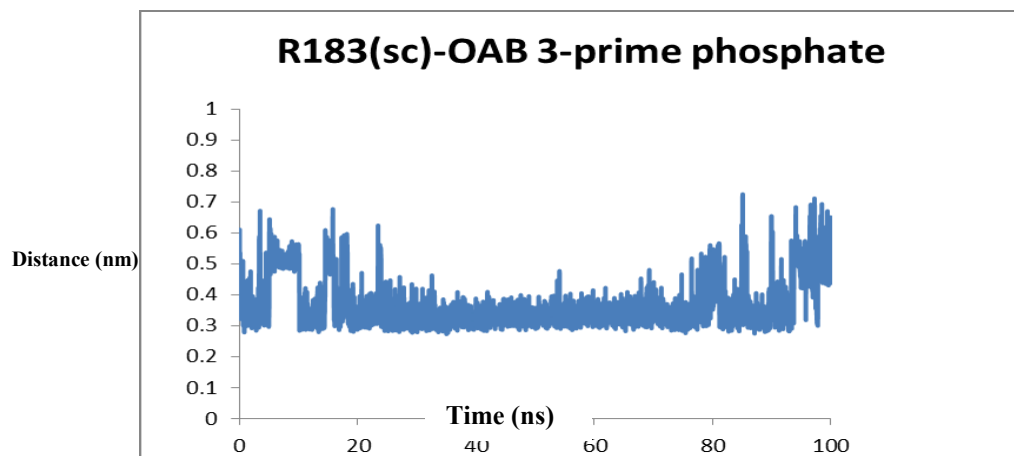
Figure S11 The electrostatic interactions of the side chain of R183 with OAB oxygen of 3'PBS region of wild type TPST-2
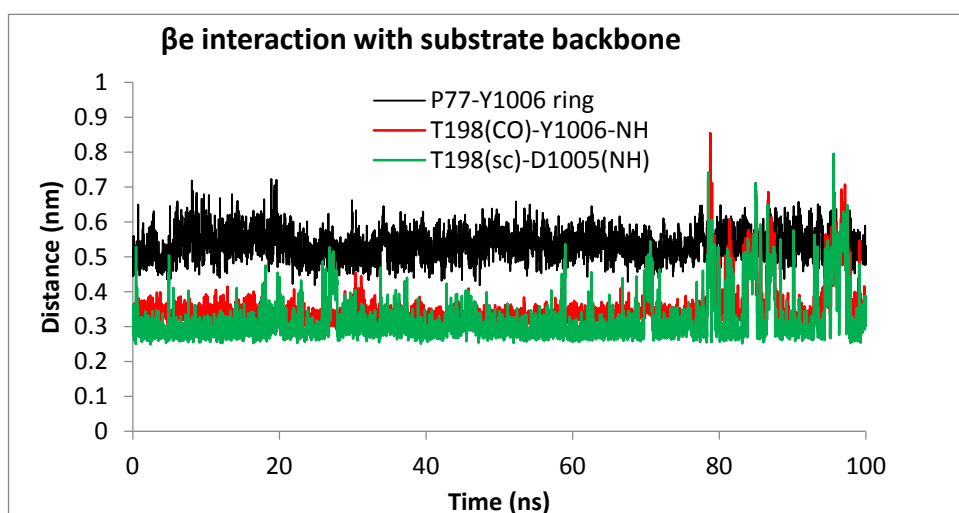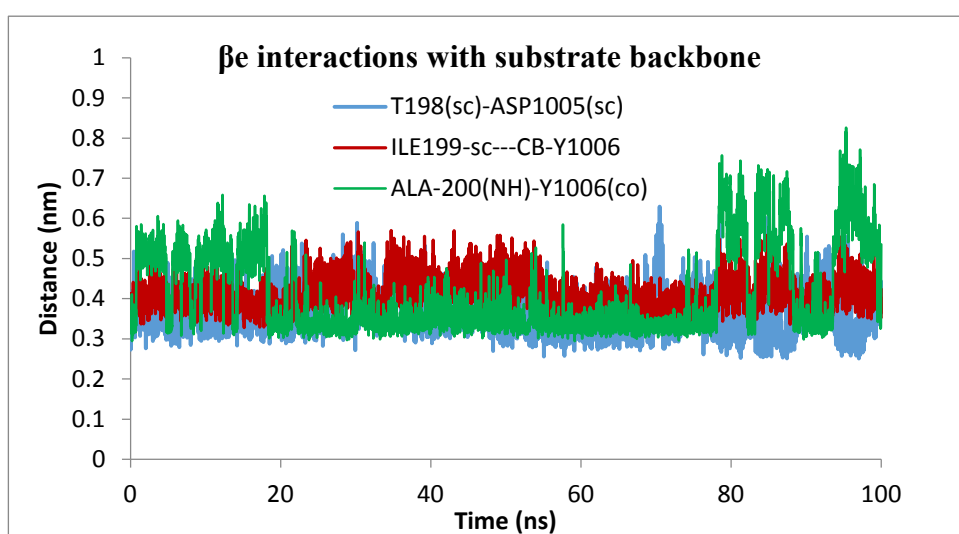
Figure S2.12 The interactions of wild type TPST-2 with the substrate tyrosine Y1006
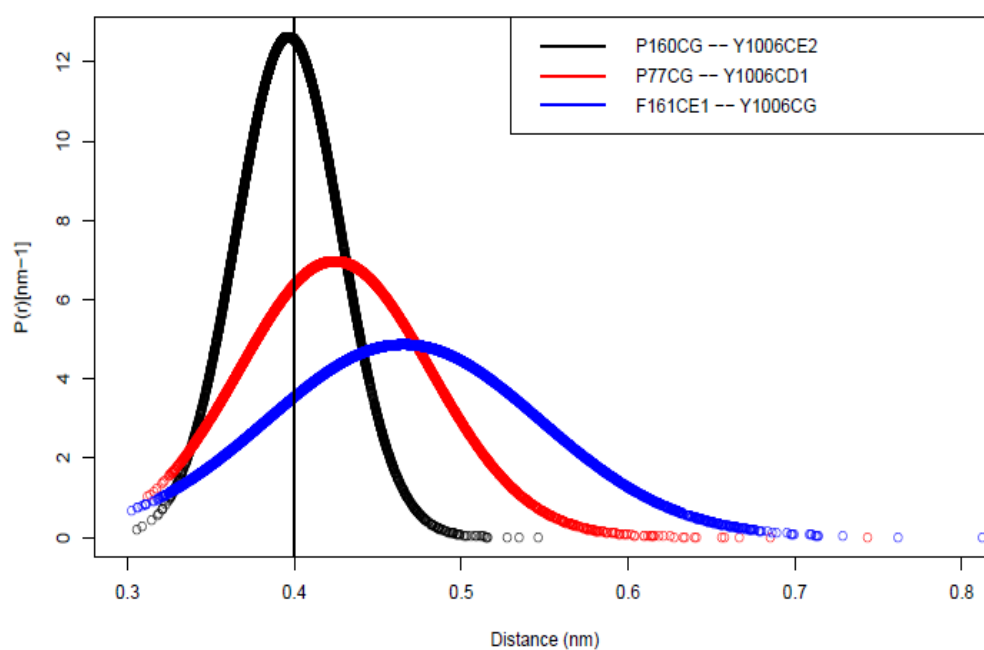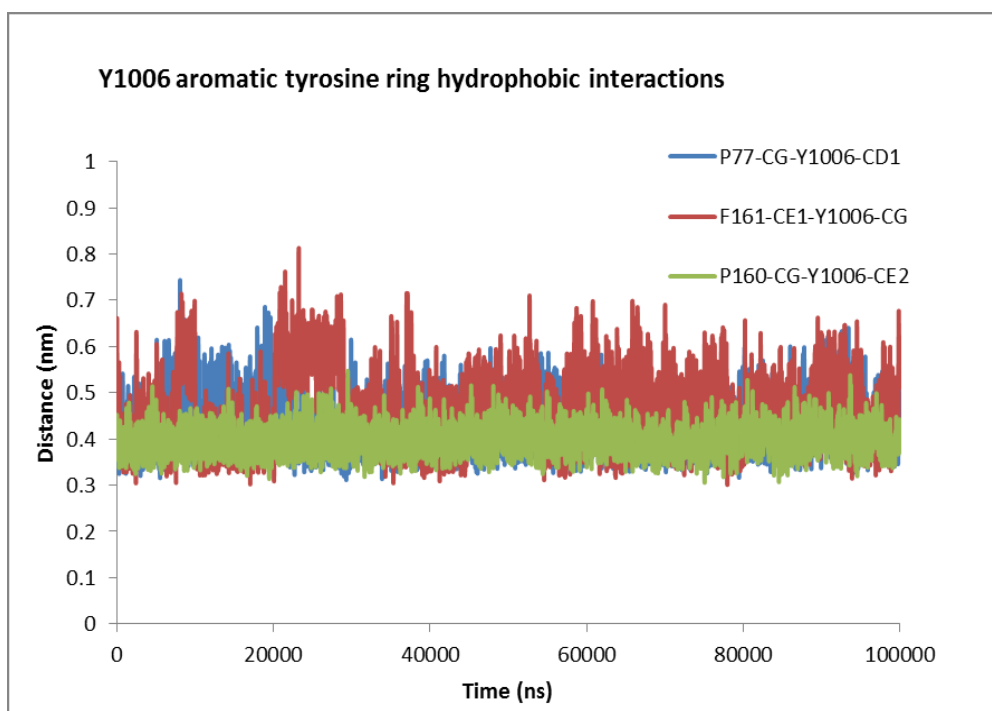
Figure S2.13 The hydrophobic interactions of wild type TPST-2 with Y1006 aromatic ring of substrate. The normalized distribution of wild type TPST-2 with Y1006 aromatic ring of substrate.
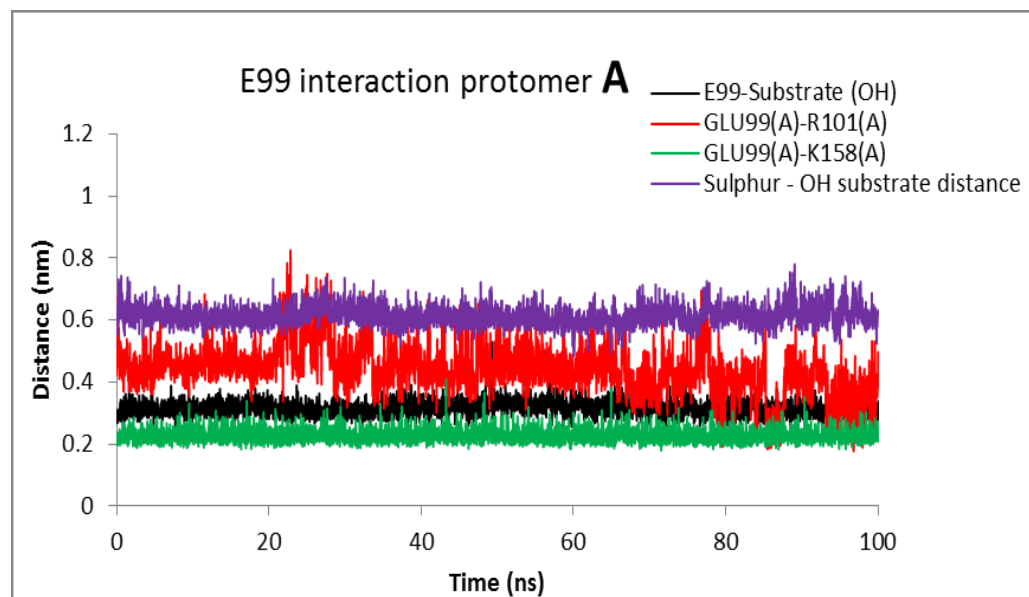
Figure S2.14 The E99 residue side chain interactions of wild type TPST-2
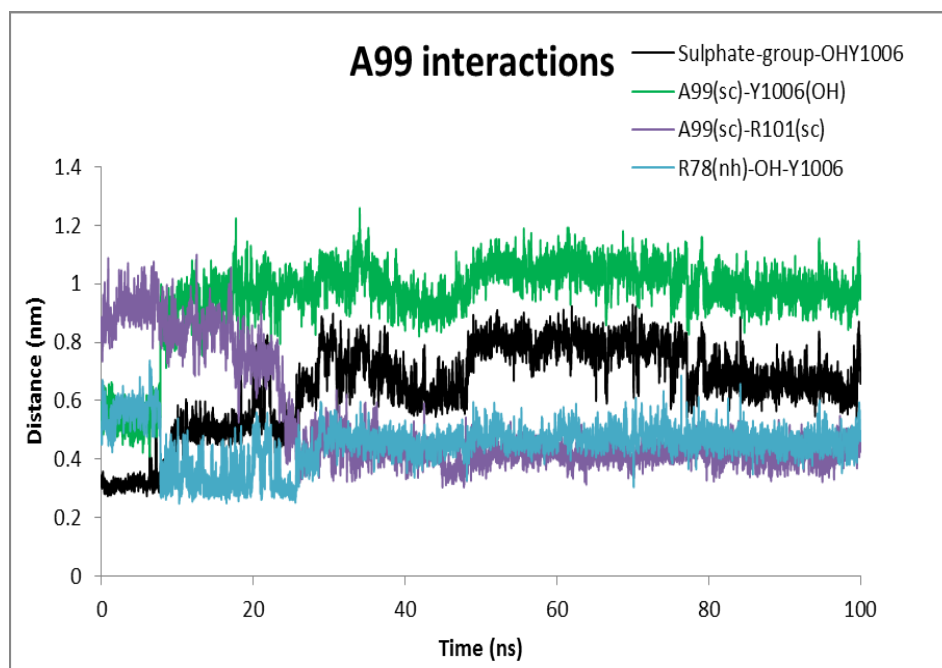
Figure S2.15 The E99A mutant interactions for 100 ns trajectory
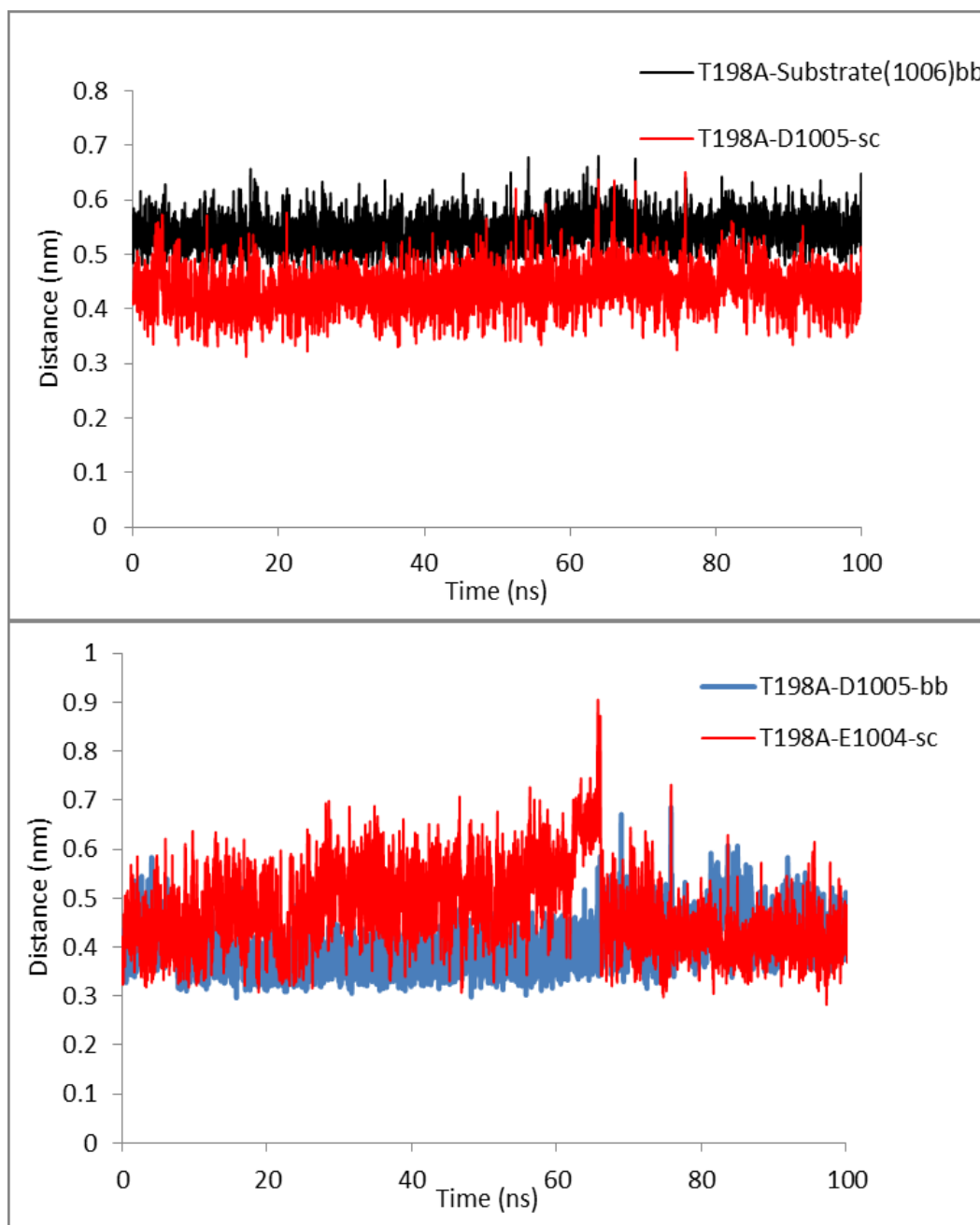
Figure S2.16 The T198A mutant interactions with residue D1005 and E1004
sidechain and backbone
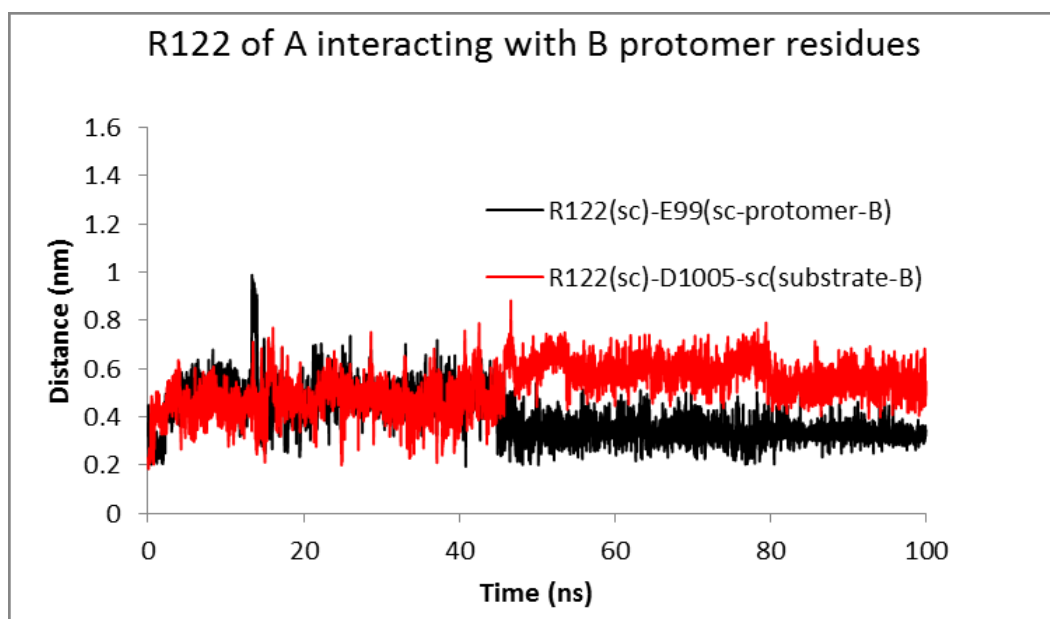
Figure S2.17 The electrostatic interactions between the side chain of R122 and E99 and D1005 of Wild type TPST-2.
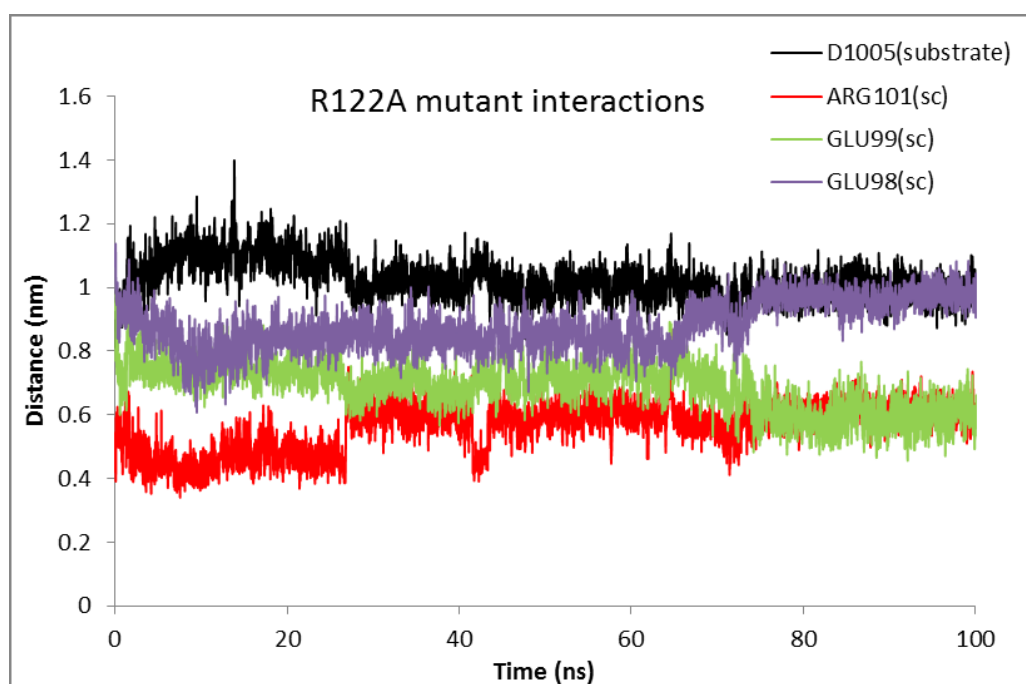
Figure S2.18 The electrostatic interaction of R122A mutant and the D1005, E99, E98 and R101 residues
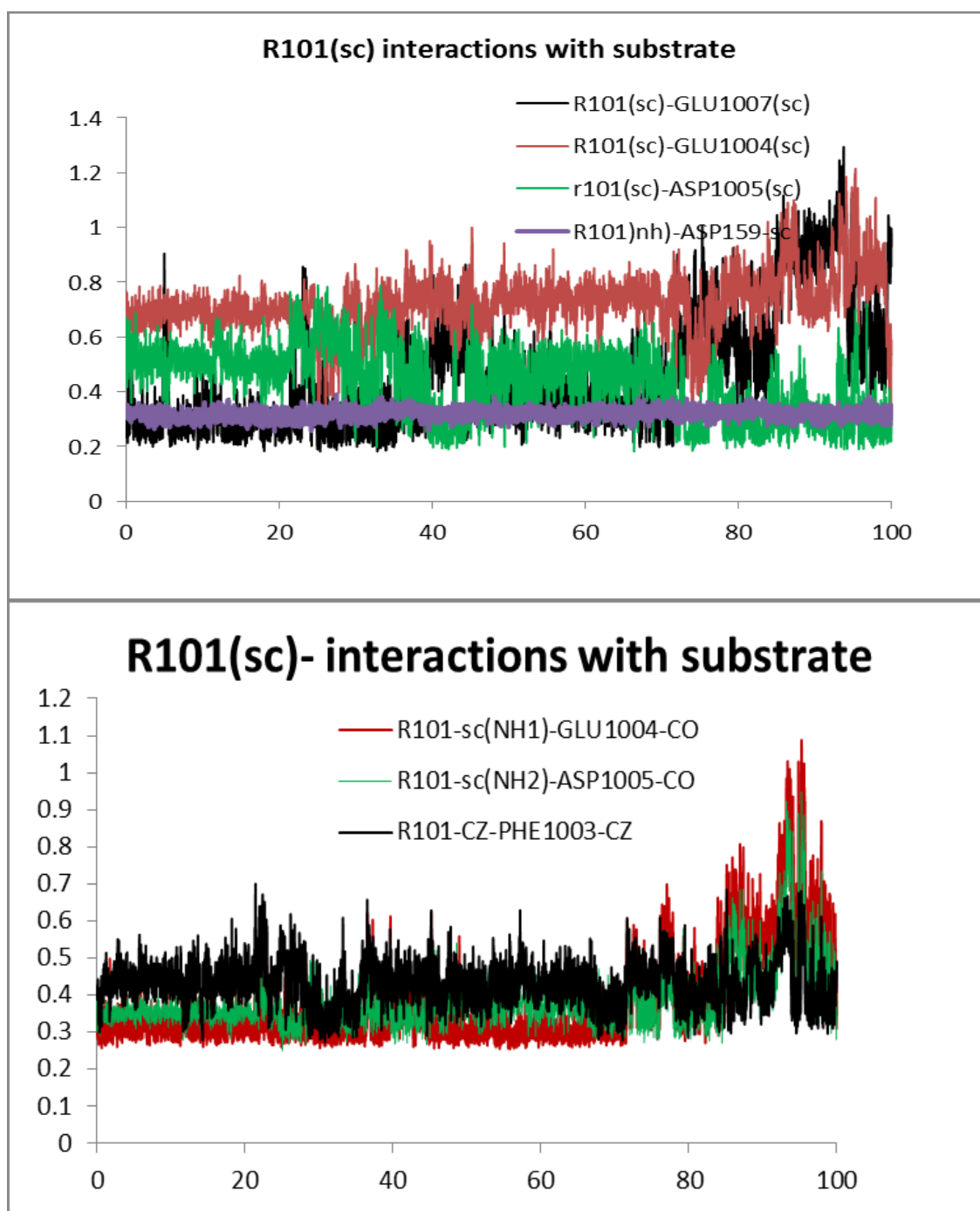
Figure S2.19 The R101 interactions of wild type TPST-2 with the substrate peptide residues
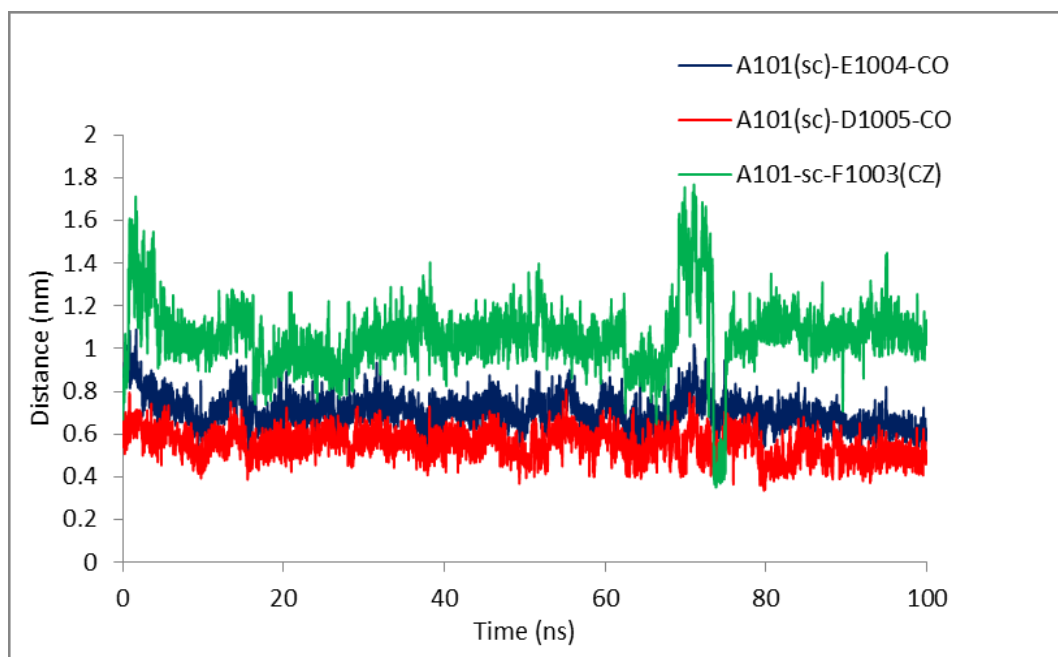
Figure S2.20 The interactions of R101A mutant with the substrate peptide
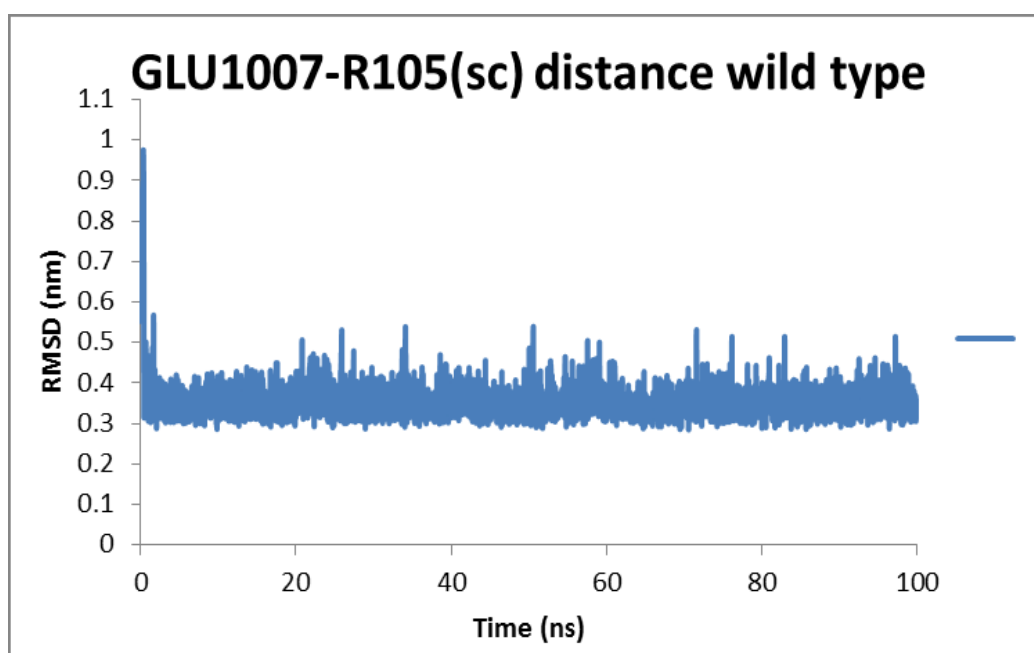
Figure S2.21 R105 residue of wild type TPST-2 interactions with substrate residue
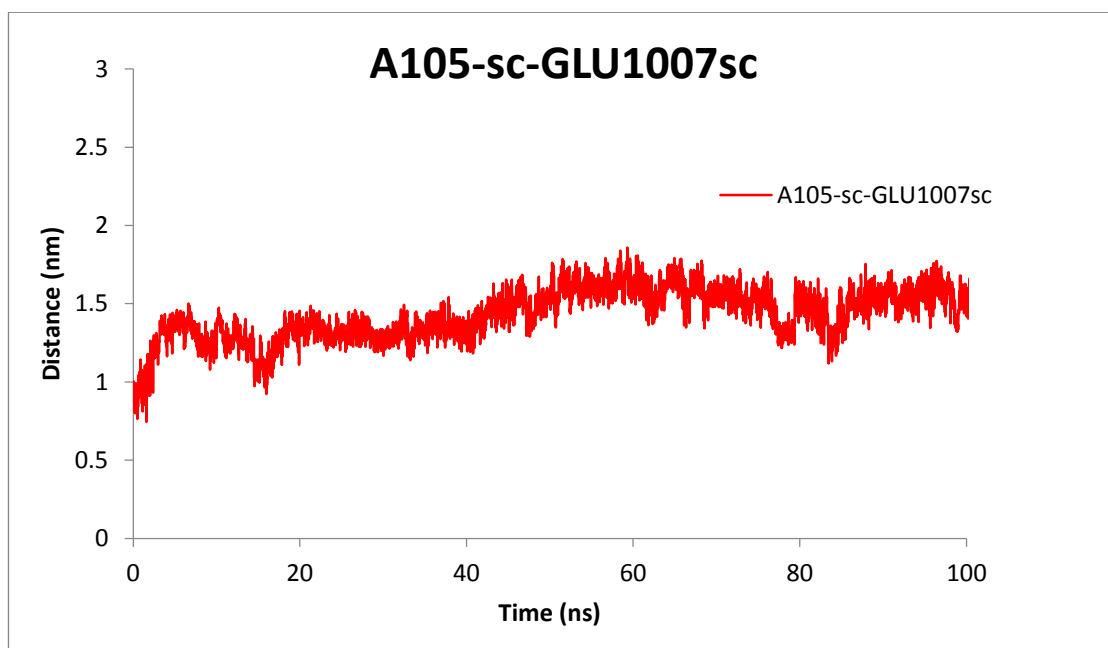
Figure S2.22 The distance between theA105 residue and GLU1007 in R105A mutant
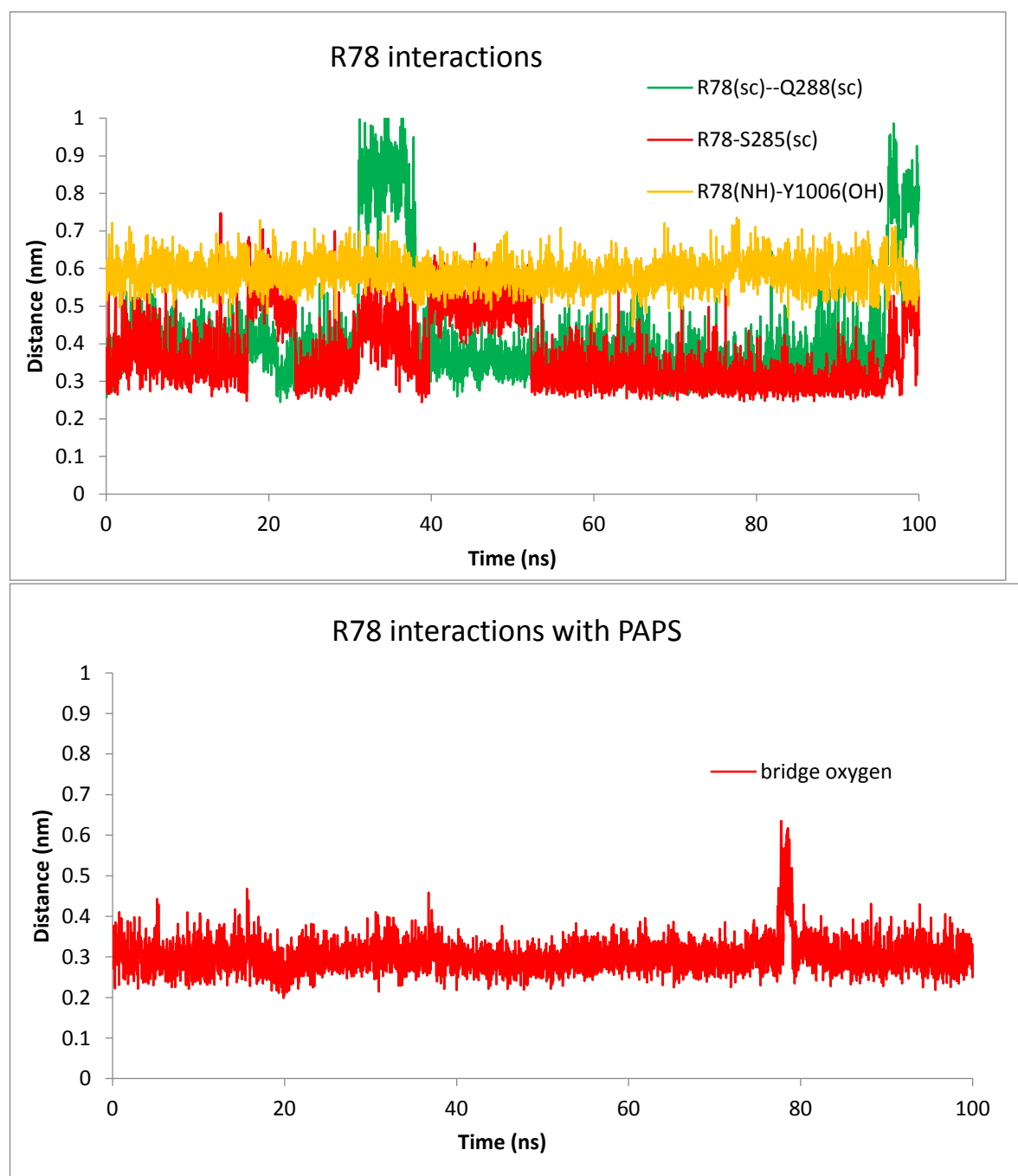
Figure S2.23 R78 residue of wild type TPST-2 interactions with the cofactor PAPS and substrate Y1006 backbone. The R78 of wild type TPST-2 interacting with the bridge oxygen of the cofactor PAPS
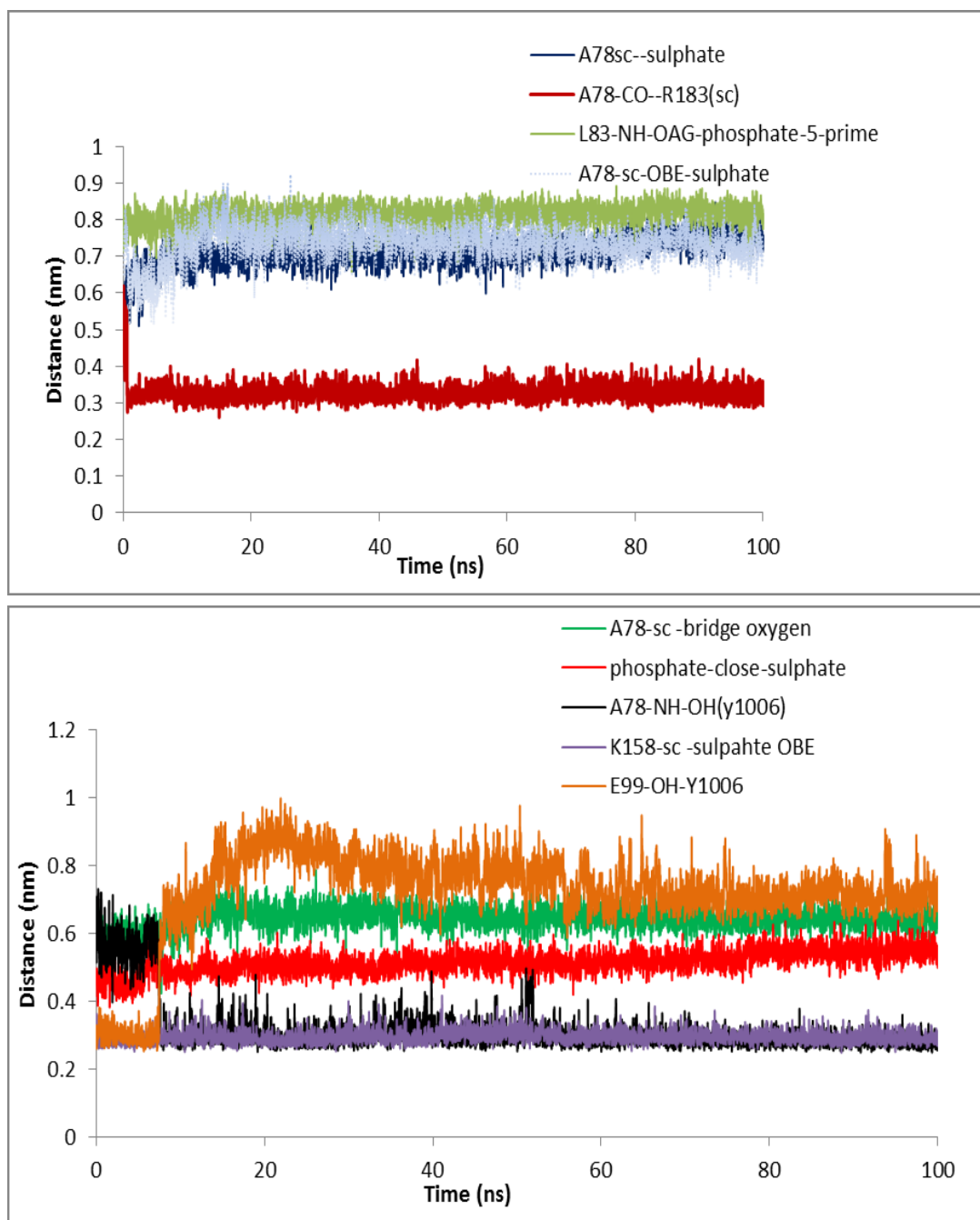
Figure S2.24 The interaction of R78A mutant for 100ns trajectory
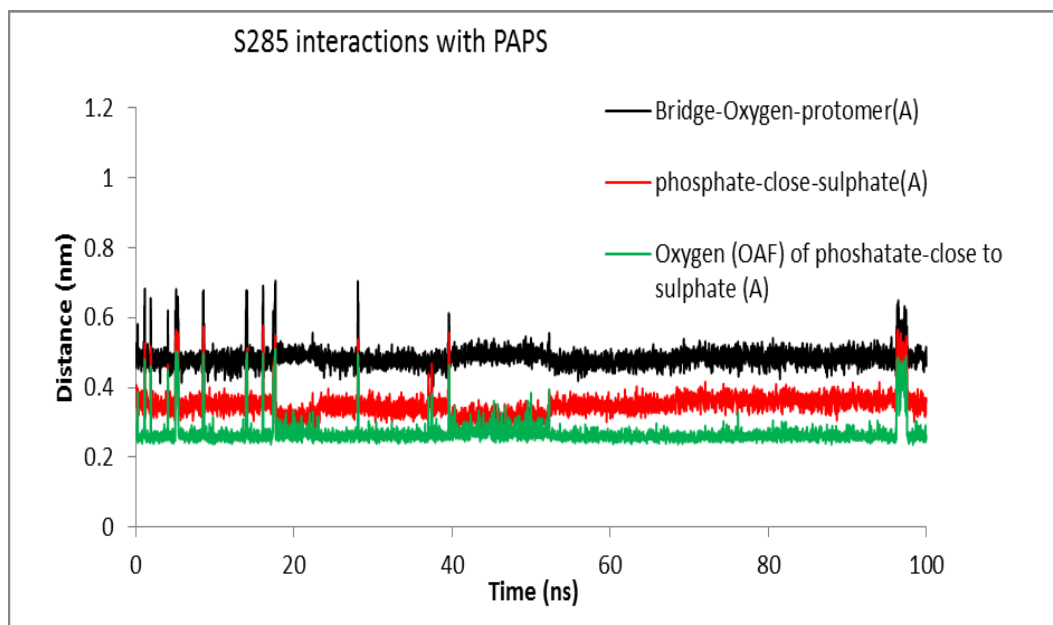
Figure S25 Interactions of S285 side chain in wild type TPST-2
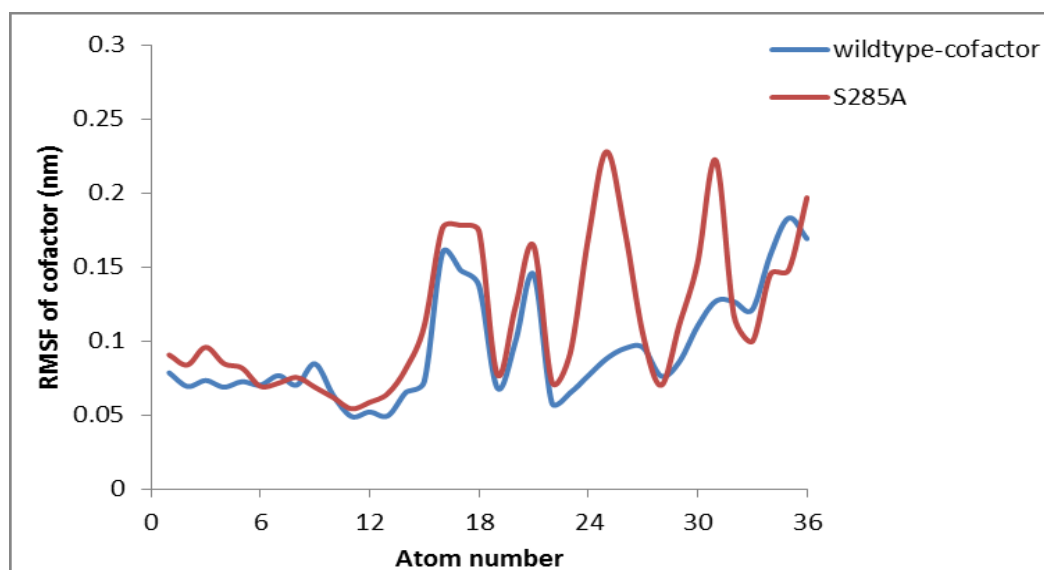
Figure S2.26 The RMSF of the cofactor of the mutant S285A and the wild type TPST-2
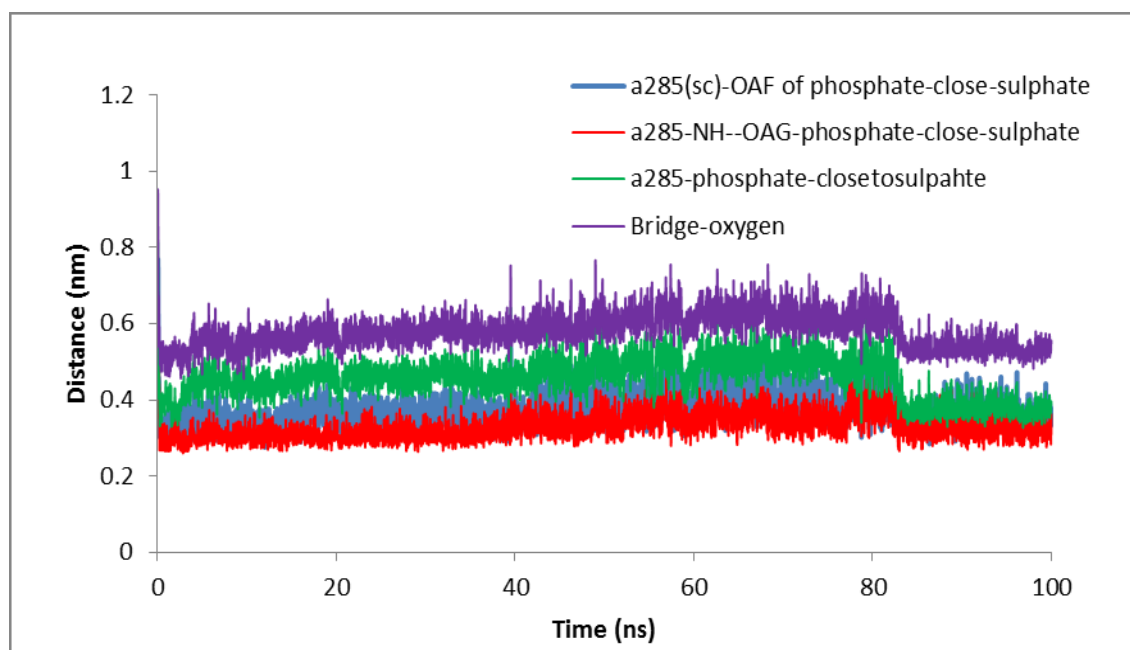
Figure S2.27 The interaction of A285 in S285A mutant for 100 ns trajectory
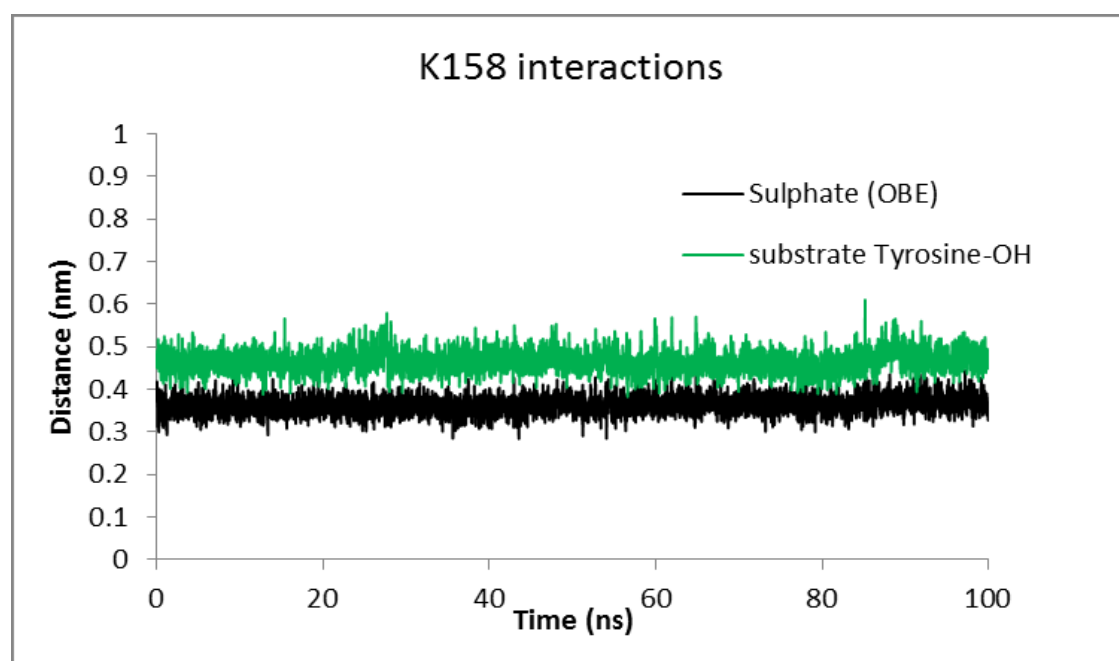
Figure S2.28 The interactions of K158 residue in wild type TPST-2 for 100 ns trajectory
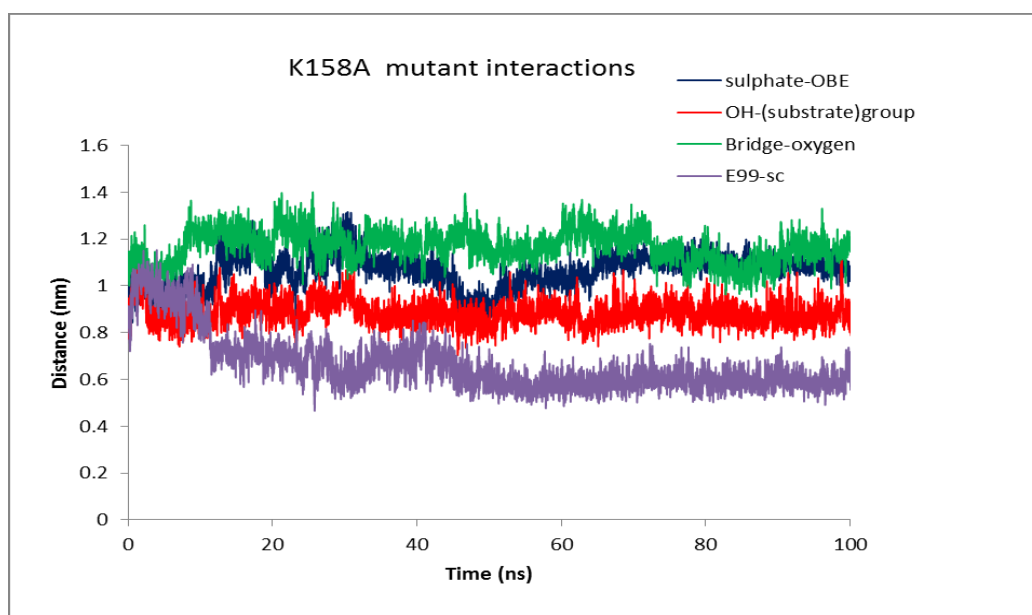
Figure S2.29 The K158A mutant interactions with the substrate and cofactor for 100 ns trajectory
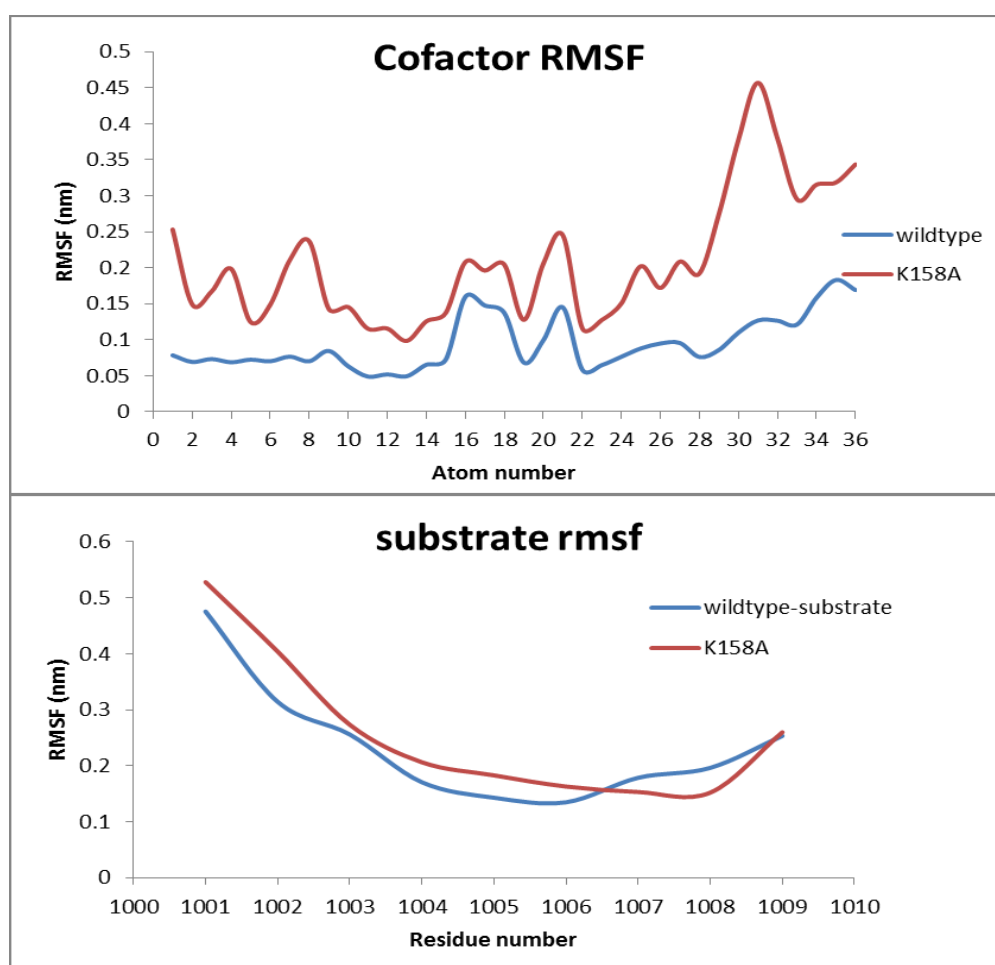
Figure S2.30 The RMSF of the cofactor and substrate of the K158A mutant and wild type TPST-2

74
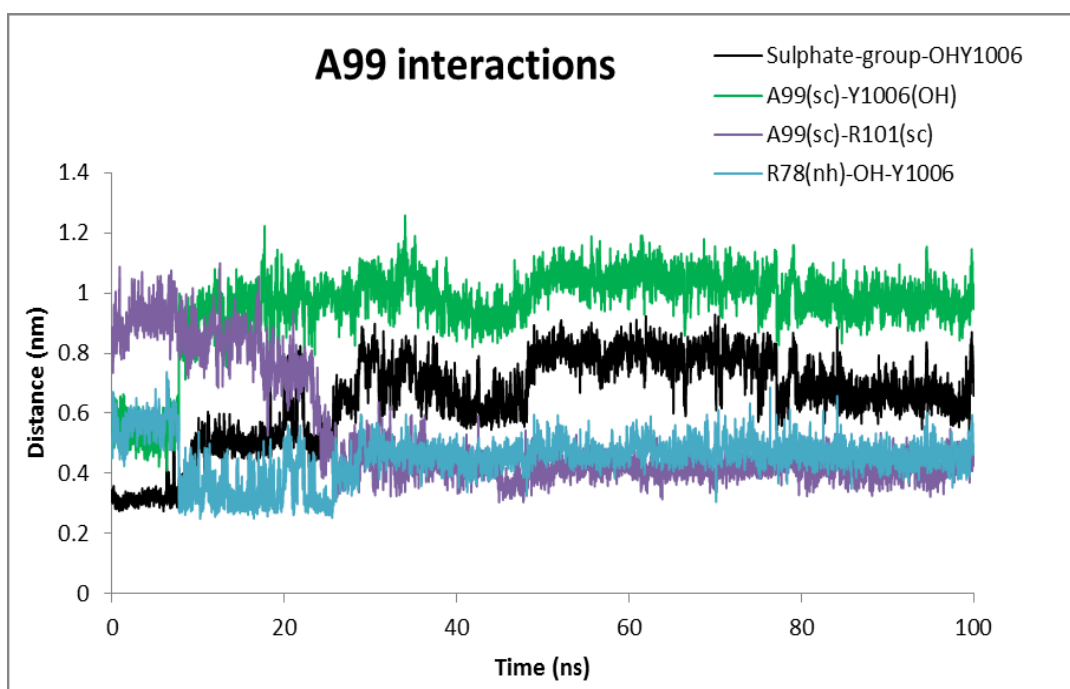
Figure S2.31 The A99 interaction of Mutant E99A for 100 ns trajectory

CHAPTER THREE   DIMERIZATION AND

LIGAND BINDING AND IN TYROSYLPROTEIN

SULFOTRANSFERASE - 2

3.1 INTRODUCTION

Sulfation of tyrosine residues of proteins is a very important post synthetic modification which is still less explored (which we review here). The reaction is catalyzed by enzyme Tyrosylprotein Sulfotransferase (TPST) [14, 71]. Tyrosine sulfation is important for cell signaling and as a target for drug design [74] [76]. Sulfated tyrosine residues are important for the interaction of HIV's gp120 protein to CCR5 in order to get entry to CD4+ T-lymphocytes [77]. Sulfated PSGL-1 is involved in the infection process of Enterovirus71 virus which is responsible for neurological diseases in children [78]. Sulfated tyrosine residues are found in approximately sixty immune system proteins [79] [80]. The sulfation of protein tyrosine residues is carried out as an enzyme catalyzed transfer of negatively charged sulfate group from the sulfate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to the hydroxyl group of a tyrosine leading to a tyrosine O4-sulfate ester and Adenosine 3',5'-diphosphate[14]. Two isoforms TPTST-1, and TPTST-2, exists in humans [81]. TPST has a type II transmembrane topology i.e. a short N terminal cytoplasmic domain, 17 residue transmembrane domain, and a luminal catalytic domain [82]. The enzyme has two N-glycosylation sites, that are four cysteine residues on the luminal oriented side of enzyme.[76] The crystal structure of the catalytic domain of human TPST-2 (denoted as TPST2ΔC18 and containing the sequence from G43 to L359), complex with PAP (3'-phosphoadenosine 5'-phosphate) and the substrate (the peptide C4P5Y3) was solved at a resolution of 1.9 Å [83] (**Figure 3.1**). The substrate peptide contains only one tyrosine residue which is the sulfate acceptor and is surrounded by six acidic residues which provide to the substrate negative charge of minus six [83]. The catalytic domain of

TPTS-2 consists from a α/β motif with a five-stranded parallel β-sheet, surrounded by α helices [83]. The TPST-2 exists as a homodimer, which the catalytically active state of the enzyme where two subunits of the dimer are designated as protomer A and B [83] and both participate in the formation of the binding site and formation of the catalytically active complex. The alpha helices α2-α4 plays an important role in dimer formation in TPST-2 and the residues R118 and R122 harboring α2-α3 also participates in the substrate binding of the other protomer and vice versa (Figure 1). The residue W113 plays key role in the dimerization process as demonstrated by experimental studies of W113A mutant [15]. The crystal structure of human TPST-2[15] provides important knowledge about the overall fold, atomistic interactions and contacts between the both protomers, but shows no information about how conformational flexibility affect the key interactions for the formation of the catalytically active dimer and the structural changes, associated with the binding of the ligands (cofactor and substrate). Molecular Dynamics is one of the most appropriate and broadly implemented method for studying of protein flexibility and dynamics and provide atomistic insights which cannot be obtained experimentally [84] [1] [28] [101] [102] [103]. In recently accepted paper [101] we comprehensively analyzed the conformational flexibility of the wild type full complex TPST2 (WTFC) which includes the enzyme, substrate and cofactor in their catalytically active dimmer form. We studied the mutations of key residues in the substrate binding site and the catalytic site and analyzed the correlated motions in the WTFC.

However, it is essential to understand the atomistic nature of the structural changes which occur upon the binding of each ligand PAPS and C4P5Y3. MD simulations can provide the needed structural insight as a fast alternative of the experimentally determined structures, as far as the only available crystal structure so far is the one of the full complex (WTFC)[15]. In addition it is very important atomistic to understand the conformational effects associated with the formation of the protein dimer, which is the catalytically active

form of TPST-2. Computational analysis of the effect of W113A mutation can provide unique atomistic insights in the dimerization process. As far as there is not available experimentally derived structure of this key mutant, atomistic MD can be reliable and fast alternative to gain the vital missing structural information. In order to complete the missing knowledge, we extended our published study [101] and made simulations to answer of the following questions: i) how the conformational flexibility influences the protein-protein interactions in the dimmer formation; ii) what is the atomistic nature of W113A which is a key mutation preventing the dimerization? iii) how the binding of the cofactor and the substrate influences the structure of TPST2 the Apoenzyme (APO), the enzyme with bound Cofactor (EC), the enzyme with bound Substarte (ES); iii) how the correlated motions in TPST2 change in the different states of the enzyme under the binding of the cofactor and the peptide substrate.
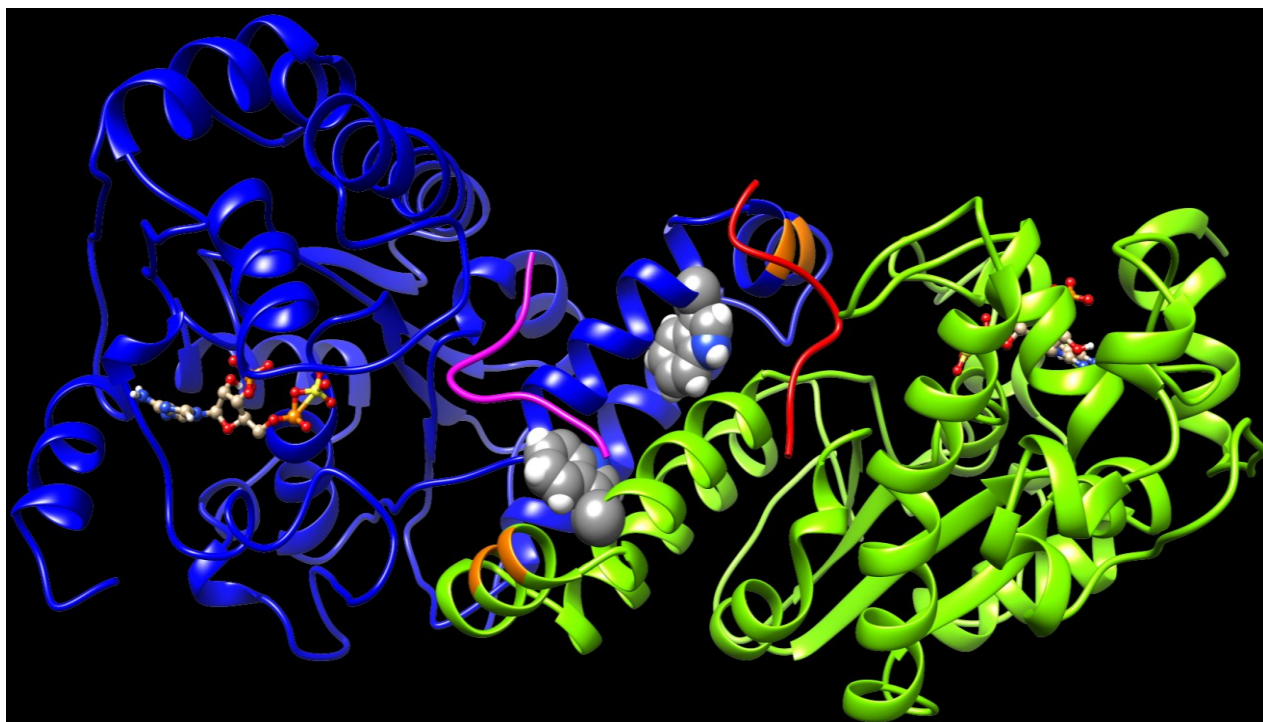
Figure 3.1   The 3D structure of human TPTS-2 (PDB code: 3AP1) enzyme drawn using UCSF Chimera. The cofactor PAPS is represented in ball and stick representation. The

Protomer A and B of homodimer TPST-2 are shown in deep-cued, round ribbon representation in blue and lime green colors respectively. The substrate peptide of protomer A and B are shown in pink and red colors respectively. The residue R118 and R112 of protomer A participating in substrate binding of protomer B are shown in orange color. The W113 residue involved in dimerization is shown in sphere representation on both protomers.

3.2 METHODS AND COMPUTATIONAL DETAILS

Conformational flexibility of the different states of the TPST-2 enzyme (APO, EC, ES and W113A mutant) were explored using atomistic Molecular Dynamics (MD) simulations for 100 ns using Gromacs 4.5.5 package [3] [4] [5] with GROMOS96 43a1 [57] force field (**Table S3.1**). The system setup and preparation was done as described previously [101]. Energy minimization in gas phase was performed by using the steepest descent [91] and conjugate gradient [92] until the maximum force was smaller than 100 $KJ/mol^{-1}/nm^{-1}$. Subsequently a periodic box was defined and periodic boundary conditions were applied. The box size was set to1.0 nm. by using Single Point Charge (SPC) [93] water model was used to solvate the system and Na^+ and Cl^- ions were used to neutralize the total charge (**Table S3.1**). New energy minimization was performed for the whole system using again the steepest descent and the conjugate gradient algorithms. After that position restrain molecular dynamics was carried out for 50 ps in NVT ensemble[94] at constant temperature of 300K with time step of 0.002 ps. The productive 100ns MD simulations were performed in NPT ensemble at constant temperature of 300K with initial velocities taken from Maxwell velocity distribution at 300K and an integration time step of 0.002 ps. The system was kept at 300K by Berendsen temperature coupling and Parrinello-Rahman pressure coupling were with constant (τ_T) of 0.1 ps and 1 bar pressure, time constant (τ_P) of 0.5 ps. The electrostatic interactions were treated by the Particle Mesh Ewald (PME) [95].

The cut off for the Coulomb interactions was 1.0 nm, the Fourier spacing was 0.135 nm. The van der Waals interactions were modelled using Lennard-Jones potential with cut off distance set to 1.4 nm (rvdw) and using switching function. The LINCS algorithm [96] was used to mainatin all the covalent bonds involving hydrogen atoms rigid. The coordinates were saved after every 20 ps.

3.2.1 Analysis of Molecular dynamics simulations

The analyses of the trajectories obtained from the simulations such as packageThe Root Mean Square Deviation (RMSD) of Cα atoms of the protein with respect to minimized crystal structure, Root Mean Square Fluctuations (RMSF), Radius of gyration (Rg), Electrostatic interactions, hydrogen bonds, Solvent Accessible Surface Area (SASA), and cluster analysis were carried out the Gromacs tools. The visualization of molecular structures and trajectories was done using VMD [97]software. The Domain Cross-Correlation Analysis (DCCA) was performed using The Bio3D package [98]. The C_{ij} matrix were used to represent a cross correlation between the ith and jth atoms and ranges from -1 to +1. To create the cross correlation C_{ij} matrix the Cα atoms of the protein, substrate and all atoms of the cofactor were included. The positive values in DCCA map reflect the correlated motions and the negative values the anticorrelated ones.

3.3 RESULTS AND DISCUSSION

3.3.1 Interactions between the Protomers and Dimer Formation

The equilibration of the MD simulations done for the subsequent analysis has been discussed extensively [101]. The TPST-2 enzyme exists as a dimer both in the crystal structure and in vivo [81]. The formation of the substrate binding site in WTFC is result from the dimerization with the participation of α2, α3 and α4 helices. During the simulations the aliphatic parts of the side chains of R110 and R105 make hydrophobic interactions with the aromatic ring of W113 (**Figure S3.1**) with average distance of 3.8 Å

and 4.1 Å respectively for 100 ns trajectory in WTFC, interactions which are result from conformational flexibility and are not seen in the crystal structure. The side chains of M109, L133 and I102 also make hydrophobic interactions with average distance of 3.8 Å, 4.0 Å and 4.4 Å respectively with aromatic ring of W113 (**Figure 3.2**). In the trajectory of the W113A (RMSD **Figure S3.2**), the residues in the vicinity of the W113A mutation site showed significantly higher RMSFs than in the WT FC. The RMSF of mutant W113A also showed significant increase in fluctuations of α2, α3 and α4 (residue 105-145) alpha helices, which are involved in the dimerization of WTFC (**Figure 3.3**). In the W113A mutant the distance between the A113 side chain and side chains of R110, R105, M109 and L133 is higher than in the WTFC during the MD trajectory (**Figure S3.5, Table S3.2**). This difference suggests indication about the roles of these interactions in the formation of the active dimer. The dimerization is very important for the formation of the substrate binding region in WTFC [15]. The RMSFs of the substrate itself and the substrate binding region in the W113A mutant are higher than in the WTFC (**Figure 3.4**). These findings indicates the importance of the dimerization for the proper formation and the stability of the substrate binding site and are in agreement with the experiments showing that the W113A mutant has reduced the enzyme activity 75 %, as compare to the wild type [83].
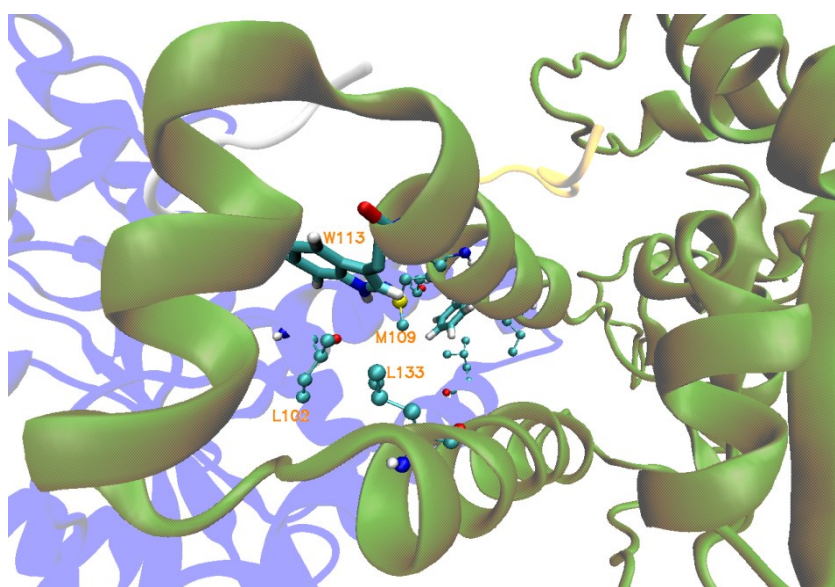
Figure 3.2 The hydrophobic interactions mediated by residue L102, L133 and M109 in wild type TPST-2 obtained from cluster analysis of WTFC for 100 ns trajectory. The Protomer A and B are represented in blue and green colour in new cartoon representation and the W113 is shown in licorice representation using VMD.

Importantly the DCCA reveals that in the mutant W113A there are more anti correlated regions than the WT FC (**Figure 3.5**). In addition in W113A mutant, the region (α2, α3 and α4) show reduced correlation with respect to α2, α3 and α4 of WT FC. These residues are involved in dimerization. Moreover residues 180-210 (3'-Phosphate binding Motif and βe i.e. substrate binding region) of protomer A show anti correlation towards the residues (α2, α3 and α4) of protomer B which explains the role of W113 in the dimerization and is in agreement with the experimental data about this mutant.
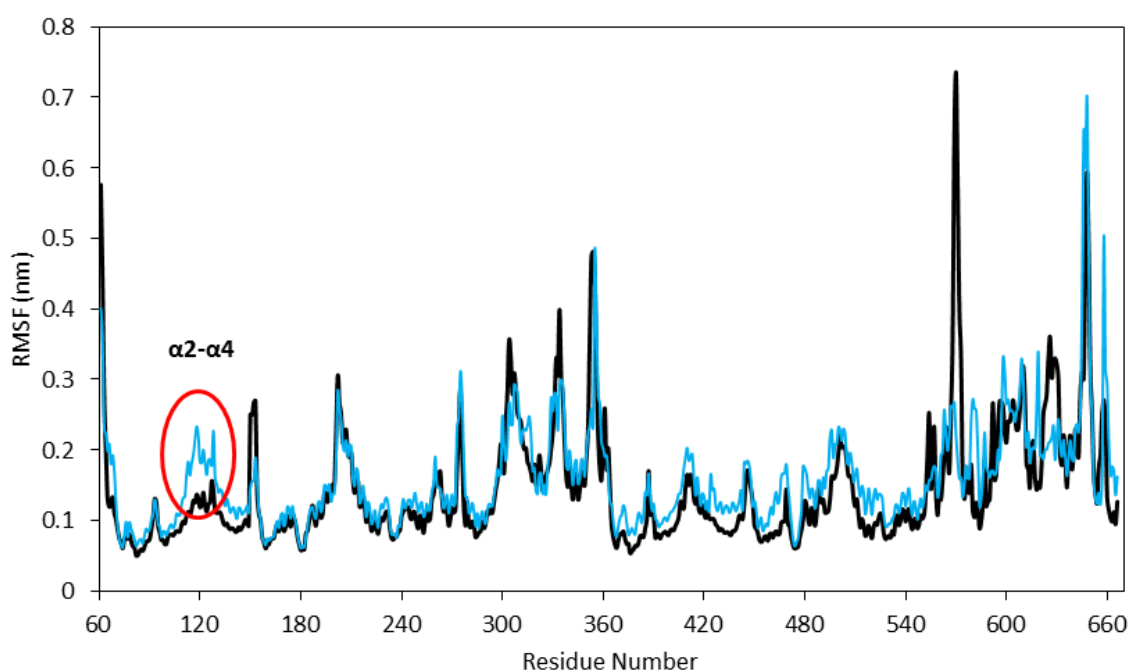
Figure 3.3 The Root Mean Square Fluctuations (RMSF) of mutant W113A (cyan) with respect to wild type TPST-2 (black) for 100 ns trajectory using C α atoms. The region α2-α4 (marked with red circle) involved in dimerization shows increase fluctuation in W113A.
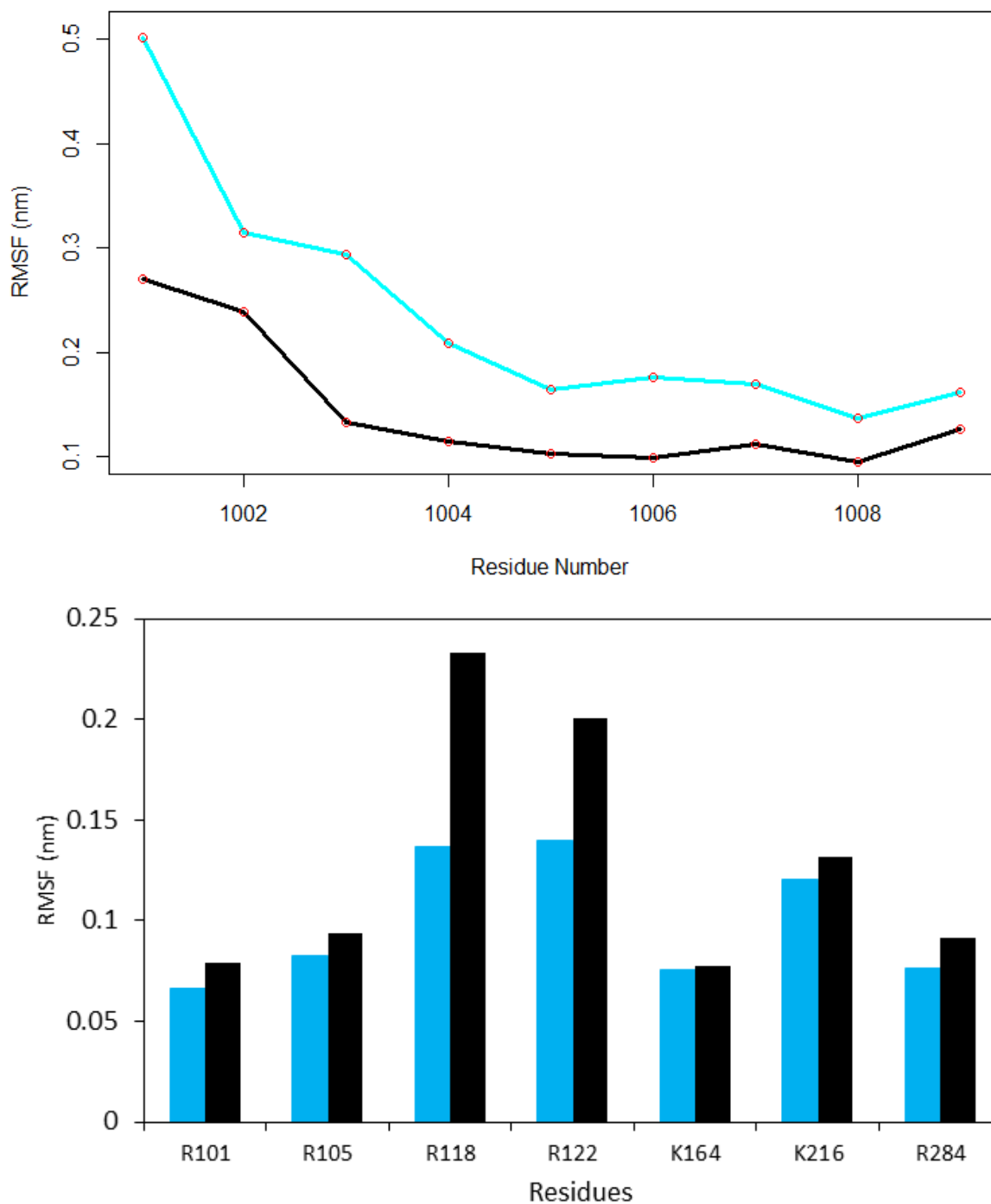
Figure 3.4 The RMSF of substrate peptide and the substrate binding region of wild type TPST-2 in comparison to W113A mutant for 100 ns trajectory using C α atoms. (A) The RMSF of the substrate peptide in WTFC (black) and mutant W113A (cyan). (B) The substrate binding residues comparison in WTFC (black) and mutant W113A (cyan).
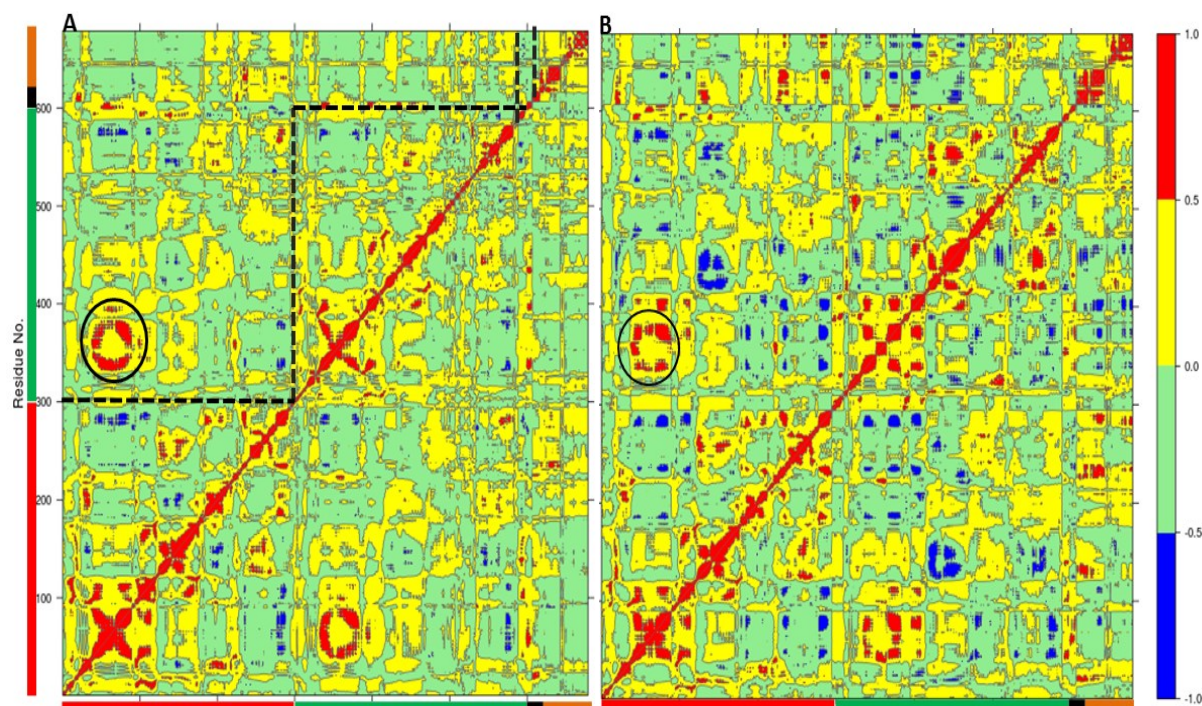
Figure 3.5 The Dynamic Cross Correlation Analysis (DCCA) of wild type TPST-2 and W113A dimerization mutant. (A) DCCA of wild type TPST-2 16, (B) DCCA of W113A for 100 ns trajectory. The α2-α4 are represented by circular region in black colour

3.3.2 Conformational Flexibility and Correlated Motions upon binding of the Cofactor and the Substrate

The MD study of the whole complex, containing the protein dimmer, substrate and cofactor (WTFC) [101] has revealed the key importance of the conformational flexibility which cannot be revealed by experimental methods. However, it is crucially important to understand the atomistic nature of the structural changes resulting from the binding of each ligand – PAPS and C4P5Y3. As far as the only available crystal structure so far is the one of the full complex (WTFC) [15] , MD simulations can provide the needed structural insight as fast alternative to the experimentally determined structures [104] . More importantly, an MD analysis of each of the APO, EC and ES in reference to the WTFC provides unique knowledge about important changes the correlated motions upon the

binding of the substrate and the cofactor which cannot be gained experimentally. In order to provide insight into the conformational changes associated with binding the cofactor PAPS and substrate peptide C4P5Y3, we simulated for 100nsec: i) the apoenzyme (APO) built from protomers A and B; ii) the apoenzyme plus the cofactor PAPS complex (EC); iii) the apoenzyme plus the substrate C4P5Y3complex (ES) and iv) the substrate in water (S) and compared them with the MD simulations of WT FC [15].

3.3.3 Stability and Flexibility of APO, EC, ES and S

The RMSDs of APO, EC, ES, and substrate in water are represented in **Figure S3.5** and **Table 3.1.** APO exhibited an average RMSD of 4.3 Å, EC 4.5Å and the ES complex 3.3 Å. The RMSD of substrate (substrate peptide in solvent) is 3.2 Å as compared to 1.51Å for substrate bound to WT FC indicating the stabilization of the substrate peptide by enzyme molecule (**Figure 3.6**). The binding of the substrate exercises the strongest contribution to the stability of the WT FC. The basal level RMSF of WT FC is 1 Å, 43 % of the RMSFs of the Cα atoms are below 1 Å and 57 % of the RMSFs above 1 Å. All APO, EC and ES trajectories have shown higher number of residues with RMSF > 1 Å (**Table 3.1**) which is largest in APO and EC with 79 % and 75 % respectively. The ES complex showed increased fluctuations in the 5'PBS (5'-Phosphosulfate binding Motif) region which indicates that the binding of PAPS has stabilizing effect on the 5'PBS. The EC shows an increase in fluctuation of R78, S285 and some part of the 3'PBS (3'-Phosphate binding Motif) region thus suggesting that substrate binding has a stabilizing effect for the cofactor and substrate binding sites in WTFC (**Figure S3.6**). The average value of Radius of Gyration (Rg) of WTFC is 28.73 Å , the Rg of EC increased to 30.2 Å and for ES is 28.8 Å (**Table S3.4, Figure S3.7**) demonstrating the effect of the substrate binding on compression of the structure of the full complex. These results show that the binding of the substrate and cofactor both contribute for the stability of the full active complex; however, the effect of the substrate binding is much more prominent. In the EC complex R78 makes

stable interaction with the 5'PBS region of cofactor as it did in WT FC (**Figure S3.8**). The S285 hydroxyl group hydrogen bonded with OAF residue of 5'PBS site during only 54% of the simulation. The hydroxyl group of S285 makes interactions with the OAD oxygen of the 3' phosphate group in contrast to the WT FC where it interacts with the OAF of 5' PBS region. The side chain of R183 in WT FC stabilizes the 3'PBS phosphate OAB oxygen atom (average distance 3.63 Å), however in EC complex it makes weaker interaction (average distance 5.1 Å). The RMSF of PAPS in the EC complex shows higher fluctuations compared to WT FC (**Figure S3.9**) which indicates the stabilizing role of the substrate for PAPS cofactor. The changes in the above atomistic interactions illustrates the key structural changes which occurs upon binding of the substrate and the cofactor and reinforce the idea of accounting for the conformational flexibility whilst modelling protein-ligand interactions (e.g. in the docking studies).
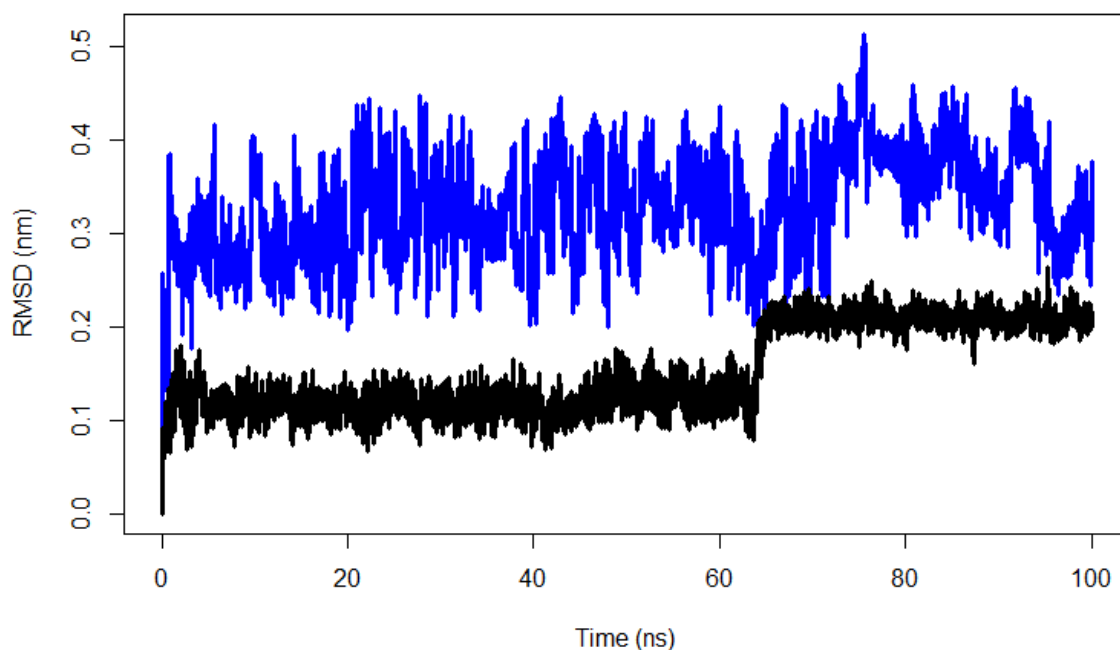
Figure 3.6 The RMSD of substrate peptide of the wild type TPST-2 (black) compared against substrate peptide only in water (blue) for 100ns using C α atoms.

Table 3.1: The RMSD values and RMSF values of different TPST-2 setups

	All C α (Å)		RMSF > 1 Å (%)	RMSF < 1 Å (%)
	Mean (μ)	SD(σ^2)		
WTFC	2.72	0.27	57	43
Apoenzyme	4.37	0.57	79	21
Enzyme + Substrate	3.32	0.35	65	35
Enzyme + cofactor	4.52	0.52	75	25
Protomer A	3.98	0.67	66	34
Protomer B	3.43	0.39	60	40
Substrate	3.28	0.56	100	0

3.3.4 Correlated Motions

An insight into the correlated motions between different parts of the enzyme molecule can be provided by Dynamic Cross Correlation Analysis (DCCA) [100, 105]. The DCCA of APO, ES and EC are shown in (**Figure S3.10–S3.12**). The positive regions in the map are represented with red and yellow colour scheme is associated with correlated motion

whereas anti-correlated motions (negative regions) are shown in light green and blue. In the APO trajectory there are less correlated and more anti correlated motions as compared to WT FC. This result suggests that the binding of the PAPS and C4P5Y3 (substrate peptide) influence not only the local structure and interactions, but also affect more globally the system of protein motions. In the ES complex, the correlated region in (Figure S11, region a) showed fewer residues involved in correlated motion with respect to the WTFC (region a) and there is not strong correlation between the substrate and the enzyme residues as in the WT FC indicating about the apparent role of the cofactor for good stabilization and orientation of the substrate. In the EC, the extent of both positive and negative correlation is also lower than WT FC again demonstrating the effect of the ligand binding more globally on the correlated motions in the molecule. The above differences with respect to the WT FC are another indication that the protein ligand-binding process is dynamical in nature, affects not only the local structure but have more delicate but sensate more global effects. The finding assert about accounting for the conformational flexibility in computational docking studies of protein-ligand interactions.

3.4 CONCLUSIONS

The present MD study reveals the key effects of the protein flexibility on the dimerization of TPST-2 and the dynamic nature of the interactions between the apoenzyme, the cofactor PAPS and the substrate peptide C4P5Y3. New hydrophobic interactions, important for the protomer-protomer interactions are revealed and the structural effect of the mutation W113A for the dimerization is explained in the absence in experimentally-derived structure. The results show that the binding of the substrate and cofactor contribute to the stability of the whole active complex, influence the local interactions in the binding site but also affect pattern of the correlated motions in the entire molecule. The study asserts about

the importance of accounting for conformational dynamics in studying protein-ligand and

protein-protein interactions.

3.5 SUPPORTING INFROMATION

Table S3.2: The MD setup for the wild type TPST-2

S.No	MD-setup	Occurrence	Ions	Timescale (ns)
1	W113A	α2	Na+ (10)	100
2	Apo-enzyme	Protomer A and B	Cl – (10)	100
3	Enzyme-substrate		Na+ (12)	100
4	Enzyme-cofactor		Cl- (2)	100
5	Protomer A		Cl- (5)	100
6	Protomer B		Cl- (5)	100
7	Cofactor			100
8	Substrate		Na+ (6)	100

Table S3.2: The cluster analysis of most populated clusters of the W133A and wild type TPST-2

	Cluster Analysis				Initial crystal structure with hydrogen
	Wild type TPST-2		W133A mutant		Wild type TPST-2
Residue	Cluster 1 (A)	Cluster 2 (A)	Cluster 1 (A)	Cluster 2 (A)	Protomer (A)
W113 (sc) – R105 (sc)	3.83	3.89	9.24	6.49	4.89
W113 (sc) - R110 (sc)	3.63	4.66	5.71	6.26	4.31
W113(sc) - M109 (sc)	3.92	3.32	6.21	6.53	3.95
W113(sc) - I102(sc)	4.27	4.13	3.56	4.36	3.72
W113(sc) - L133(sc)	3.65	3.31	4.41	4.54	3.90

Table S3.3: The RMSD values and RMSF values of different TPST-2 setups

	All C α (Å)		RMSF > 1 Å (%)	RMSF < 1 Å (%)
	Mean (μ)	SD(σ²)		
WTFC	2.72	0.27	57	43
Apoenzyme	4.37	0.57	79	21
Enzyme + Substrate	3.32	0.35	65	35
Enzyme + cofactor	4.52	0.52	75	25
Protomer A	3.98	0.67	66	34
Protomer B	3.43	0.39	60	40
Substrate	3.28	0.56	100	0

Table S3.4: The radius of gyration and solvent accessible surface area of different setups of TPST-2

	Radius of gyration in (Å)		SASA (Å)	
	Mean	sd	Mean	sd
WT	28.73	0.19	2700.14	34.97
Apoenzyme	29.36	0.26	2638.35	44.57
Enzyme + Substrate	28.80	0.20	2663.38	52.21
Enzyme + cofactor	30.23	0.22	2694.05	34.83
Protomer A	18.61	0.21	1385.39	36.37
Protomer B	18.59	0.24	1426.67	31.77
Substrate	6.42	1.13	127.92	9.98

Table S3.5 The average properties of the multiple runs of W113A and important distances

for 100ns

Name	Mean (Å)	S.D (Å)	S.E.M (Å)	A99(sc) – Y1006(sc0	A99(sc) – K158 (sc)	Substrate angle
Run 1	2.42	0.25	0.006	9.5	4.3	103.5°
Run 2	2.99	0.30	0.006	8.2	4.0	98.9°
Averaged all	2.7	0.27	0.006	8.8	4.1	101.2°
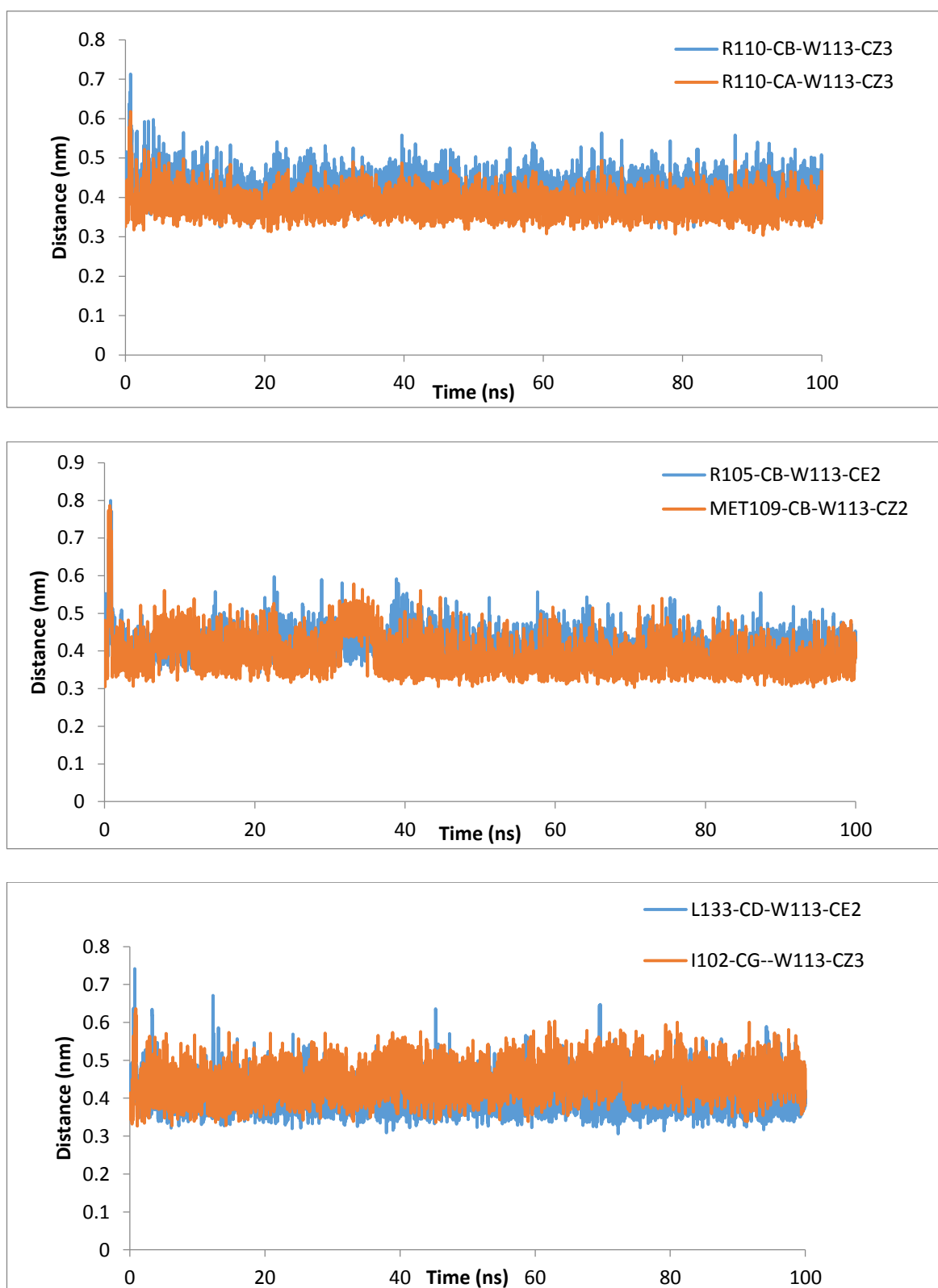
Figure S3.1 The Hydrophobic interactions of residues of wild type TPST-2 with W113 residue for 100 ns trajectory.
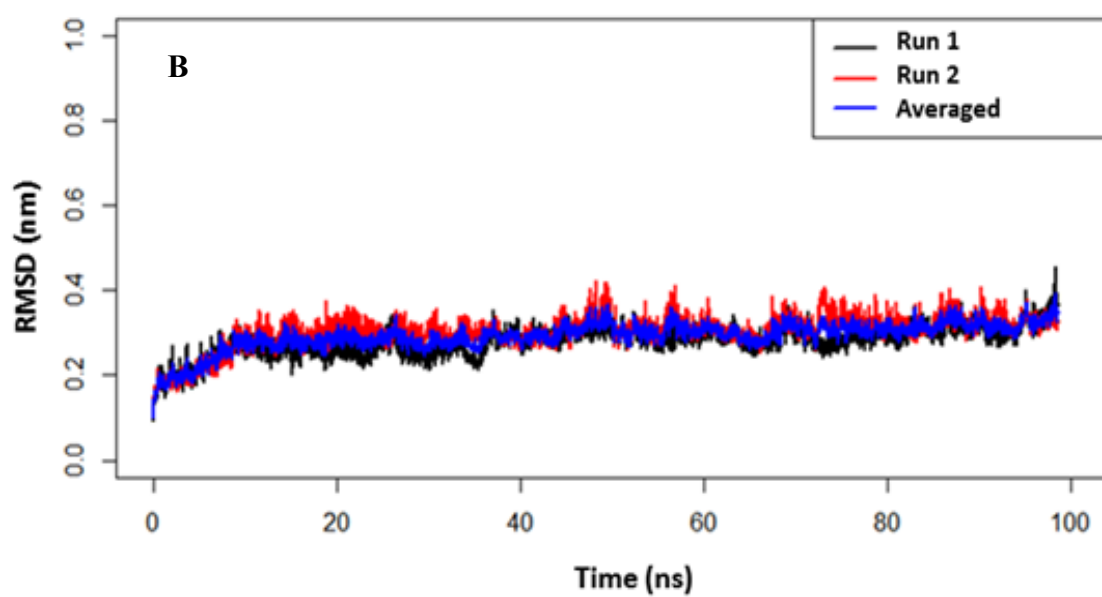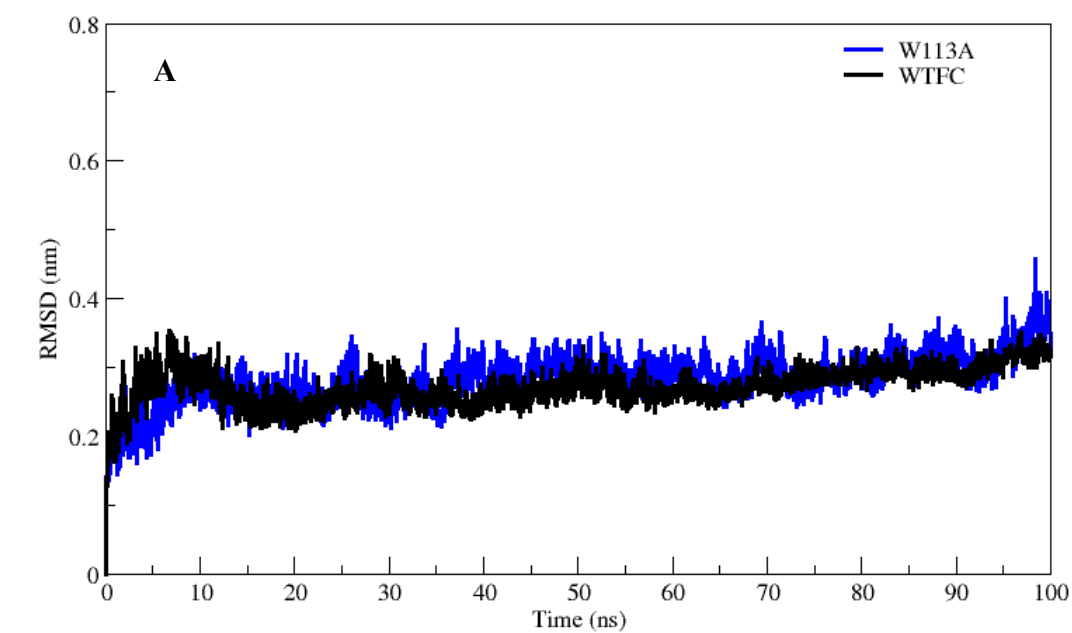
Figure S3.2 The Root Mean Square Deviation (RMSD) of wild type TPST-2 and W113A mutant using C α atoms. (A) The RMSD of WTFC and W113A. (B) The replica runs of the W113A mutant along with average run (100 ns) from two trajectories.
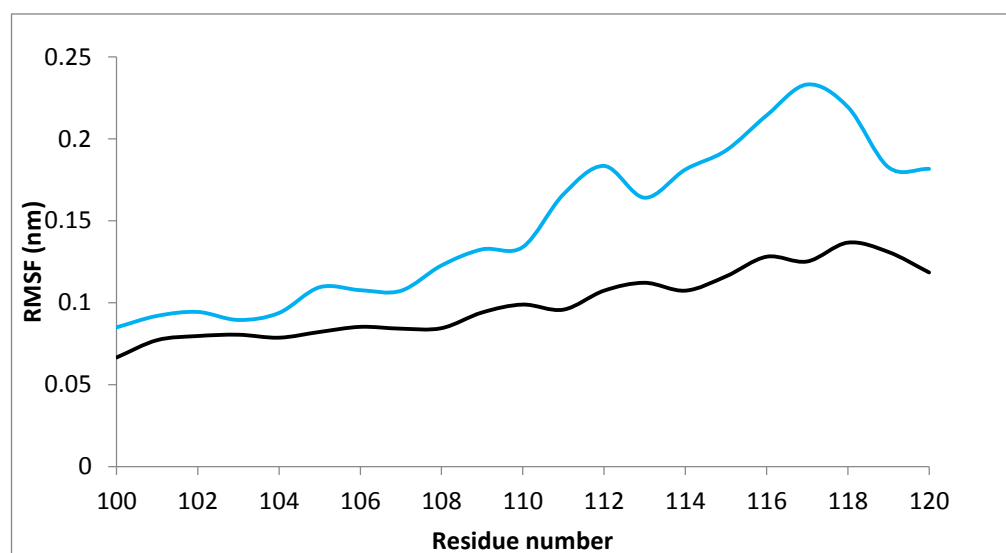
Figure S3.3 RMSF plot of residues in vicinity of W113 residue in wild type TPST-2 (WTFC, black colour) and the mutant W113A (cyan colour) for 100 ns trajectory.
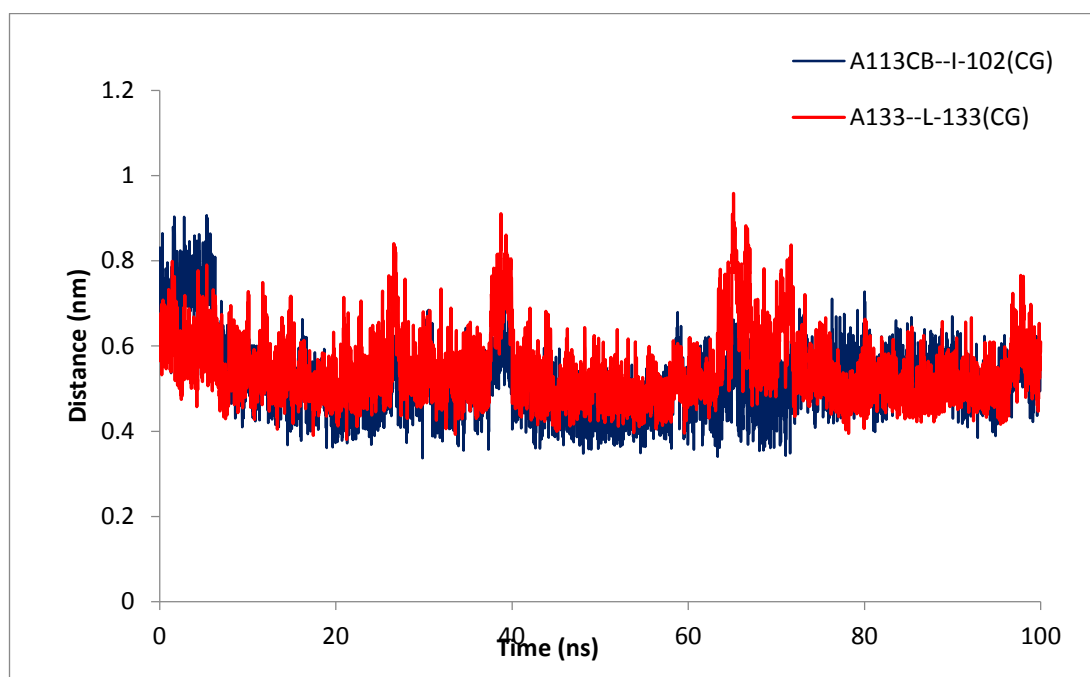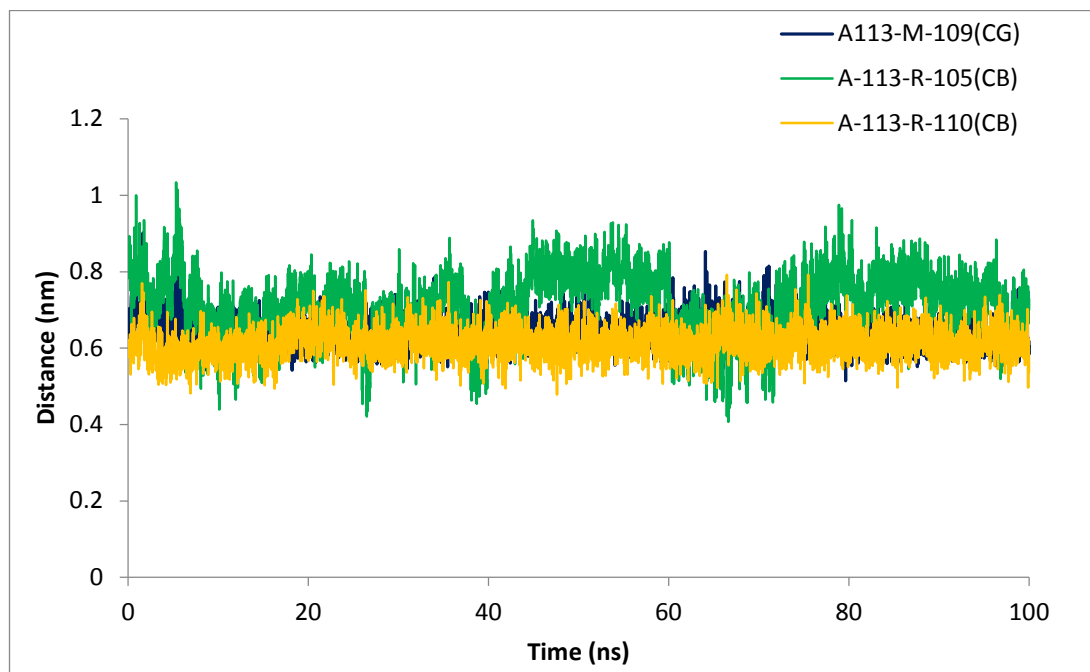
Figure S3.4 The interactions of A113 of W113A mutant for 100 ns trajectory.
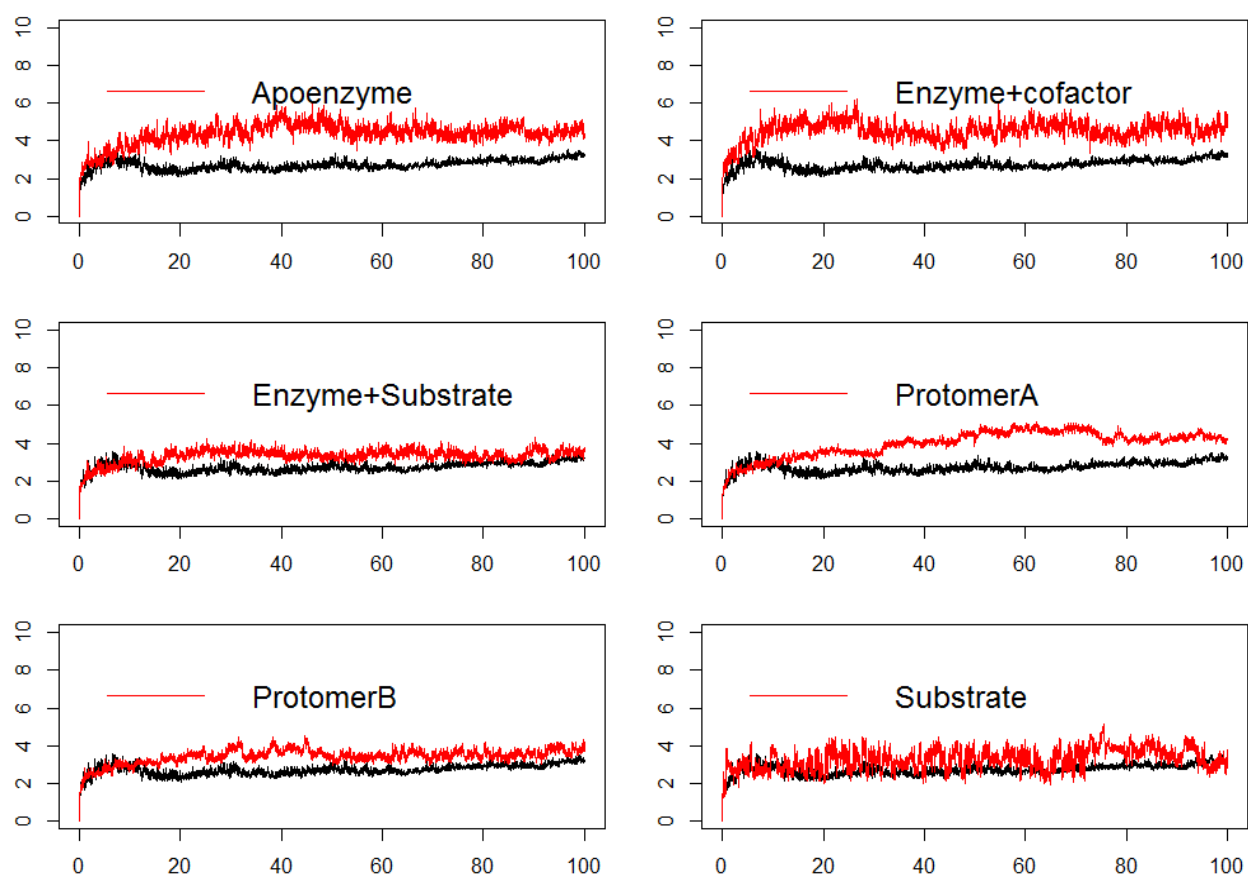
Figure S3.5 The RMSD different setups of wild type TPST-2 using C α atoms.
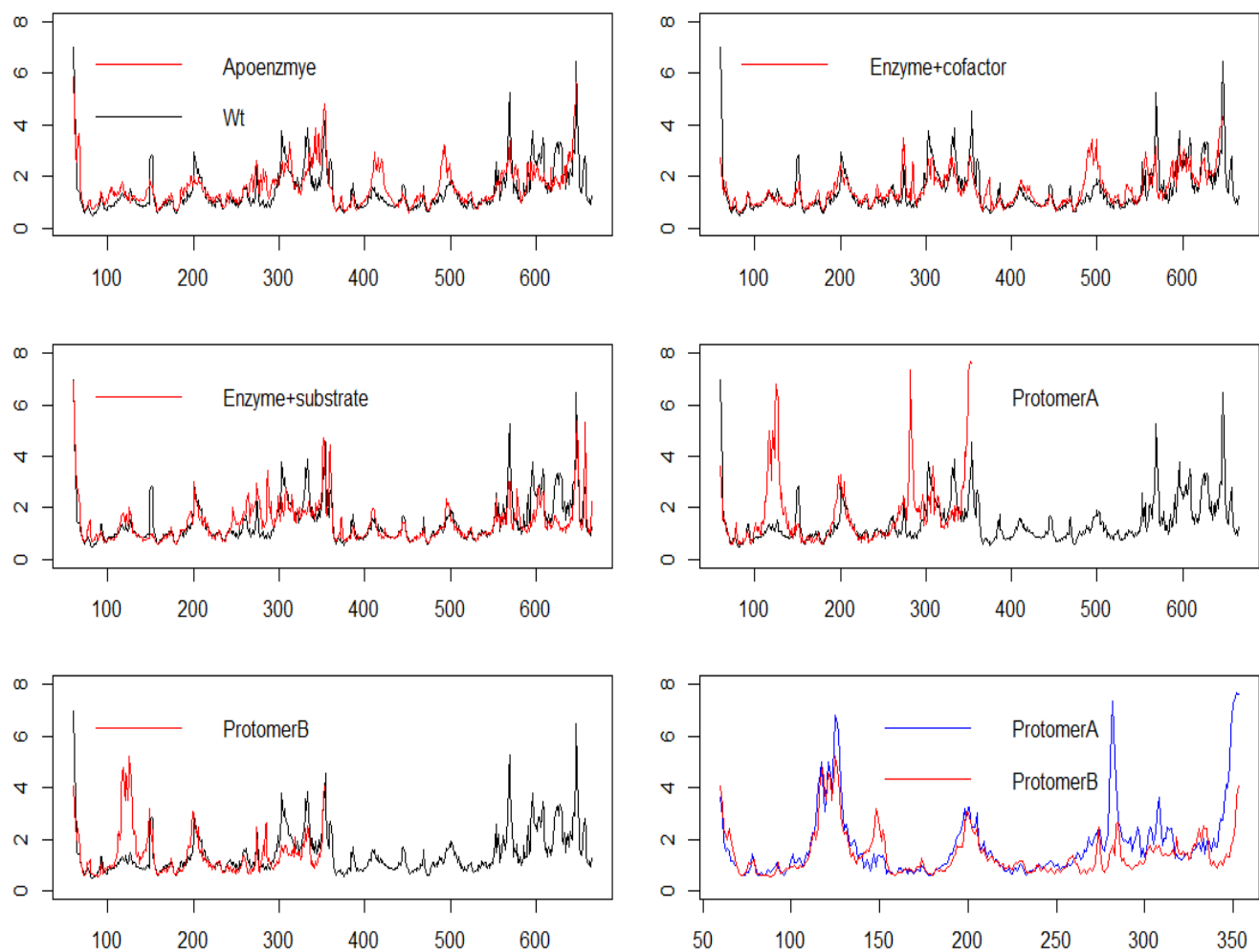
Figure S3.6 The RMSF of the different TPST-2 setups for 100 ns trajectory
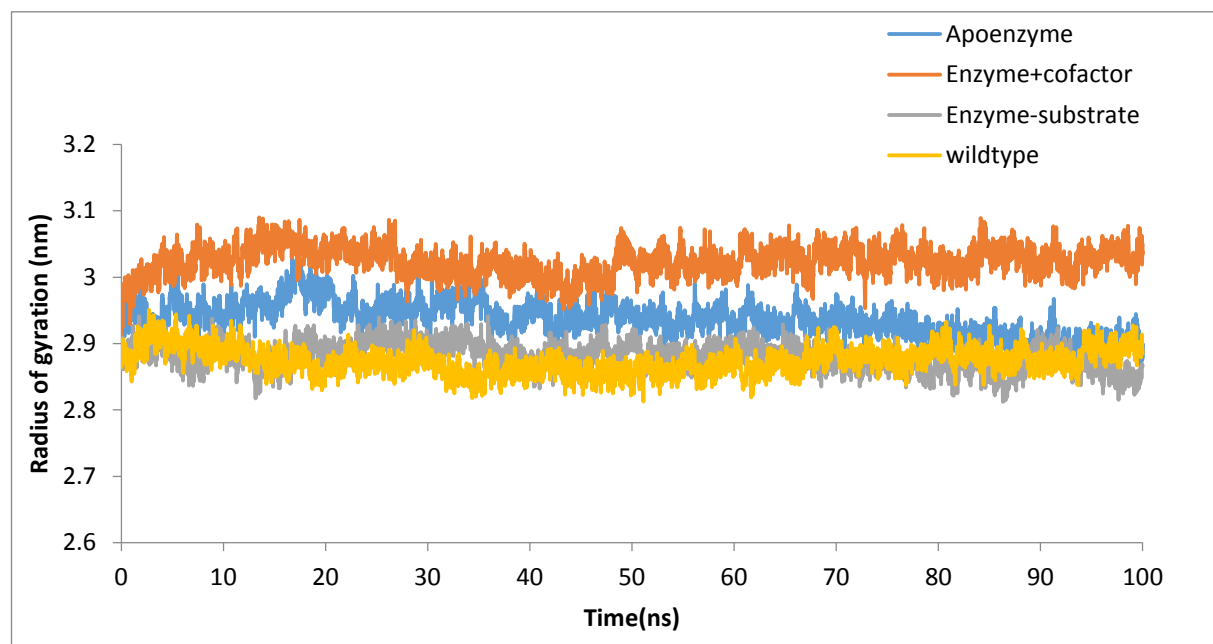
Figure S3.7 The radius of gyration of WTFC, APO, EC and ES for 100 ns trajectory
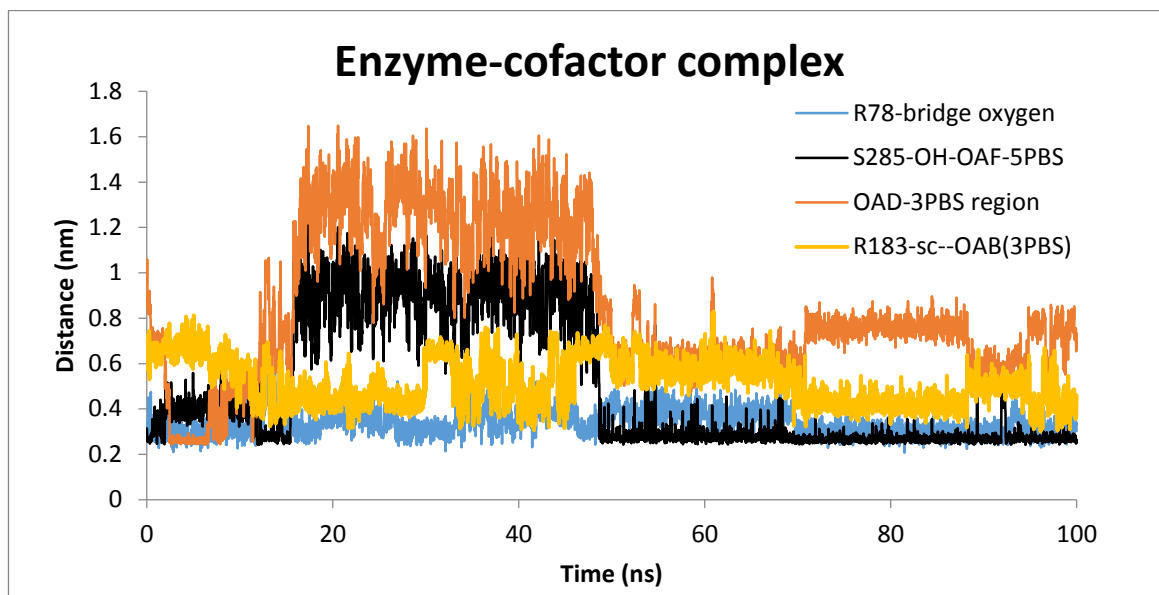
Figure S3.8 The enzyme cofactor setup (EC) and important distances between the protein and cofactor for 100 ns trajectory
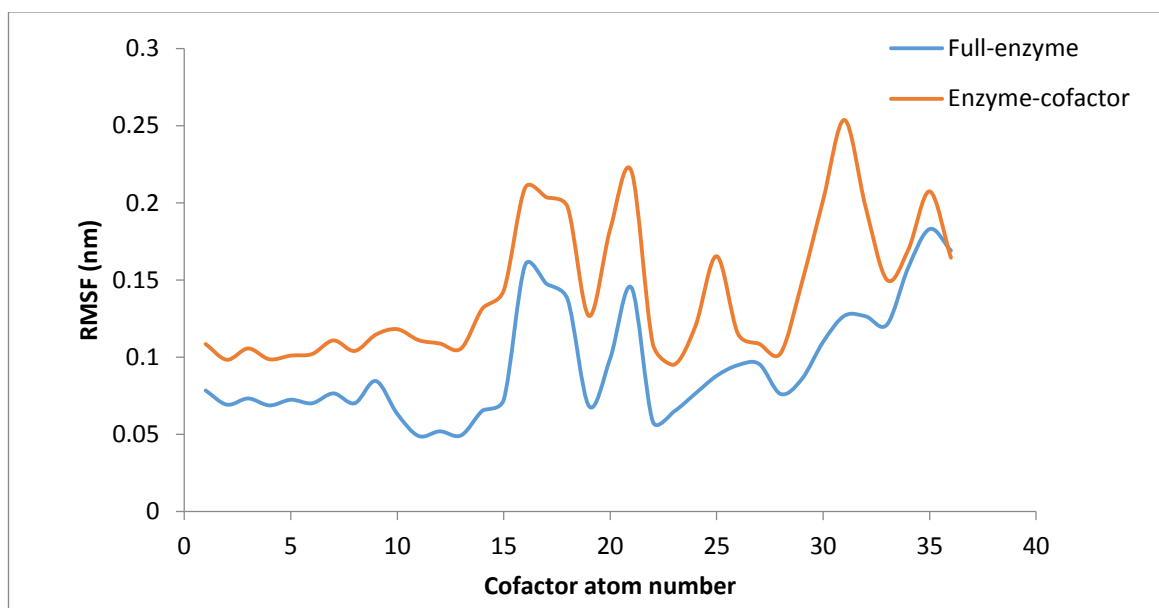
Figure S3.9 The cofactor RMSF in wild type full complex (WT FC) and enzyme-cofactor complex (EC) for 100 ns trajectory
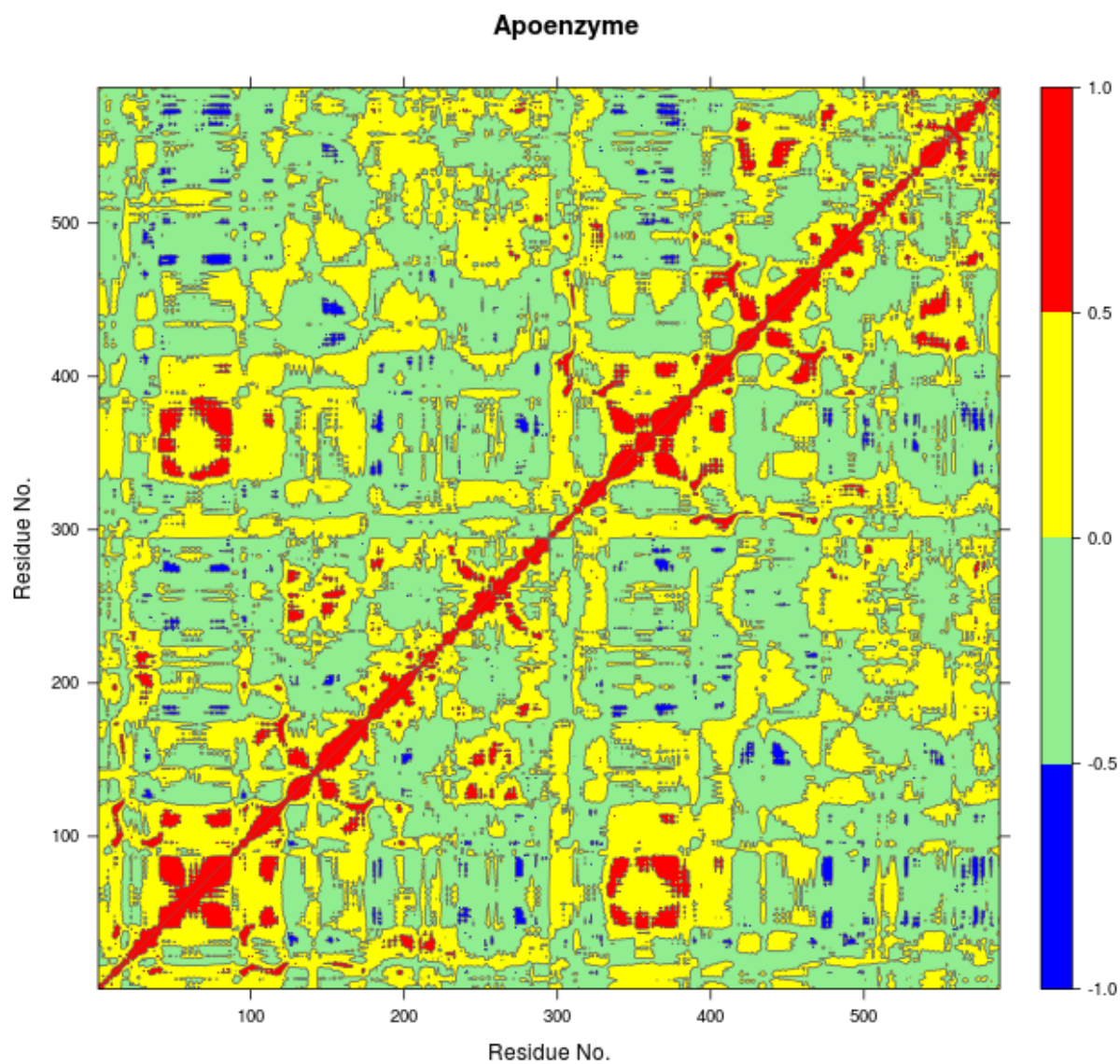
Apoenzyme

Figure S3.10 Dynamic Cross Correlation Analysis of Apoenzyme (Apo) for 100 ns trajectory
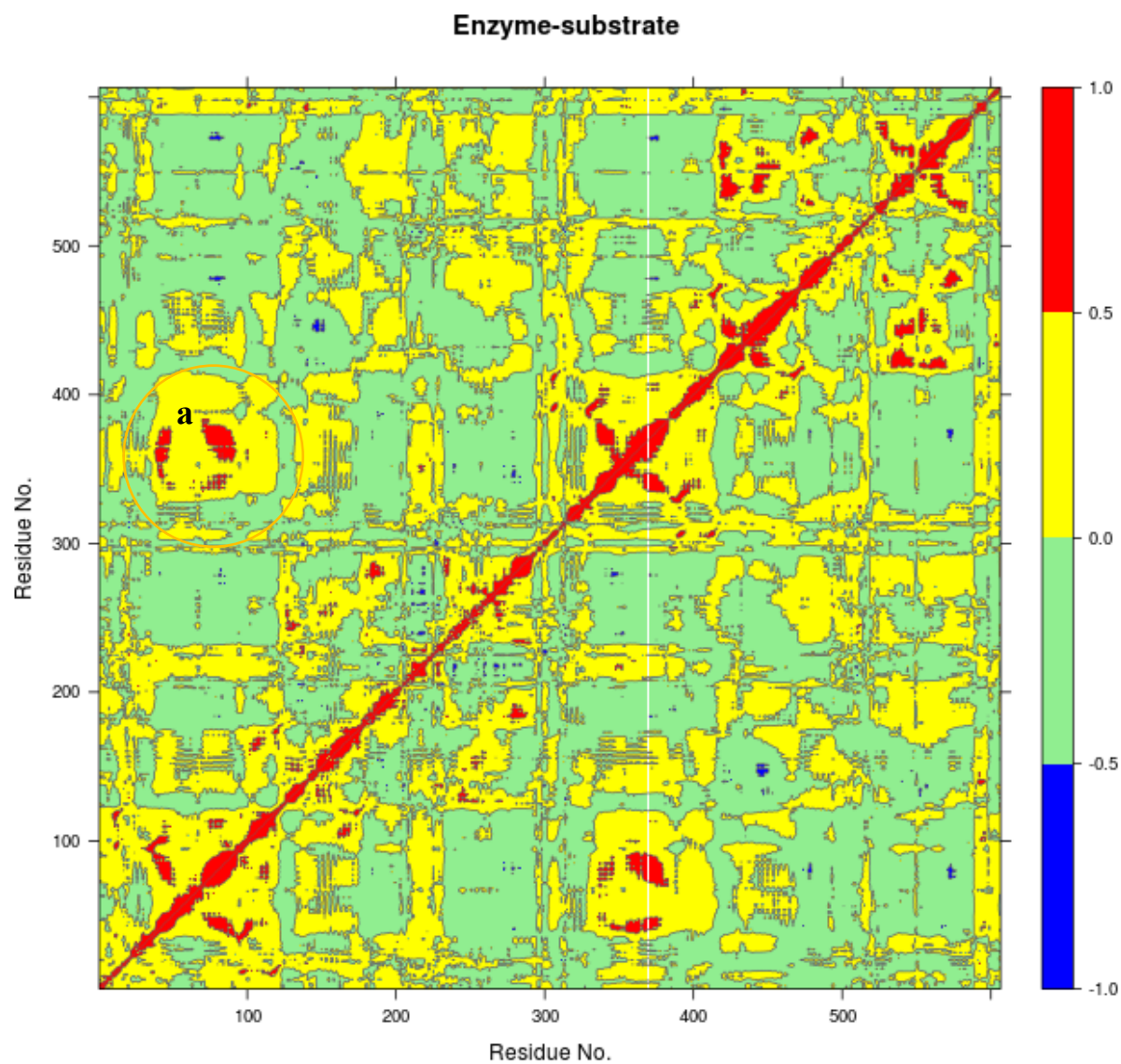
Figure S3.11 Dynamic Cross Correlation Analysis of Enzyme + substrate complex (ES)
for 100 ns trajectory
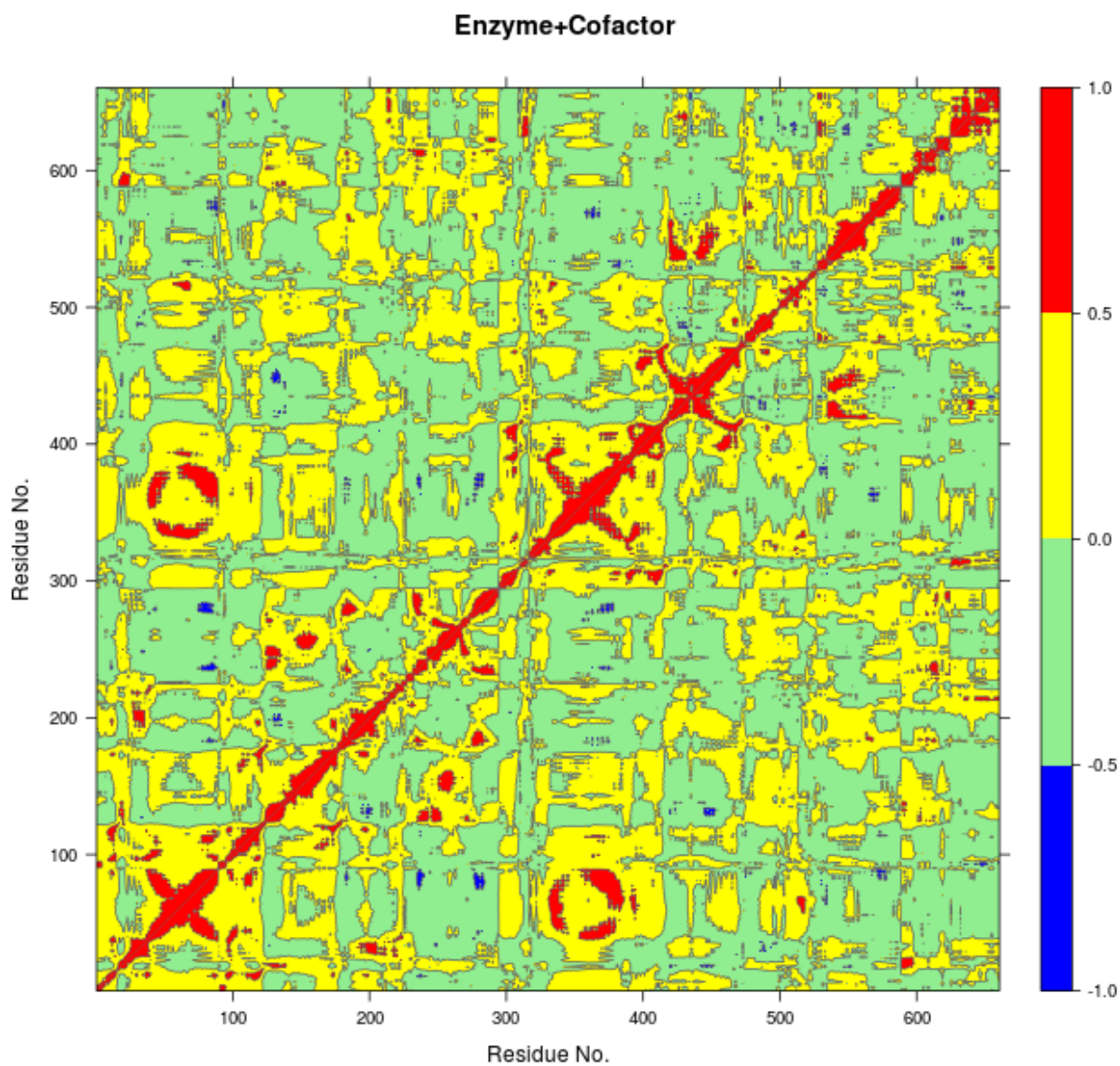
Figure S3.12 Dynamic Cross Correlation Analysis of Enzyme + cofactor (EC) for 100 ns trajectory

# CHAPTER FOUR   NIRE ENZYME AND MUTANTS

4.1 INTRODUCTION

Uroporphyrinogen III (uro'gen III or UP2) act as a common scaffold for the synthesis of diverse tetrapyrroles such as chlorophylls [106], cobalamins, siroheme, phytochromobilin, heme d1, and coenzyme F430 [107] [108]. Heme d1 is an iron-containing dioxo-isobacteriochlorin which acts as a cofactor for cytochrome cd1 nitrite reductase enzyme [109]. Cytochrome cd1 nitrite reductase is the only enzyme in which heme d1 is a cofactor and where it functions as a site for nitrite reduction [110] i.e. reduction of nitrite to nitric oxide and water [111, 112]. This denitrification process is a respiratory mechanism for many bacteria including human pathogen *Pseudomonas aeruginosa*, and thus represents a potential drug target [113] [114]. The direct inhibitors targeting specific enzymatic processes of *P. aeroginonsa* have distinct advantages over general antibiotics; they reduce the selection pressure towards antibiotic resistance in bacterial populations and can be used in combination with antibiotics to increase effectiveness and lower dosage requirements. With the rise of more and more antibiotic resistant bacteria it has become imperative to design new therapies that reduce the likelihood of creating multiple antibiotic resistant bacterial strains [115]. The synthesis of heme d1 proceeds via precorrin-2 which is the product of two methyl group transfer to uro'gen III [116]. The methyl groups are transferred from S-adenosyl-L-methionine (SAM) to uro'gen III by SAM-dependent Uro'gen III MethylTransferase (SUMT) called NirE [109, 116]. The crystal structure *of P. aeruginosa* NirE in complex with its substrate uro'gen III and the reaction by-product SAH (S-Adenosyl-L-homocysteine) was solved recently [117] (**Figure 4.1 and 4.2**). The NirE enzyme is a homodimer; each monomer consists of two domains A and B which are connected by a shorter linker region of four residues length. The active site pocket of NirE

is located between the two domains of each monomer but residues contributed by both protein monomers for each active site [117].

Figure 4.1

Uroporphyrinogen III   S-adenosyl-L-methionine
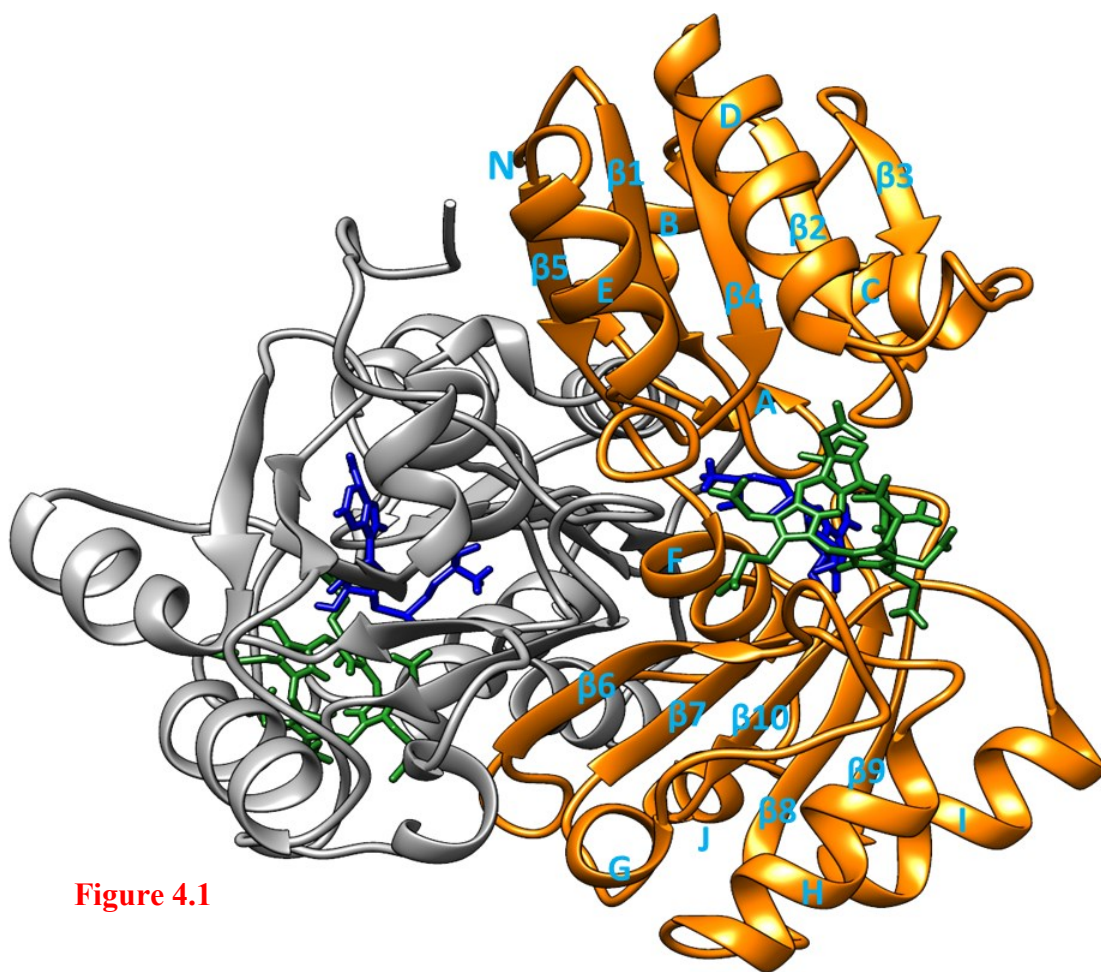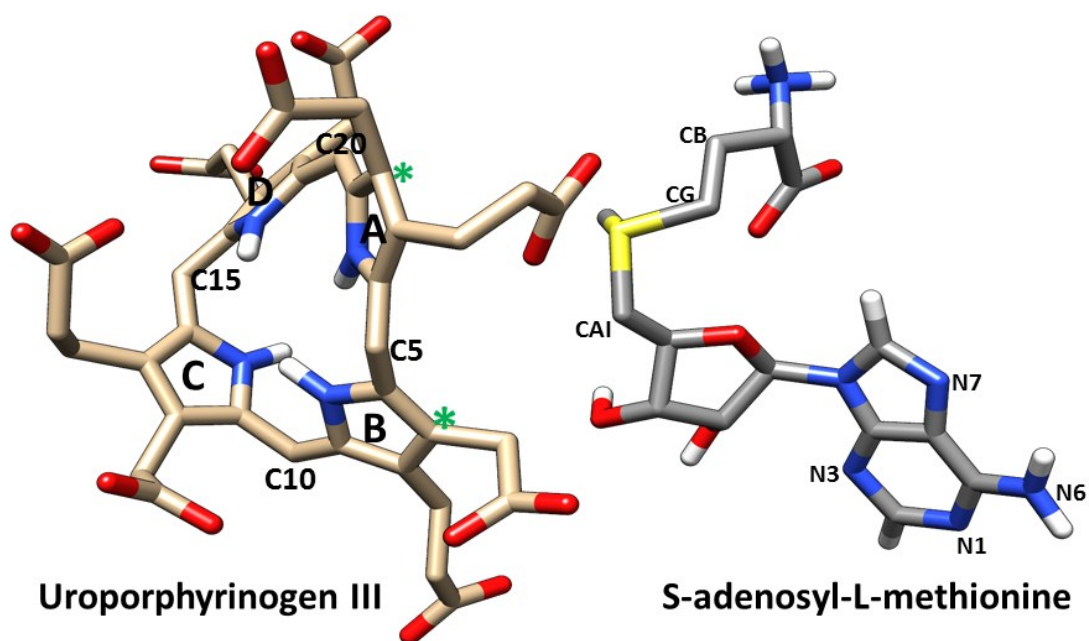
Figure 4.2

Figure 4.1 The 3D X-ray crystal structure (PDB code: 2YBQ)[117] of NirE enzyme complexed with Uro'gen III substrate and SAH cofactor using UCSF Chimera [118]. (A) The homodimer NirE enzyme is represented in monomer A and B in orange and grey new cartoon representation respectively. The spent methyl donor product SAH and substrate Uro'gen III are shown in sticks representation in dark green and blue colours respectively. The loop of monomer B interacts with the active site of monomer A especially with substrate UP2 and vice versa.

Figure 4.2 The substrate Uro'gen III pyrrole ring system from A to D, the C20 carbon atom is a potential proton abstraction site on UP2 and the asterisks represents potential methyl group acceptor sites in the substrate from SAH. The images were drawn using UCSF Chimera [118] in sticks representation.

NirE is subject to substrate inhibition at high concentrations of uro'gen III and product inhibition at high concentrations of SAH [119] [120]. The crystal structure describes the detailed binding of UP2 indicating that it is exposed to solvent and bound loosely in the active site. However SAH binds tightly and is located in the deep interior pocket of the active site [117].

Proteins are large flexible molecules and conformational dynamics is their fundamental property which correlates proteins structure and functions [1]. Although the crystal structure of NirE reveals the important atomistic details of the enzyme, its ligand binding and possible catalytic mechanism, it does not relate how conformational flexibility and dynamics and the effects of mutations might influence key interactions with substrate and cofactor. In order to understand how structural plasticity might influence the structure-function relationships of NirE, we performed 50 ns atomistic (AT) molecular dynamics (MD) simulations on the wild-type full complex NirE, containing the apoenzyme, the

cofactor and the substrate (WT FC), its mutant forms, the apoenzyme (APO), built from monomer A and B, the complex between the enzyme and the cofactor (EC), and the complex with the substrate (ES).

4.2 METHODS

The coordinates of the wild type NirE were obtained from Protein Data Bank (PDB) [85] (PDB ID 2YBQ) [117]. The missing residues were added using PyMOL [121] to the wild type NirE [117]. The methyl group was added to the cofactor molecule SAH using GaussView 5.0 [87]. The resulting SAM molecule was used as a cofactor for the molecular dynamics simulation of NirE. There were nine experimental single amino acid mutants [117] used in this study (Table S1). Mutant models were prepared by using What IF server [88]. The force field parameters for cofactor and the substrate were calculated using PRODRG [89]. The Molecular Dynamics simulations were performed using Gromacs 4.5.5 package [3] [4] with GROMOS96 43a1 force field [57] . The hydrogen atoms were added to the protein molecule by using pdb2gmx utility in Gromacs. The energy minimization in vacuum were performed initially on the protein structure by using steepest descent [91] until the maximum force was smaller than 100 KJ/mol^{-1}/nm^{-2}. The protein molecule was placed in the cubical box with cut-off distance of at least 1.0 nm between the protein and the box boundaries using editconf command. The minimized protein structure was then solvated using Single Point Charge [93] (SPC) water model and periodic boundary conditions were then applied to treat all the parts of the system equally both at its interior and edges. The system was neutralized by adding the Na$^+$ and Cl$^-$ to various mutants and the wild type enzyme. This was followed by energy minimization of the system by using first the steepest descent and then the conjugate gradient algorithm until the maximum force was smaller than 100 KJ/mol^{-1}/nm^{-1}. The energy minimized

structure was then subjected to position restrain dynamics for 50 ps. The simulation was performed in NVT ensemble [94] (constant Number (N) of particles, Volume (V), and Temperature (T)) at constant temperature of 300K with time step of 0.002 ps. The productive 50ns MD was carried out using NPT (constant number of particles (N), system pressure (P) and temperature (T)) ensemble at constant temperature of 300K and the initial velocities for MD simulation were drawn from Maxwell velocity distribution at 300K. The MD was performed with an integration time step of 0.002 ps. The Particle Mesh Ewald (PME) method [95] was used for electrostatic interactions, the van der Waals interactions was treated using Lennard-Jones potential. The cut-off distance for van der Waals interaction was set to be 1.0 nm. The coordinates were saved after every 20 ps from multiple MD trajectories. The analyses of the trajectories obtained from the simulations were performed using tools from the Gromacs software package and VMD [97]. The visualization of MD trajectories and the structures were performed using VMD software [97]. The Bio3D package [98] in R [122] was used to perform dynamic cross correlation Analysis .The positive value represents the correlated motion and the negative values represent the anti-correlated motion. The R [122]and Xmgrace were used to prepare plots and analysed data in this study.

4.3 RESULTS AND DISCUSSION

4.3.1 Overall Stability and Flexibility of NirE Structures

Overall stability of the WTFC structure was assessed by considering the Root Mean Square Deviation (RMSDs) of $C\alpha$ atoms during the MD simulations (**Figure 4.3A**). The system equilibrated after 12ns, multiple runs of the WT FC trajectory showed an average RMSD value of 2.8 Å (**Table S4.2, Figure S4.1**). We considered the relative flexibility of secondary structural elements by class (α helices, β sheets and loops); beta sheets maintained the lowest RMSD value of 1.5 Å along the 50 ns trajectory compared to alpha helix (2.14 Å) and loops (2.9 Å) (**Figure S4.2**). The high RMSD values of the loops region

are common feature of many proteins such as TPST-2 [101] (Tyrosylprotein Sulfotransferase) where loops displayed highest RMSD values as compared to alpha helix and beta sheets. The RMSD profile of individual domains of monomer A indicates that the domain B is more stable compared to domain A; even though both domains have the same secondary structural topology (**Figure S4.3**). The reason for the higher stability of domain B in respect to domain A is that majority of the residues which participates in stabilizing interactions with the substrate and cofactor such as R*149, H161, M186 etc. are located in domain B. The RMSD profile of the linker region (domain A and domain B, here of monomer A) showed stable RMSD; however there was overall increase in the RMSD after 30 ns which could be attributed to the change of Cα angle from 135 ° to 140° between the residues (129, 132 and 134) of the linker region (**Figure S4.4**). The RMSD profile of the mutants equilibrate around 12 ns apart from E114Q and H161F mutants which equilibrated around 18 ns and 10 ns respectively (**Figure 4.3A**). The average RMSD value of all Cα atoms of WT FC is ~ 3.2 Å, whereas in mutants its ranges from 2.4 Å in E114Q and R*149 deprotonated to 3.1 Å in the R111K mutant (**Table S4.3, Figure 4.3A and Figure S4.5**).

The Root Mean Square Fluctuations (RMSF) of WT FC (**Figure 4.3B**) showed that the loop between β3 and the D alpha helix (residues 70-80) has high fluctuations. This loop was missing from the reported crystal structure [117] indicating it to be a particularly flexible component of the protein despite its proximity to substrate UP2 in the crystal structure. The loop was modelled for MD simulations. The short loop region between E alpha helix and β5 (residue 121-125) also display increased flexibility compared to other residues in the RMSF plot. The loop region and G alpha helix encompassing residues 163-179 between β6 and β7 showed high flexibility. The residues forming this loop make important interactions especially with the side chain of the UP2 substrate during the MD simulation. The residues 214-220 that form the loop between β8 and β9 makes interactions with the adenine ring of SAM also revealed greater flexibility than the majority of other

residue in WT FC. The loop between β9 and I alpha helix also showed increased fluctuations in WT FC. The basal level fluctuation of the WTFC was considered to be 1.1 Å, 53 % of RMSFs in WTFC are < 1.1 Å and 47 % of the residue's RMSFs are > 1.1 Å. The mutants RMSF > 1.1 Å range from 29 % (E114Q) to 45 % (R*149 de-protonated) (**Figure S4.6**). The detailed analysis of the residues of the mutants which showed increased or decreased RMSF's relative to WT FC are represented in **Table S4.4**. The substituted residues in the mutants used in this study are all located in close proximity of the substrate UP2 and are mainly involved in substrate binding apart from M186 which is involved in both substrate and cofactor binding. It was evident from the RMSFs of all the mutants in comparison to WT FC that the substrate binding region was clearly affected. In particular there were increased fluctuations of substrate binding residues (160-170), located on the loop between β6 and G helix in every mutant (**Figure S4.7** and **Table S4.4**). This loop showed high flexibility in WT FC as well and was missing from the crystal structure [117] because of its high flexibility. The mutants R51K, R*149K, M186K, G189K and R111K show increased fluctuations of the residues located on the C alpha helix and loop between C alpha helix and β3. This region is also involved in substrate binding in WT FC. The residues (70-80) of the loop located between β3 and D alpha helix showed increased flexibility in all the mutants in comparison to WT FC and these residues were also involved in the substrate binding and were missing in the crystal structure [117] (**Figure S4.8**). The mutations increased the local fluctuations of the substrate binding residues in NirE. Many substrate binding residues (R51, R111, H161, M189, R*149 and G189) in the WT FC were mainly located on the loop region in the enzyme and the methyl donor binding (SAM) residues with the expectation of (M186, D105, I108 and 217) were mainly present on the alpha helix and beta sheets on the protein, further clarifying the relatively high flexibility of the substrate binding residues.
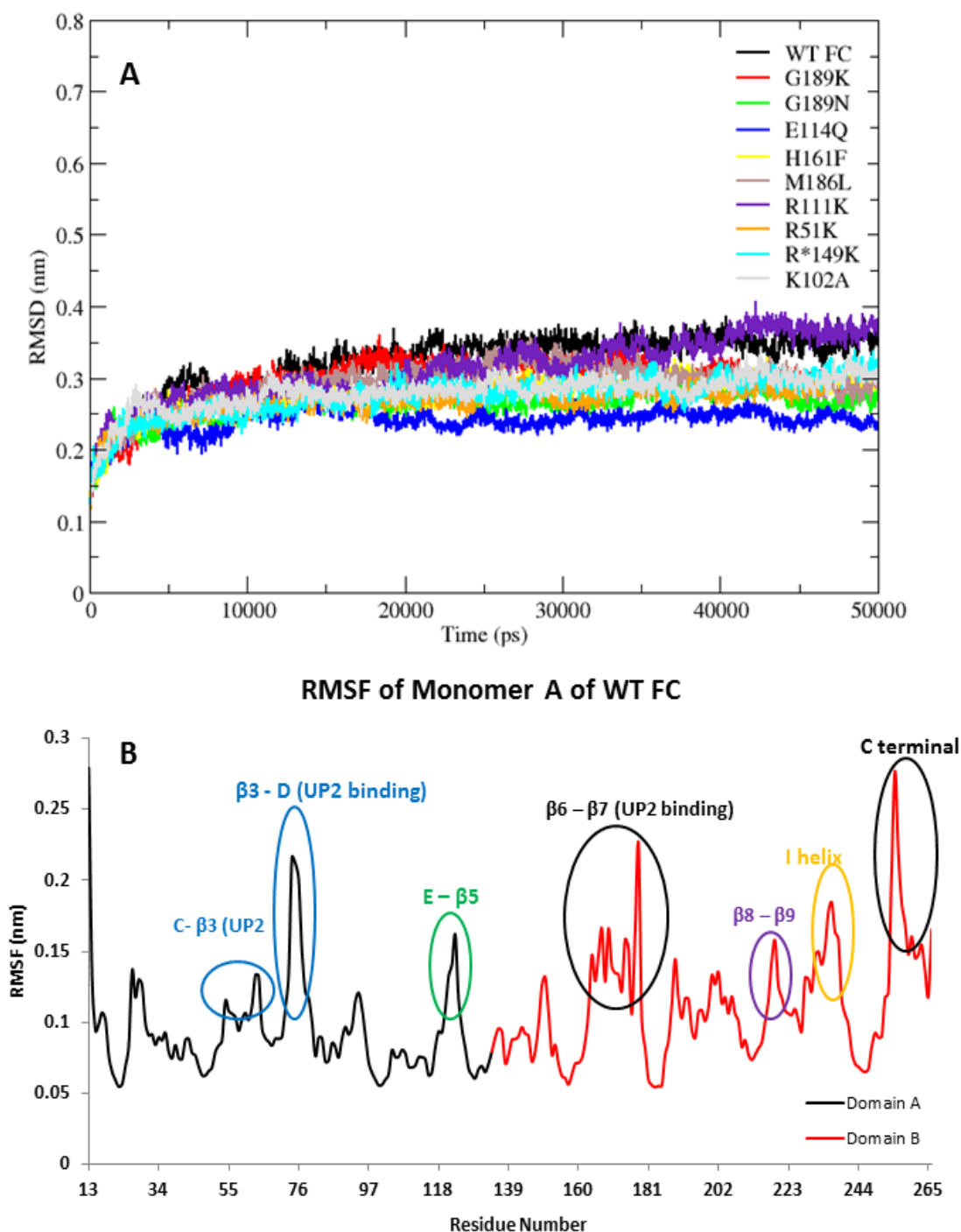
Figure 4.3 (A) The Root Mean Square deviation (RMSD) of all C α atoms of wild type NirE (WTFC) and the mutants for 50 ns trajectory, (B) RMSF of all C α atoms of residues of WT FC for 50ns trajectory. The RMSD image was drawn using Xmgrace plotting tool.

4.3.2 Cofactor Interactions with Substrate and Protein

The time evaluation of interactions between the methyl donor (SAM) and substrate (UP2) were analysed for WTFC. The propionate side chain of ring A of UP2 formed electrostatic interactions with the positively charged sulphur atom of SAM for the entire length of simulation (average distance of 3.1 Å, **Figure S4.9**). The distance between the methyl group on the SAM and the potential methyl acceptor sites on ring A and B of the substrate UP2 (**Figure 4.2**) showed average distances of 6.0 and 6.3 Å respectively. The angle between the sulphur, methyl group and methyl acceptor sites on the substrate were sampled over the 50 ns trajectory to be on average 93.7° (**Figure S4.10**). The addition of the transferred methyl group to the cofactor (SAM) resulted in the appearance of a new hydrophobic cluster which was not present in the crystal structure of NirE [117]. This newly formed hydrophobic cluster consisting of M186, Y185 and F109 residues stabilized the methyl group of the SAM with average distance of 4.3, 4.2 and 4.0 Å respectively (**Figure S4.11**, **Table S4.7**). These interactions likely play a significant role in maintaining the correct orientation of the methyl group in the active site of NirE for the methyl transferase reaction.

We considered that our inclusion of SAM in these simulations would reveal a different pattern of interaction of the methyl donor than those evident with the spent donor SAH in the crystal structure. The hydrogen bonds between the backbones of I108 and D105 with the amino group (NH^{3+}) of the SAM existing in the crystal structure [117] are stable during the MD simulation of WT FC. However the hydrogen bonds between the side chain of D105 and the amino group (NH^{3+}) of SAM were not maintained after 30 ns in the simulation due to an increase in distance (**Figure S4.12**). The crystal structure [117] describes the presence of hydrogen bonds between the carboxylate of SAM with sidechain of T133 and between the backbone of A134 and D105. In this MD study, the carboxylate group of the SAM is hydrogen bonded to the side chain of T133 as described in the crystal

structure [117]; however a new interaction between the side chain of Y185 and carboxylate of SAM, not presented in the crystal structure appeared. The backbone of A134 made weak interactions with the carboxylate of SAM and with the Y185 sidechain. This interaction positions the side chain of Y185 to establish its own interaction with the carboxylate group of the SAM. The backbone of D105 does not interact with the carboxylate group of SAM, in contrast to the SAH-bound crystal structure (**Figure S4.13**). The side chain of Y185 interacts with the dihydroxyoxolan ring of SAM with an average distance of 3.2 Å. In our simulation the carboxylic side chain of the SAM forms an intramolecular interaction with positively charged sulfur atom (**Figure S4.14**). The crystal structure [117] suggests stabilization of the adenine ring by hydrogen bonds with G215 and P27. In our MD simulations neither residue interacts with the adenine ring, instead P27 makes hydrophobic interactions to A134 and L32 residues. In our simulation a completely new set of interactions of residues C138, P242, V212, Q214, Q217 and M186 stabilized the adenine ring by hydrogen bonds (**Figure S4.15** and **S4.16**). L244 (3.3 Å, average distance) also interacts with the adenine ring of SAM (Table S5) (**Figure S4.17**). The side chain of P242, A134 and Y185 made hydrophobic interactions with the adenine ring of the cofactor as in the crystal structure (**Figure S4.18**).

4.3.3 Substrate interactions with Protein

The crystal structure [117] shows that the substrate UP2 adopts a twisted two-up and two-down conformation of the pyrrole rings, however in the MD simulation three out of four pyrrole rings move down and one ring takes an upward orientation. Initially, ring A and C were pointing downwards and ring B and D pointing upwards as described in the crystal structure. However during the course of simulation ring B which was in upward conformation now points downward. The ring D remains in the upward conformation as described in the crystal structure. The dihedral angle for the four rings is plotted in **Figure S4.18** and changes little over time with an average fluctuation of -10°. There are two water

molecules localized in the vicinity of sulfur of the cofactor and the propionate side chain of ring A indicated by two distinct peaks around distance of 3 Å and 4 Å (**Figure S4.19**) which might be related to catalysis. The crystal structure [117] shows that the side chain of R*149 makes polar contacts with the guanidino group of R111, however in MD studies the R111 makes electrostatic interactions with the side chain of acetate group of ring A. The side chain of R*149s drift away on average 8.9 Å from the side chain of the R111 during the MD simulation due to positive charge repulsion of guanidine group (**Figure S4.20**). The R51 side chain also makes very brief contact with the sidechain of the acetate group of ring A of UP2. Interestingly the aliphatic side chain of R51 makes hydrophobic interactions with the propionate aliphatic side chain (ring A) and therefore stabilizes it to make stable interactions with the sulfur of the cofactor SAM (**Figure S4.21**). The crystal structure [117] indicates the hydrogen bond between the backbone of G189 and propionate side chain of ring B which was not maintained in the trajectory (**Figure S4.22**). The side chain of R51 also does not interact with the acetate side chain of ring B of UP2 (**Figure S4.23**). The imidazole ring of H161 in the crystal structure [117] interacts with the pyrrole ring C of UP2 by hydrophobic interactions which is not stable in the MD. The propionate side chain of ring C made only one hydrophobic interaction with the Q163 side chain with an average distance of 3.9 Å. The side chain of Q163 stabilized the propionate side chain of ring D by hydrogen bond with average occupancy of 44 % during simulation (**Table S4.5**) which was not mentioned in the crystal structure (**Figure S4.24**, **Figure S4.25**) and suggests that a possible role of Q163 in substrate binding and stabilization. The new set of hydrophobic interactions involving T159 and L162 sidechain emerged during the simulation to stabilize the propionate side chain of ring D with an average distance of 4.0 and 4.4 Å respectively (**Figure S4.25**, **Table S4.6**). The crystal structure [117] reveals that the side chains of R111 and R*149 make hydrogen bonds with the acetate side chain of the ring D of substrate, while in MD the R*149 they did not make any stable interactions with

any of the side chain of the ring D. In the simulation the backbone of R111 along with G110 made hydrogen bonds with the acetate side chain of ring D. However due to the high flexibility of acetate side chain the interactions of the R111 and G110 to the acetate side chain of ring D, the hydrogen bonds broke after 15 to 20ns (**Figure S4.26**). In the crystal structure [117] M186 with its sulfur makes a hydrogen bond with the NH group of the pyrrole rings A, and C, which was not maintained in the MD trajectory (**Figure S4.27**); however the side chain of the M186 did participate in the hydrophobic interactions with the acetate side chain of ring D.

4.3.4 Conformational Effects of R111 and R*149 Deprotonations

In the crystal structure [117] the arginine residues R111 and R*149 in their deprotonated states, were proposed to be involved in proton abstraction from the C20 (potential proton abstraction site) of the substrate that initiates the methyl transfer reaction and E114, that presented near R111, was also proposed to stabilize R111 for catalysis. In order to explore the effect of deprotonation of R111 and R*149, and their interactions with the potential proton abstraction site of the substrate; we performed MD simulations with each of the two arginines in their deprotonated states and compared the results to the protonated forms. The criteria adopted for R111 or R*149 to act as a catalytic base was based upon its proximity to the C20 site of UP2. In the MD simulation of the WT FC with both R111 and R*149 protonated, these residues are on average 5.8 Å and 7.4 Å away from the C20 site compared to 5.5 Å and 3.8 Å respectively in the minimized crystal structure (**Figure S4.28**). R111 makes stable interaction with the C20 of UP2 whilst the interactions of R*149 are highly flexible in WT FC. R111 is also involved in stabilization of acetate side chain of ring A and D via salt bridges and hydrogen bond respectively, however E114 made no significant interactions with either R111 or R*149 residues in WTFC (Figure S29). In the MD simulation where R111 was deprotonated (R*149 kept protonated), its distance to C20 atom of UP2 was not dramatically affected. The average distance from

deprotonated R111 side chain to C20 is 5.5 Å, compared to 5.8 Å in WT FC (**Figure 4.4 A**), however there was a change in the distance of R*149 (in its protonated state) to C20 to 3.6 Å and this interaction was stable throughout the 50ns trajectory. The side chain of R*149 (in its protonated state) also made hydrogen bonds with the acetate side chain of ring D and with E114 which were not seen in the WT FC **(Figure S4.30)**.

In the MD simulation where R*149 was deprotonated (with R111 in its protonated state) R*149 makes a stable interaction with an average distance of 3.5 Å to C20 of UP2 (**Figure 4.4B**). However the distance between the R111 side chain and the C20 has increased in this simulation to an average of 7 Å as compared to 5.5 Å in R111 deprotonated setup. This increase in distance was due to the fact that R111 residue made stable interactions with the E114 residue (**Figure S4.31**). The mutation of R*149 completely abrogated the enzyme activity indicating an indispensable role in catalysis. However the mutation of R111 reduces the enzyme activity of wild type NirE by ~ 94 %. Based on the simulations with deprotonated forms we suggest that R*149 might act as a potential base for proton abstraction from the C20 of UP2. This conclusion was further tested by simulating a second run of R*149 in its deprotonated state with the average distance R*149 to UP2 C20 being 3.8 Å (**Figure S4.32**, **Table S4.7**).
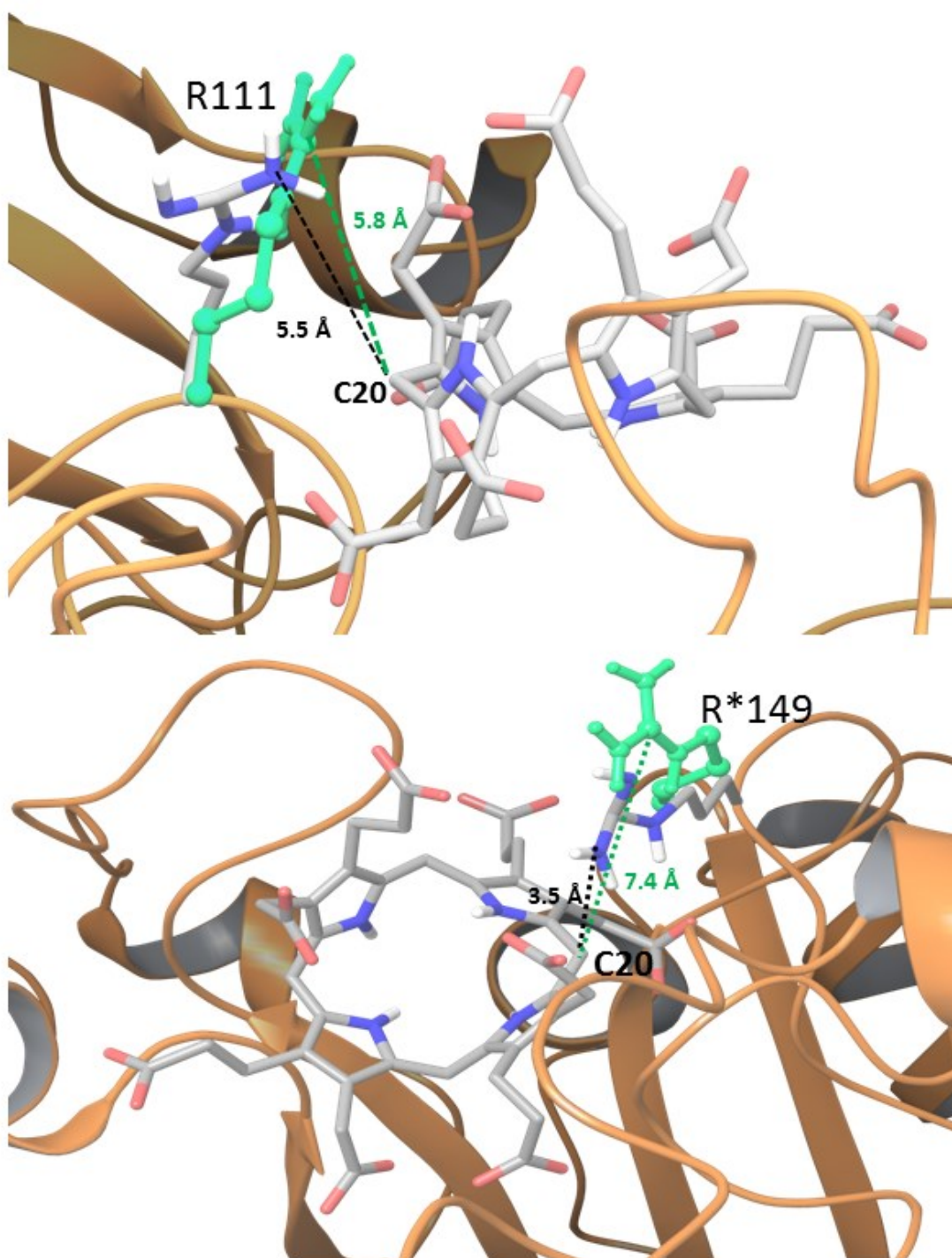
Figure 4.4 (A) The deprotonated R111 in comparison to protonated R111 to C20 of UP2, (B) The deprotonated R*149 in comparison to protonated R*149 to C20 of UP2. The distances shown in the figure are averaged over the entire length of 50 ns trajectory.

4.3.5 Effects of Mutations of Residues involved in Binding and Catalysis

The mutation R51K reduced the enzyme activity by 40 % , SAM binding by 25 % which led to conclusion the R51 was mainly involved in substrate UP2 coordination [117]. In our MD studies, R51 was mainly involved in stabilization of the propionate side chain of ring A by hydrophobic interactions with an average distance of 4.1 Å. In the mutant R51K, the side chain of K51 makes stable electrostatic interactions with the acetate side chain of ring A, however there were no hydrophobic interactions and the average distance moved out to 5.7Å between the side chain of K51 and the propionate side chain of ring A (**Figure S4.33**). Interestingly the conformations of the rings of UP2 are all in upward orientation in contrast to WT FC where three rings were downwards and one ring upwards. The RMSF of the loop region between the β2-C and β4-E showed increased fluctuations in R51K mutant in respect to WT FC. The residues in vicinity of K51 (defined as ±10 residues away) in R51K mutant showed increased fluctuations in comparison to WT FC. Interestingly the residues which are involved in cofactor binding and substrate binding present on the loop between β4-E (103 -115) also showed increased fluctuations, indicating the long range effect of mutation on distant secondary structures in NirE (**Figure S4.34**).

R111K reduces the enzyme activity by 94 % and SAM binding by 25 % with respect to wild type NirE [117]. In the MD studies, the side chain of R111 stabilizes the UP2 ring A acetate side chain by electrostatic interactions. The mutant R111K however made very unstable interactions with an average distance of 5Å between the acetate of ring A of UP2 (**Figure S4.35**). The RMSFs of the residues in the vicinity (defined as ±10 residues away) of K111 in the mutant R111K showed a significant increase in fluctuations. The RMSF of residues present on the C helix and the loop between the C-β3 also showed considerably higher flexibility in comparison to WT FC (**Figure S4.36**).

The conserved E114 residue is proposed to provide a microenvironment in the active site of WTFC to promote the arginine residues to act as bases initiating catalysis. The mutation E114Q resulted in dramatic loss of enzyme activity (reduced by 86 % relative to WT FC), but had no effect on SAM binding [117]. In this mutant there was increase in the RMSDs of the UP2 substrate (3.5 Å) in respect to the WT FC (2.3 Å), validating a role of E114 in substrate stabilization. The RMSF profile of UP2 in E114Q also showed higher fluctuations with respect to WT FC (**Figure S4.37**).

M186 side chain interacts with both SAM and UP2, the mutation M186L affects both the enzyme activity (reduced by 84 %) and SAM binding (reduction by 50 %) indicating an important role of M186 residue in stabilization of both SAM and UP2 [117]. In the mutant M186l, there were no hydrophobic interactions (average distance 8.2 Å) observed between the side chain of L186 and methyl group of SAM. The backbone of L186 also moved away from the N3 of SAM and there was no hydrogen bonds observed due to increased distance of 7.2 Å between two groups (**Figure S4.38**). The L186 side chain no longer stabilizes the UP2 ring A and D with hydrogen bonds. The RMSD profile of UP2 in M186L also shows increased values in respect to WT FC (**Figure S4.39**) indicating the stabilizing role of M186 on the substrate UP2.

The mutations G189N and G189K significantly increased activity by 150 % and 170 % relative to WT FC [117]. In the WT FC, the G189 residue made no interactions with either SAM or UP2. The average distance of G189 from ring B propionate side chain is 7.7Å.However in mutant G189N, N189 residue both side chain and the backbone participates in hydrogen bonds (average distance of 2.3 and 2.1 Å respectively) with the propionate side chain of ring B of UP2, hence stabilizing the substrate more than the WT FC causing an increase in activity (**Figure S4.40**). The stability of UP2 in the G189N mutant was also described by very stable and lower RMSD values in comparisons to

WTFC indicating the correlation of hydrogen bonding and stabilization of the UP2 substrate in the G189N mutant (**Figure S4.41**). The propionate group of ring B is not stabilized at all in WT FC; therefore extra stability of this group could contribute to enhanced enzyme activity. In the G189K mutant, the side chain of K189 made electrostatic interactions with the propionate side chain of ring B with an average distance of 4.0 Å (**Figure S4.42**). The backbone of K189 made hydrogen bonds with the propionate side chain of ring B of UP2. The RMSD analysis of UP2 in mutant G189K also showed more stability with respect to the WT FC (**Figure S4.43**). The electrostatic interactions between K189 and the side chain of ring B (propionate) stabilize to greater extent than in G189N hence the increased activity in G189K in comparison to G189N.

The crystal structure [117] described the involvement of K102 in maintaining the overall architecture of the active site by participating in a hydrogen bonding network with residues E115, G110, R100, Q81 and D50 (**Figure S4.44**, **Figure S4.45**). In WT FC MD simulation, K120 made electrostatic interactions with the side chain of E115 with an average distance of 3.7 Å. The backbone of K120 also made hydrogen bonds with R100. The side chain of R100 and Q81 stabilized the side chain of E115 by hydrogen bonding with an average distance of 3.2 and 2.9 Å respectively. Interestingly, the side chain of Q81 also made hydrogen bonds (average distance 3.4 Å) with the side chain of R111 during MD simulation and this interaction was not noted in the crystal structure [117]. The backbone of E115 made hydrogen bonds with the backbone of R111 residue; the E115 residue is located in vicinity of E114 and R111 residues which were demonstrated to be involved stabilization of UP2 in WTFC, the mutations *E114Q* and *R111K* reduce the enzyme activity.

The side chain of K102 in WT FC also made hydrogen bonds with the backbone of G110 which in turn stabilized the acetate side chain of ring D of UP2 by hydrogen bonding and

also is present in vicinity of R111 residue. The mutation *K102A* completely negates the enzyme activity and reduces the binding of SAH by 75 % in contrast to WTFC (**Figure S4.46**). The A102 in K102 mutant no longer made electrostatic interactions (average distance 8.3Å) with the side chain of E115. This destabilization of side chain of E115 in vicinity of E114 and R111 also caused the Q81 (side chain) not to make hydrogen bonds with the side chain of R111 and backbone of E115 no longer made hydrogen bonds with backbone of R111. This destabilization of R111 side chain resulted in a lack of interactions (average distance 8.1Å) with the acetate side chain of ring A of UP2. The side chain of E114 in K102A mutant made hydrogen bonds with the side chain of Q81 and shifted away from the active site. There were no hydrogen bonds formed between the backbone of G110 and side chain of A102 as the average distance was 12.1 Å. As a result the backbone of G110 no longer stabilized the acetate side chain of ring D of UP2. The active site was dramatically perturbed due to a loss of many interactions in the K102 mutant, even though the K102 residue was not directly involved in binding with either UP2 or SAH. K102 stabilizes the active site in wild type NirE through a long range electrostatic effect.

The H161F mutant is described in the crystal structure [117] to reduce activity by ~ 48 % and has no effect on the SAM binding. The aromatic ring of F161 in H161F mutant made π-π stacking interactions with the ring system of the UP2 (**Figure S4.47, 4.48**). However there were no hydrogen bonds formed by the side chain of F161 as compared to H161 in WT FC with the NH group of the pyrrole ring C of UP2. The local RMSF of residues in the vicinity of the F161 (±10 residues) in H161F also showed significant increase in respect to WT FC (**Figure S4.49**).

4.3.6 Conformational Flexibility and Ligand Binding

The RMSD profile of the APO, ES and EC equilibrated around 12 ns and were stable throughout the simulation (**Figure S4.50**). However the RMSD trajectory of EC complex showed a slight shift in RMSD profile close to 30 ns in comparison to WTFC. It was evident from the simulation that binding of the substrate (UP2) has most stabilizing effect on WTFC. The binding of both cofactor and substrate have resulted in increased structural deviation in WT FC (3.2 Å) in contrast to APO (2.4 Å). The RMSFs of EC complex show high fluctuation relative to APO, ES and WTFC (**Figure S4.51**, **Table S4.4**). The loop between β3 and the D alpha helix (residues 70-80) and residue 163-179 between β6 and β7 which are mainly involved in stabilizing UP2 substrate including R51 and R111 in WTFC showed significantly high fluctuations in EC. The data clearly indicates that the residue in vicinity of substrate UP2 have showed high fluctuations in absence of substrate in EC complex and indicate there importance in substrate stabilization. The normalized distribution of center of mass distance between domain A and B in WTFC (monomer A) (**Figure S4.52**) showed an average increase of 29.7 Å with respect to APO (27.5), ES (28.1) and EC (27.7). This data clearly indicates that the two domains open up or extended conformation during the simulation possibly to accommodate cofactor (SAH) and large substrate UP2. The binding of cofactor and substrate also increased the solvent accessible surface area of WTFC in contrast to APO enzyme (**Figure S4.53**) which also points out towards more opened conformation of enzyme when the substrate and cofactor binds to it.

Dynamic Cross Correlated Motions

The cross correlation analysis allows the exploration between correlated and anti-correlated motions in proteins [123]. The N terminal region (residues 13-22) of the WT FC (**Figure S4.54**) showed correlated motion against the β4 and the loop between β4 and D α-helix (90-102 residues). The β4 sheet which is in close proximity to the cofactor SAM also

showed correlated motion towards the residues 110 to 115 which belong to a loop between β4 and E α-helix and participles extensively in substrate binding in WT FC. The residues 37- 47 of B α- helix and loop between B α-helix and β2 showed correlated motion against residue of β3 and loop region between β3 and D α-helix. The residues 37 to 47 also showed correlated motion against the residues 90 to 102 (loop between β4 and D α-helix). Interestingly the residues of monomer A (32 to 42) showed correlated motion towards residues of B α-helix and loop between B α-helix and β2 of monomer B. This loop contains residues which make interactions with the adenine ring of the cofactor SAM. The cross correlation motion map of APO showed a similar pattern of correlated motion compared to WT FC however the extent of correlated motion was decreased overall in particular in the region between β4 and D α-helix (90-102 residues) compared to the WT FC. There was also increased anti-correlation motion in APO in contrast to the WT FC (**Figure S4.55**). The greatest increase in anti-correlated motion was seen in the EC with respect to WT FC. The residues of the loop between β3 and D helix show anti-correlated motion towards residues of β7 and loop residues between I α-helix and β10. The residues of loop β7 and I α-helix make interactions with the substrate in the WT FC (**Figure S4.56**, **S4.57**). The ES complex showed extremely low anti-correlated motion and the profile was very similar to that of the WT FC, however the extent of the correlated motions was overall decreased (**Figure S4.58**). The cross correlation analysis indicates that complex molecular motions are involved in the binding of the substrate and cofactor and that their binding influences the pattern of the correlations.

4.4 CONCLUSIONS

By applying a large number of Atomistic Molecular Dynamics Simulations on different forms of the SAM-dependent Methyltransferase NirE (Full, complex, ES complex, EC complex, Apoenzyme and nine experimental mutants), we have provided an insight into the enzyme structure-function relationships which cannot be gained experimentally. We

complemented the crystal structure information with information at atomistic levels about the impact of dynamics on enzyme structure and enzyme interactions with its substrate, UP2, and cofactor, SAM. Importantly, we discriminated between the two candidate arginine residues for the initial proton abstraction step indicating that R*149 is the likely proton abstractor. We provided atomistic structural insight into the effects of the mutations of nine important residues in the binding and catalytic sites, explaining the experimentally measured effects on both binding and catalysis. The study asserts the importance of understanding protein dynamics in addition to the crystallographic, biochemical, and kinetic studies, thus way providing synergistic insight into understanding of NirE structure-function relationships. The results provide a basis for further investigation of the enzyme mechanism using QM/MM methods, free energy of binding as well for applications in chemical biology and biotechnology.

4.5 SUPPORTING INFORMATION

Table S4.1 The setup of the NirE enzyme and the mutants for Molecular Dynamics simulations

S.No	MD-setup	Occurrence	Timescale (ns)
1	Wt NirE (WT FC)	Chain A + B + UP2 + SAH	50
2	ES	Chain A + B + UP2	50
3	EC	Chain A + B + SAH	50
4	APO	Chain A + B	50
5	Deprotonated R111	β4/E helix	50
6	Deprotonated R*149	F helix/β6 of chain B interacts with chain A	50
7	Deprotonated R111, R*149	β4/E helix, F helix / β6	50
8	E114Q	E helix	50
9	G189K	H helix	50
10	G189N	H helix	50
11	H161F	β6/G helix	50
12	M186I	β7/H helix	50
13	R111K	β4/E helix	50
14	R*149K	F helix / β6	50
15	R51K	β2/C helix	50
16	K102A	β4	50

Table S4.2 The multiple run trajectory of the wild type NirE (WTFC) for 50 ns.

Name	Mean (Å)	S.D* (Å)	S.E.M* (Å)
Run 1	3.2	0.41	0.0083
Run 2	2.9	0.27	0.0055
Run 3	2.6	0.28	0.0056
Averaged all	2.9	0.30	0.0064

* S.E.M Standard Error of Mean, S.D Standard Deviation

Table S4.3: The average and standard deviation of RMSD of WT FC and the mutants along with RMSF analysis

Name	Mean (Å)	S.D(Å)	RMSF > 1.1 Å (%)	RMSF < 1.1 Å (%)
WT FC	3.2	0.41	47	53
APO	2.4	0.15	31	69
ES	2.4	0.20	25	75
EC	2.9	0.42	48	52
E114Q	2.4	0.18	29	71
G189K	2.9	0.34	41	59
G189N	2.6	0.22	26	74
H161F	2.7	0.27	27	73
M186L	2.9	0.30	40	60
R*149K	2.7	0.29	42	58
R51K	2.6	0.22	40	60
R111K	3.1	0.43	38	62
R111K –Deprotonated	2.7	0.26	40	60
R*149K – Deprotonated	2.4	0.26	45	55
K102A	2.7	0.25	37	63

R111, R*149K – Deprotonated	2.5	0.15	30	70

R*149 represents the residue of monomer B interacting with residue of monomer A

Table S4.4: Local structural RMSF of WTFC and mutants of NirE, the region is chosen to increase or decrease if there is an average change in RMSF of > 0.3 Å in at least 50% of its residue

Name	Increased		Decreased	
	Residue span	Name	Residue span	Name
R*149K	55-63, 75-78, 160-170, 188-205, 228-233	C C-β3, β3-D, β6-G, H H-β8, I	175-178, 218-221	G-β7, β8-β9
H161F	73-78, 162-170	β3-D, β6-G	26-36,134-143,188-194,226-237	β1-A A-B, F F-β6, H, β9-I
M181L	55-63, **73-78**, 82-92, **162-171**	C C-β3, β3-D, D, β6-G,	26-36, 134-150, 173-176	β1-A A-B, F F-β6, G
G189N	75-77, 162-168	β3-D, β6-G	25-37, 148-151, 190-193, 232-238	β1-A A-B, F-β6, H, I
G189K	55-59, 72-90, 110-125,188-194	C, β3-D D, E E-β5, H	25-35,134-140,172-176	β1-A A-B , F, G
R111K	53-64, **79-85**, 110-117, **163-168**	C C-β3, β3-D D, E, β6-G	27-29, 133-144, 235-237	β1-A, F F-β6, I-β10
E114Q	72-78, 80-87, 163-171	β3-D, D, β6-G	26-39, 133-144, 172-175, 188-193, 231-238	A B, F F-β6, G, H, I-β10
R51K	50-55, **74-87,** 103-115, **161-168**	β2-C, β3-D D, β4-E E, β6-G	25-34, 143-150, 173-177,	β1-A A-B, F-β6, G
K102A	52-55, 70-85, 110-113, 160-167,231-239	β2-C, β3-D D, β4-E, β6-G, I- β10	27-37, 171-180	β1-A, G G- β7

126

			25-36, 133-145, 169-179, 196-206, 228-237	β1-A A-B, F F-β6, G G-β7 , H H-β8, I-β10
ES	71-86, 110-120	B3-D D, E	25-36, 133-145, 169-179, 196-206, 228-237	β1-A A-B, F F-β6, G G-β7 , H H-β8, I-β10
EC	51-61, 73-86, 109-115, 133-136, 162-173, 190-201	β2-C C-β3, β3-D D, E, F, β6-G, H	231-237	I-β10
APO	71-73, 163-168, 187-203	β3-D, β6-G, β7-H	32-35, 121-124,148-152, 174-177	A-B, E-β5, F-β6, G

* Regions underlined are the loops between the alpha helix and beta sheets.

Table S4.5 The hydrogen bonding* Interactions of substrate (UP2) and cofactor (SAM) with protein in WTFC

Donor	Acceptor	Distance (Å)	Probability (%)	Minimized Crystal structure
D105(bb)	SAM (NH3$^+$)	3.2	32.7	2.8
D105(sc)	SAM (NH3$^+$)	3.5 (30 ns)	21.5	5.5
I108(bb)	SAM (NH3$^+$)	2.9	76.9	2.8
Y185(sc)	SAM(COO$^-$)	2.2	85.6	3.5
T133(sc)	SAM(COO$^-$)	2.9	80.0	2.6
A134 (bb)	Y185 (sc)	2.8	32.4	3.6
SAM (N6)	C138 (sc)	3.4	26.2	4.0
SAM(N6)	P242 (bb)	3.3	3.0	5.0
SAM(N6)	V212(bb)	3.5	29	6.2
SAHM(N6)	Q214(bb)	3.6	45.7	4.5
Q217 (sc)	SAM (N7)	3.5	71.4	4.6
M186 (bb)	SAM (N3)	2.9	36.3	6.4
Ring C (NH)	H161(bb)	2.8	12	3.5
Q163 (sc)	Ring D Propionate (sc)	3.2	44.0	4.1
R111 (bb)	Ring D acetate (sc)	2.4	39	3.2
G110 (bb)	Ring D acetate (sc)	2.5	70	2.7

*Hydrogen bond criteria: Angle > 120° and donner acceptor distance < 3.5 Å

Sc = side chain, bb = backbone, NH3$^+$ (amino group), COO$^-$ (Carboxylate group)

Table S4.6 The important interactions observed in WTFC in context of cofactor SAM and substrate UP2

Side chain	Side chain	Distance (Å) in md	Crystal structure
ring A *	Methyl	6.0	5.9
ring B *	Methyl	6.3	5.8
Methyl (SAM)	M186 (sc)	4.3	3.9
Methyl (SAM)	Y185	4.2	5.0
Methyl (SAM)	F109	4.0	3.8
P242 (sc)	SAM (adenine)	3.8	3.7
A134 (sc)	SAM (adenine)	3.6	3.6
Y185 (sc)	SAM (adenine)	3.6	3.9
L52 (sc)	SAM (CAI)	4.1	4.7
R111 (sc)	R*149 (sc)	8.9	3.6
R51 (sc-ali)	Ring A Propionate (ali)	4.1	5.7
ring A propionate (sc)	Sulphur	3.1	5.9
R111 (sc)	Ring A Acetate (sc)	4.3	4.8
D105 (sc)	NH3$^+$ (SAH)	4.0	5.6
H161(sc ali)	Ring C acetate (sc)	4.4	3.7
Q163 (sc ali)	Ring C Propionate (sc) ali	3.9	5.5
T159 (sc ali)	Ring D Propionate (sc) ali	4.0	5.1
L162 (sc ali)	Ring D Propionate (sc) ali	4.4	5.3

Ali = aliphatic side chain of amino acids, Cofactor (SAM)

Table S4.7 The average and multiple runs of R*149 deprotonated

Name	Mean (Å)	S.D (Å)	R*149 (sc) – C20 UP2
Run 1	2.4	0.25	3.5
Run 2	2.8	0.27	4.2
Averaged all	2.6	0.24	3.8
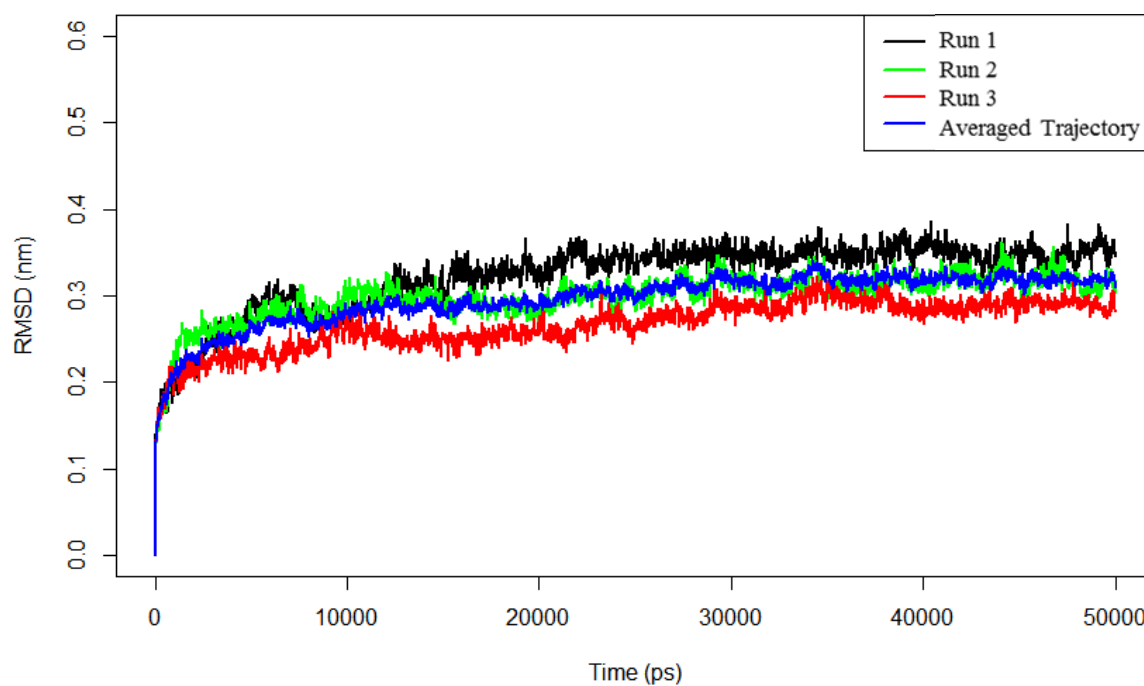
Figure S4.1 The multiple runs of the wild type NirE (WTFC) enzyme for 50 ns. The average trajectory of all the runs is represented in blue colour
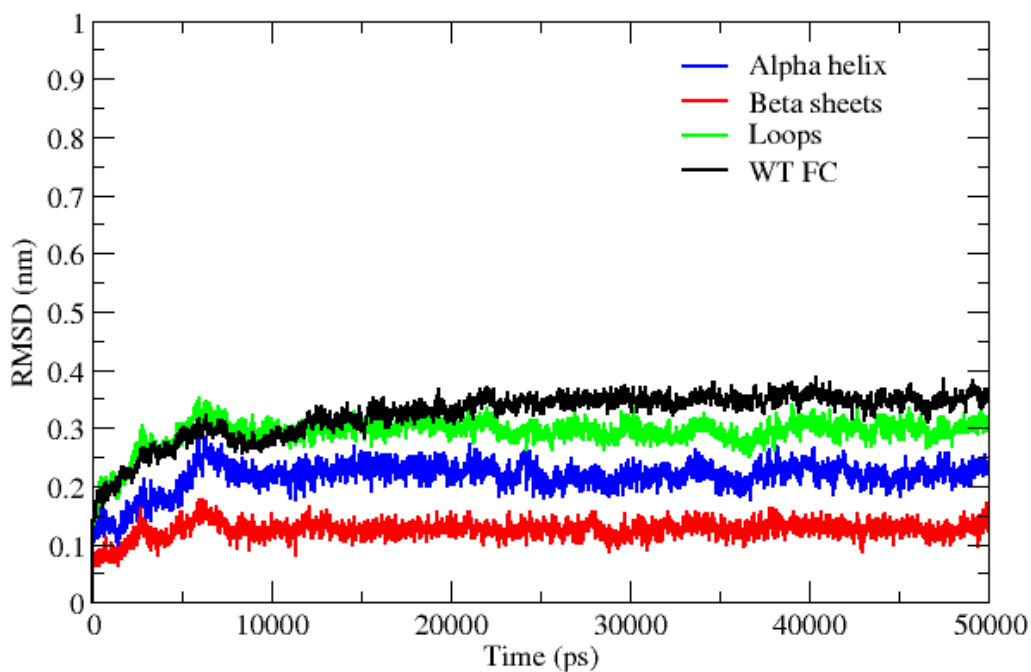
Figure S4.2 The RMSD decomposition of WT FC which include alpha helix beta sheets and loops for 50ns using C α atom
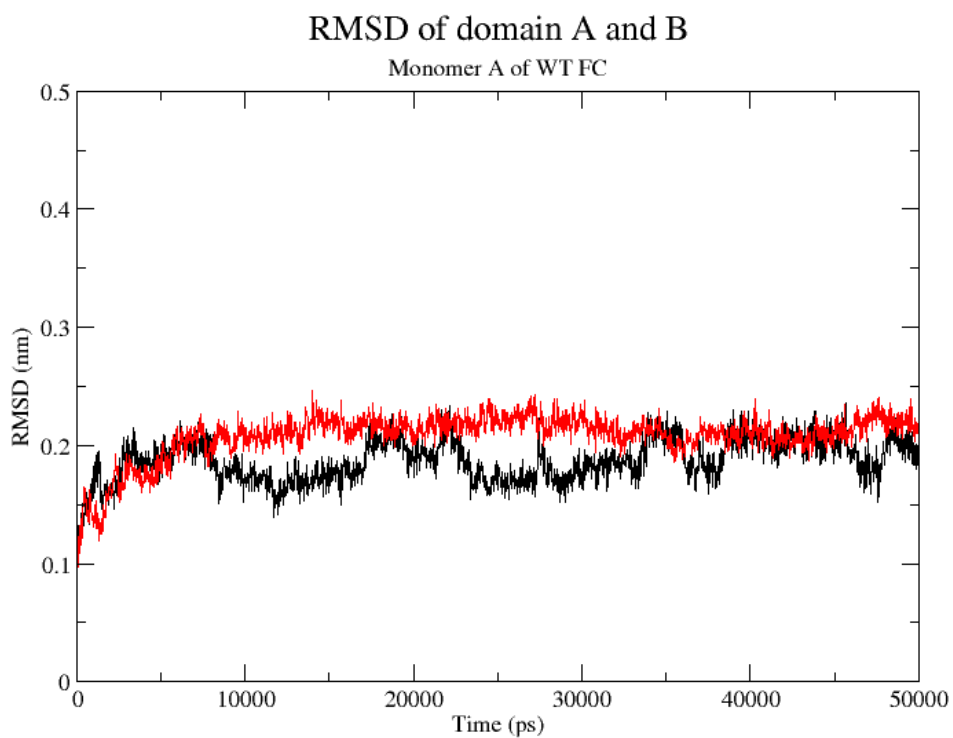
Figure S4.3 The RMSD of domain A and B of monomer A of WTFC for 50 ns.

Linker RMSD and Angle
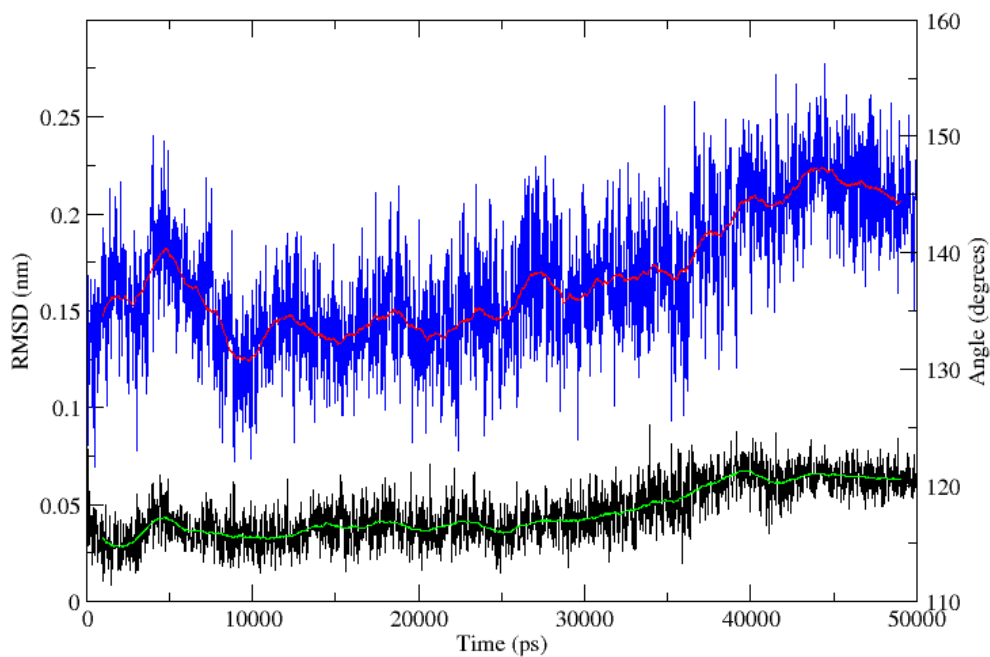
Figure S4.4 The RMSD of the linker region of WTFC and angle of the linker region residues (Cα of 129, 132 and 134) as function of time for 50 ns.
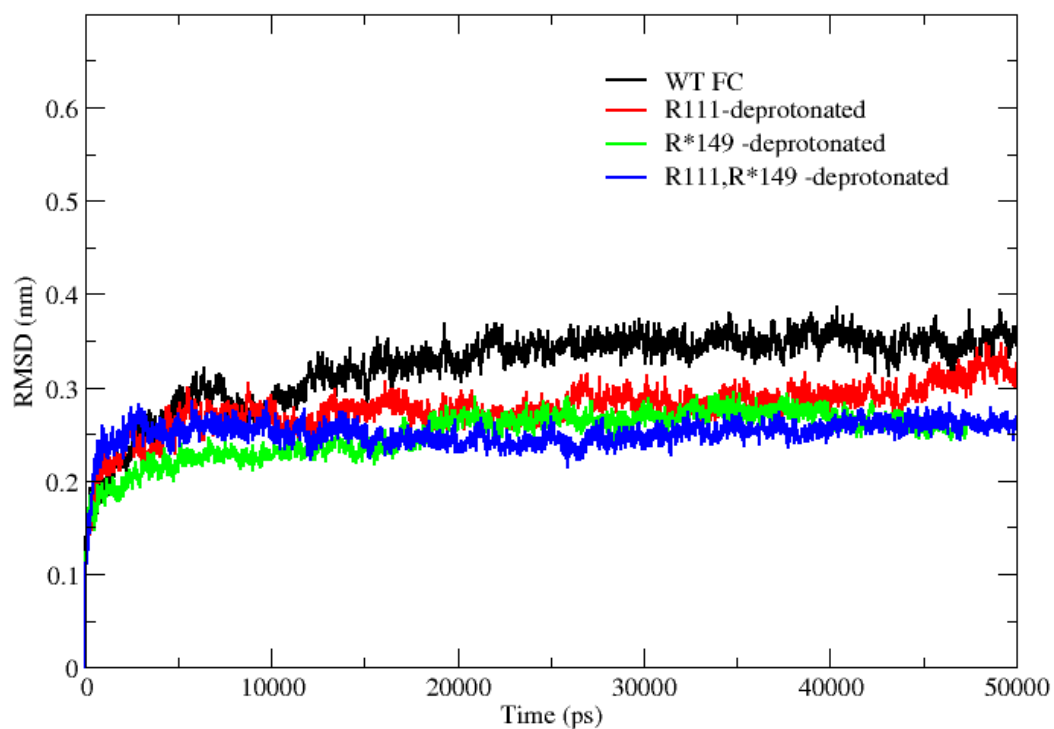
Table S4.5 The RMSD of the deprotonated arginine residue of the NirE enzyme for 50 ns.
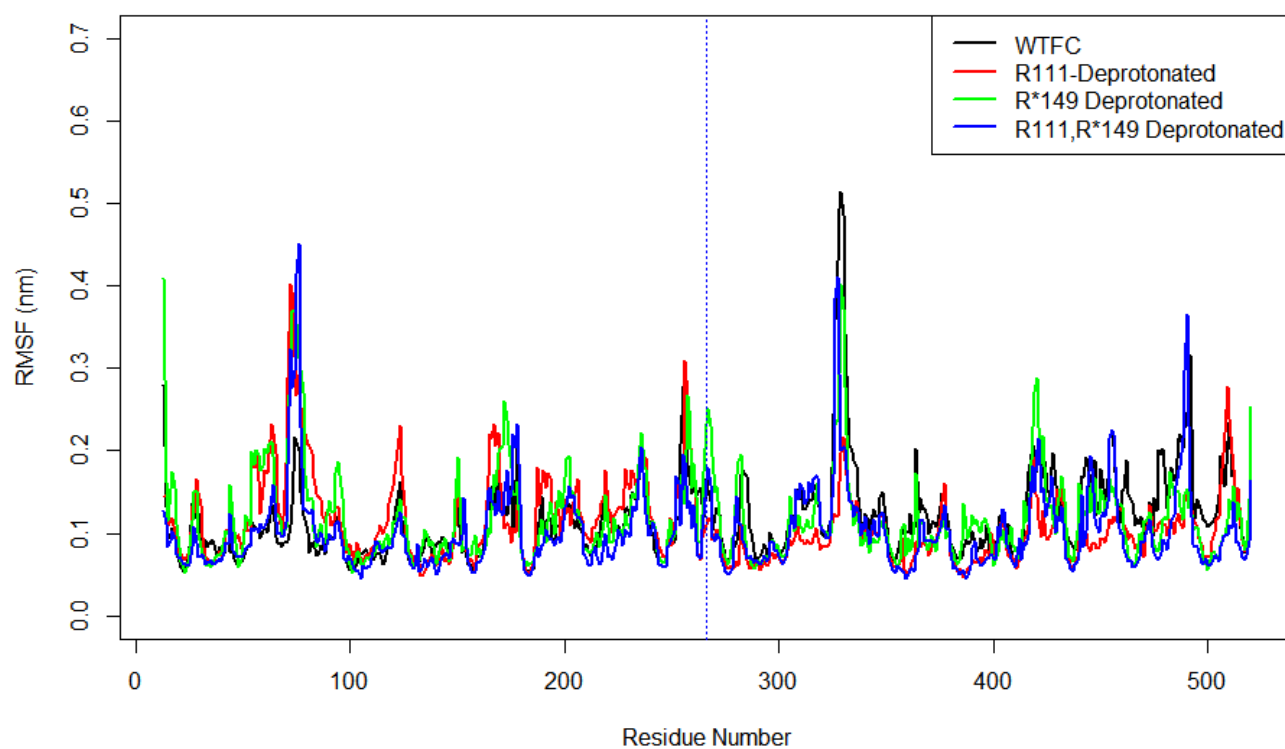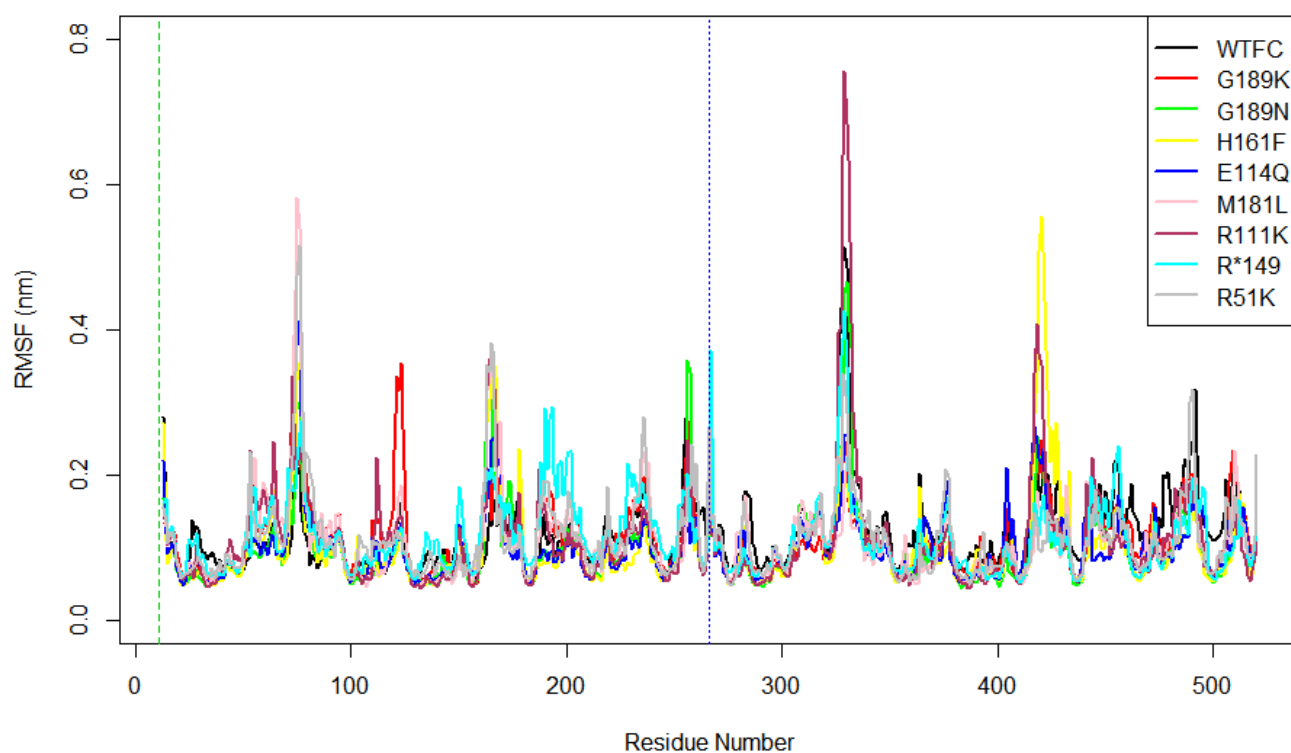
Table S4.6 The RMSF of the WTFC and the mutants for both monomer A and monomer B. The blue line separates the monomer A and monomer B. The green line in the start of the graph show that the residue starts according to the canonical naming in the crystal structure from residue number 13.
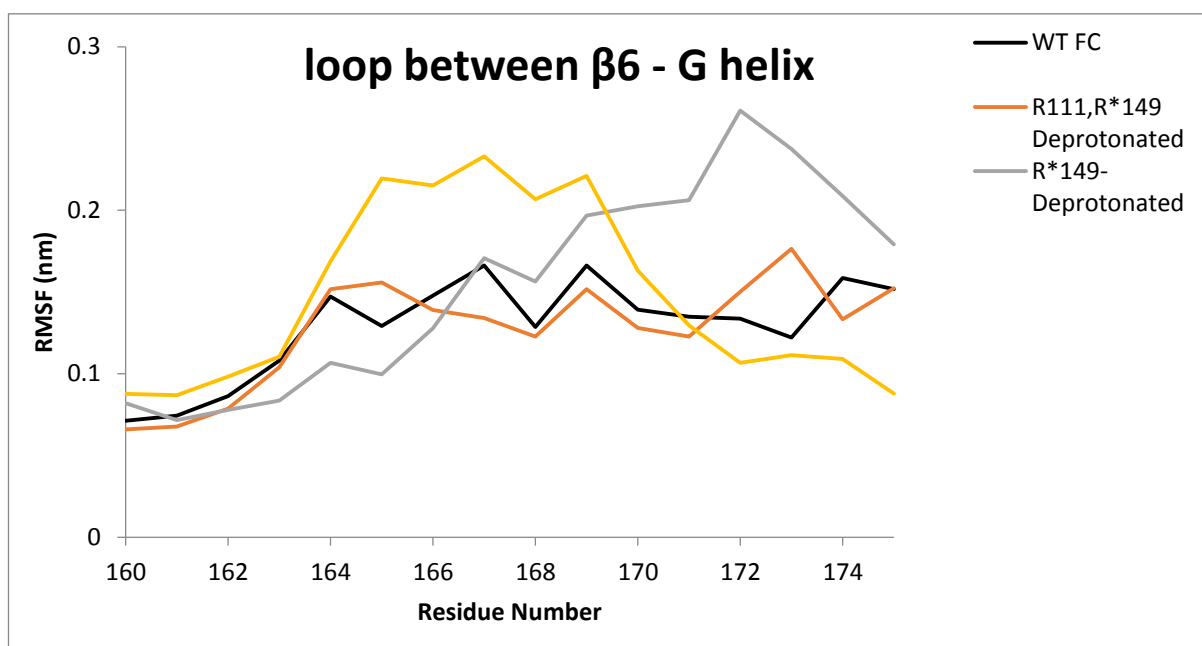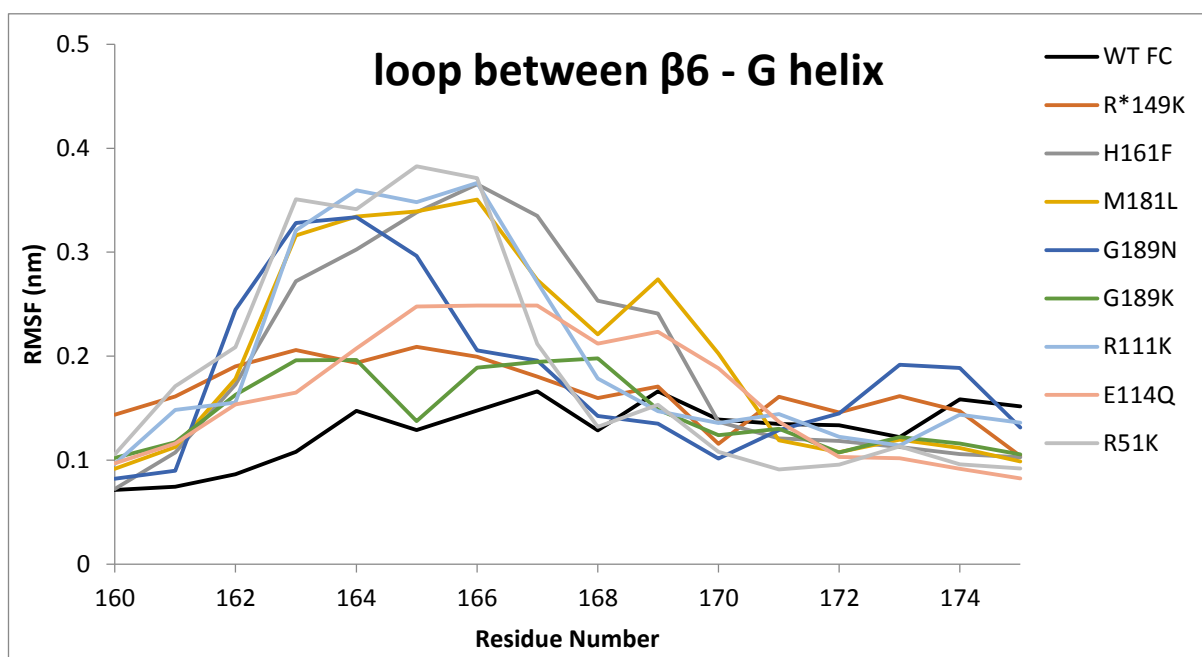
Figure S4.7 The RMSF of the mutants and WT FC of residues between loop β6-G helix
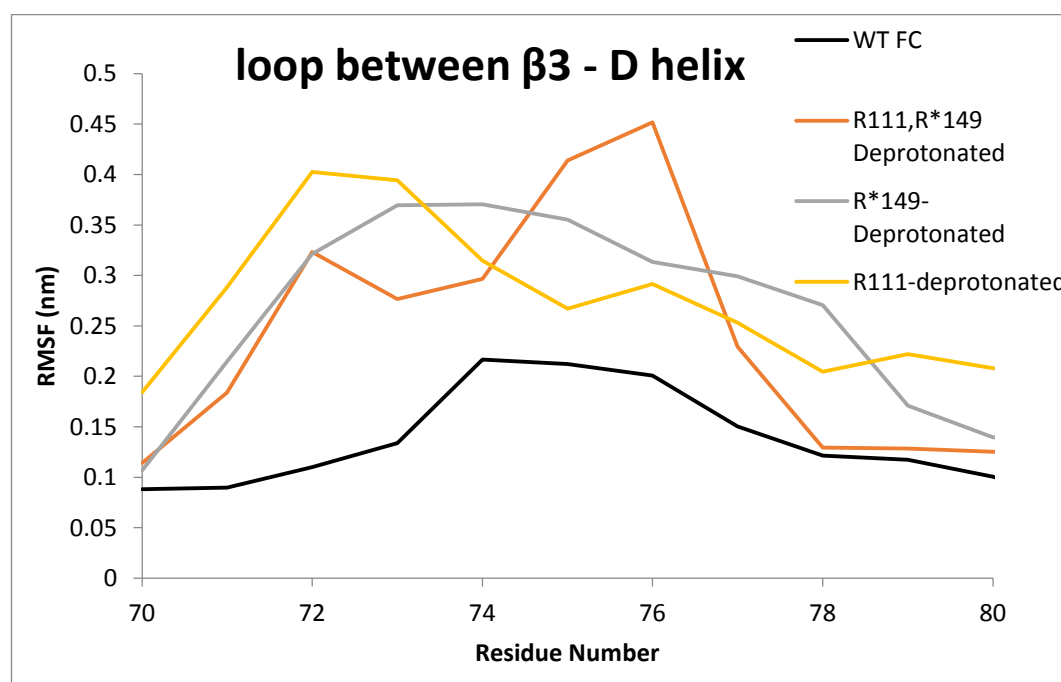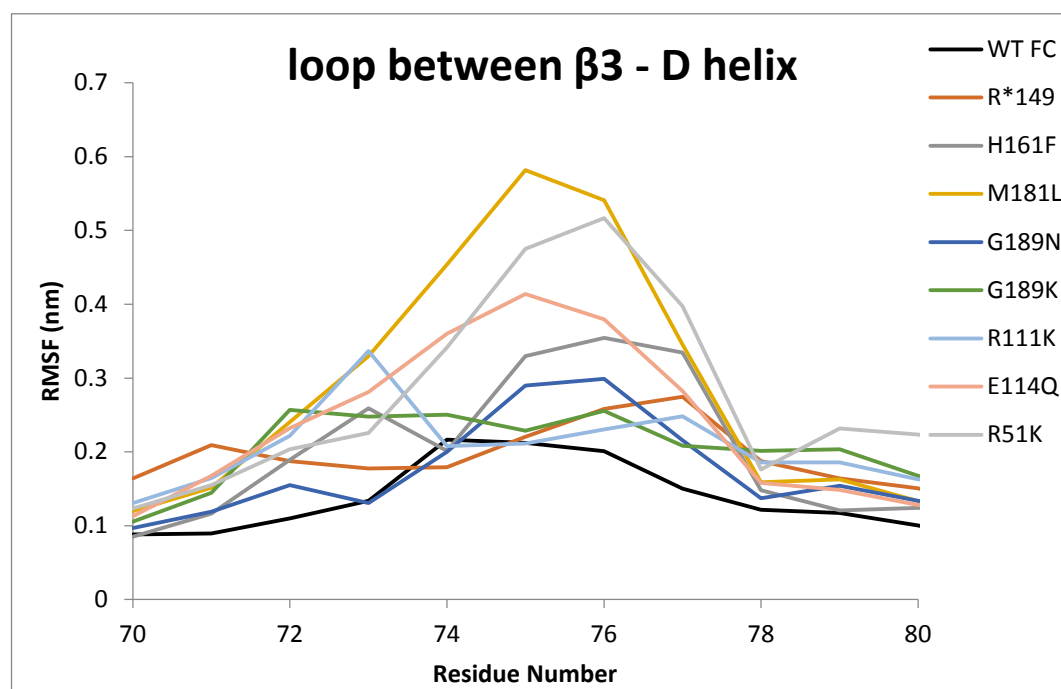
Figure S4.8 The RMSF of the mutants and WT FC of residue between β3- D
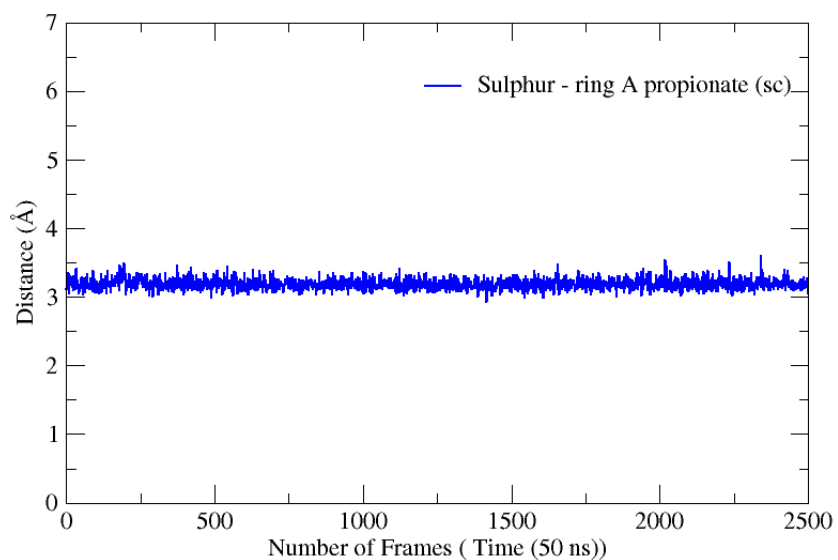
Figure S4.9 The sulphur atom of cofactor (SAM) electrostatic interactions with the propionate side chain of ring A of the UP2

Poteintial methyl transfer sites on substrate
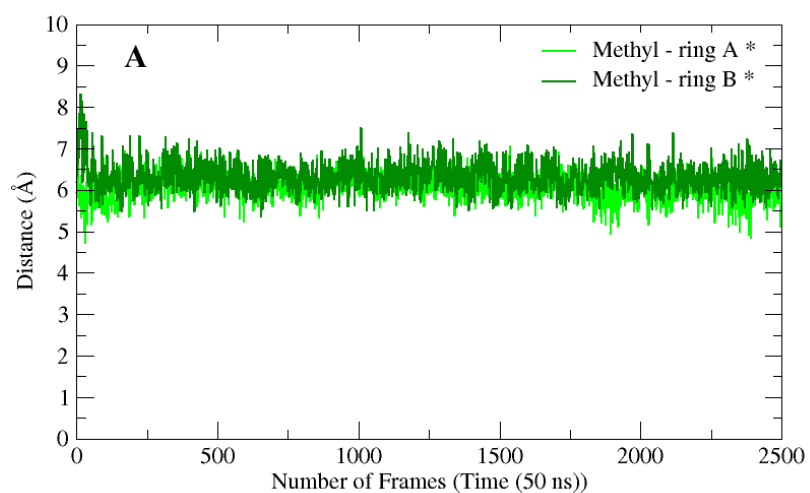
Angle between Sulphur - Methyl and Potential site on UP2

135
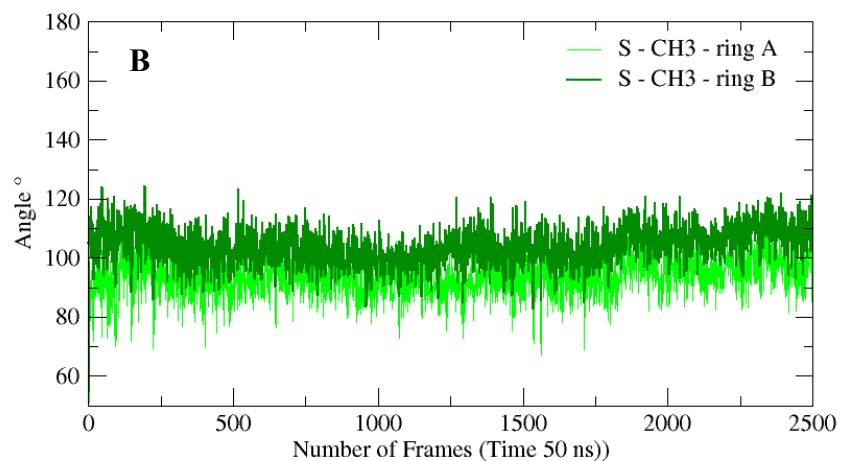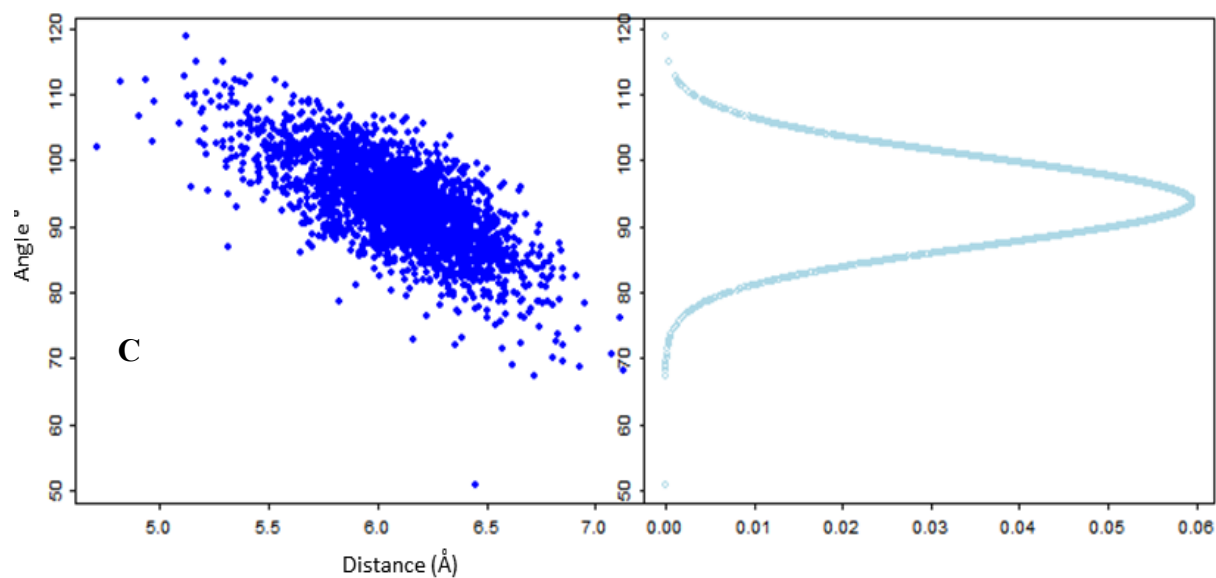
Figure S4.10 The methyl distance and angle between Sulphur, methyl group and methyl acceptor sties for the nire wild type for 50 ns simulation. (A) The methyl distance from the cofactor to its potential transfer site on the substrate , (B) The measurement of the angle between the sulphur, methyl and potential methyl acceptor site on the substrate, (C) The distribution of distance and angle
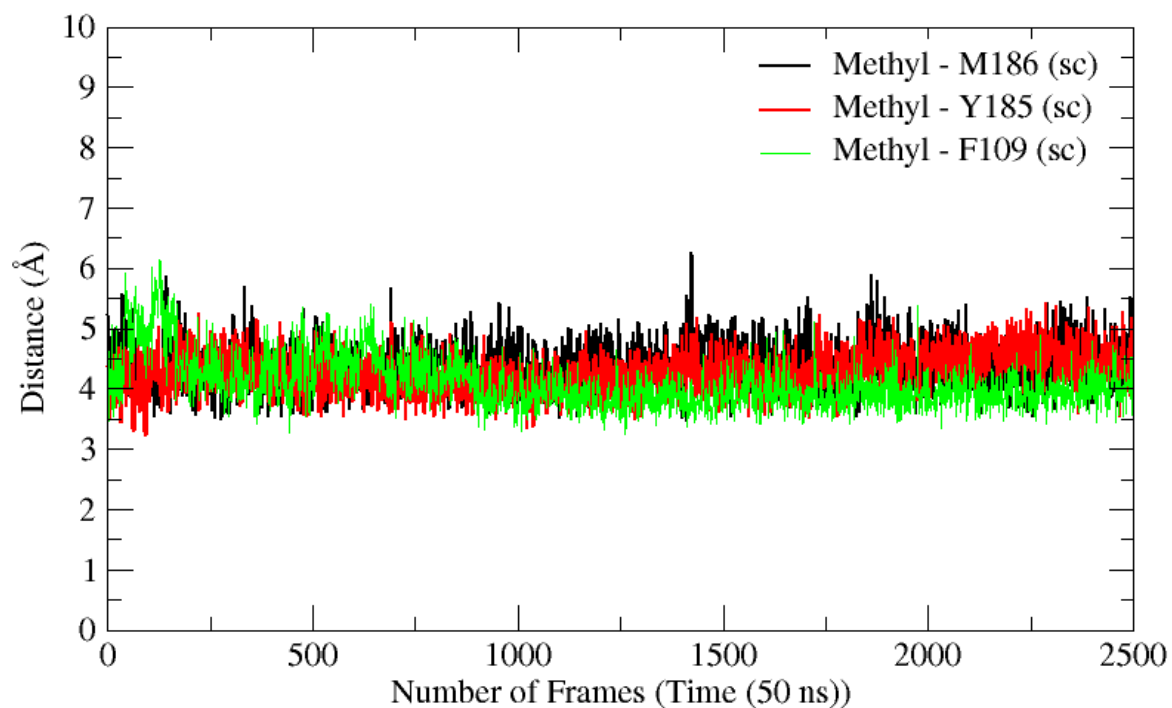
Figure S4.11 The interaction of newly added methyl group of cofactor SAM with the

WTFC residues.
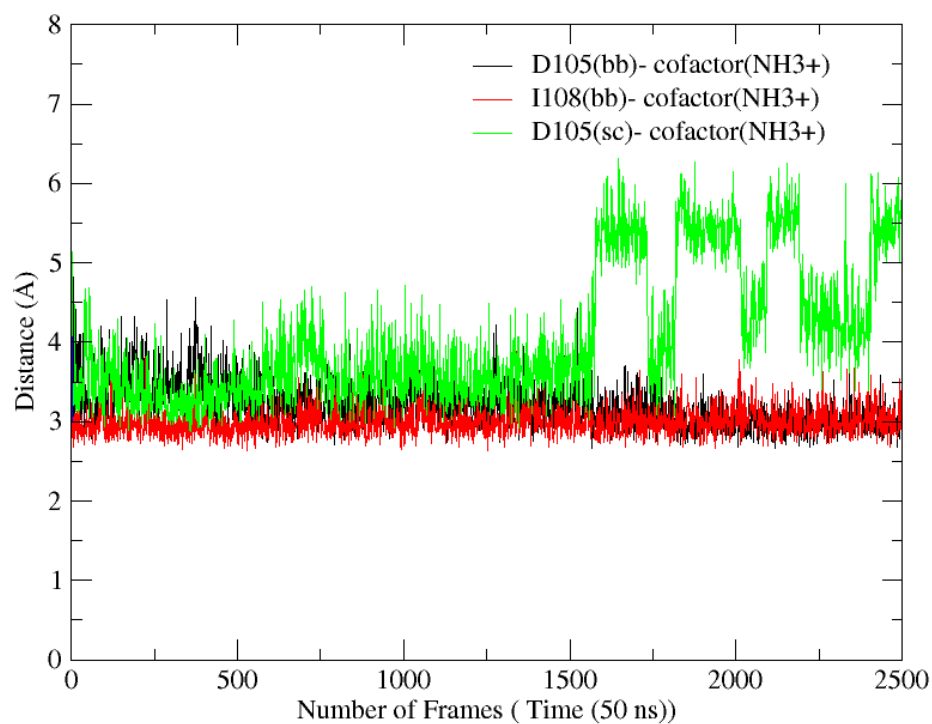
Figure S4.12 The interactions of WTFC residues with the N terminal group (NH^{3+}) of the

cofactor SAM
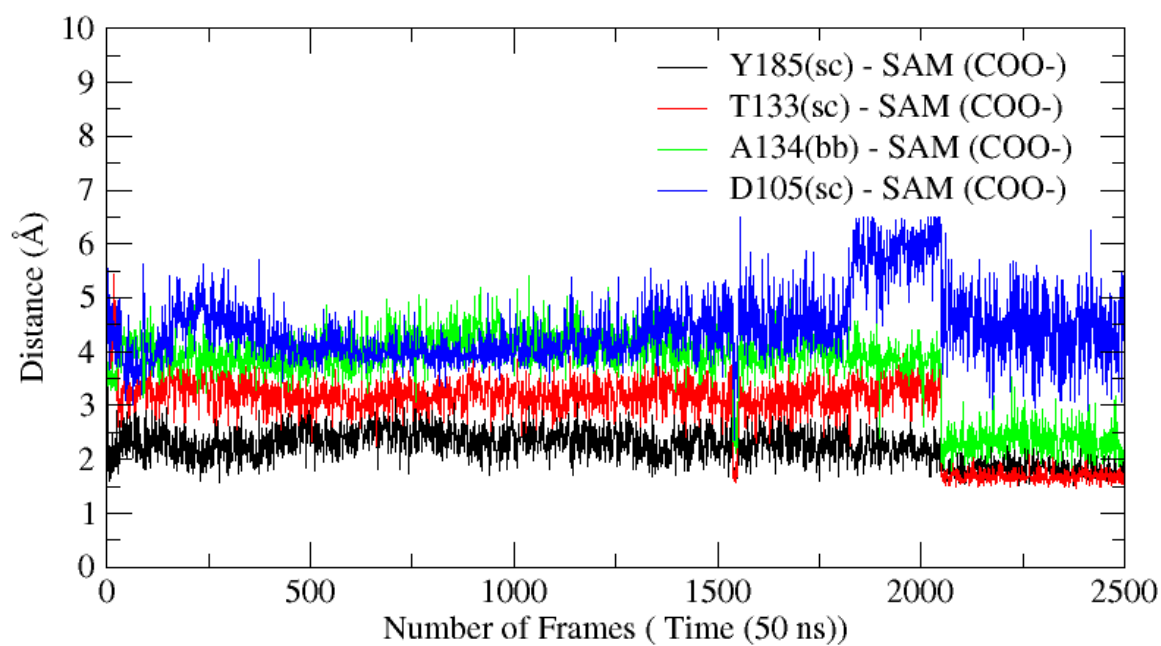
Figure S4.13 The interaction of the carboxylic group of the cofactor SAM with WTFC residues
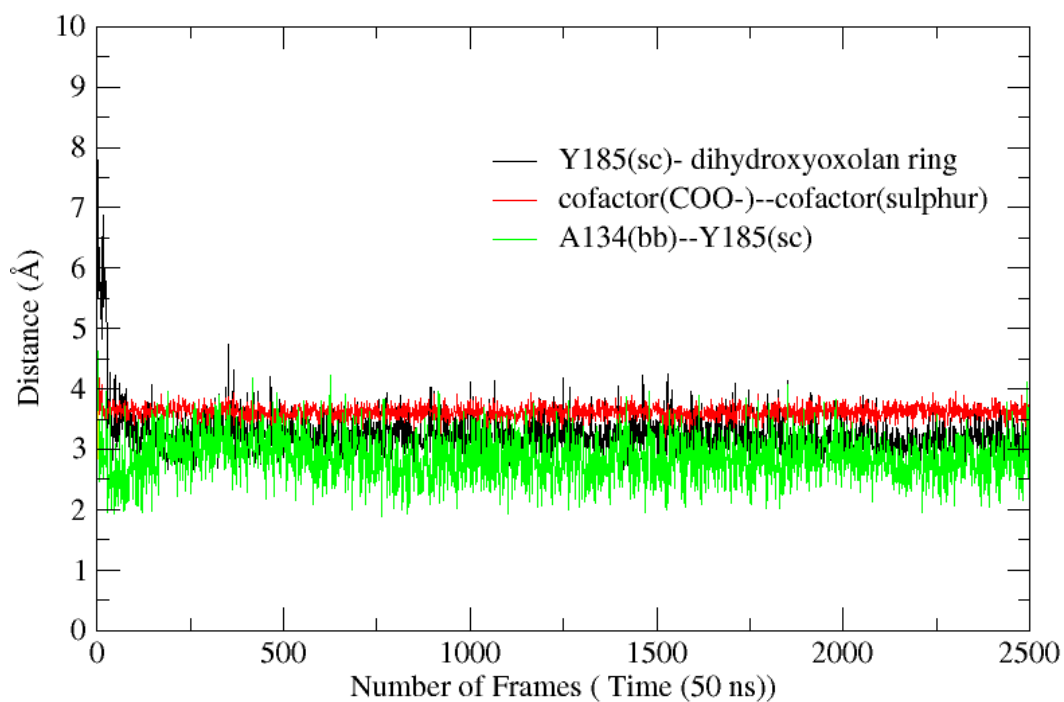
Figure S4.14 The interactions of the protein resdiue (WTFC) with the carboxylic group of the cofactor SAM.
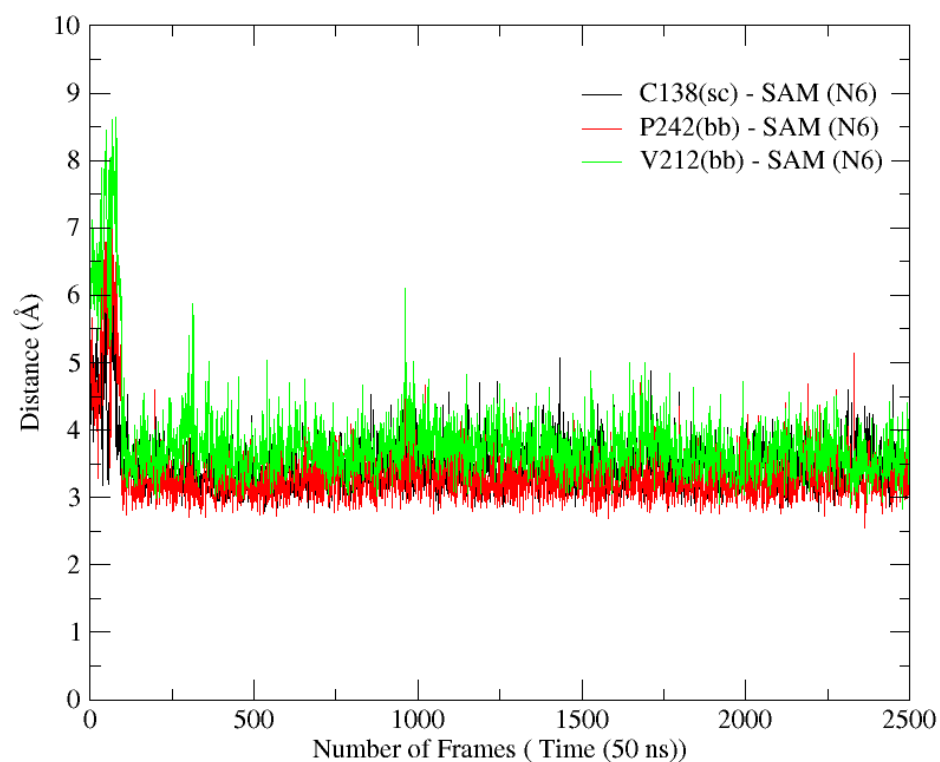
Figure S4.15 The adenine ring interactions of cofactor SAM with WTFC residues
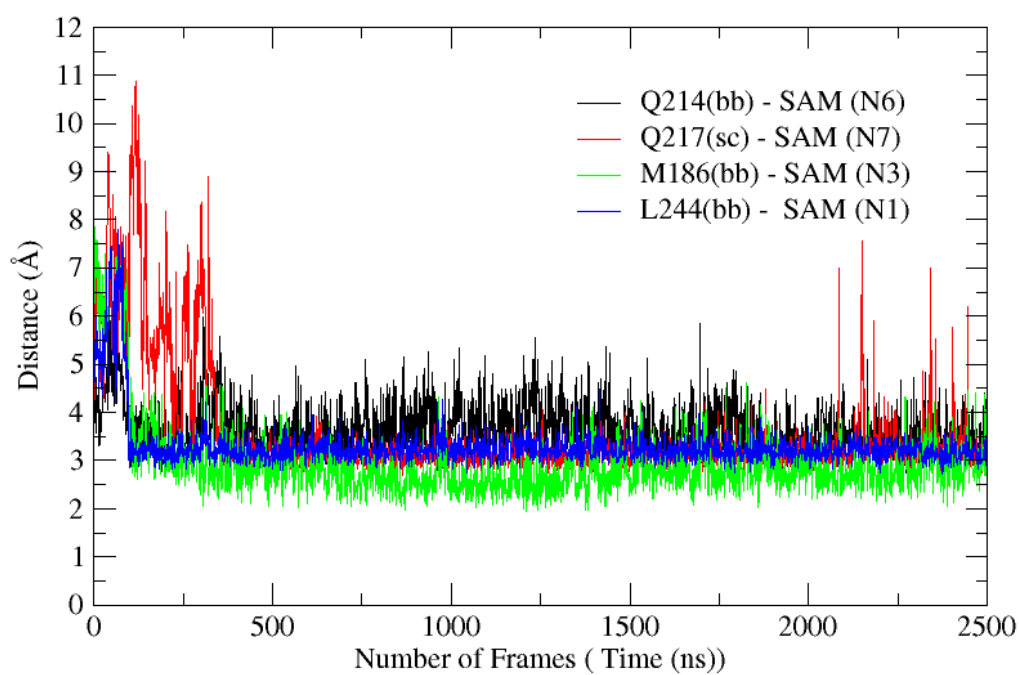
Figure S4.16 The adenine ring interactions of cofactor SAM with WTFC residues
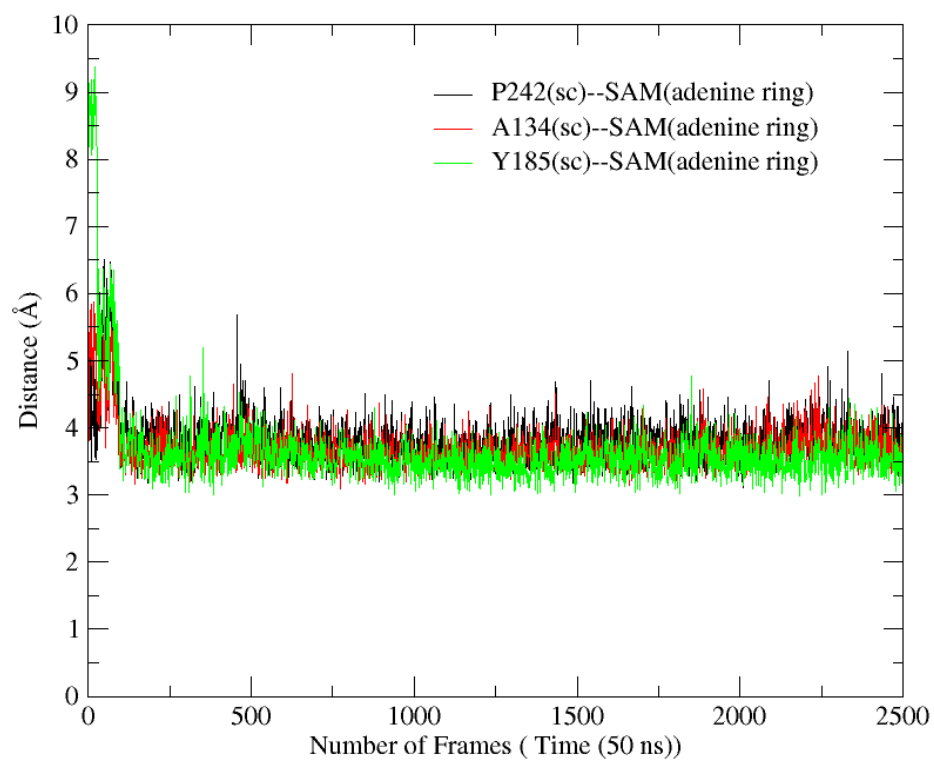
Figure S4.17 The adenine ring of the cofactor SAM and the hydrophobic interaction it makes with WTFC residues

Dihedral angle of pyyrol rings
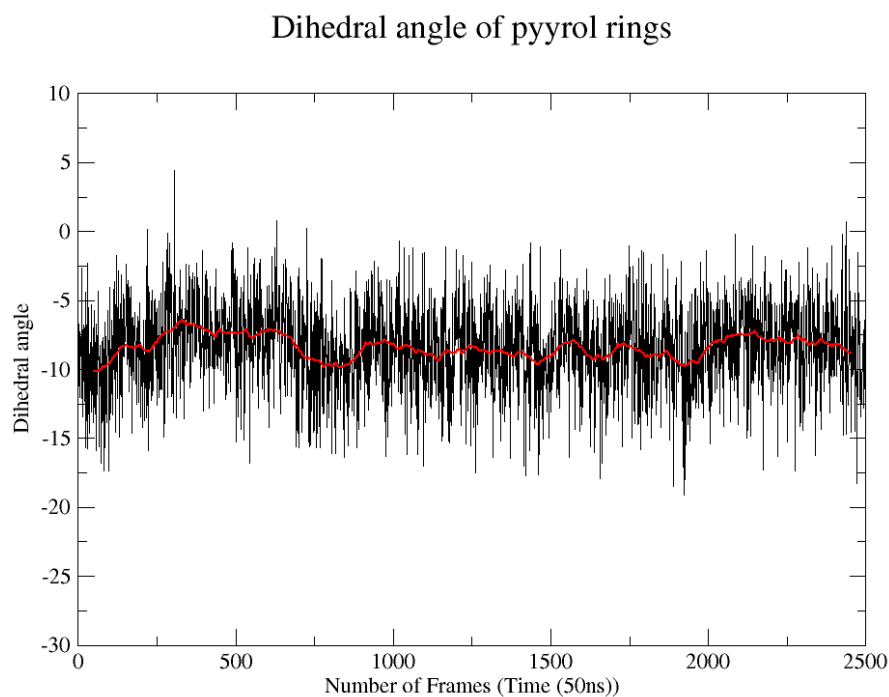
Figure S4.18 The dihedral angle of four pyyrole ring of the substrate UP2 in WTFC
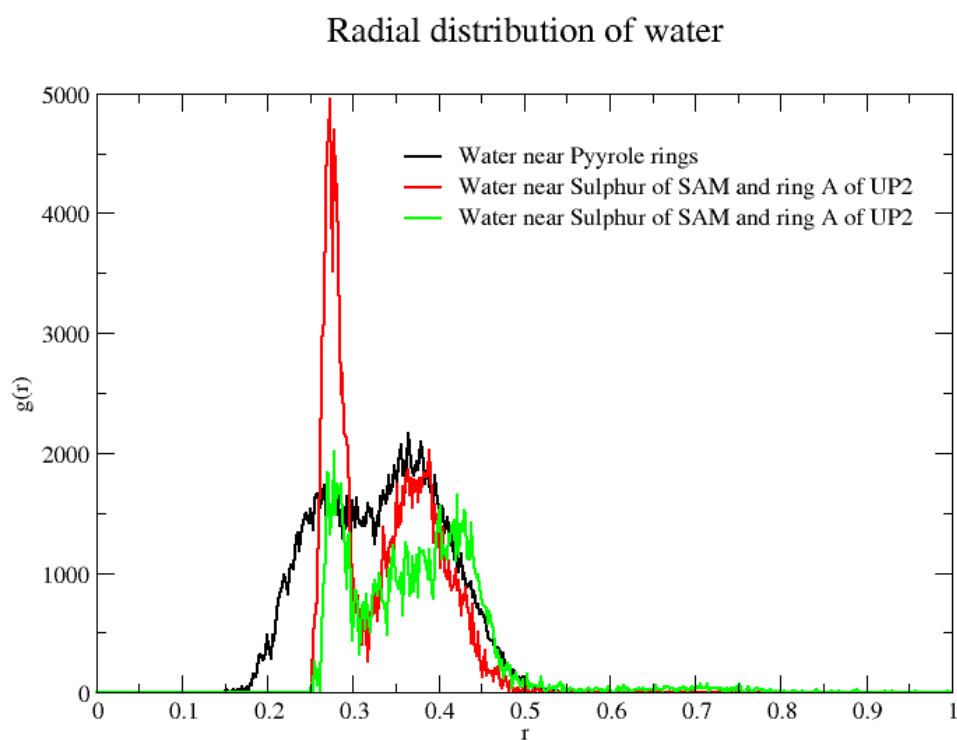
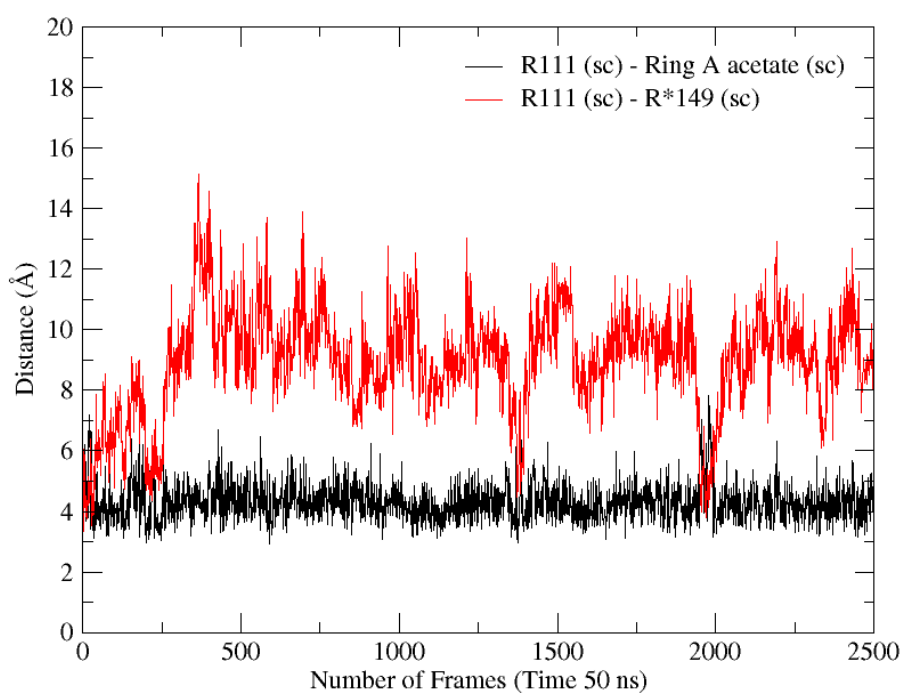Figure S4.19 The radial distribution of solvent molecules near the UP2 and SAM of WT FC

Figure S4.20 The interactions of the ring A acetate side chain with the residues of WTFC.

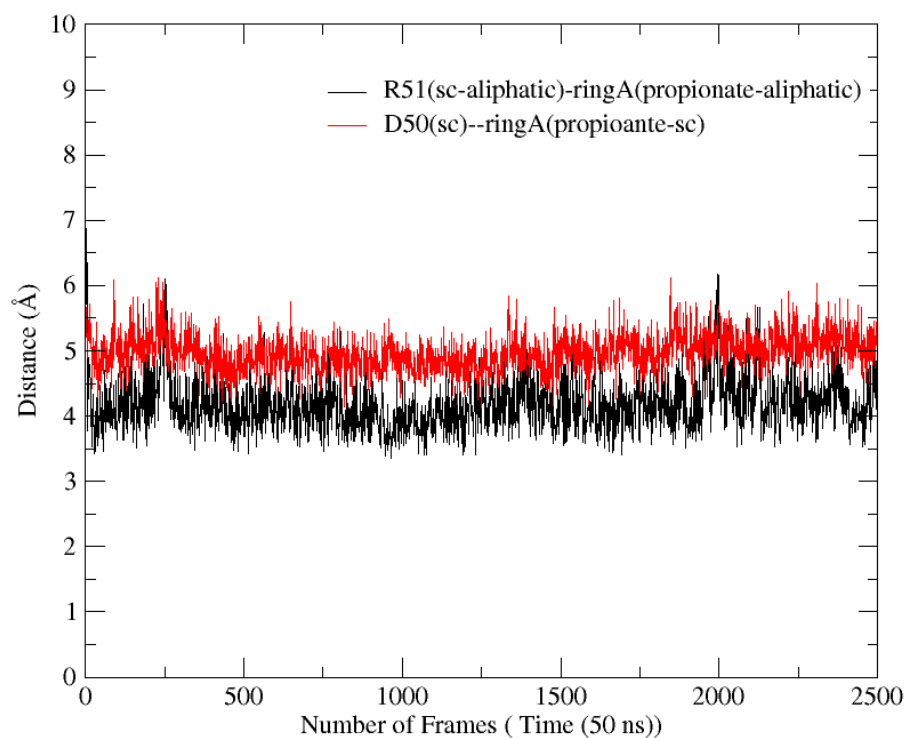The side chain of R111 distance with sidechain of R*149 of WTFC for 50 ns simulation

Figure S4.21 The interactions of the acetate side chain of ring A of UP2 with protein residue of WTFC
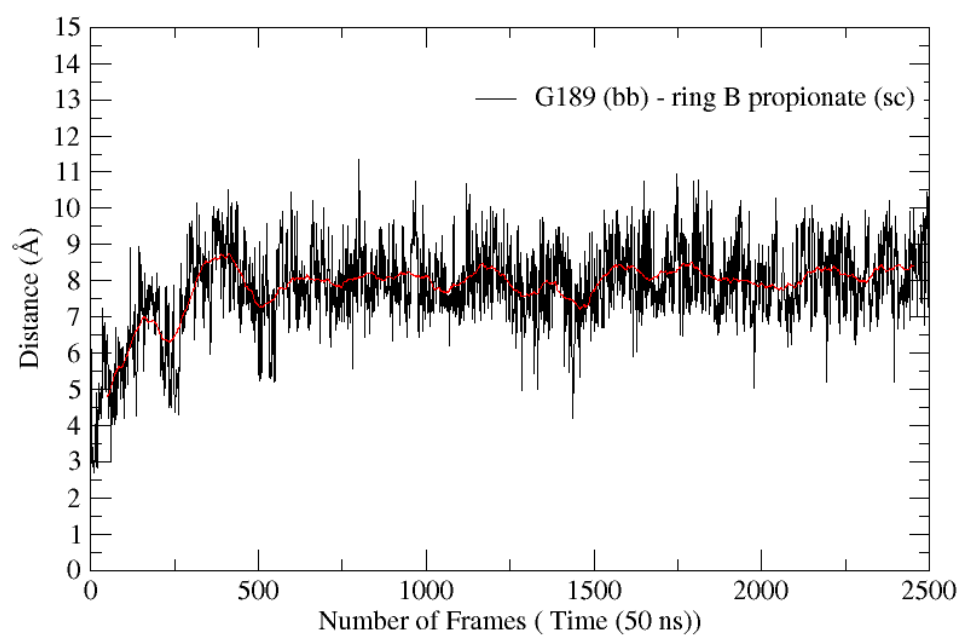

Figure S4.22 The interactions of the propionate side chain of ring B of the substrate UP2 with the G189 residue of WTFC
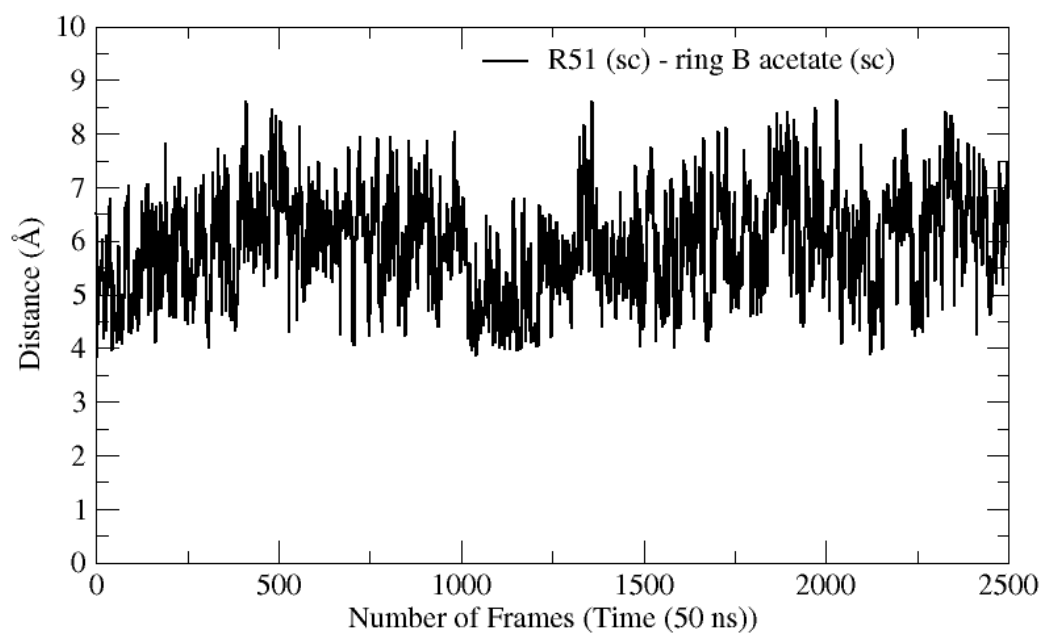
Figure S4.23 The side chain of the acetate of ring B of substrate interactions with the R51 residue of the WTFC
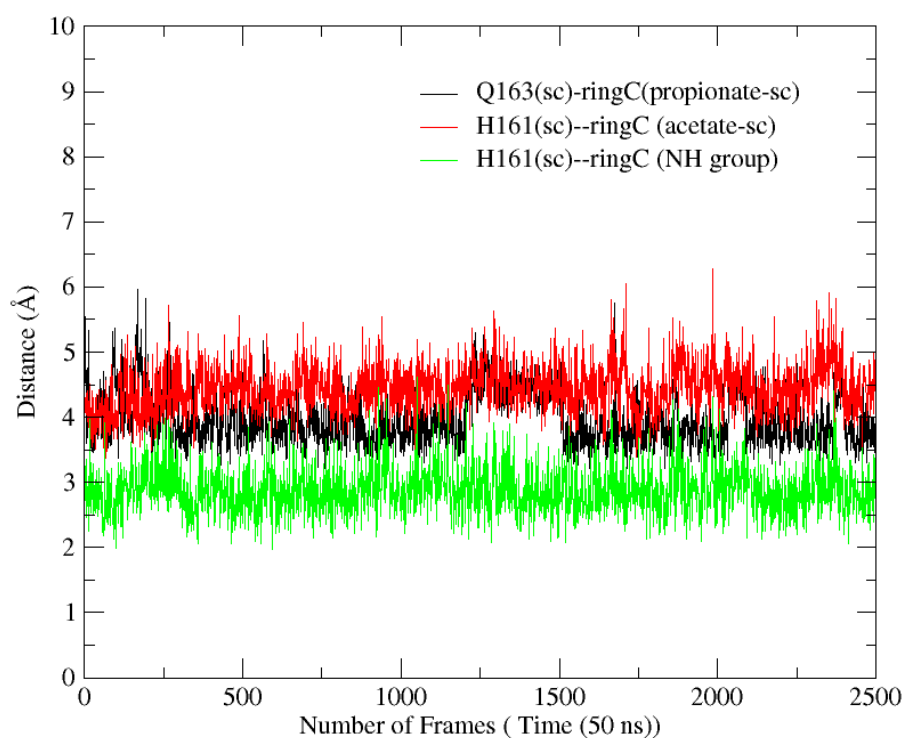

Figure S4.24 the interactions with the ring C of the substrate of substrate UP2 in WTFC
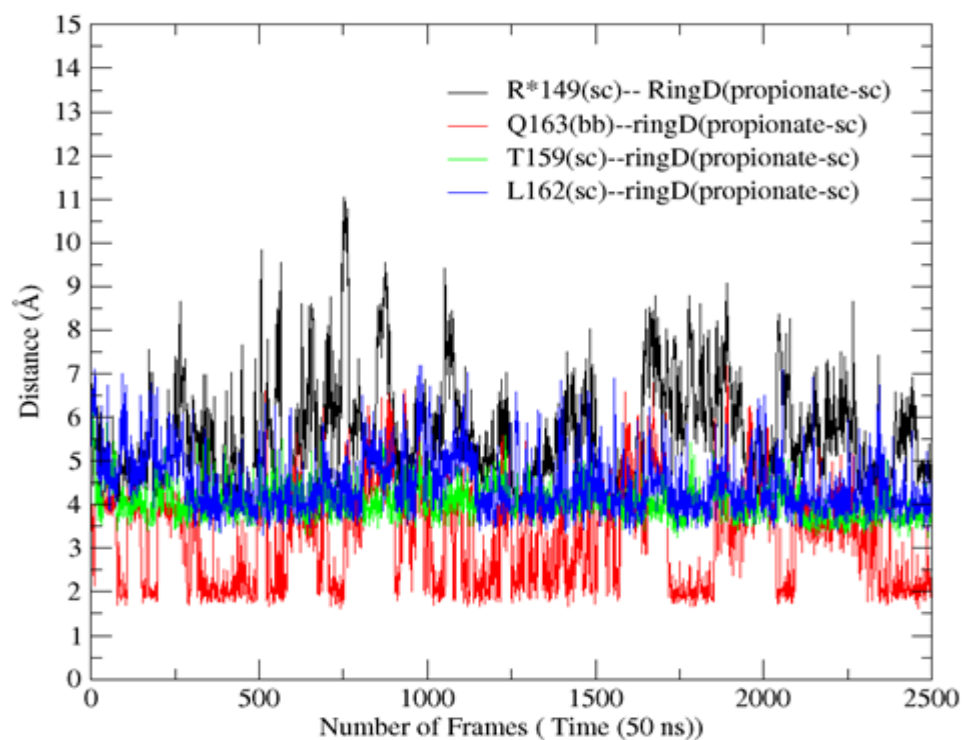
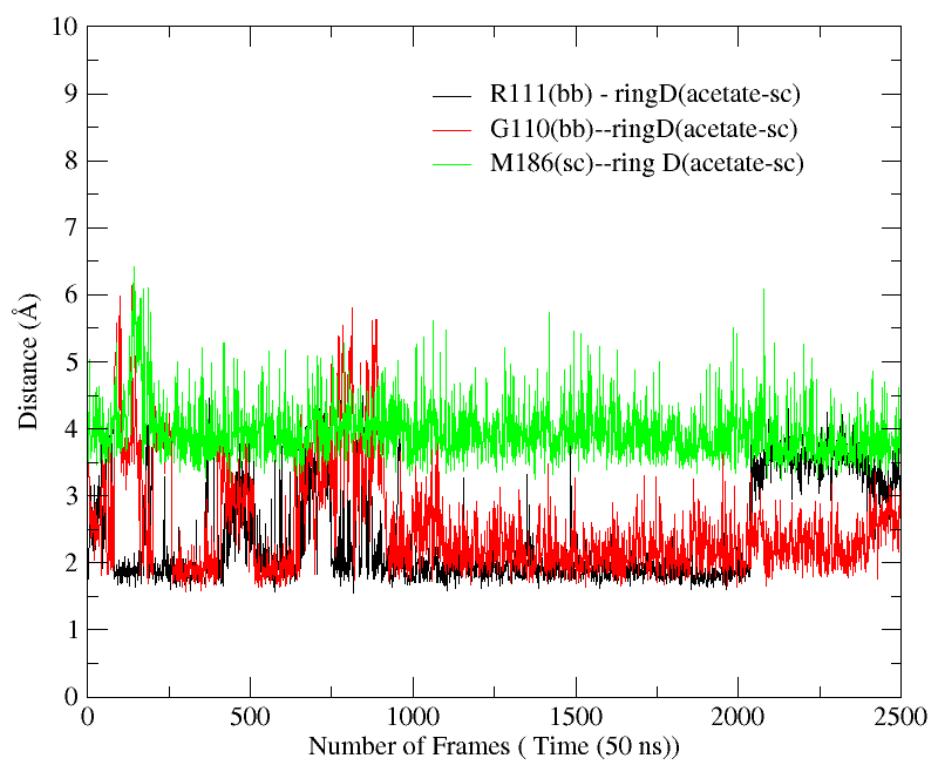Figure S25 The interactions of the side chain of ring D proipionate of substrate UP2 in WTFC

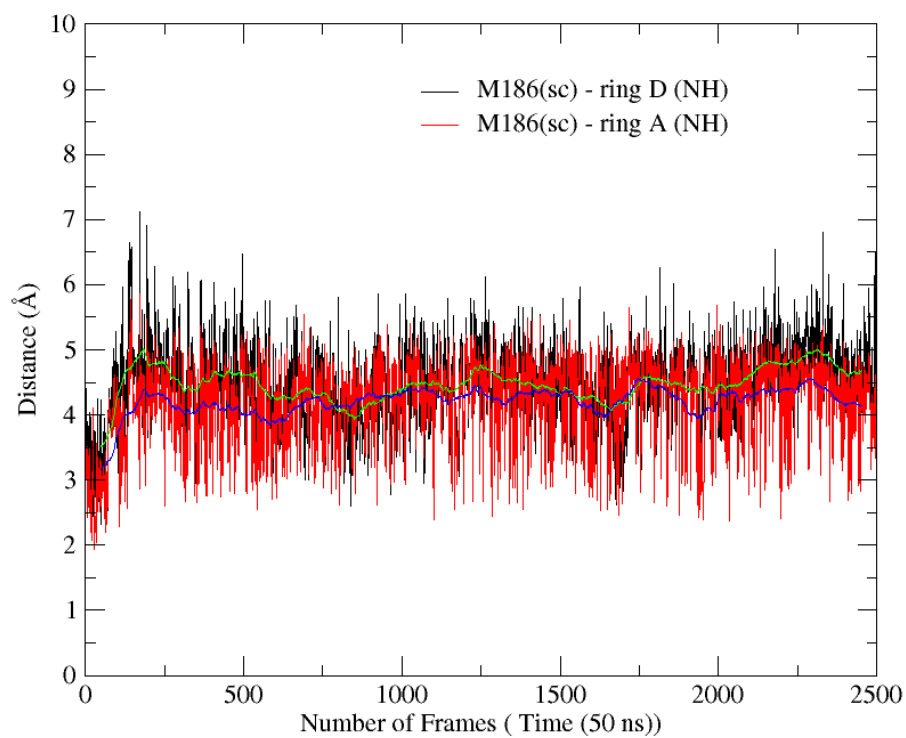Figure S4.26 The interactions of ring D of substrate UP2 with residue of WTFC

Figure S4.27 the hydrogen bonding of the M186 residue with the ring NH of substrate

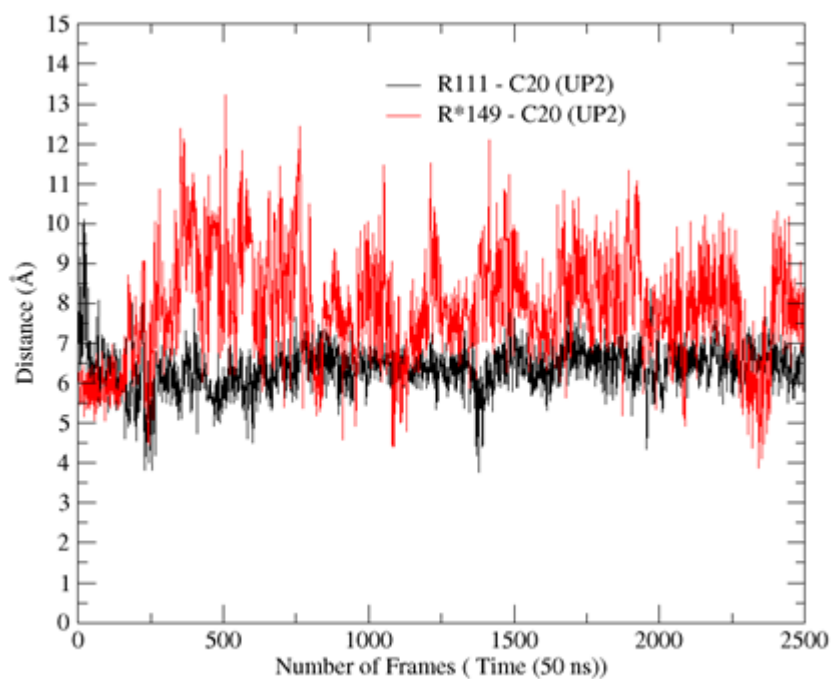Figure S4.28 The interactions of the arginine R111 and R*149 of wild type NirE (WTFC) with the C20 potential proton abstraction site on substrate UP2.
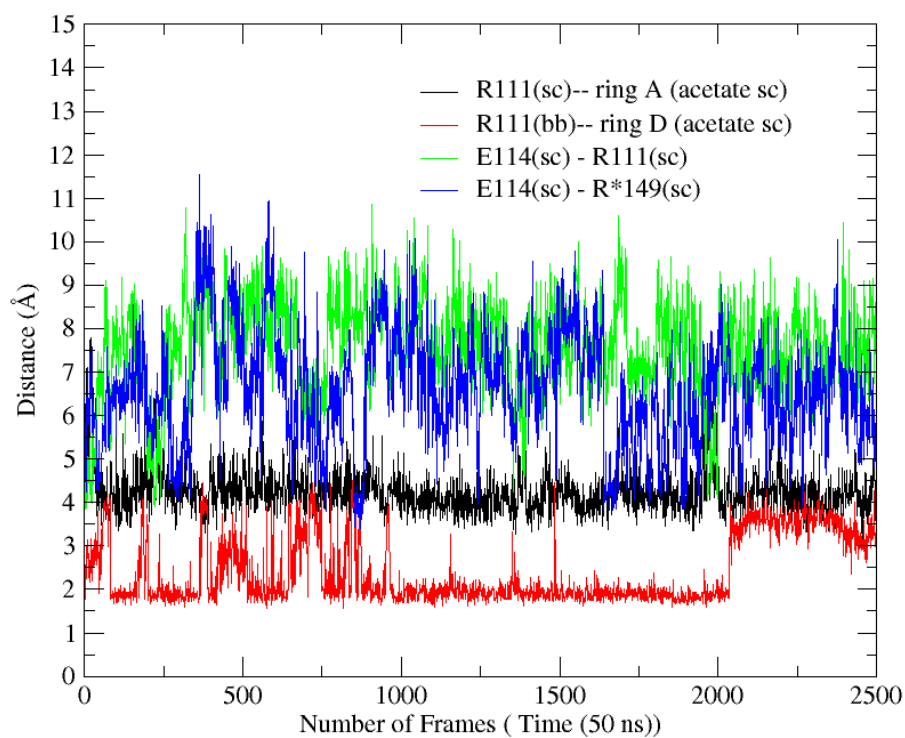
Figure S4.29 The interactions of R111 and E114 residues in the WT FC for 50 ns trajectory
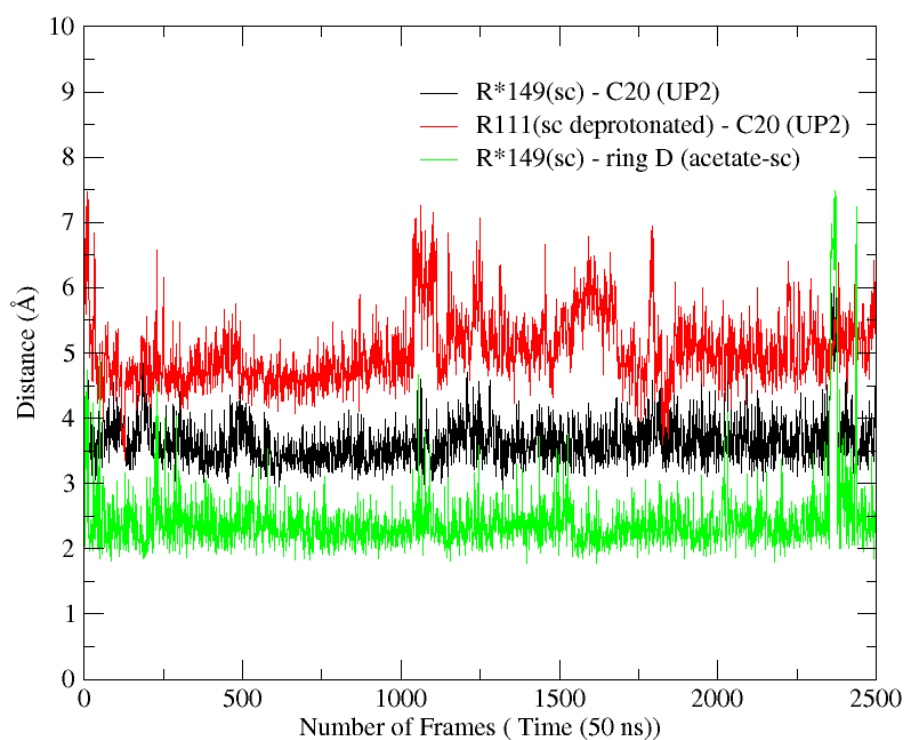
Figure S4.30 The MD simulation of R111 deprotonated setup and its important interactions

with substrate UP2. Note here R*149 is in its normal state which is protonated (+1 charge).
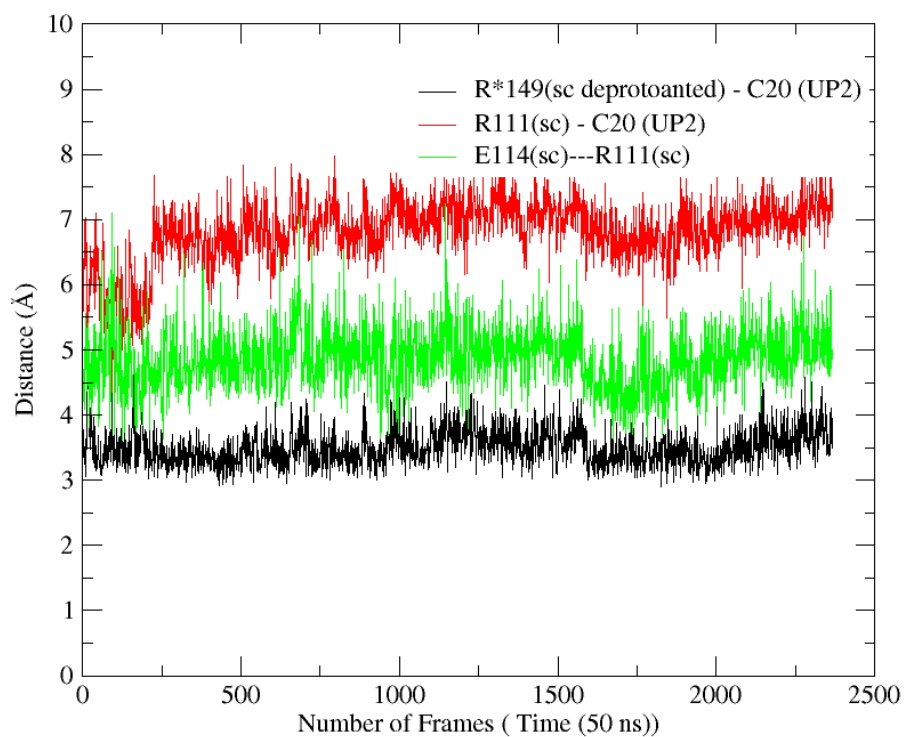
Figure S4.31 The MD simulation of R *149 deprotonated setup and its important interactions with substrate UP2. Note here the R111 is in its protonated normal state with chare of +1.
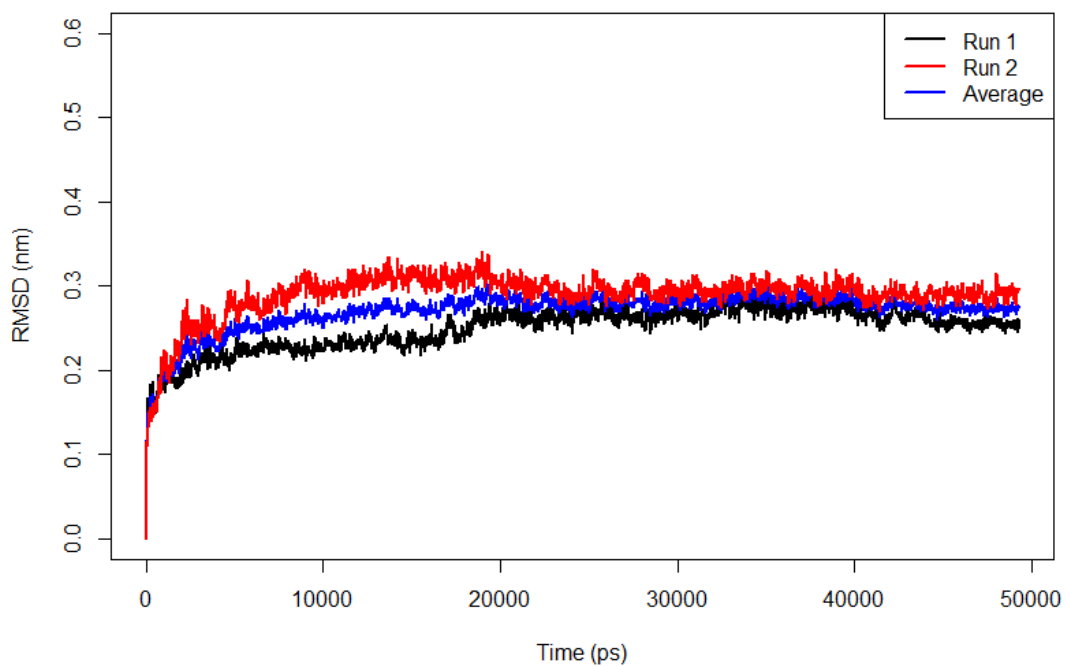
Figure S4.32 The deprotonated R*149 trajectory with average run for 50 ns. The average run is shown in blue colour.
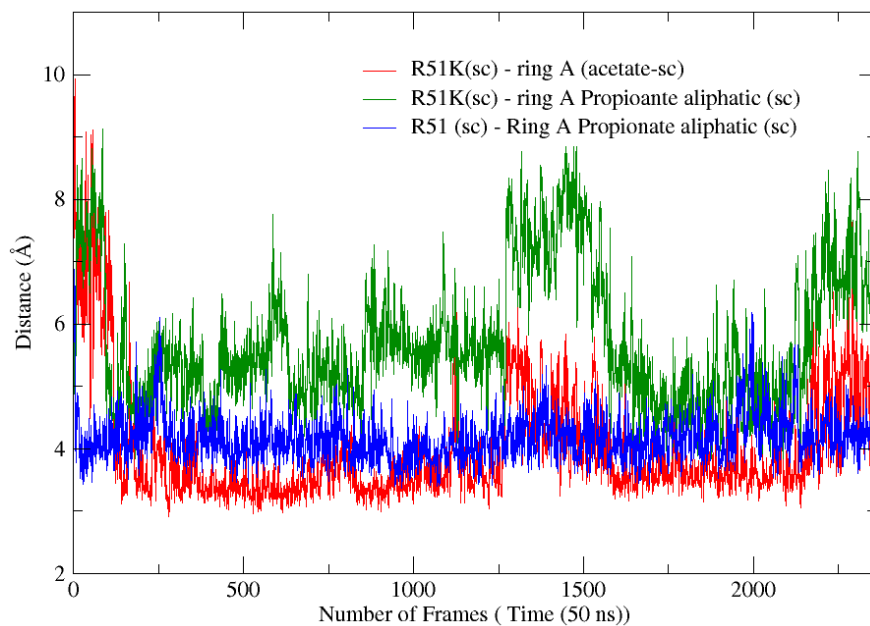
Figure S4.33 The interaction of mutant R51K with the substrate UP2
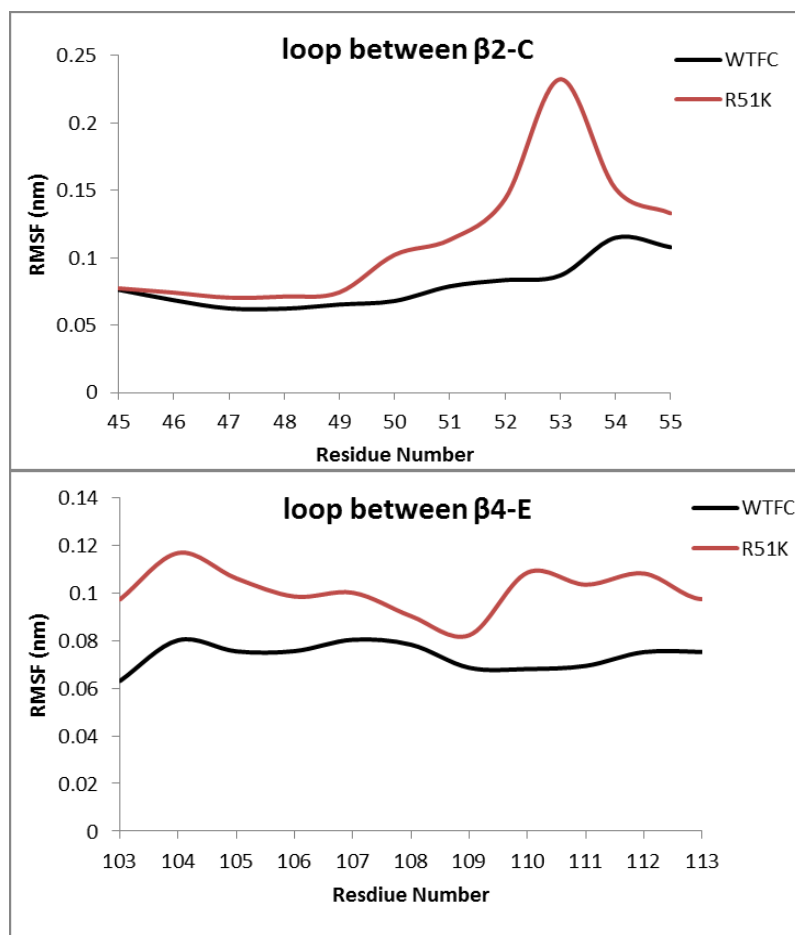
Figure S4.34 The RMSF value of the mutant R51K and the wild type NIRE
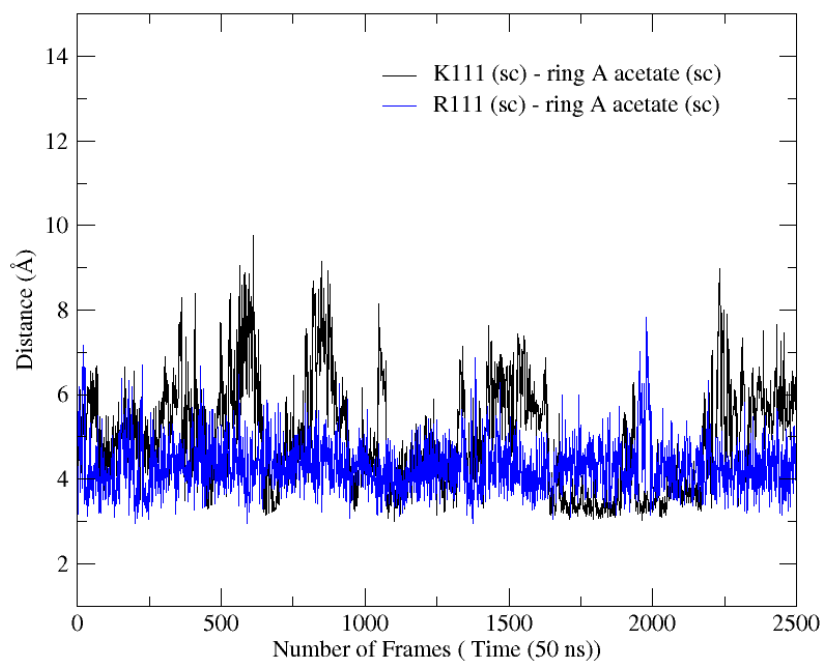
Figure S4.35 The interactions of the R111K mutant in comparison to wild type NIRE
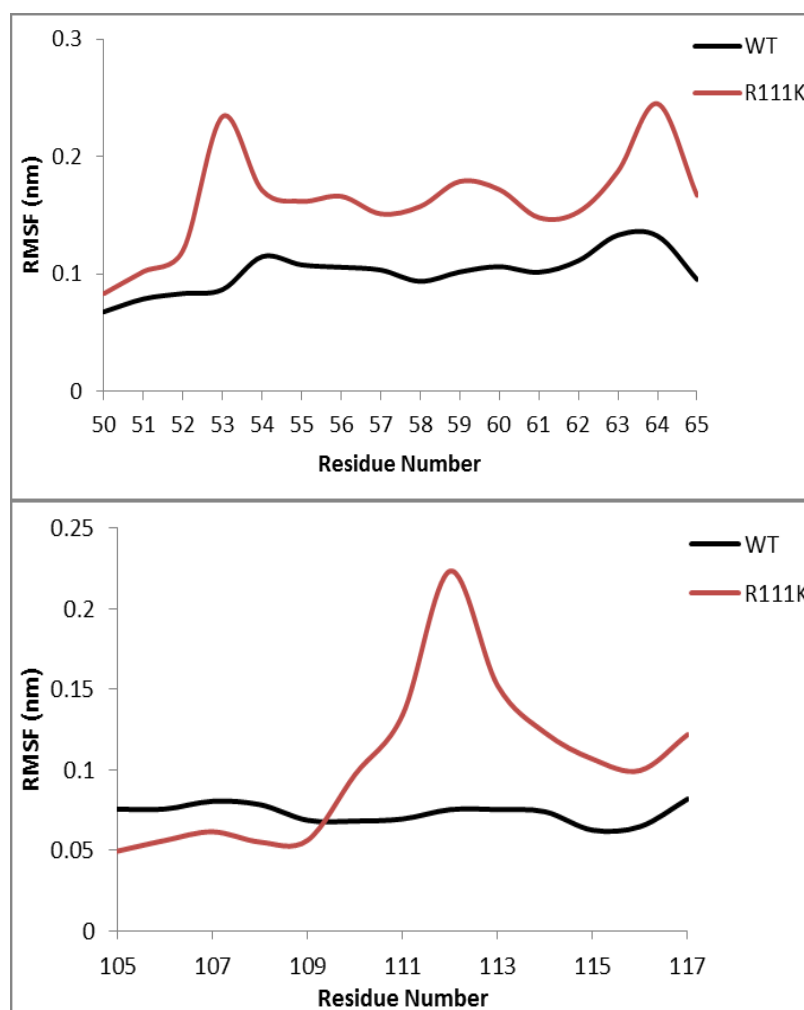
Figure S4.36 The residue show high RMSF in R111K mutant in contrast to WT FC
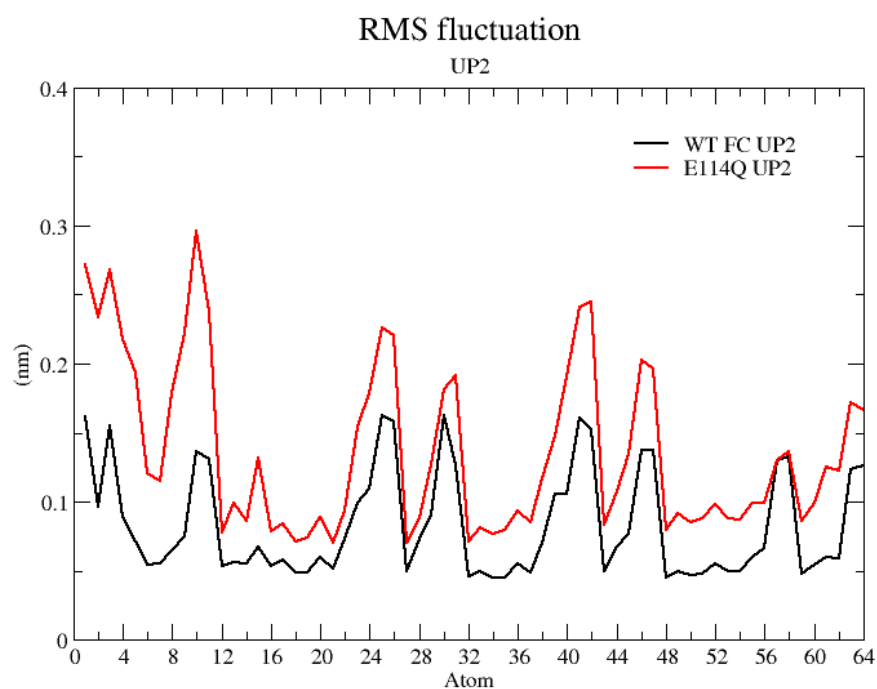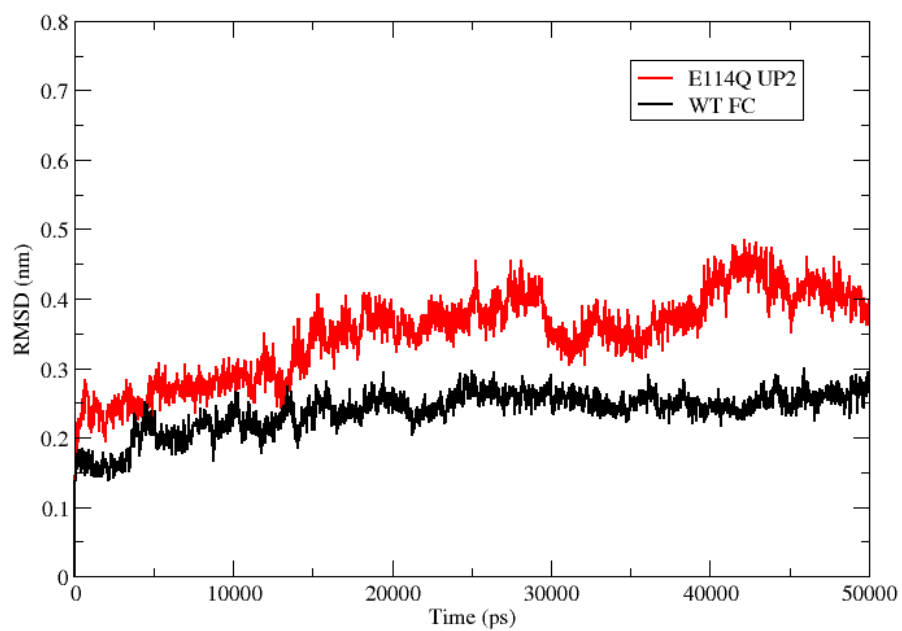
Figure S4.37 The RMSD and RMSF of substrate UP2 in mutant E114Q and WTFC
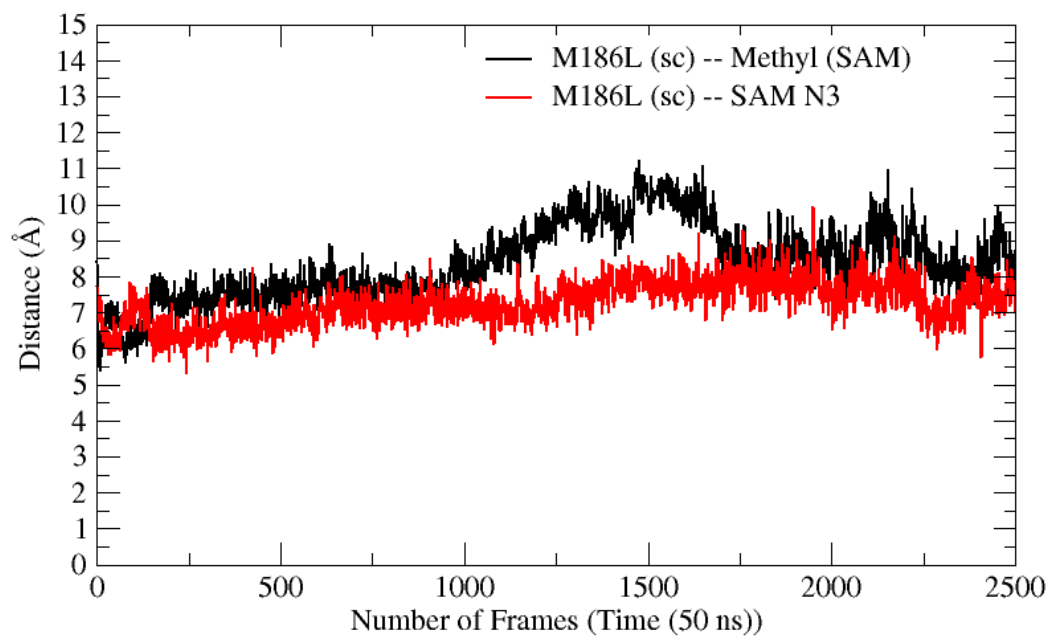
Figure S4.38 The interactions of the M186L mutant with cofactor SAM
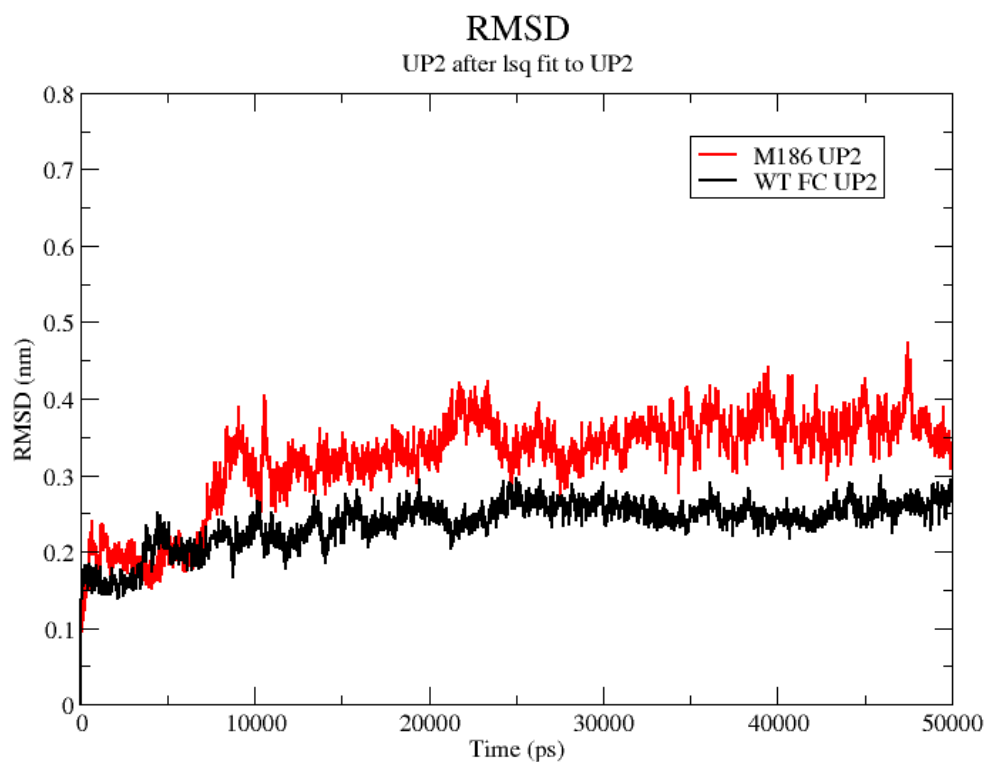
Figure S4.39 The RMSD profile of UP2 substrate using in M186L mutant
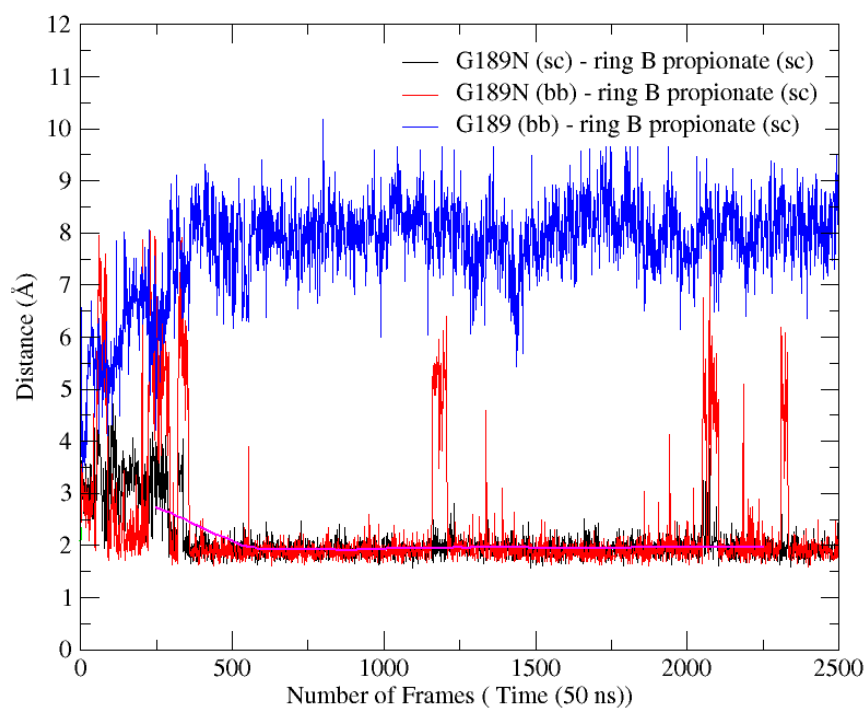
Figure S4.40 The interactions of the G189N mutant and comparison with WTFC G189 residue
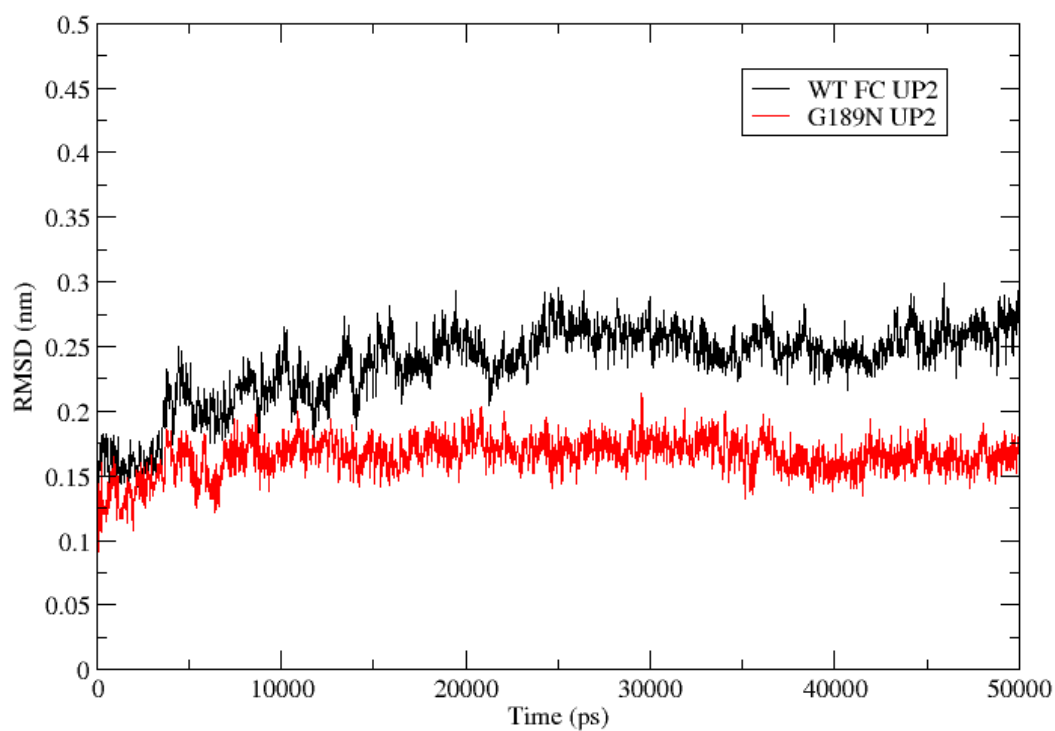
Figure S4.41 The RMSD of substrate UP2 in WTFC and mutant G189N for 50 ns trajectory
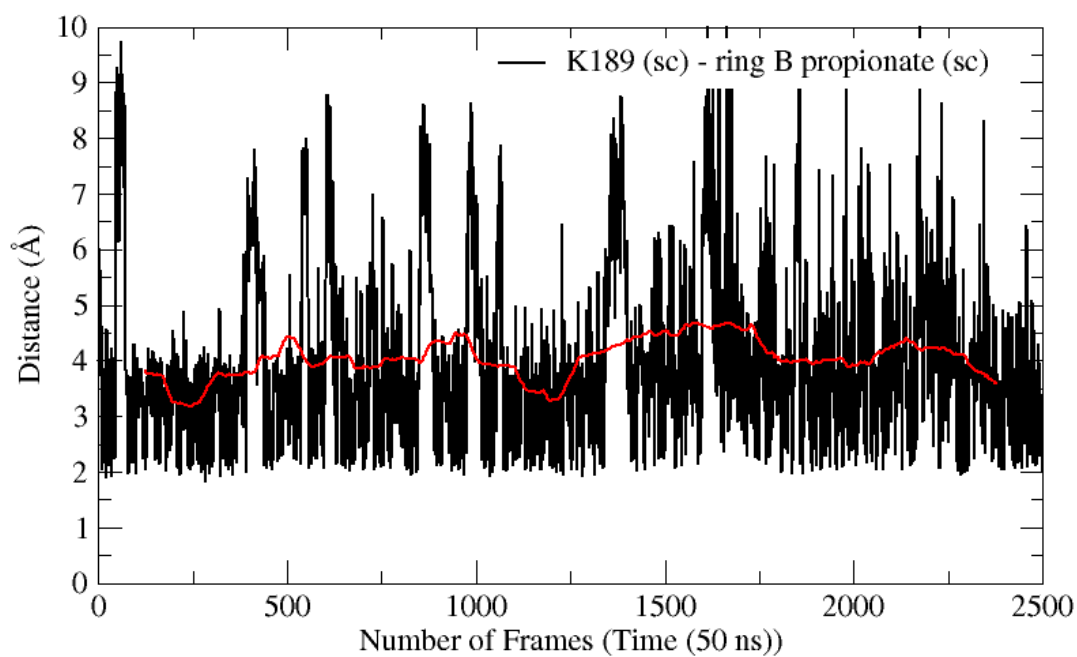
Figure S4.42 The electrostatic interaction of K189 in G189K mutant with the side chain of

ring B of propionate of UP2
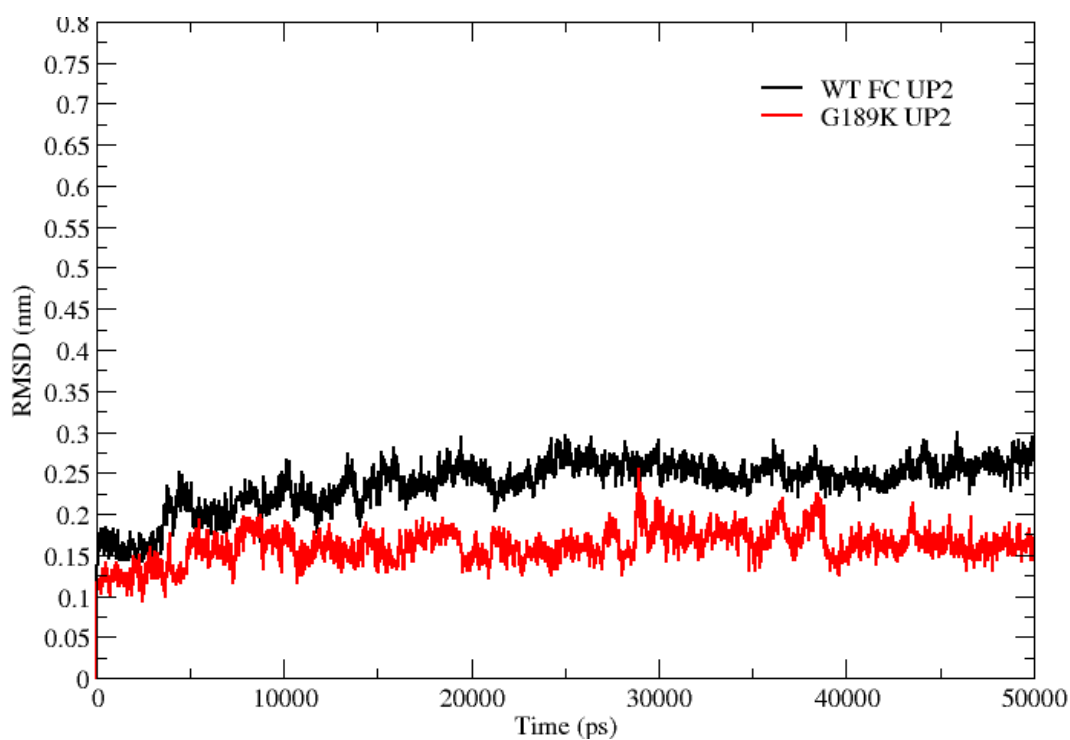
Figure S4.43 The RMSD of substrate UP2 for the WTFC and the G189K mutant
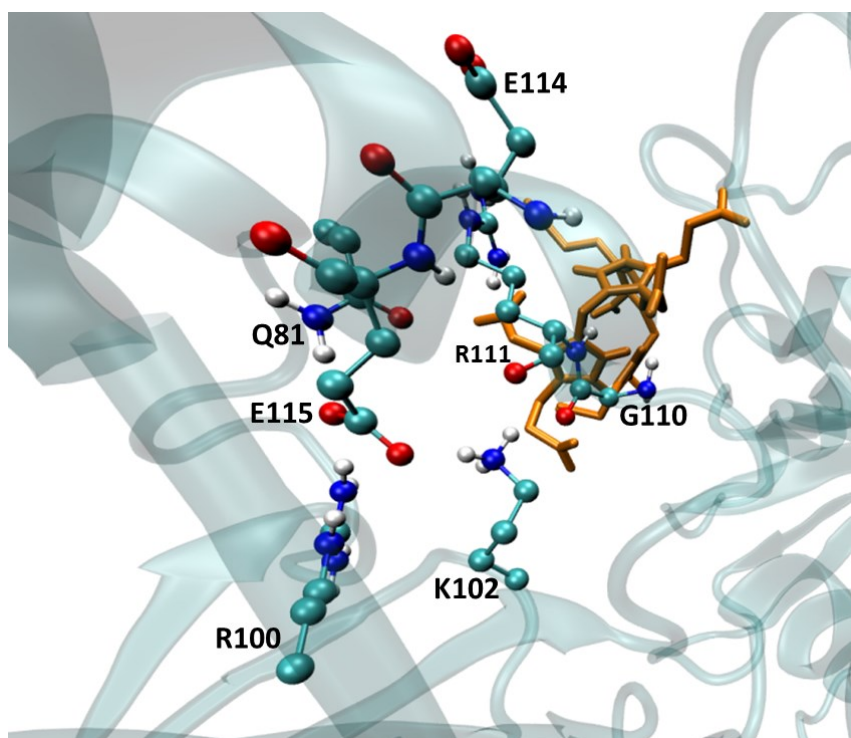
Figure S4.44 The interactions of K102 in the active site of the NirE wild type (WTFC) during MD simulation studies.
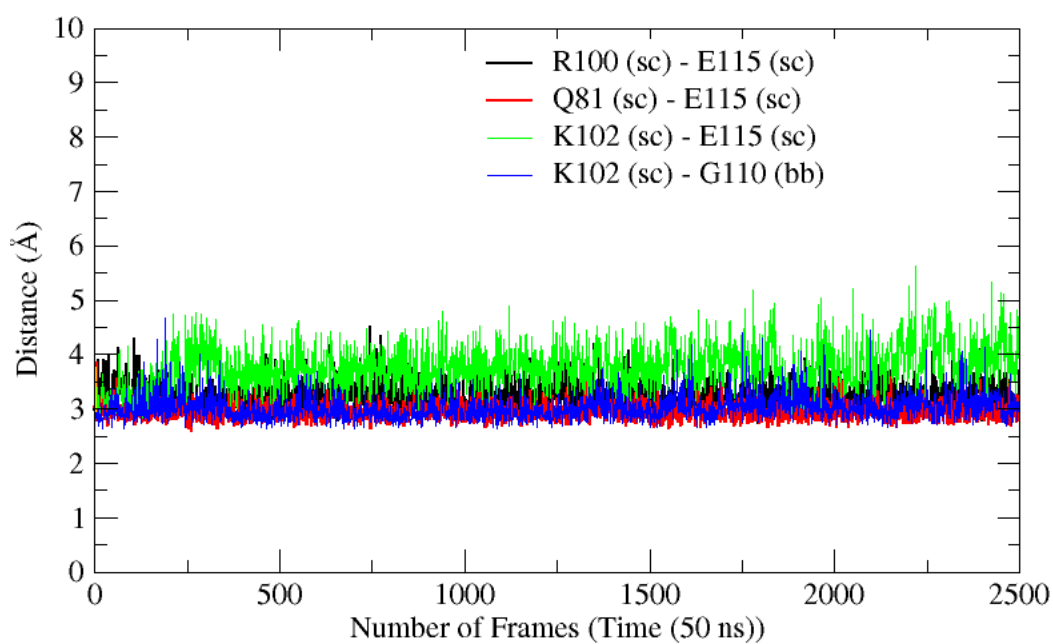
Figure S4.45 The interactions of K102 with the residues in vicinity of active site in WTFC

K102A mutant
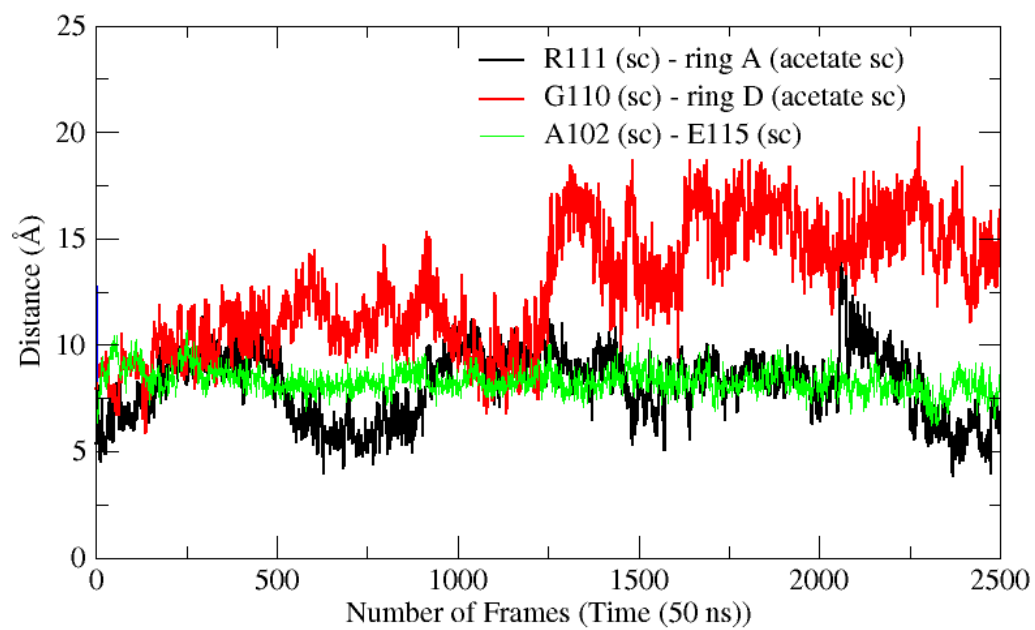
Figure S4.46 The interactions of K102A mutant.
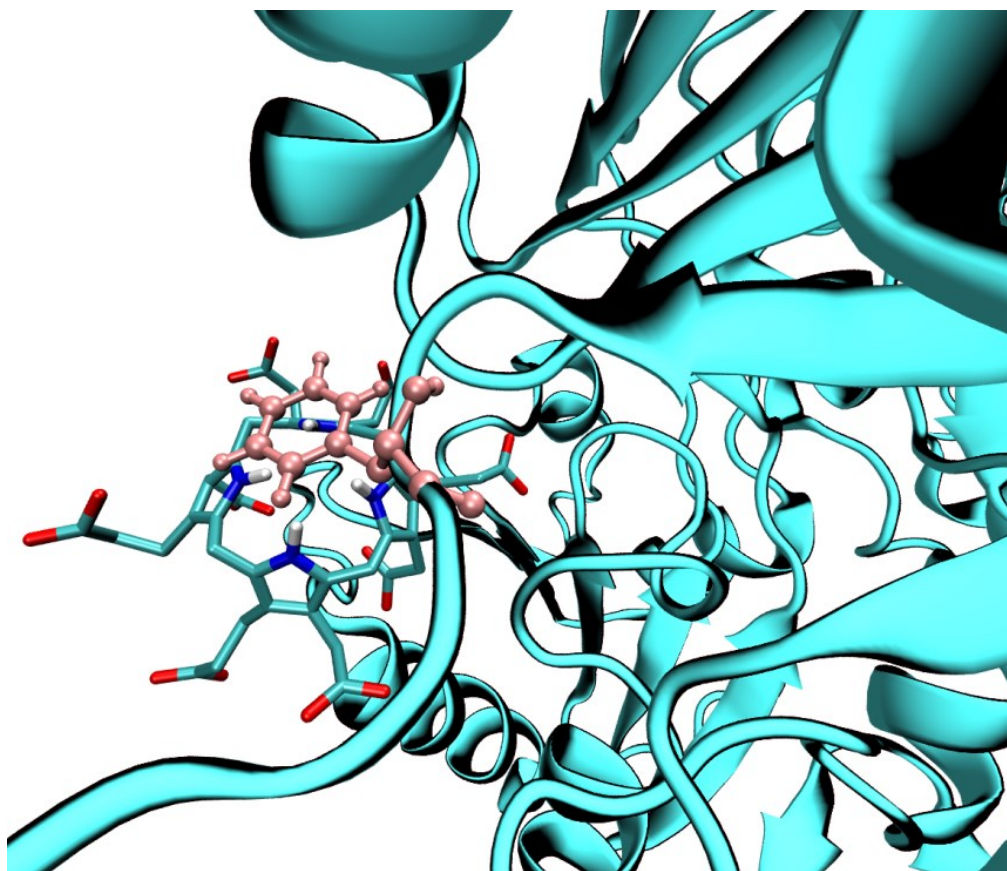
Figure S4.47 The head on overlap of the F161on UP22 in H161F mutant
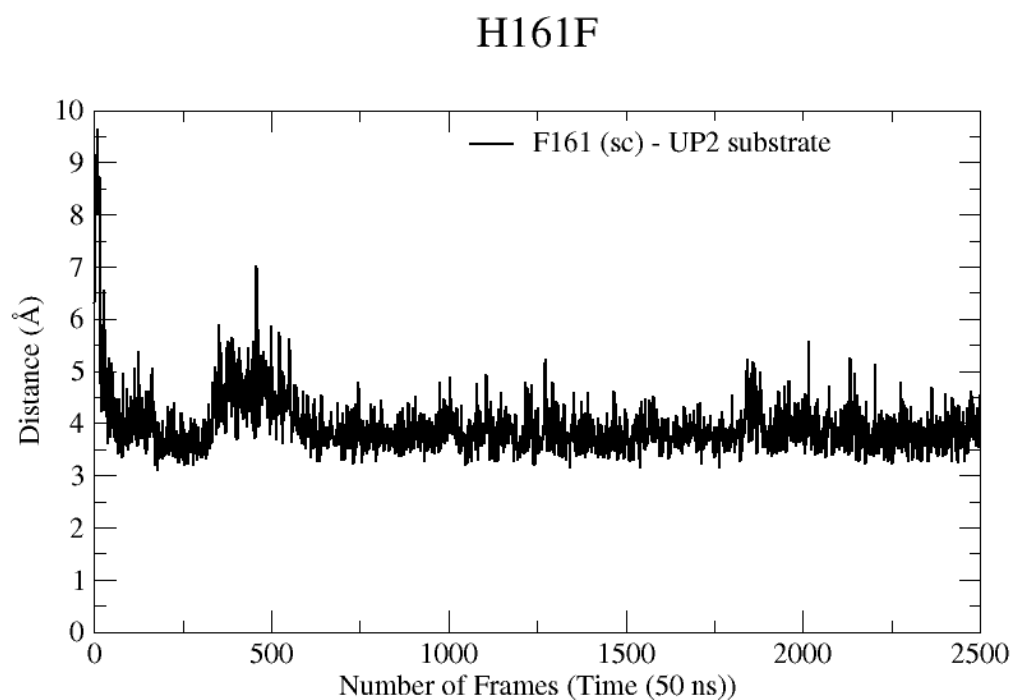
H161F

Figure S4.48 The pi pi stacking of the aromatic ring of the F161 in mutant with the UP2 ring system.
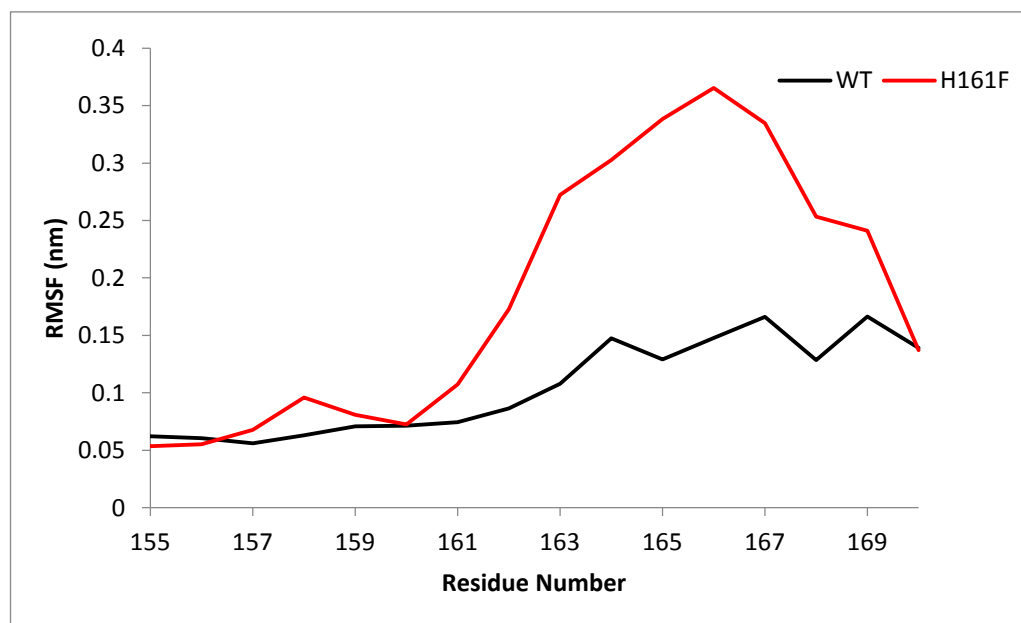
Figure S4.49 The RMSF of the local residue in vicinity of the mutant H161F

RMSD
C-alpha after lsq fit to C-alpha
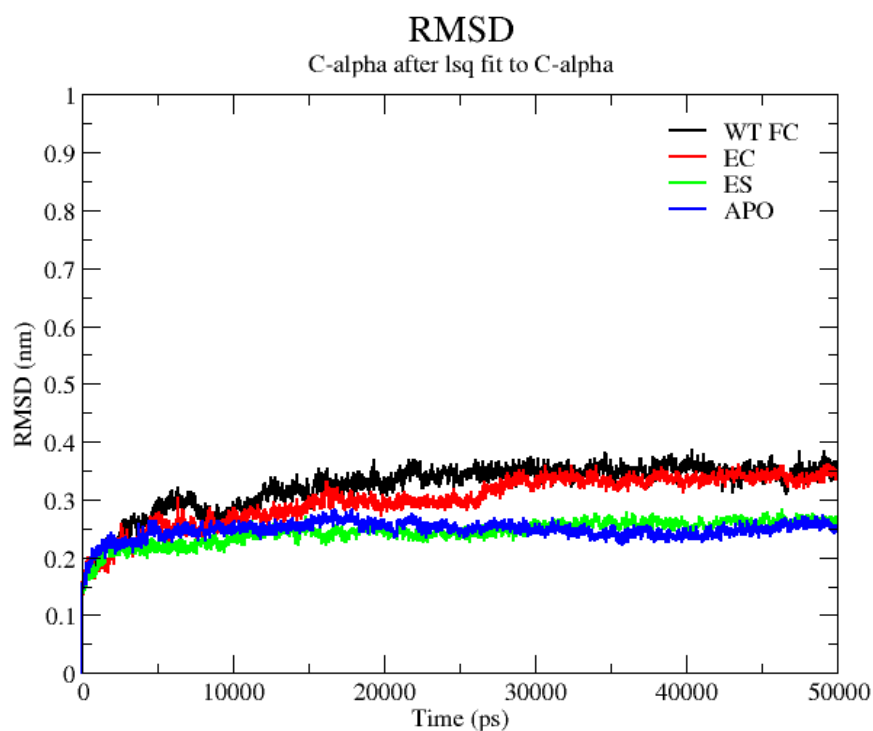
Figure S4.50 The RMSD of wild type NirE along with Apoenzyme, Enzyme substrate and Enzyme cofactor
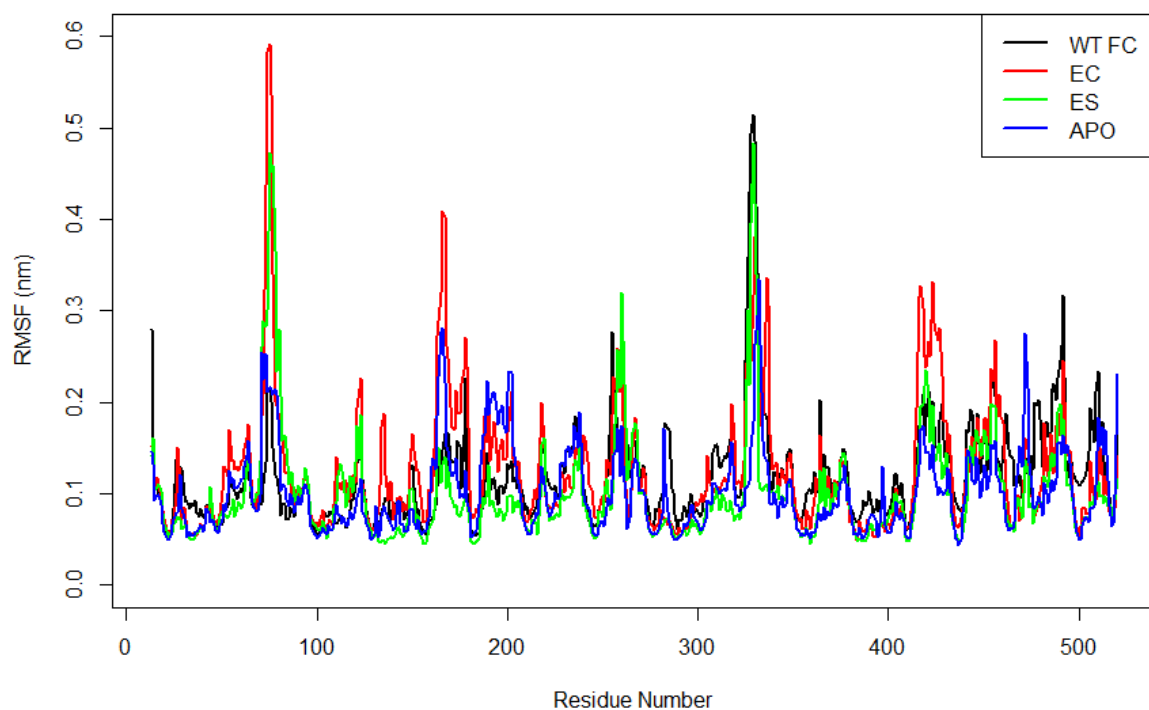
Figure S4.51 The RMSF plot of ES, EC, APO in comparison to WTFC using C α atoms for

50 ns simulation
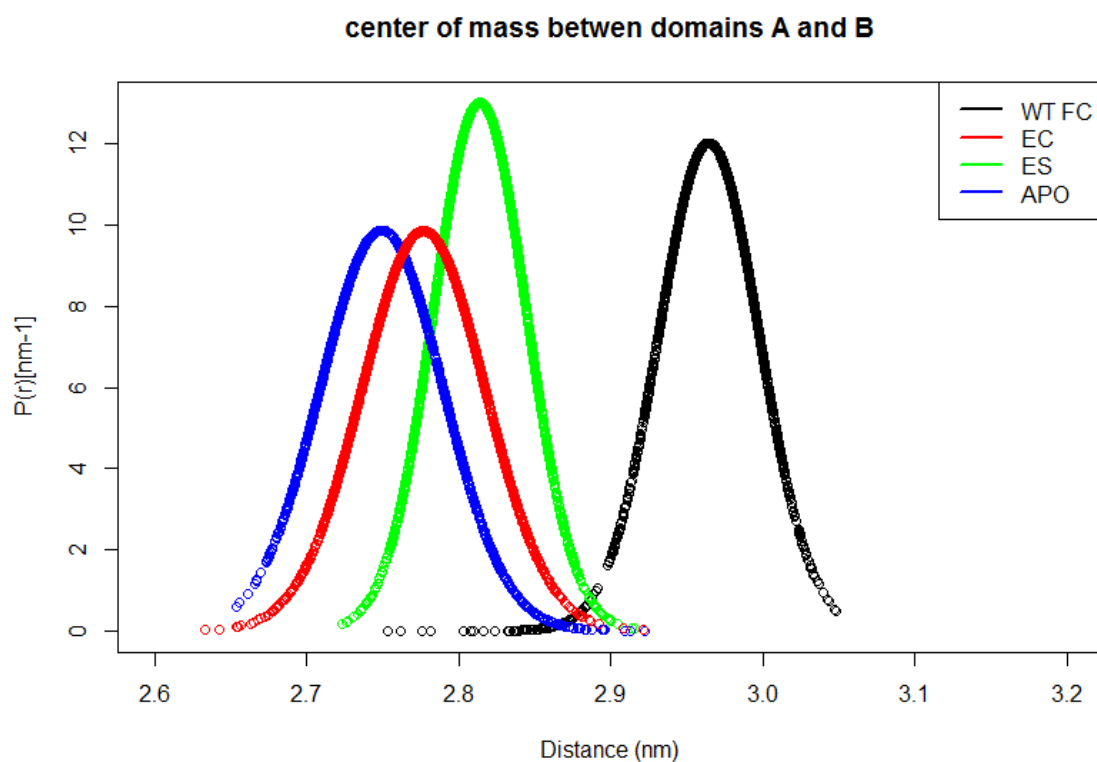
Figure S4.52 The normalized distrubtion of the centre of mass of domain A and B of monomer A of WTFC and the that of APO, ES and EC setups for 50 ns trajectory.
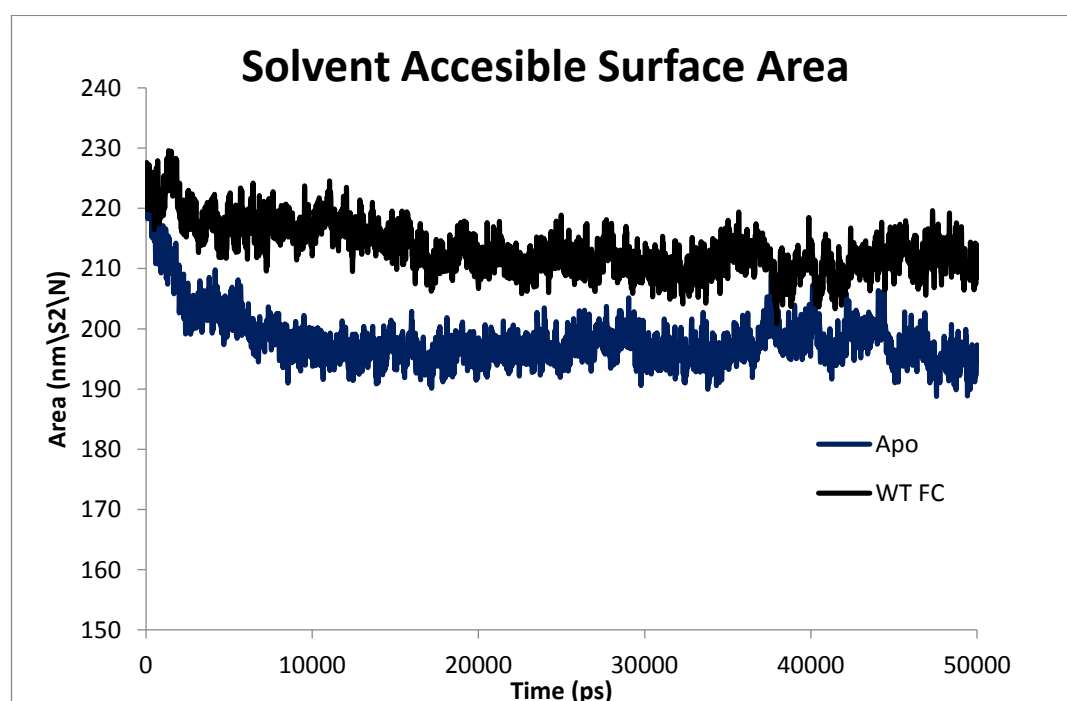
Figure S4.53 The solvent accessible surface area of the Apoenzyme and WTFC for 50 ns.
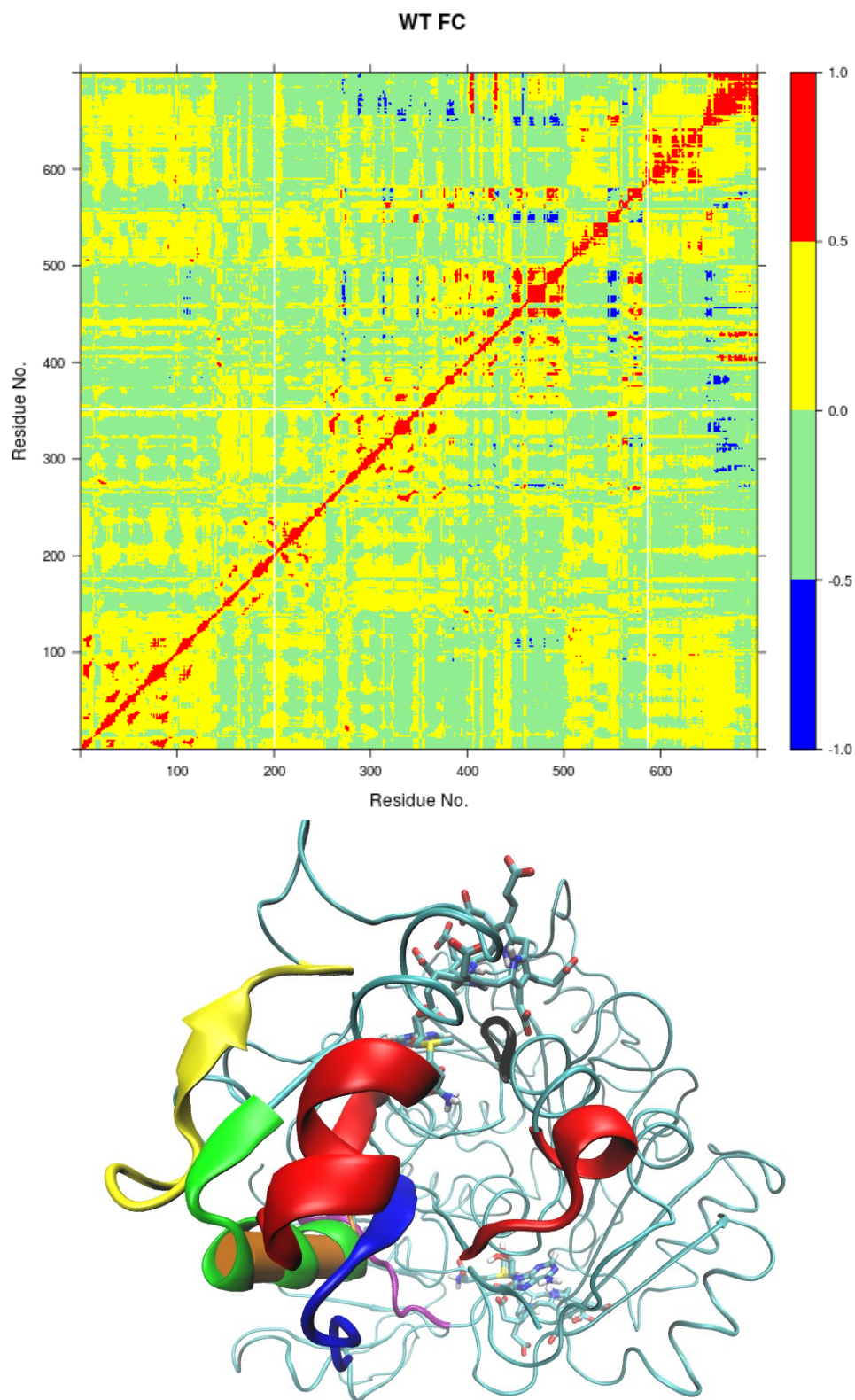
Figure S4.54 The Dynamic Cross Correlated motion analysis of the WT FC NirE for 50 ns.
The correlated motion is color coded (1) blue is correlated to red color, (2) green to yellow
color, (3) orange to purple,4 black color represent correlated motion towards the substrate
UP2 atoms especially the side chains. The plots were made using Bio3D package in R
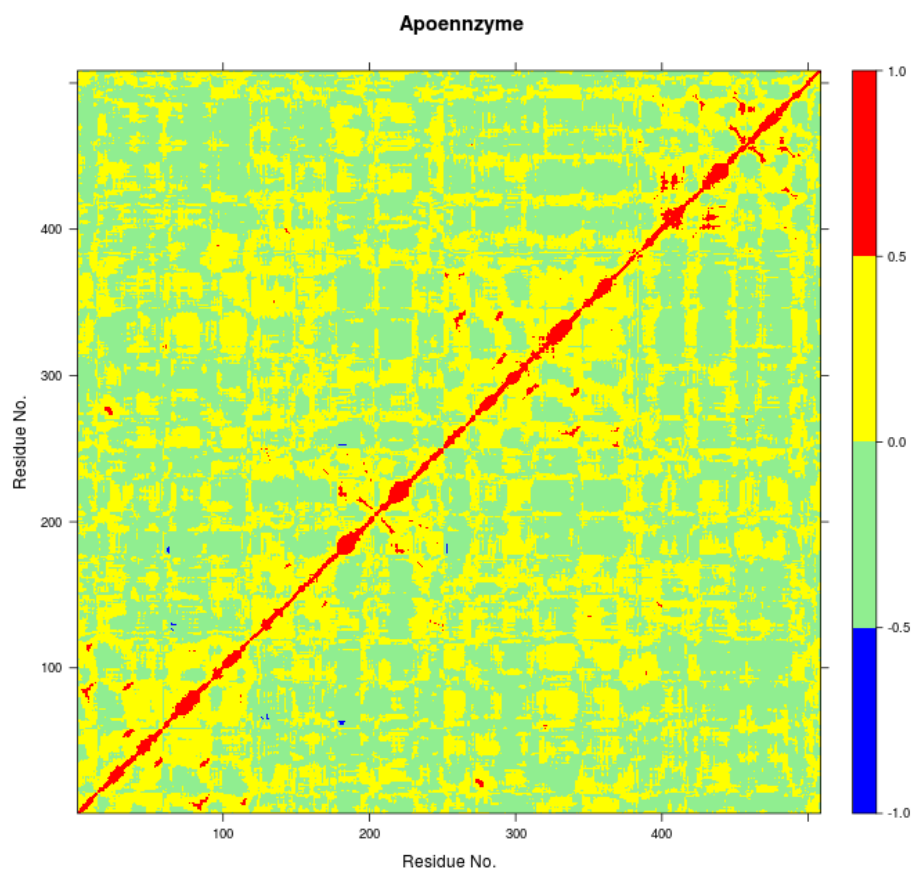
Figure S4.55 The Dynamic Cross Correlated motion analysis Apo enzyme for 50 ns
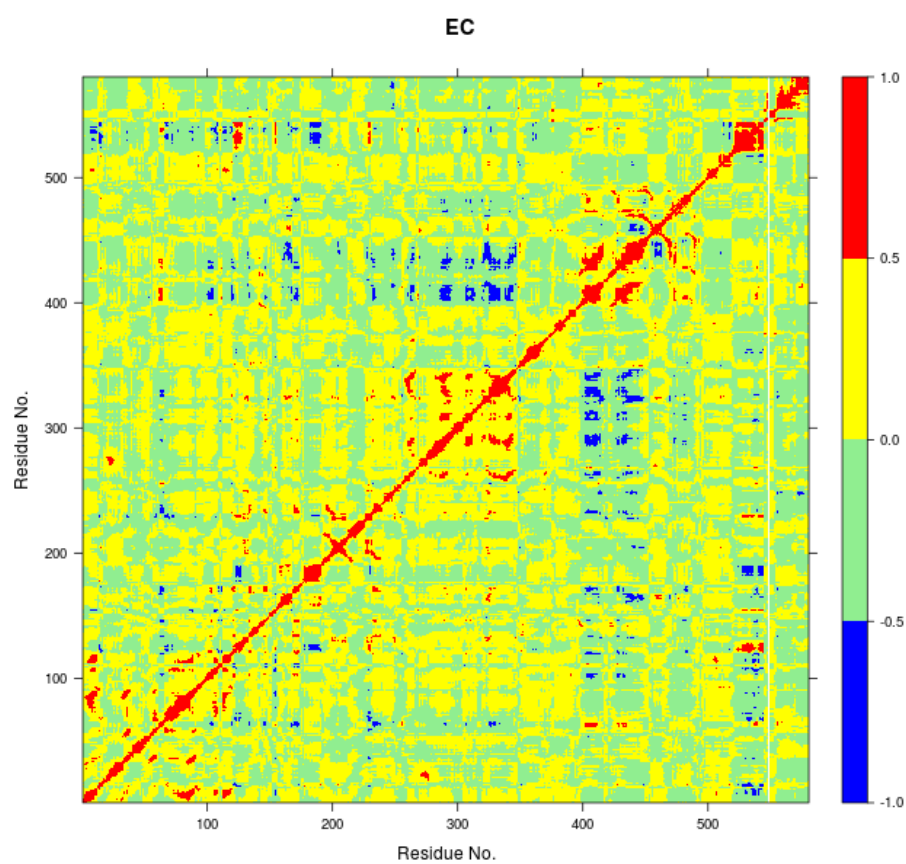
Figure S4.56 The Dynamic Cross Correlated motion analysis of EC complex

Figure S4.57 The anti-correlated motion is colour coded (1) blue is correlated to red colour, (2) red to red colour, (3) green to purple colour
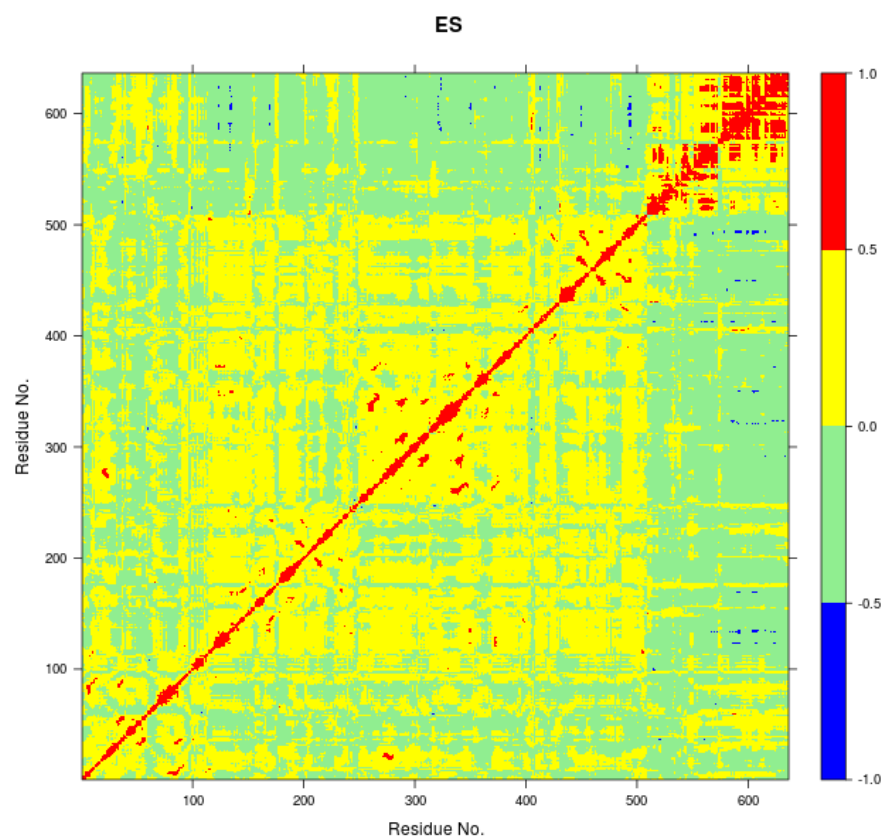
Figure S4.58 The Dynamic Cross Correlated motion analysis of ES complex

CHAPTER FIVE MATRIX METALLOPROTEINASE-1 AND LINKER REGION

5.1 INTRODUCTION

One of the primary components of the extracellular matrix (ECM) is collagen, which is also the most abundant protein in mammals[124]. There are at least 29 different types of collagens that occur in vertebrates. The collagen molecule consists of three polypeptide strands (α chains) that organize themselves in a ropelike triple-helix conformation, stabilized by inter-chain hydrogen bonding. The degradation of collagen is a key process in normal development and homeostasis; however, unbalanced collagenolysis contributes to numerous pathologies such as cancer, arthritis, and cardiovascular and neurodegenerative diseases. Amongst the enzymes capable of collagen catabolism in vertebrates are matrix metalloproteinases (MMPs). MMP-1, also known as interstitial collagenase or fibroblast collagenase, is a zinc and calcium dependent endopeptidase located in the ECM in vertebrates [125] [126] [127] [128]. MMP-1 degrades interstitial (types I-III) collagen into ¾ and ¼ fragments [129] [130]. The three-dimensional X-ray crystallographic structure of catalytically inactive (E200A) MMP-1 complexed with a collagen-model triple-helical peptide (THP) has been solved by Manka *et al.* (**Figure 5.1**) [18]. MMP-1 in this structure consists of the *N*-terminal catalytic (CAT) domain, the *C*-terminal hemopexin-like (HPX) domain, and the linker region connecting the CAT and HPX domains. The THP, with three strands designated as leading (L), middle (M), and trailing (T), is approximately 115 Å in length and makes extensive interaction with both the CAT and HPX domains [18]. In concert with the X-ray crystallographic "snapshot" of MMP-1•THP, dynamic information about MMP-1 interaction with triple-helices was obtained through a series of NMR spectroscopic structures and molecular docking calculations [131].

Molecular dynamics (MD) simulations have been applied to study the MMP family of enzymes. Diaz and coworkers observed a conformational shift from a closed X-ray crystallographic MMP-2 structure to an elongated structure using atomistic MD simulations on a 100 nsec time scale [132]. MMP-2 can adopt extended conformations in solution prior to collagen hydrolysis [133]. MD has provided insight into the degradation of type I collagen fibrils by MMP-8 [134], but limited efforts have been made to study MMP-1. Presently, the largest collection of experimental data for MMP-catalyzed collagenolysis has been obtained using MMP-1 [135] [136] [18] [137] [131] , yet there is still much to be learned as to how conformational flexibility and dynamics influences MMP-1 structure-function relationships, and specifically to provide insight on the effects of dynamics on the interactions between the MMP-1 CAT and HPX domains and with collagen triple-helices. In addition, it is important to understand how the composition and the flexibility of the linker region influence the interactions between the MMP-1 CAT and HPX domains and with collagen. In order to provide insight into MMP-1-catalyzed collagenolysis, we herein performed a set of long-range atomistic (AT) MD simulations on an MMP-1•THP complex. Effects of the linker region on inter-domain interactions and interactions with the substrate were studied by MD simulations of *in silico* engineered MMP-1 with the linker of MMP-14/MT1-MMP or MMP-13.

Figure 5.1. (A) The 3D structure of human MMP-1 (E200A) complexed with a THP (PDB 4AUO [18]) drawn using UCSF Chimera [118]. MMP-1 (displayed in silhouette round ribbon) consists of the CAT domain (orange), inter-domain linker (red), and HPX domain (grey). The L, M, and T strands of the THP are shown in tube representation in cyan, green, and red, respectively. The zinc and calcium ions bound to the enzyme are shown in spherical representation in blue and purple, respectively. (B) 2D topology diagram of (E200A) MMP-1 made using Pro-origami tool using STRIDE algorithm for secondary structure prediction [138]. The color gradient of blue and red are used for the N- and C-

termini of the protein. The β-sheets are represented numerically and the α-helices alphabetically.

5.1 METHODS

5.1.1 Initial structure preparation

The coordinates of the wild type MMP-1 bound to a THP model of the type II collagen MMP cleavage site were obtained from the Protein Data Bank [85] (PDB: 4AUO [18]). SwissPDBViewer [86] was used for adding missing atoms and selecting one from the alternative side chain orientations [86]. The active form of MMP-1 was made by substituting the A200 with E200 using the Modeller [139] program. The linker of MMP-14 and MMP-13 were modelled into the linker region of MMP-1 by using the Modeller [139] program. The numbering of the THP in the MMP-1•THP complex and subsequence simulations was assigned as 763-795, based on the sequence numbering within the triple-helical region of type II collagen, instead of 963-995 as in the X-ray crystallographic structure.

5.1.2 Molecular dynamics simulations

In order to explore the dynamic properties of MMP-1, we performed an extensive set of MD simulations for 300 nsec (wild type MMP-1) and 100 nsec for the modified linker systems (**Table S5.1**) by using Gromacs 4.5.5 package [3] [4, 5] with GROMOS96 43a1 [57] force field. The protonation states of His residues in the protein molecule were assigned based upon the optimal hydrogen bonding conformation performed in Gromacs using pdb2gmx. However H149, H164, H199, H203, and H209 were protonated at their delta and H177 at the epsilon position according to their local environment in the vicinity of Zn^{2+}. The Zn^{2+} and Ca^{2+} ions were restrained in MD simulation according to their X-ray crystallographic structure distance using harmonic potential. *In vacuo* energy minimization was performed to remove steric clashes in the crystal structure first by using the steepest descent [91] and then by using conjugate gradient [140] until the maximum force was found smaller than 100 kJ/mol^{-1}/nm^{-1}. The editconf command was used to define the dimension of the cubical box and the protein molecule was placed in the box. The periodic

boundary conditions were then applied to treat all the parts of the system equally both at its interior and edges. The box size was set to ensure a distance of at least 1.0 nm between the protein and the box boundaries. The energy minimized protein structure was then solvated by using Single Point Charge (SPC) [141] water model in the cubical simulation box by using Genbox command. The system was neutralized by adding Cl^- to MMP-1 and the linker-modified enzymes (Table S1). In order to relax the solvent molecule and remove constrains from the entire system, the energy minimization of the whole system was performed using first the steepest descent and then the conjugate gradient algorithm until the maximum force was found smaller than 100 $kJ/mol^{-1}/nm^{-1}$. The energy minimized structure was then subjected to position restrain dynamics for 50 psec. The simulation was performed in NVT ensemble (constant Number of particles, Volume, and Temperature) [94] at constant temperature of 300 K with time step of 0.002 psec. The Berendsen temperature coupling and Parrinello-Rahman pressure coupling were used to keep the system at 300 K and 1 bar pressure during the simulation procedure. The productive MD was carried out using NVT ensemble and the initial velocities for MD simulation were drawn from Maxwell velocity distribution at 300 K. The MD was performed with an integration time step of 0.002 psec. The Particle Mesh Ewald (PME) [95] method was used for electrostatic interactions with Coulomb cut off of 1.0 nm, Fourier spacing of 0.135 nm tolerance of 1e-5 and an interpolation order of 4. The Lennard Jones potential was employed for the treatment of van der Waals interaction with cut off distance set to 1.4 nm. The LINCS algorithm [96] was utilized to keep all the covalent bonds involving hydrogen atoms rigid.

5.1.3 Analysis of Molecular dynamics simulations

The analyses of the trajectories obtained from the simulations were performed using tools from the Gromacs software package. The Root Mean Square Deviation (RMSD) of Cα atoms of the protein with respect to minimized crystal structure, Root Mean Square

Fluctuations (RMSF), electrostatic interactions, hydrogen bonding, Solvent Accessible Surface Area (SASA), and cluster analysis were performed. The visualization of MD trajectories and the structures were performed using VMD software [97]. The Bio3D [98] package in R was used to produce principal component (PC) analysis and domain cross correlation. Center of mass was calculated using Cα atoms of the CAT and HPX domains over simulation time using g_dist in Gromacs and also CALCOM [142] to verify the results obtained from the g_dist in Gromacs.

5.1.4 Dynamic Cross Correlation Analysis

Correlated and collective atomistic motions play an important role in the functionality of dynamic biomolecular systems [105]. The Cα atoms of the protein and THP were used to create a cross correlation C_{ij} matrix using MD trajectory. The cross correlation between the *i*th and *jt*h an atom were represented by a *Cij* matrix and was extracted by the following equation:

$$C_{ij} = \langle \Delta r_i . \Delta r_j \rangle / \langle \Delta r^2_i \rangle^{1/2} \langle \Delta r^2_j \rangle^{1/2}$$

The Δr_i and Δr_j are the displacement vectors corresponding to *ith* and *jth* atom and the angle brackets denote an ensemble average. The C*ij* matrix ranges from +1 to -1, where the positive value represented the correlated motion and the negative values represented the anticorrelated motion.

5.3 RESULTS AND DISCUSSION

5.3.1 Conformational Dynamics of MMP-1•THP Complex

The overall stability of the MMP-1•THP complex and changes associated within the internal structure and geometry of the individual CAT and HPX domains and linker region were assessed by computing the RMSD of the backbone Cα atoms with respect to the minimized X-ray crystallographic structure (PDB 4AUO). MMP-1•THP equilibrated after 10 nsec with an average RMSD value of 4.7 Å (**Figures 5.2, S5.1A**, and **S5.1B**, **Table**

S5.2). This RMSD is comparable to the RMSD of 4.2 Å of the collapsed (closed) form of MMP-1 in complex with the THP obtained from NMR studies [131]. Three different runs of MMP-1 (**Figure S5.1B**) were performed in order to enhance the conformation sampling and avoid statistical errors, with an average value of 5.1 Å obtained (**Table S5.3**). Full length MMP-2 in a compact structure (PDB 1CK7) exhibited an average RMSD value of 4 Å during the first 70 nsec of simulations performed by Diaz and co-workers [132]. The MD simulation of MMP-2•fTHP-5 performed by Díaz and coworkers showed on average RMSD of ~6 Å [143]. The MMP-1 CAT domain equilibrated at ~10 nsec and was stable throughout the simulation with RMSD less than 2 Å (**Figure S5.1**). The HPX domain showed a drift at ~60 nsec with an average value of 2.6 Å and then stabilized for the entire length of the simulation (**Figure S5.1**). The linker region was also stable with an RMSD of 2 Å and showed equilibration at ~10 nsec (**Figure S5.1**). The linker region underwent limited structural deviation indicating that the inter-domain arrangement in MMP-1 in solvent is maintained close to the X-ray crystallographic structure throughout the simulation. The RMSD of the CAT and HPX domains and the linker of MMP-1 were comparable to the MD results from the simulation of the CAT domain (2.9 Å), HPX domain (2.4 Å), and linker region (2.1 Å) of MMP-2 complexed with THPs [144]. The RMSD data suggested that the HPX domain shows greater structural deviation in comparison to the CAT domain. The simulations of MMP-2 also showed higher structural deviation of the HPX domain (1.7 Å) in comparison to CAT domain (1.1 Å) after 200 nsec [133]. The CAT and HPX domains in MMP-2•fTHP-5 showed RMSD values of ~2 and ~4 Å, respectively, and the HPX domain showed greater structural deviation in contrast to CAT domain [143].

The THP exhibited a high RMSD value (7 Å) relative to components of MMP-1 (**Figure S5.1**) indicating a significant structural change in comparison to the X-ray crystallographic structure. The THPs in the MMP-2 complex also exhibited a higher RMSD (on average 4.7

Å) [144]. In order to evaluate the compressing effect of the enzyme on the triple-helix dynamics, free THP in water was simulated. The average RMSD value of the THP in water was 22 Å with respect to 7 Å in the MMP-1•THP complex for 50 nsec simulation, which implied a restraining effect of the enzyme on the THP and also highlighted the extent of conformational flexibility and dynamics of the THP in solution. Comparison of the averaged structures of the THP in water (based upon cluster analysis of the most populated cluster) and in the enzyme•substrate complex indicated considerable conformation changes (**Figure 5.3**), although the triple-helical nature of the THP in water was maintained by consistent interchain hydrogen bonding. Enhanced conformational flexibility of THPs during MD has been reported previously [145] [146].

The flexibility of MMP-1 and THP residues was assessed using RMSF analysis (**Figures 5.4A and S5.2A**). Average RMSF values were 1.14 Å for the CAT domain, 2.15 Å for the linker, 1.40 Å for the HPX domain, and 2.0 Å for the THP. The linker had fluctuations which peaked at 3.1 Å while the HPX domain displayed greater fluctuations relative to the CAT domain. The RMSF value of linker region in the compact conformation of MMP-2 was approximately 2.4 Å [147]. The RMSF values of residues belonging to the HPX domain of MMP-2 also exhibited greater fluctuations in contrast to its CAT domain and linker [133, 144]. The loop region between parallel β-strands β2, β3, and β4 presented high fluctuations in the MMP-1 CAT domain. The αC of the CAT domain and the loop connecting the linker to β6 of the HPX domain showed increased flexibility. Increased fluctuations in the HPX domain was seen in the region encompassing residues of β17 and the loop between β17 and β18. The basal level fluctuations for MMP-1 were considered to be 1.4 Å (based upon the mean value), where 43% of the residues showed fluctuations greater than 1.4 Å and ~56.9% of the residues showed fluctuations less than 1.4 Å (**Table S5.2**).

The highest peak in the RMSF plot was from the *N*- and *C*-termini of the THP (**Figure 5.4A**). The RMSF profile of the scissile bond (G775-L776) also showed high flexibility for the M and T strands with respect to the L strand (**Figure S5.2A**). In the MD simulations of MMP-2 the *N*- and *C*-termini of the THP bound to MMP-2 also exhibited high RMSF in the range of 8 to 16 Å, respectively [144].

PC analysis can provide information on the overall dynamics of the THP in water. PC1 and PC2 accounted for ~90% of the overall motion of the THP in solvent (**Figure S5.2B**). The maintenance of triple-helical structure and PC analysis of the THP in water shows good agreement with the fluorogenic THP simulation in aqueous solution where the first PC accounts for the twist motion of the triple-helix and the triple-helix is maintained throughout 50 nsec trajectory [148]. The projection of PC1 on the THP indicated bent or twist conformation as has been described by cluster analysis. The dynamics cross correlation analysis of the THP chains showed significant correlated motion through the 50 nsec trajectory (**Figure S25.C**). The extent of both positive and negative motion were in accordance with the maintenance of triple-helical structure during simulation.

The L strand of the THP approached closest to the active site in the simulation and exhibited the largest bending out of all of the THP strands (**Figure S5.3**). Experimental studies on MMP-1[131, 137] [136] suggested that the active site of MMP-1 could not accommodate the 15 Å diameter of the collagen triple-helix. Therefore, the sequential hydrolysis of single collagen strands occurs through the binding of the triple-helix to the HPX domain and allowing inter-domain conformational flexibility through the linker to facilitate the reorientation of CAT domain to position itself to perform catalysis. The reaction mechanism of collagen hydrolysis by MMPs requires the carbonyl group of the scissile peptide bond to coordinate with the catalytic Zn^{2+} in active site [147, 149]. The scissile bond between residues G775 and L776 of the L strand is located 8.8 Å from the

catalytic Zn^{2+} in the CAT domain; therefore, the X-ray crystallographic structure of MMP-1 bound to the THP is unproductive due to absence of the scissile bond near the S_1-S_1' subsites in the CAT domain. In the MD simulations, the bond distance decreased on average to 6.7 Å (**Figure S5.4**) but still does not represent the active complex.
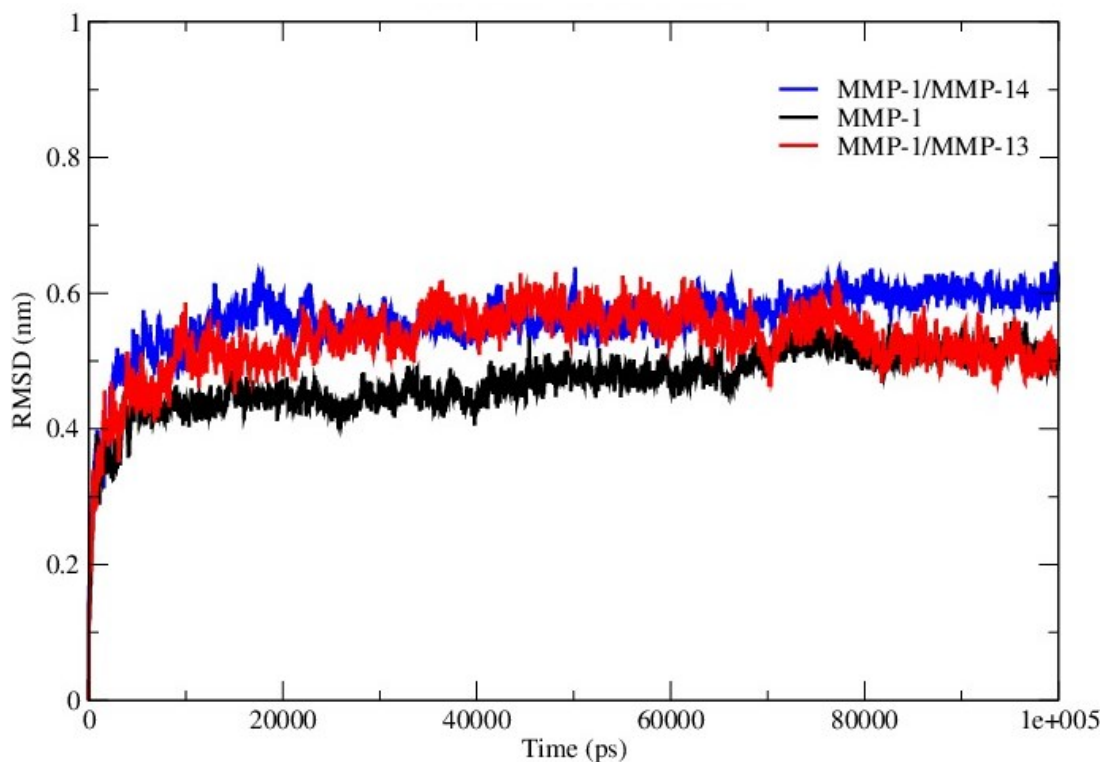
Figure 5.2. The RMSD of all Cα atoms of the MMP-1•THP complex. The RMSD of MMP-1 is compared with MMP-1 with the linker of MMP-14 (MMP-1/MMP-14) and MMP-1 with the linker of MMP-13 (MMP-1/MMP-13).
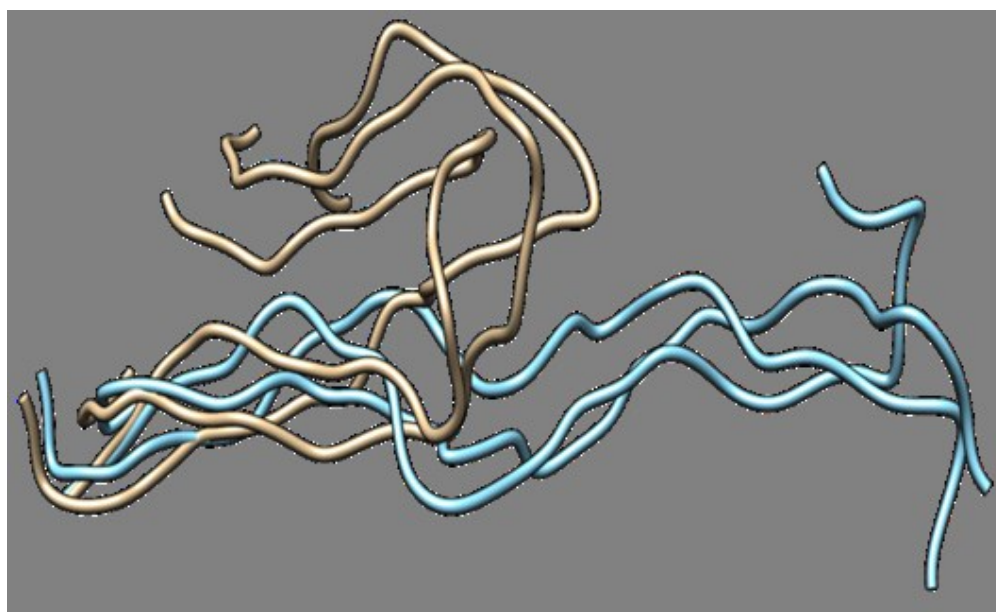
Figure 5.3. Superposition of the THP in water versus the THP in the MMP-1•THP complex. The structure of the most populated cluster of THP in water (brown) was aligned against the structure obtained from the most populated cluster of THP in the MMP-1•THP complex (cyan).

5.3.2 Triple-Helix Interactions with the CAT and HPX Domains

It has been demonstrated experimentally that specific residues present in blades 1 and 2 of the HPX domain interact with the collagen triple-helix [131] [18, 135] [136]. In the present simulation specific residues within blade 1 (residues 250-300) belonging to β6-β9 (**Figure 5.1B**) interacted with the THP (**Figure S5.5**). New sets of interactions emerged between the L strand of the THP and the HPX domain. The backbone of L295 makes hydrogen bonds with the side chain of R780 with an average distance of 3.5 Å. The side chain of N296 participated in hydrogen bonding with the backbone of G781 and hydrophobic interaction with the aliphatic side chain of R780. The guanidino group of R780 had a close interaction with the carboxylic group of E294 (average distance of 4.5 Å) with a weak electrostatic interaction between the residues.

The X-ray crystallographic structure revealed a hydrophobic cluster formed by V300 and F301 side chains with V783 of the L THP strand [18]. This interaction was lost in the MD simulation and a new hydrophobic interaction between the side chains of P303 and I782 was established (**Figure S5.6**). A hydrogen bond also emerged between the side chain of Q335 and G784 (L strand) (**Figure S5.6**). R780 to G784 are located in the THP strands which are most bent during the simulation. R780 makes extensive interactions with the HPX domain and could be a reason why the THP is so bent during the simulation. The M strand of the THP made most of the interactions with the HPX domain in similar fashion to the X-ray crystallographic structure; nonetheless, new interactions emerged from the

simulation studies, specifically involving T270 and N296 (**Figure S5.7**). R272 made hydrogen bonds with the backbone and side chain of O786 (M strand) and R789 (M strand), respectively, as described in the X-ray crystallographic structure. The aliphatic region of R272 was involved in formation of a hydrophobic cluster with the side chains of R789 and L785 of the M THP stand (**Figure S5.8**).

We previously demonstrated experimentally that the MMP-1 HPX domain binds to the THP with 1:1 stoichiometry [131]. This suggested that the HPX domain binds to more than one THP chain. The MMP-1 X-ray crystallographic and NMR structures [18] [131] showed the interaction of L and M strands of the THP with residues of blades 1 and 2 of the HPX domain. Our MD simulation also showed the existence of interactions between the HPX domain and two THP strands (L and M).

The CAT domain mainly interacts with the L strand of the THP (**Table S5.4**). The X-ray crystallographic structure indicated that the Y221 backbone formed a hydrogen bond with the side chain of Q779 (L strand); however, the simulation results revealed that the side chain of Q779 (L strand) moves away with an average distance of 5.5 Å from the backbone NH group of Y221 and makes a close interaction in the S_1 subsite of the CAT domain (**Figure S5.9**). The MD studies confirm the presence of the Q779 side chain (L strand) in the S_1 subsite with an average distance of 4.7 Å (**Figure S5.10**) further indicating the absence of the scissile bond near the active site. Our MD shows no significant interactions been made by a residue of the CAT domain with residues in vicinity of scissile bond, thus confirming the unproductive orientation in the X-ray crystallographic structure.

5.3.3 Flexibility and Domain Interactions in MMP-1 with Linker Regions from MMP-14 or MMP-13

The linker regions in MMPs have been proposed to play important roles for the mutual orientation of the domains and their interactions with collagen [136] [137] [131]. In order to explore the effects of the nature of the MMP-1 linker region (both the length of the linker and its amino acid sequence) on modulating domain interactions, it was exchanged with the linker region from MMP-14 (which contains 35 residues compared with 17 residues in MMP-1) or MMP-13 (which consists of the same number of residues as the MMP-1 linker but with a different sequence) [150]. The trajectory of MMP-1 with *in silico* MMP-14 linker (MMP-1/MMP-14) showed that the system equilibrated at ~20 nsec (**Figure S5.11**). The average RMSD value of MMP-1/MMP-14 was 5.6 Å compared to 4.7 Å for MMP-1 (**Figure 5.2**). Comparison of the RMSD of the CAT and HPX domains of MMP-1/MMP-14 with MMP-1 indicated that the CAT domain of MMP-1/MMP-14 showed slightly higher structural deviation in comparison to MMP-1 (**Figure S5.12**). The average RMSD values of the linker region are 5.6 Å for MMP-1/MMP-14 and 2.0 Å for MMP-1 (**Figure S5.13**) indicating greater structural deviation in the former case. These differences could have significant impact on interactions between the CAT and HPX domains and would affect the conformational dynamics of both domains.

The basal level fluctuation for MMP-1 was considered to be 1.4 Å. In MMP-1/MMP-14 57% of the residues showed fluctuations greater than 1.4 Å (**Table S5.2**) indicating an overall increase of 14% higher fluctuations in MMP-1/MMP-14 compared to the RMSF of MMP-1 (**Figure S5.14**). This increase could be attributed to the larger size of the MMP-14 linker (35 amino acids) compared to the 17 amino acid residue linker of MMP-1. The RMSFs for both domains of MMP-1/MMP-14 showed on average >10% overall increase in comparison to MMP-1 (**Table S5.5**). The THP complexed to MMP-1/MMP-14 showed 20% increased fluctuations with respect to the THP complexed to MMP-1 (**Table S5.5**).

The overall effect of the MMP-14 linker in MMP-1 was an increase in the flexibility of the enzyme•substrate complex. Residues 171-180 encompassing the loop between β4 and β5 exhibited increased fluctuations in MMP-1/MMP-14. These residues are in vicinity of the structural Zn^{2+} in the CAT domain. Residues 188-191 encompassing the loop between β5 and αB also showed increased fluctuations in MMP-1/MMP-14. The adjacent αB helix contains H199, E200, and H203, which are involved in coordination with the catalytic Zn^{2+} in the CAT domain. The residues of the loop between α-helices αB and αC showed increased flexibility. This loop contains H209, which is coordinated to the catalytic Zn^{2+}. There was significantly more fluctuation in the HPX domain compared to CAT domain of MMP-1/MMP-14 (**Figure S5.14**).

The linker region in MMP-13 is the same size as the linker in MMP-1, but with a different sequence [150]. The RMSD profile of MMP-1 with the linker of MMP-13 (MMP-1/MMP-13) showed that the system equilibrated at ~20 nsec (**Figure 5.2**) and indicated that MMP-1/MMP-13 had greater structural deviation with average value of 5.3 Å in comparison to MMP-1 (4.7 Å). The RMSD profile of the linker in MMP-1/MMP-13 showed a similar trend to the MMP-1 linker, but exhibited higher flexibility (**Figure S5.15**). The RMSF plot of MMP-1/MMP-13 showed overall increased fluctuations in comparison to MMP-1 (**Figure S5.16**). In MMP-1/MMP-13 68% of the residues exhibited fluctuations greater than 1.4 Å (**Table S5.2**) indicating an overall increase of 25% higher fluctuations compared to the RMSF of MMP-1. The RMSF of the two domains of MMP-1/MMP-13, in contrast to MMP-1, showed that the CAT domain of MMP-1/MMP-13 contains on average 64% residues with RMSF >1.4% with respect to 31.6% in MMP-1. The HPX domain of MMP-1/MMP-13 contained 58.5% of residues with fluctuations >1.4% as compared to 46.8% in MMP-1. Comparison of the RMSF of the linker region residues showed an increase in comparison to MMP-1 that is slightly higher than in the linker of MMP-1/MMP-14 (**Figure 5.4B**). An increase in the flexibility of residues in the CAT domain of

MMP-1/MMP-13 was seen in the loop connecting αB and αC. The αC is directly connected to the linker region of MMP-1/MMP-13.

In order to understand how the inter-domain distance between CAT and HPX domain changes as a function of time we computed the distance between centers of mass of the CAT and HPX domains. The average inter-domain distances for MMP-1, MMP-1/MMP-14, and MMP-1/MMP-13 were 37.8, 37.0, and 37.7 Å, respectively, and remained relatively stable throughout the simulation (**Figure S5.17**). The inter-domain distance of MMP-1/MMP-14 did show a decrease after 80 nsec simulation (**Figure S5.17**). The center of mass between the CAT and HPX domain in MMP-2 also showed similar trends with an average distance of 39.6 Å and remained stable throughout the entire length of simulation [133] [144]. The average distance between the scissile bond of the THP L strand and the CAT domain (catalytic Zn^{2+}) was 6.7 Å in MMP-1, 8.4 Å in MMP-1/MMP-14, and 8.3 Å in MMP-1/MMP-13 (**Figures S5.18** and **S5.19**), illustrating that changing the linker does not lead to a more productive conformation, but rather the opposite.

5.3.4 Triple-Helix Interactions with MMP-1/MMP-14 and MMP-1/MMP-13

The average bending angle during the MD simulation for the THP leading strand was 115.8° in MMP-1/MMP-14 and 86.4° in MMP-1 (**Figures S5.3** and **S5.20**). The bending angle was also increased in MMP-1/MMP-13 compared with MMP-1 (**Figure S5.21**), suggesting that the triple-helix is accommodated in different orientations in MMP-1/MMP-14 and MMP-1/MMP-13 in comparison to MMP-1. The backbone of Y221 in MMP-1 formed hydrogen bonds with the backbone of R780 of the L THP strand (**Table S5.4**). However, in MMP-1/MMP-14 this interaction was not maintained during the simulation (**Figure S5.22**). The side chain of Q779 of the THP L strand no longer accessed the active site of MMP-1/MMP-14 and was on average 10.3 Å away, while in MMP-1 it was 4.7 Å from the active site.
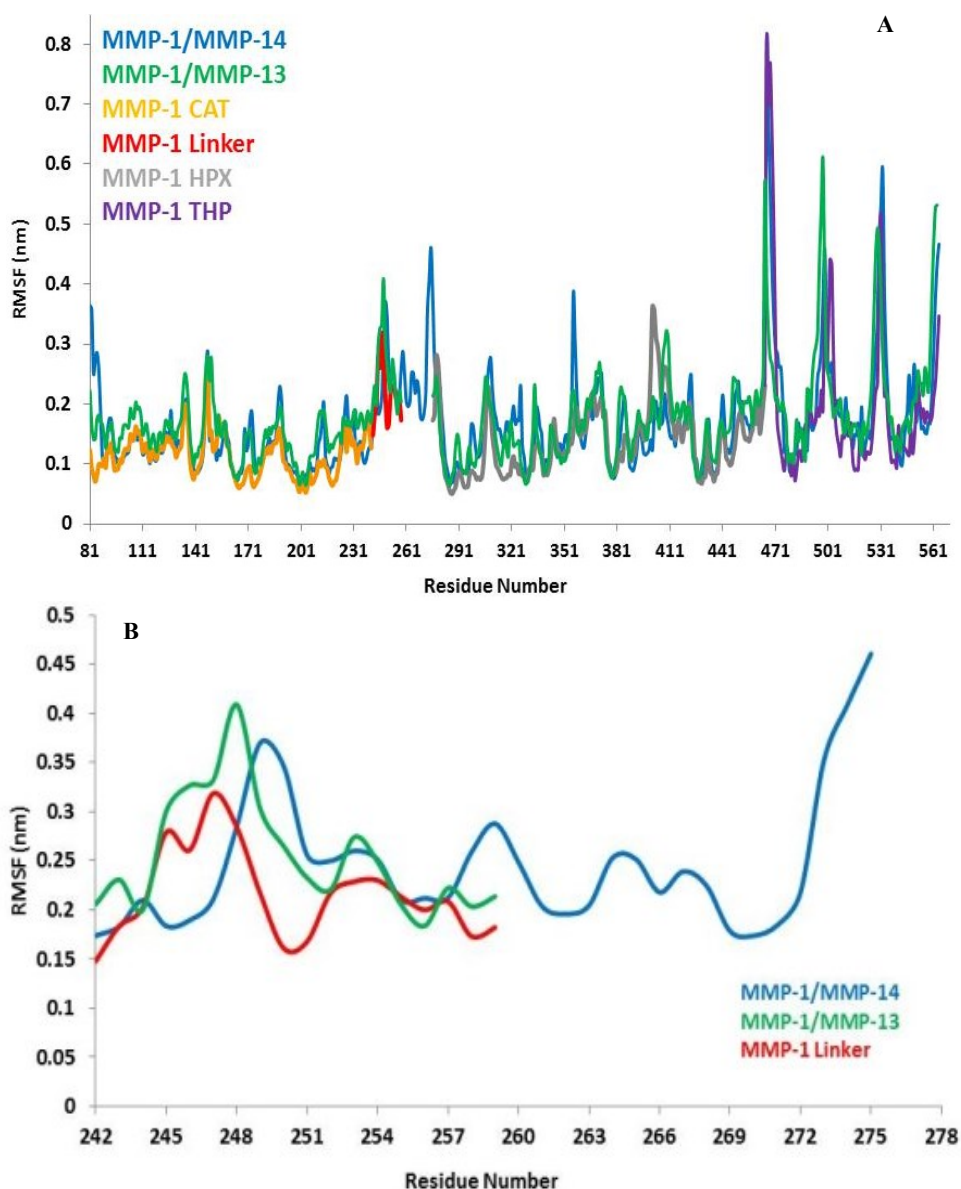
Figure 5.4. (A) RMSFs of MMP-1, MMP-1/MMP-14, and MMP-1/MMP-13 using Cα atoms. (B) RMSF of Cα atoms of the linker region of MMP-1 with respect to MMP-1/MMP-14 and MMP-1/MMP-13 averaged over 100 nsec MD trajectory. The numbering of residues is according to the 4AUO X-ray crystallographic structure [18].
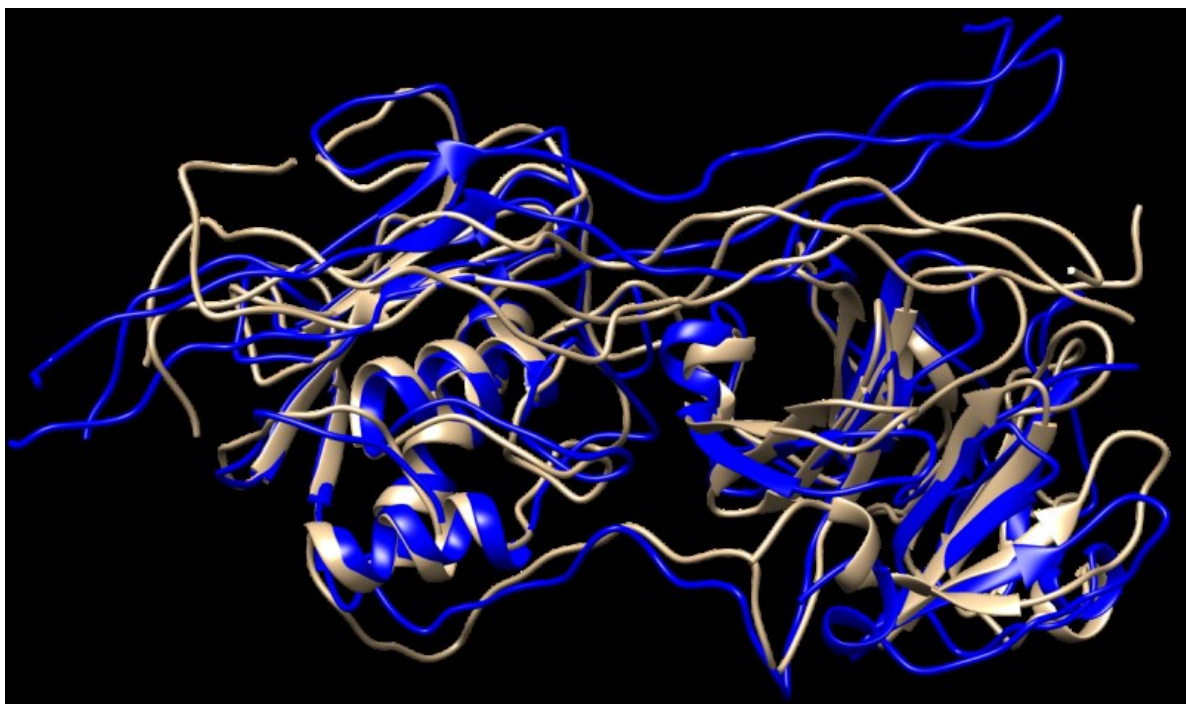
Figure 5.5. Structures obtained from the most populated cluster of MMP-1 (brown) and

MMP-1/MMP-13 (blue).
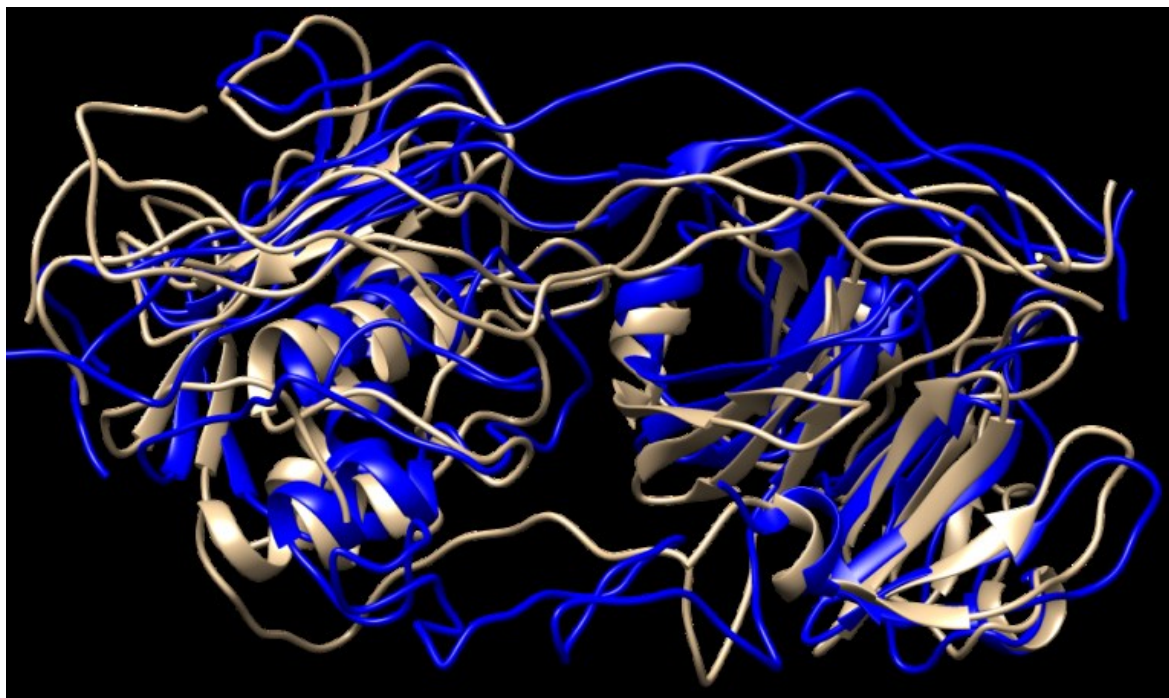
Figure 5.6. Structures obtained from the most populated cluster of MMP-1 (brown) and

MMP-1/MMP-14 (blue).

The backbone of S220 formed hydrogen bonds with the side chain of Q779 after 40 nsec simulation in MMP-1 (Table S4), while there were no hydrogen bonds formed between the S220 backbone and the side chain of Q779 in MMP-1/MMP-14 (**Figure S5.23**). R780 of the THP L strand interacts with the MMP-1 HPX domain, while all of the interactions of R780 seem to be very unstable in the MMP-1/MMP-14 complex throughout the entire length of the simulation (**Figure S5.24**). R272 of the HPX domain was extensively involved in hydrogen bonding with the L785, O786, and R789 residues of the THP M strand in MMP-1. In MMP-1/MMP-14 the side chain of R272 makes a stable interaction with the backbone of O786 for 40 nsec and disappears as the simulation evolved over time (**Figure S5.25**).

The change in atomistic interactions in MMP-1/MMP-14 in comparison to MMP-1 pinpoints the role of the linker in modulating the interactions between the enzyme and substrate. The secondary structure of the linker region of MMP-1 appears relatively stable as a function of simulation time (**Figure S5.26**). However, the linker of MMP-1/MMP-14 showed changes in secondary structure during the simulation (**Figure S5.27**).

In MMP-1/MMP-13, the Q779 side chain of the THP L strand accessed the active site S_1 pocket with an average distance of 3.3 Å compared to 4.7 Å in MMP-1. The backbone of Y221 and P219 in MMP-1/MMP-13 made a hydrogen bond with R780 of the THP L strand in a similar manner as in MMP-1. The side chain of Q167 also made hydrogen bonds with the O771 side chain of the L strand as it did in MMP-1 (**Figure S5.28**). Residues of the HPX domain such as N296, T270, and R272 showed a similar pattern of interaction as MMP-1/MMP-14 (**Figure S5.29**). These results indicated that overall the interactions of the CAT domain with the THP in MMP-1/MMP-13 is closer to MMP-1, while the HPX domain interactions of MMP-1/MMP-13 showed a pattern which is closer to MMP-1/MMP-14 than MMP-1. The structures obtained from the most populated

clusters from cluster analysis of MMP-1, MMP-1/MMP-13, and MMP-1/MMP-14 showed significant differences in orientation and binding of the THP to both the CAT and HPX domains (**Figures 5.5 and 5.6**). The linker clearly modulates important interactions in both the CAT and HPX domains.

5.3.5 Correlated Motions of the HPX and CAT Domains: Influence of the Linker

The main aim of cross correlation analysis was to understand correlated motions between the MMP-1 HPX and CAT domains and to see how both domains are related to each other in the presence of triple-helices (**Figure 5.7**). Residues 105-115 of the CAT domain αA showed correlated motions with residues 175-195 of β5. The region 175 to 195 contains H164 and H177, which are coordinated to the structural Zn^{2+} and are located on a β-strand. The region from 182 to 194 belongs to loop which connects this β-sheet to an α-helix containing H199 and E200 which are coordinated to the catalytic Zn^{2+}. The residues ranging from 140-150 of β3 and the loop joining β4 showed correlation towards residues 170-180 of β5 of the CAT domain. The β3, β4, and β5 are arranged in parallel conformation to each other in the X-ray crystallographic structure (Figure 1). These β-sheets also surround the structural Zn^{2+} of the CAT domain. Residues 117-125 of αA showed correlation to residues 235-240 of αC, which is directly connected to the linker region. There was also correlation observed between residues 105-115 of αA of the CAT domain and residues 287-294 of the loop region between β8 and β9 of the HPX domain. Residues 180-195 of the loop region between β5 and αB of the CAT domain showed correlation towards residues 220-230 of the loop region between αB and αC of the CAT domain. The loop encompassing residues 220-230 is in very close proximity to the THP L strand. Residues 180-195 show correlation towards HPX domain residues 280-290. Residues 180-195 also show correlation to THP residues 780-785. This is the region of the THP showing local destabilization. HPX domain residues 315-320, belonging to the loop connecting two β-sheets together, are correlated to the ends of the THP strands. Overall,

there was very limited correlated motion seen between the HPX and CAT domains suggesting very limited influence of the HPX domain on the conformational dynamics of the CAT domain via the linker.
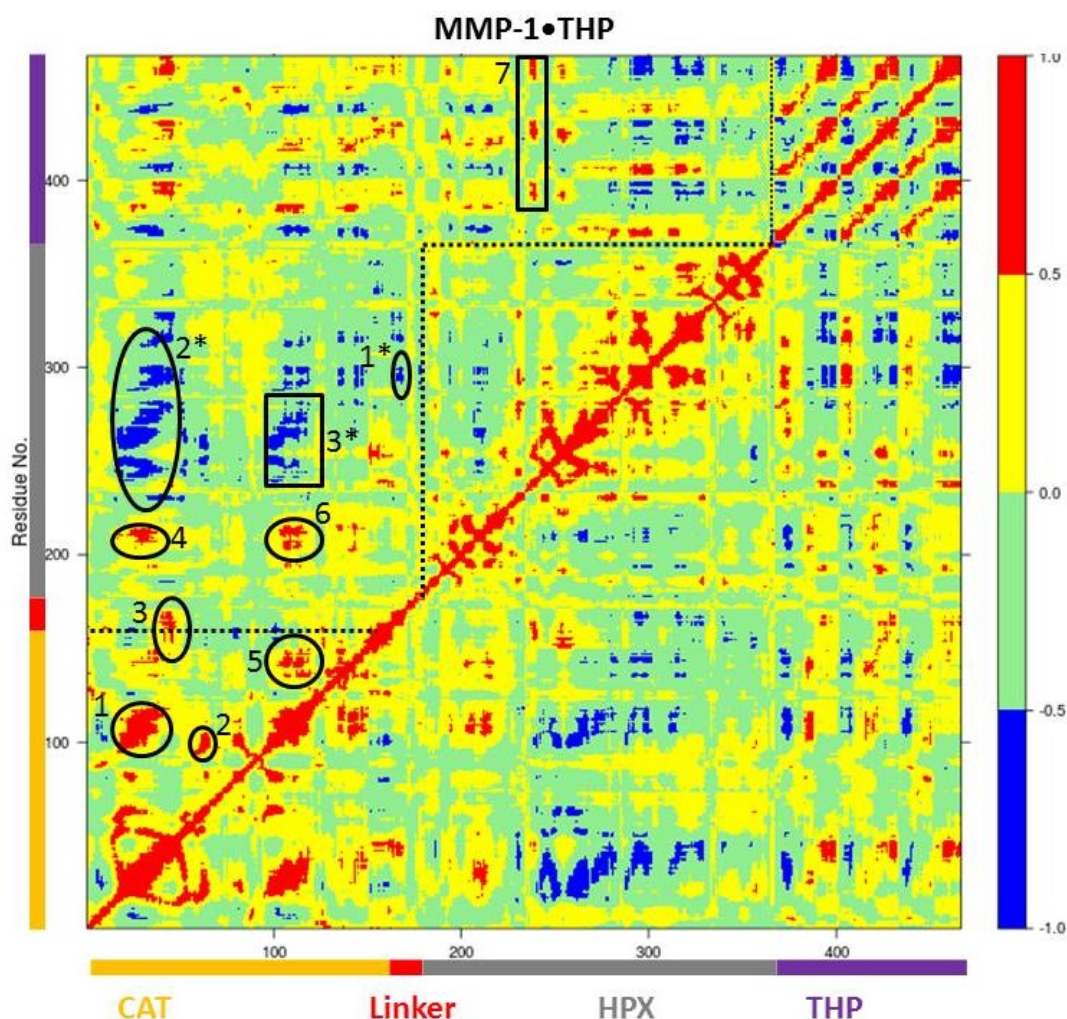
Figure 5.7. Dynamic cross correlation analysis of MMP-1•THP. Intensities are indicated by the bar on the right. Numbering of correlations corresponds to specific interactions as follows:

1. The CAT domain residues 105-115 show correlation towards residues 175-195.

2. Residues 140-150 show correlation towards residues 170-180.

3. Residues 117-125 show correlation to residues 235-240.

4. CAT domain residues 105-115 show correlation towards HPX domain residues 287-294.

5. CAT domain residues 180-195 show correlation towards CAT domain residues 220-230.

6. Residues 180-195 show correlation towards HPX domain residues 280-290. Residues 180-195 also show correlation to THP residues 780-785.

7. HPX domain residues 315-320 are correlated to the ends of the THP strands.

1*. Residues 160-165 show negative correlation to HPX domain residues 275-280.

2*. CAT domain residues 100-120 show negative correlation towards HPX domain residues 310-380.

3*. CAT domain residues 175-185 show negative correlation towards HPX domain residues 330-350.

Cross correlation analysis of MMP-1/MMP-14 and MMP-1/MMP-13 was subsequently performed (**Figures 5.8** and **5.9**, respectively). The residues of the linker region of MMP-1/MMP-14 showed correlation with the loop region of the *C*-terminus (residues 439-447) of the HPX domain. Residues 422-427 of β22 of the HPX domain also showed correlation towards the residues of the linker region. The residues in the loop connecting the *C*-terminus of the HPX domain showed positive correlation toward the linker region residues. Residues 409-413 of β20 showed correlation towards residues 262-267 located in vicinity of the linker region in the HPX domain. The residues adjacent to the linker region (200-205) showed correlation towards β10 residues 312-317 of the HPX domain. One important observation was the low degree of correlated motions observed between the THP and the CAT and HPX domains of MMP-1/MMP-14 in comparison to MMP-1. Interestingly, there was significant positive correlation between the residues of the linker region and the HPX domain in MMP-1/MMP-14. Overall, MMP-1 has more correlated and anti-correlated motion than MMP-1/MMP-14.

The residues of the MMP-1/MMP-13 linker region showed correlated motion towards the *C*-terminal residues (436-444) and β21 of the HPX domain. The residues of the loop region between β8-β9 (288-296) showed correlation towards the *C*-terminal residues (436-444) of

the HPX domain. Residues of the linker region of MMP-1/MMP-13 and MMP-1/MMP-14 have positive correlation towards the HPX domain, which was not observed with the natural linker of MMP-1. The greater extent of positive and negative correlation in MMP-1/MMP-13 in contrast to MMP-1 indicated that even changes in the sequence of the linker produces significant differences in the motion of residues.
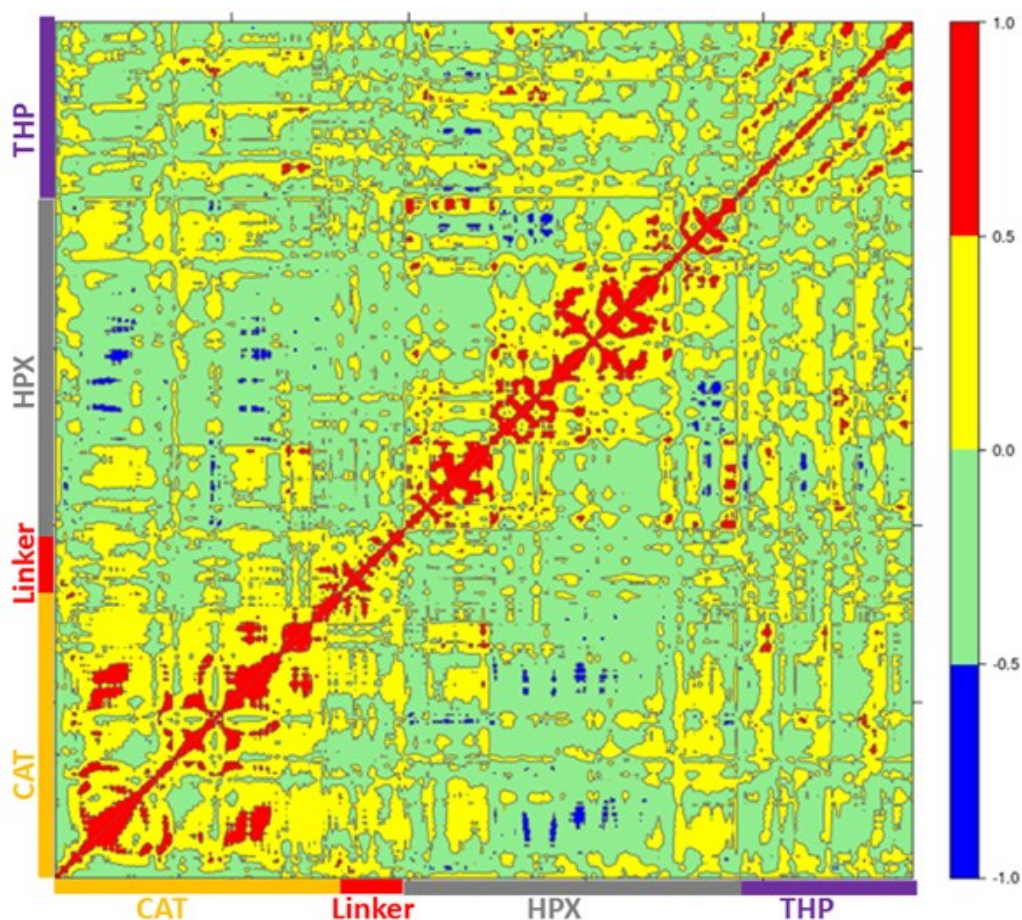
Figure 5.8. Dynamic cross correlation analysis of MMP-1/MMP-14•THP. Intensities are indicated by the bar on the right.
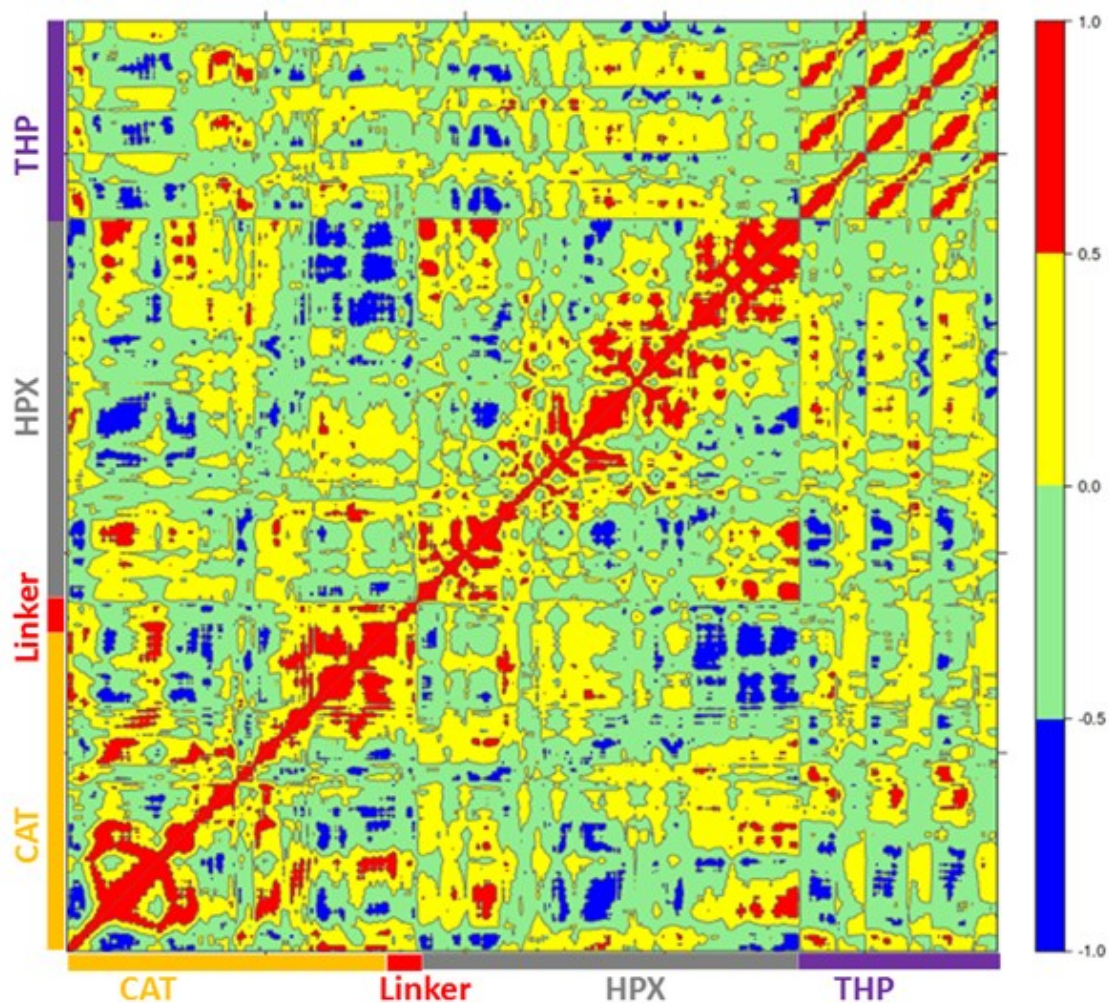
Figure 5.9. Dynamic cross correlation analysis of MMP-1/MMP-13•THP. Intensities are indicated by the bar on the right.

5.4 CONCLUSIONS

MD simulations of the X-ray crystallographic structure of MMP-1 [18] revealed that conformational changes exercise an important role and influence the interactions between the enzyme HPX and CAT domains and revealed the existence of extensive correlated motions. The interactions between the THP chains and the enzyme also have a flexible nature. Importantly, we confirm prior NMR experimental results [131] that the X-ray crystallographic structure does not represent the functional enzyme•substrate complex. Although we used this non-productive enzyme-substrate complex as a starting point for the MD simulations, as opposed to the productive complex reported from NMR spectroscopic

studies, it is important to note that the X-ray crystallographic structure of MMP-1•THP is the only structure that defines the specific molecular interactions between the enzyme and substrate, as the NMR-derived structure is based on docking of residues. The MD simulation results are relevant to MMP-1 catalyzed collagenolysis, as the X-ray crystallographic and NMR structures of MMP-1•THP agree on the majority of THP interactions with the HPX domain and agree on the "closed" conformation of MMP-1 during collagenolysis. The significant point of departure in the two structures is the position of the THP with regards to the CAT domain, where the NMR-derived structure represents a productive complex while the X-ray crystallography-derived structure does not. The effects of these differences on MMP-1 dynamics and correlations can be explored in subsequent studies using MD simulations.

The linker plays a key role in modulating the interactions between the domains and formation of the active complex. The MD simulations with exchanged linker regions from MMP-14 and MMP-13 reveal the important role of the linker region to influence overall enzyme flexibility, the pattern of the correlated motions between the HPX and CAT domains, and the interactions between the triple-helical substrate and the enzyme. Prior experimental studies examined the effects of MMP linker exchange on collagenolysis. A chimeric MMP-8 whose linker region (17 residues) was replaced with the corresponding MMP-3 sequence (26 residues) lost activity towards collagen [151]. In similar fashion, MMP-1/MMP-3 chimeras possessing the MMP-3 linker were not active towards collagen [152] [153]. The linker has been proposed to be critical for proper orientation of the CAT and HPX domains [154] [155] [131]. The present study furthers the role of the linker, as it was found to facilitate correlative interactions between MMP domains and MMP domains and triple-helices. Overall, examination of MMP-1•THP demonstrates the potential of MD simulations to provide additional insight into experimentally-derived data.

5.5 SUPPORTING INFORMATION

Table S5.3. The MD setup for the wild type MMP-1 and *in silico* modified linkers.

#	MD-setup	Ions	Timescale (nsec)	Replica Runs
1	Wt-MMP-1	Cl⁻ (10)	300	4
2	Apo MMP-1	Cl⁻ (4)	100	2
3	MMP-1/MMP-13	Cl⁻ (7)	100	2
4	MMP-1/MMP-14	Cl⁻ (11)	100	2
5	Apo MMP-1/MMP-14	Cl⁻ (5)	100	2
6	Collagen in water	Cl⁻ (6)	50	2

Table S5.2. The average RMSD and the RMSF of MMP-1 and linker variants.

Enzyme	Mean (Å)	SD (Å)	> 1.4 Å (%)	< 1.4 Å (%)
MMP-1	4.7	0.44	43.1	56.9
MMP-1/MMP-14	5.6	0.46	57.02	42.9
MMP-1/MMP-13	5.3	0.50	68.4	31.6

Table S5.3. Analysis of MMP-1 for 100 nsec trajectory.

Name	Mean (Å)	S.D (Å)	S.E.M
MMP-1	4.7	0.44	0.00636
Run 1	5.2	0.45	0.00644
Run 2	5.5	0.46	0.00660
Run 3	5.1	0.39	0.00557
Averaged trajectory	5.1	0.38	0.00548

Table S5.4. The hydrogen bonding profile of MMP-1 for 300 nsec trajectory.

Donor	Acceptor	Distance (Å)	Probability (%)
Y221 (bb)	R780 (bb, L)	2.0	99.1
Y221 (sc)	Q779 (bb, L)	3.5	46.8
I782 (sc, L)	Y221 (sc,)	3.0	96.3
R780 (bb, L)	P219 (sc)	2.0	99.6
R780 (sc, L)	Y218 (sc)	3.1	91.5
Q167 (sc)	O771 (bb, L)	2.9	99.6
Q167 (sc)	E85 (sc)	3.2	99.1
S220 (bb)	Q779 (sc, L)	3.3	64.2
H203 (bb)	H199 (bb)	3.5	47.8
H203 (sc)	L207 (bb)	2.8	97.0
Q774 (sc, L)	E85	4.0 *	42.5
Q774 (sc, L)	S208	4.5 *	59.5

Bb = backbone, sc = side chain, L = leading strand.
* Distance fluctuates throughout the trajectory

Table S5.5. The RMSF decomposition analysis of the individual regions of MMP-1 in comparison to MMP-1/MMP-13 and MMP-1/MMP-14.

Enzyme	CAT domain		HPX domain		THP	
	> 1.4 Å (%)	< 1.4 Å (%)	> 1.4 Å (%)	< 1.4 Å (%)	> 1.4 Å (%)	< 1.4 Å (%)
MMP-1	21.8	78.2	46.8	53.2	61.2	38.8
MMP-1/MMP-14	31.6	68.4	59	41	80.6	19.4
MMP-1/MMP-13	64.5	35.5	58.51	41.4	86.7	13.3
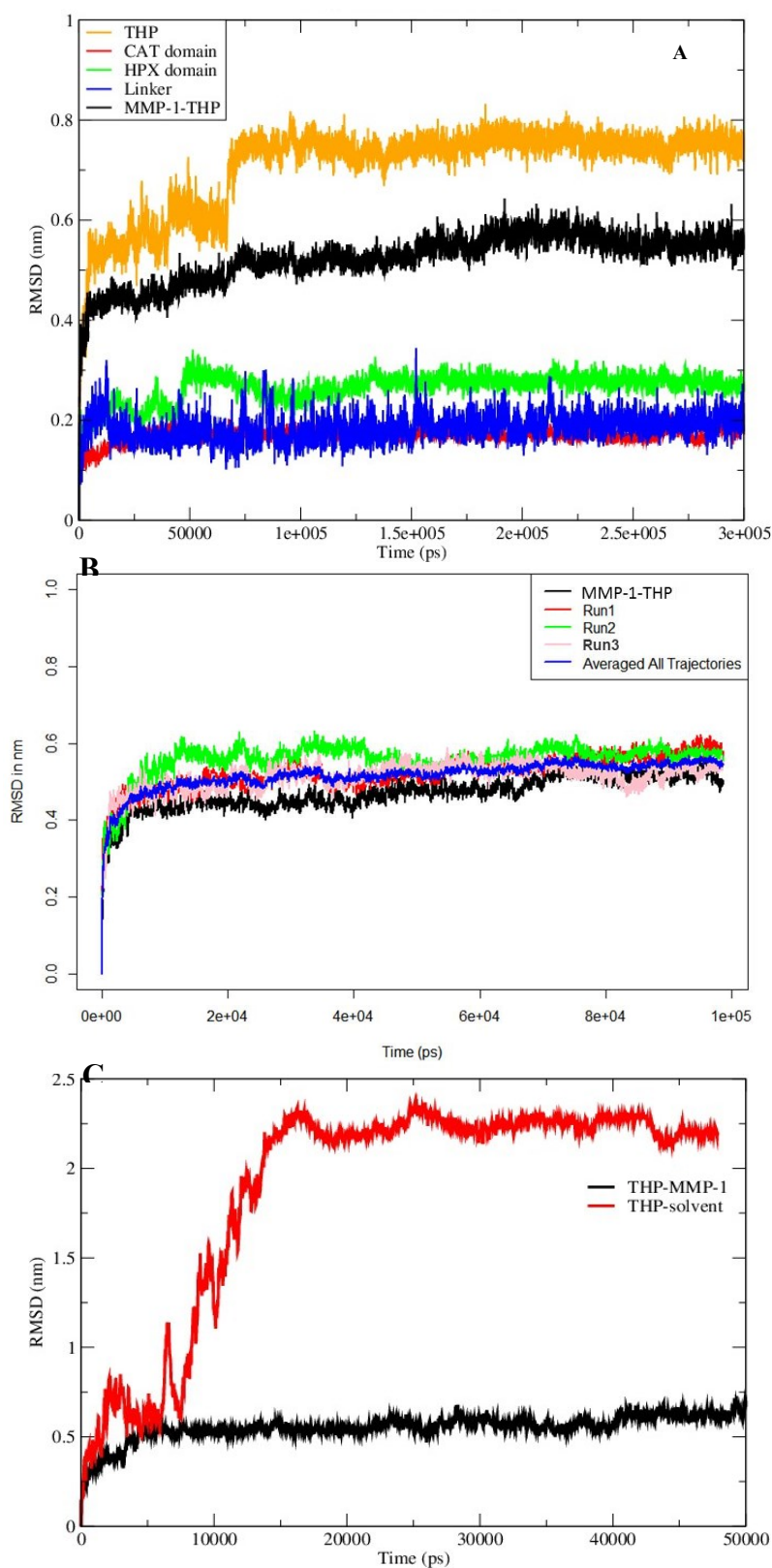
Figure S5.1. The RMSD of Cα atoms of MMP-1•THP and THP in water. (A) The CAT domain, HPX domain, and linker region RMSD are calculated individually for the 300 nsec simulation. (B) The average run of MMP-1•THP (blue) from individual runs performed using different initial velocities to run MD simulations. (C) The THP bound to MMP-1 and isolated in solvent.

A

B
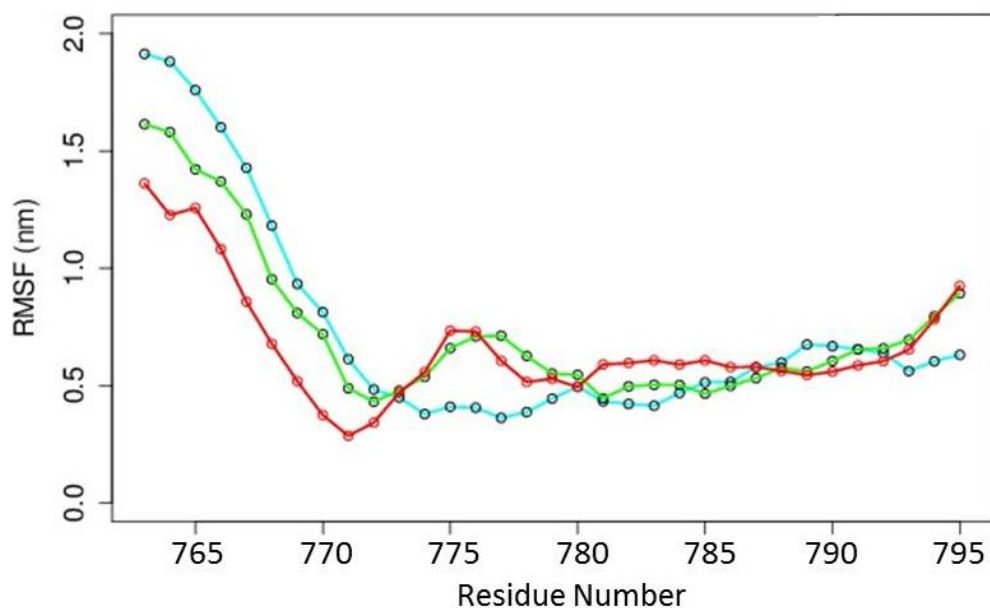
C
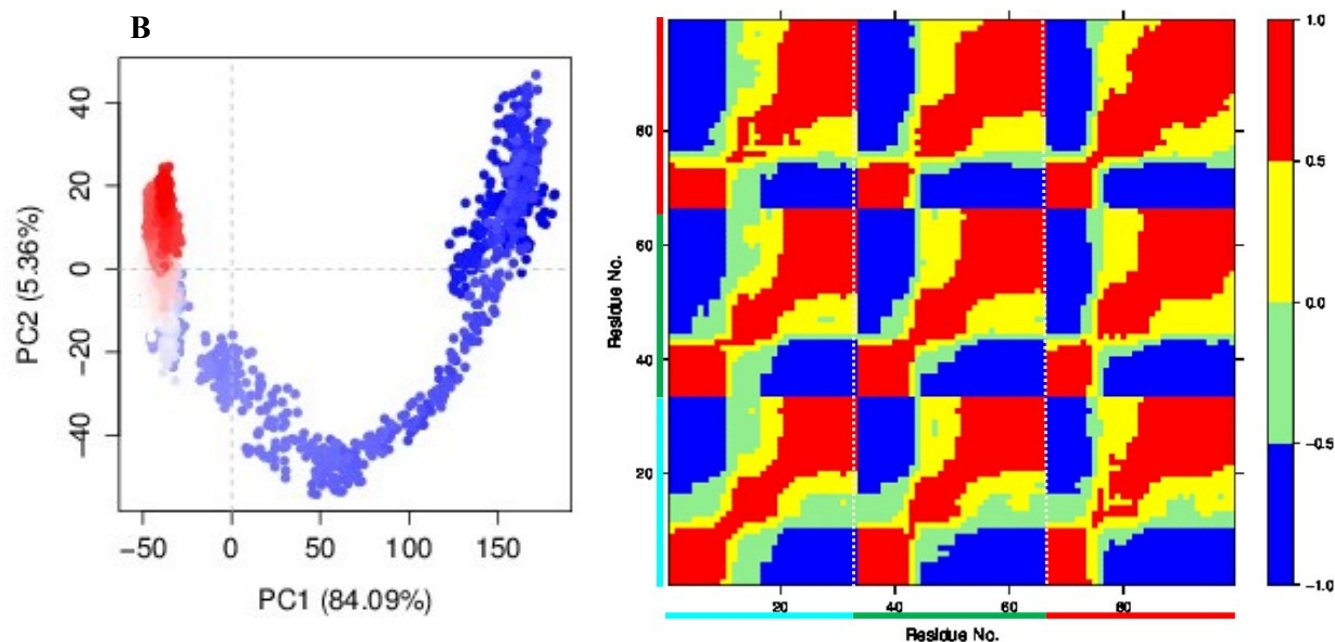
Figure S5.2. Analysis of THP in water performed using Gromacs tools and Bio3D package in R using the Cα atoms. (A) The RMSF of individual strands of THP (leading (cyan), middle (green), and trailing (red)) in aqueous solution for 50 nsec trajectory. (B) The principal component (PC) analysis of THP in water and projection of PC1 versus PC2 along with projection of PC1 motion on the THP chains. The trajectory frames colored blue to red in order of time. (C) Dynamic domain cross correlation analysis of THP in water for 50 nsec in which positive correlation is shown by red and anti-correlated motion by blue colour. The chains of the THP are according to the color notation mentioned above.
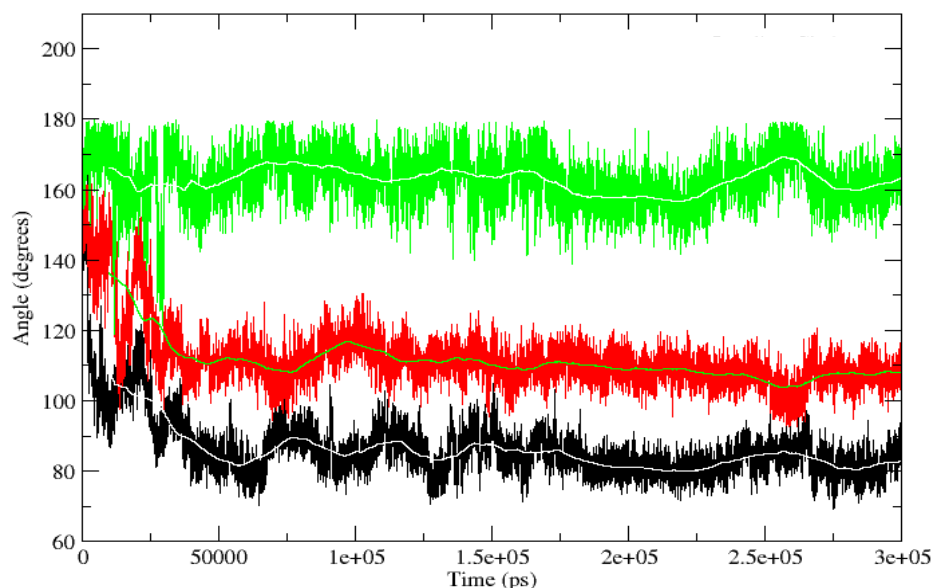
Figure S5.3 The angle of bend measured along the 300 nsec trajectory for Cα atoms of residues 778, 781, and 784 of the THP leading (black), middle (red), and trailing (green) strands.
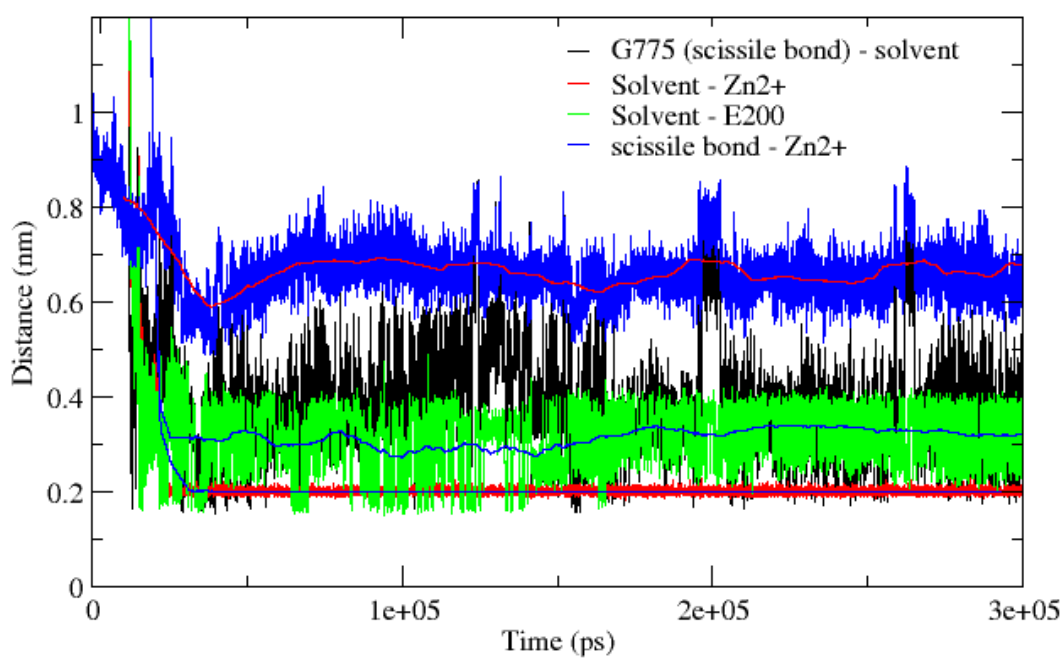
Figure S5.4. The role of solvent in scissile bond distance to MMP-1 for 300 nsec trajectory. There were two solvent molecules in close vicinity of the active site of MMP-1•THP during MD simulation and participates in the hydrogen bonds with E200 and the scissile peptide bond of MMP-1•THP.

191
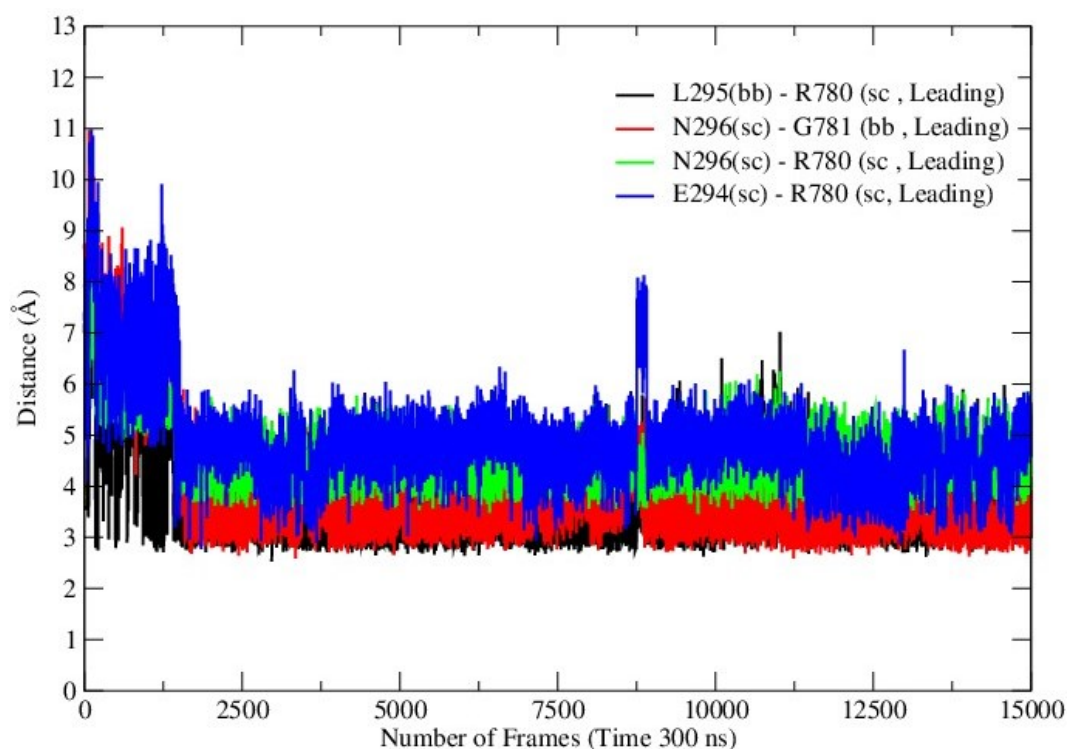
Figure S5.5. MMP-1 HPX domain residue interactions with the THP leading strand. Side

chain = sc, backbone = bb.
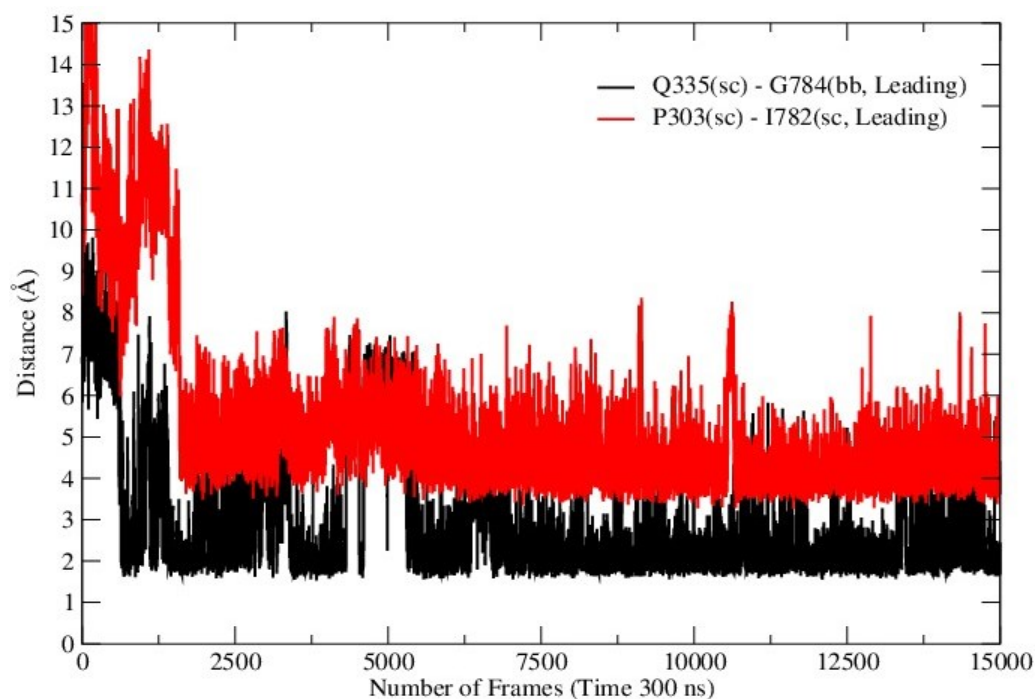
Figure S5.6. MMP-1 HPX domain residue interactions with the THP leading strand. Side

chain = sc, backbone = bb.
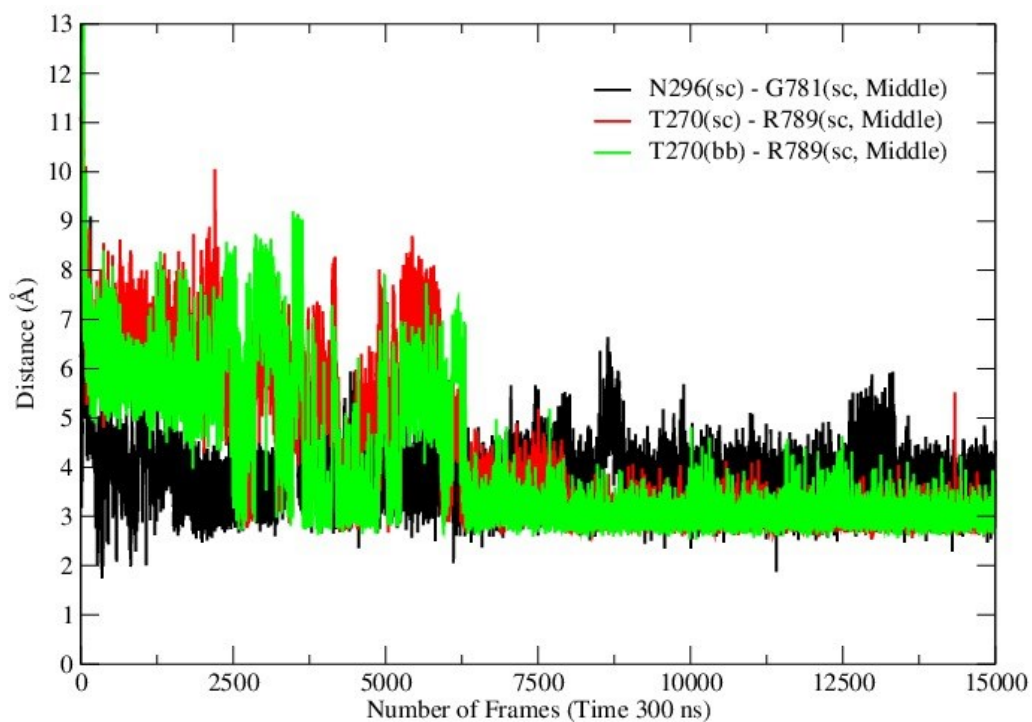
Figure S5.7. MMP-1 HPX domain residue interactions with the THP middle strand. Side chain = sc, backbone = bb.
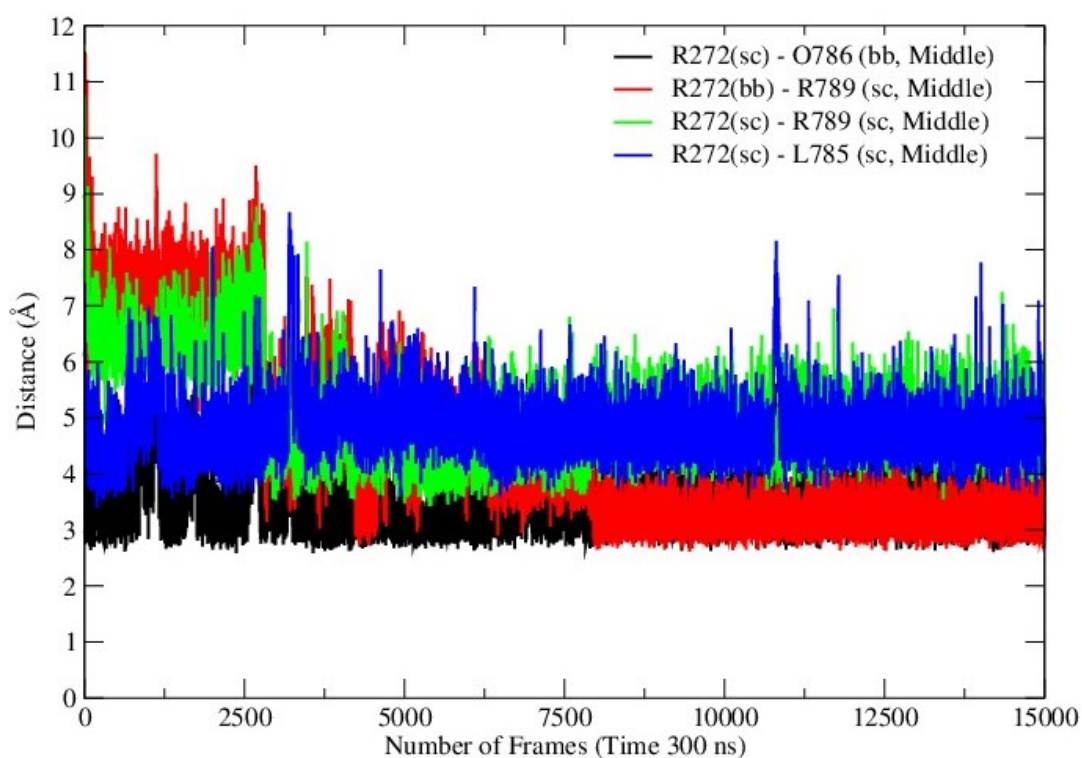
Figure S5.8. Interaction of MMP-1 R272 with the THP middle strand. O = 4-hydroxy-L-proline. Side chain = sc, backbone = bb.
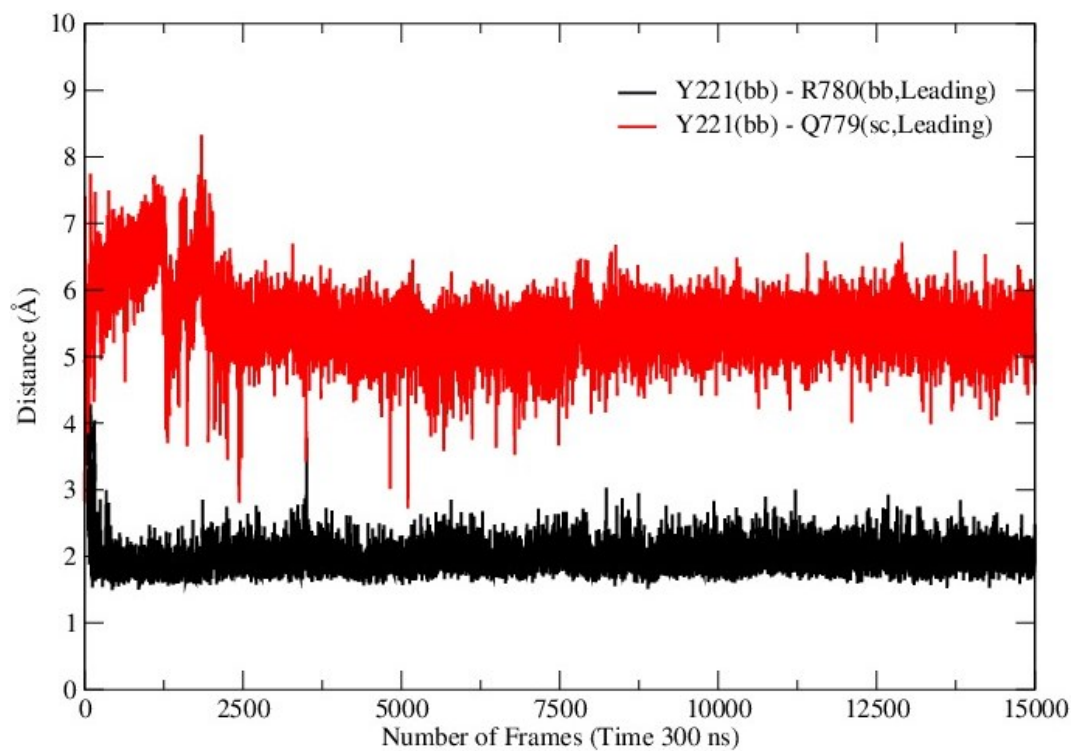
Figure S5.9. Interactions of MMP-1 CAT domain Y221 with the THP leading strand. Side

chain = sc, backbone = bb.
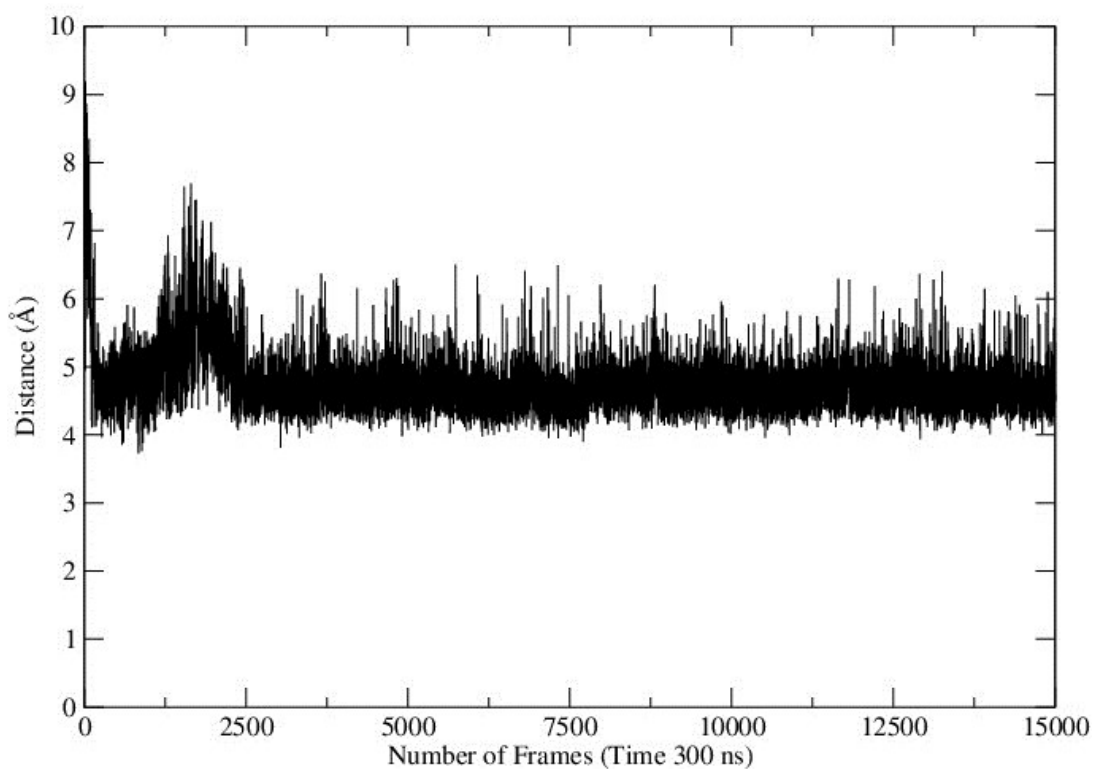
Figure S5.10. The distance of the THP leading strand Q779 side chain to the S_1 pocket of

the MMP-1 CAT domain.
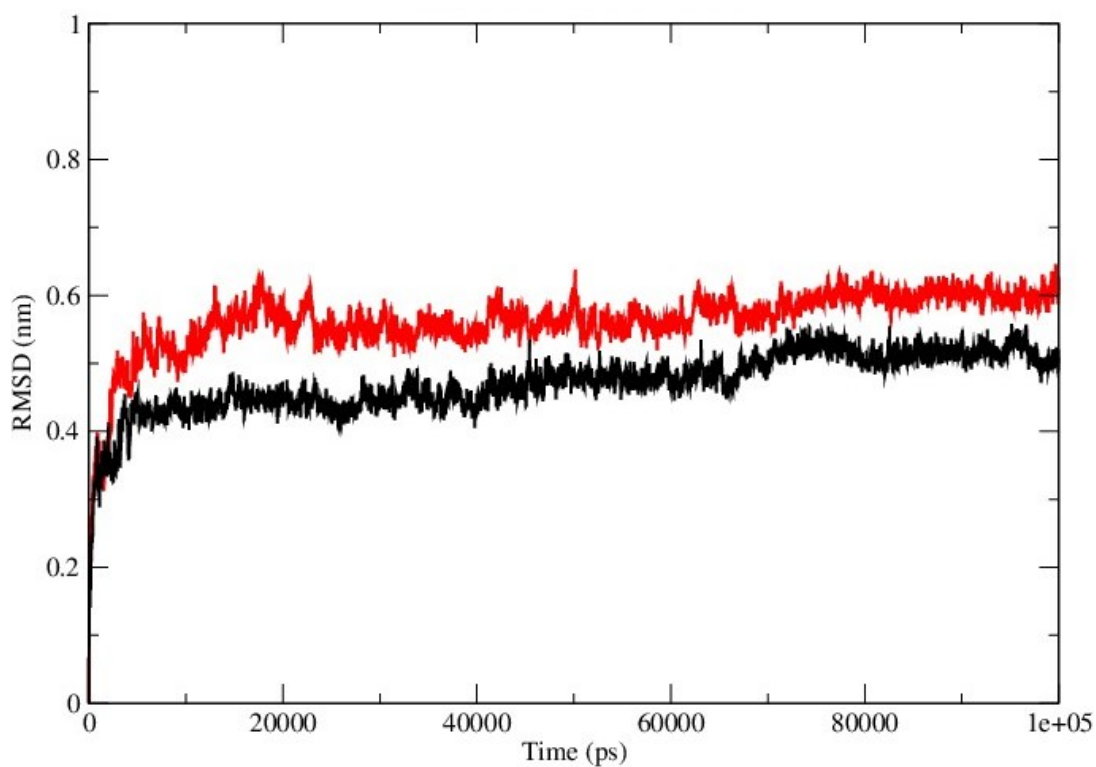
Figure S5.11. The RMSD of MMP-1 (black) and MMP-1/MMP-14 (red).
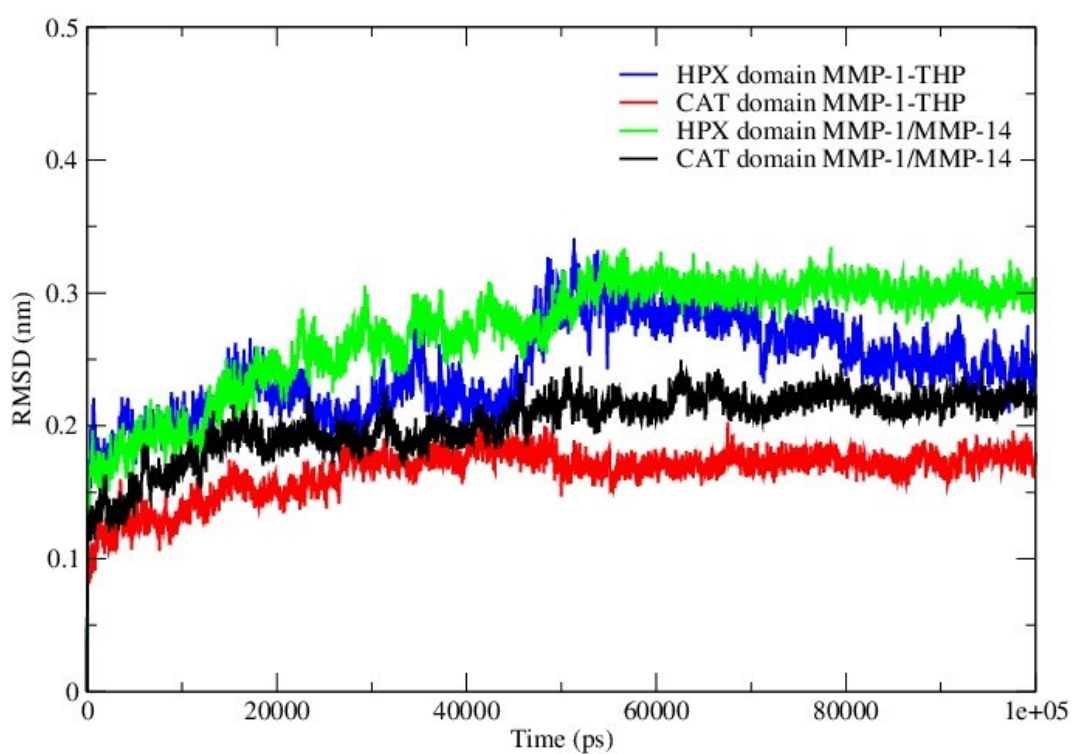
Figure S5.12. The RMSD of CAT and HPX domains of MMP-1 compared to MMP-1/MMP-14.
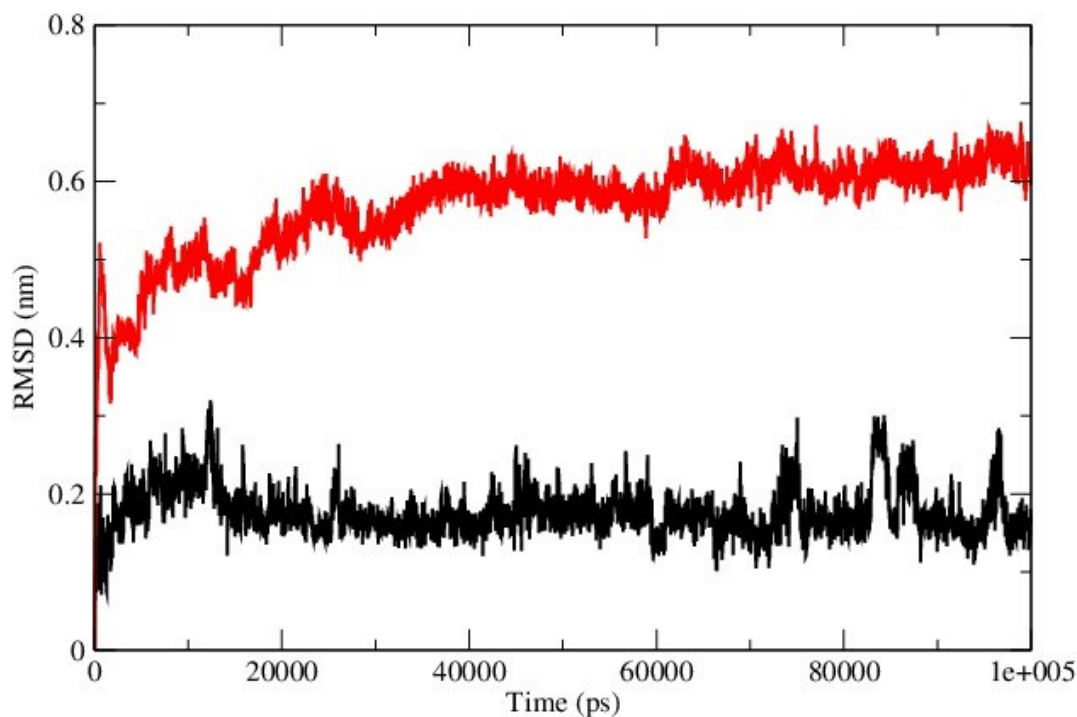
Figure S5.13. Comparison of RMSD of Cα atoms of the linker regions of MMP-1 (black) and MMP-1/MMP-14 (red).

Figure S5.14. Comparison of the RMSF of MMP-1 with MMP-1/MMP- 14. The MMP-1/MMP-14 RMSF is plotted in cyan throughout and consists of an extra 18 residues in the linker region as compared to linker of MMP-1. Thus, the comparison is only made between

the CAT and HPX domains and THP of MMP-1 and MMP-1/MMP-14. The CAT domain,

linker region, HPX domain, and THP are plotted in gold, red, grey, and maroon,

respectively. There is a gap in the graph because the linker of MMP-1/MMP-14 is 18

residues greater than the linker of MMP-1.

Figure S5.14. Comparison of the RMSF of MMP-1 with MMP-1/MMP- 14. The MMP-

1/MMP-14 RMSF is plotted in cyan throughout and consists of an extra 18 residues in the

linker region as compared to linker of MMP-1. Thus, the comparison is only made between

the CAT and HPX domains and THP of MMP-1 and MMP-1/MMP-14. The CAT domain,

linker region, HPX domain, and THP are plotted in gold, red, grey, and maroon,

respectively. There is a gap in the graph because the linker of MMP-1/MMP-14 is 18

residues greater than the linker of MMP-1.

Figure S5.15. The RMSD profile of the linker region of MMP-1/MMP-14 (green), MMP-1/MMP-13 (red), and MMP-1 (black).
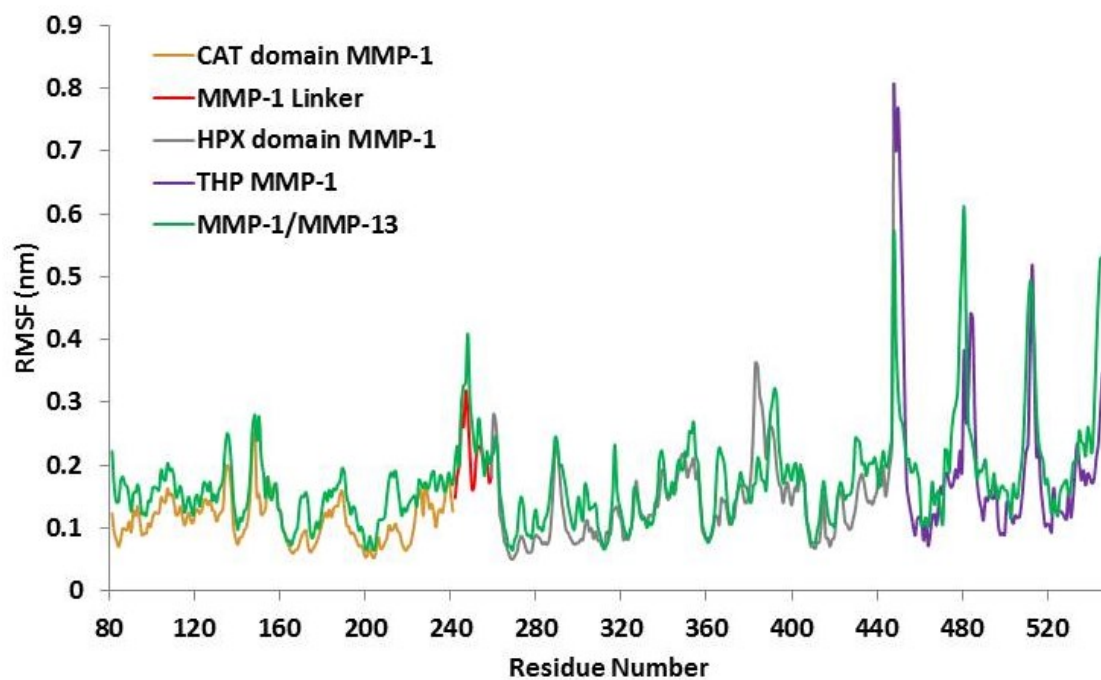
Figure S5.16. The RMSF profile of MMP-1/MMP-13 in comparison to MMP-1.
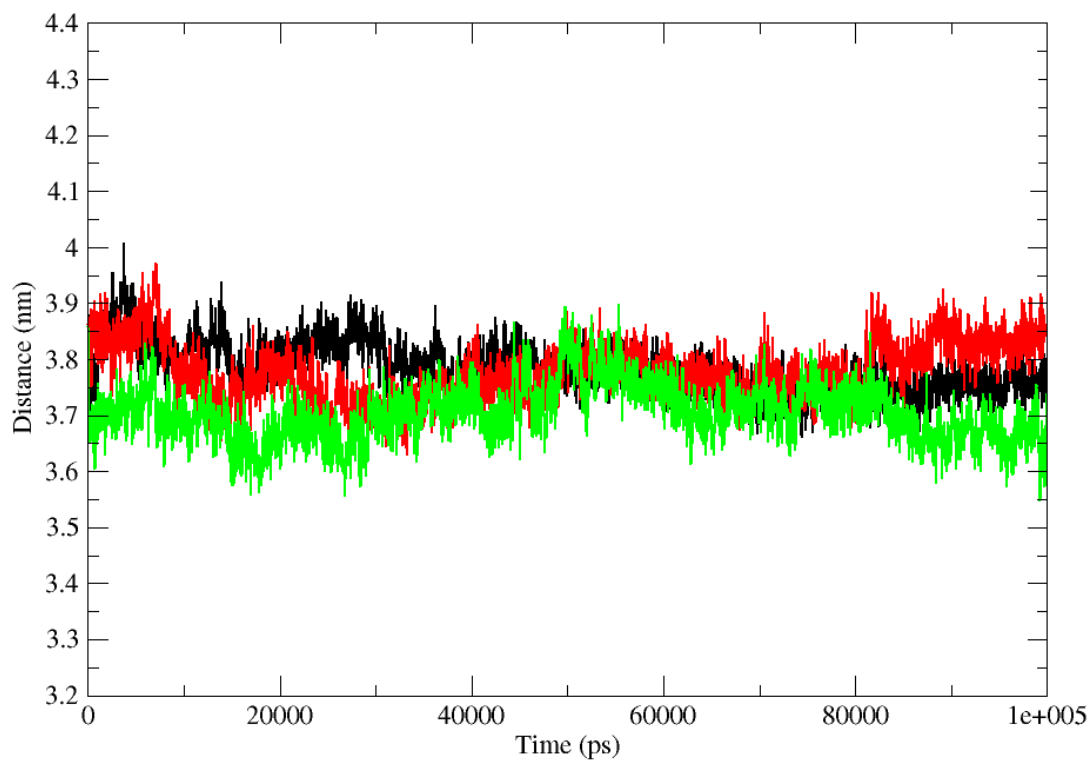
Figure S5.17. The center of mass of the CAT and HPX domains of MMP-1 (black), MMP-1/MMP-14 (green), and MMP-1/MMP-13 (red).
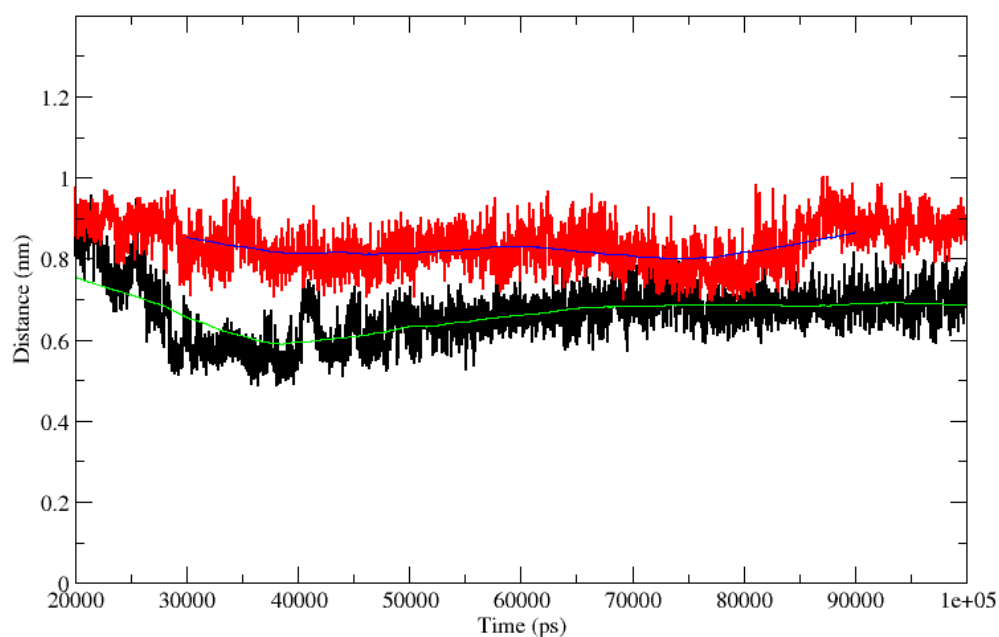
Figure S5.18. The scissile peptide bond distance measured from the catalytic zinc in MMP-1 (black) and MMP-1/MMP-14 (red).
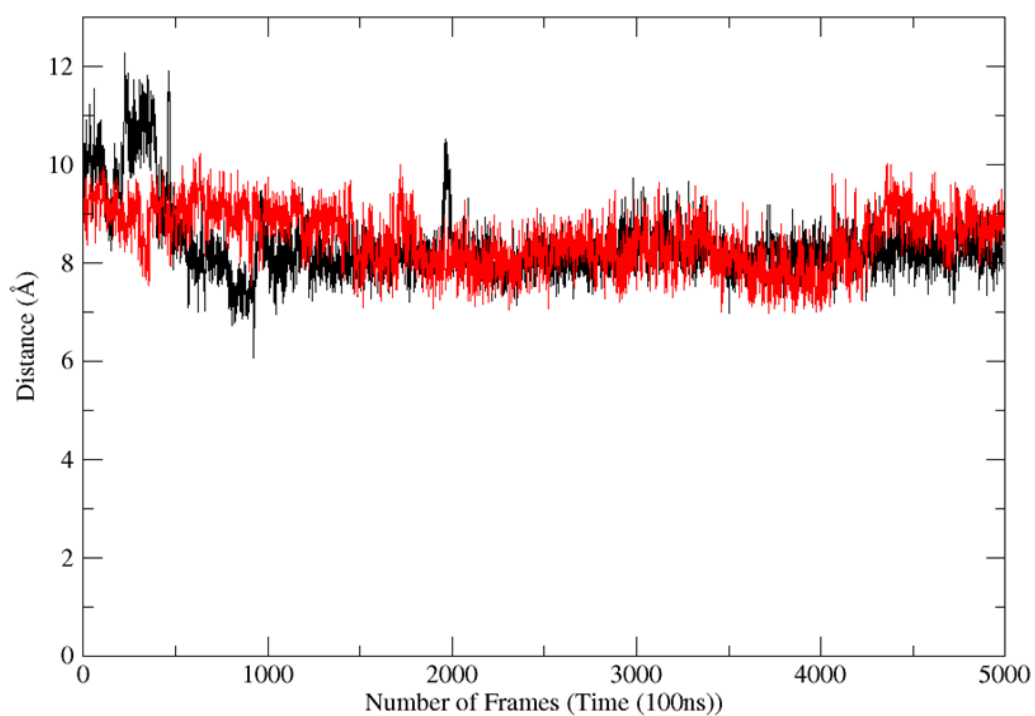
Figure S5.19. The scissile bond distance between MMP-1/MMP-13 (black) and MMP-1/MMP-14 (red).
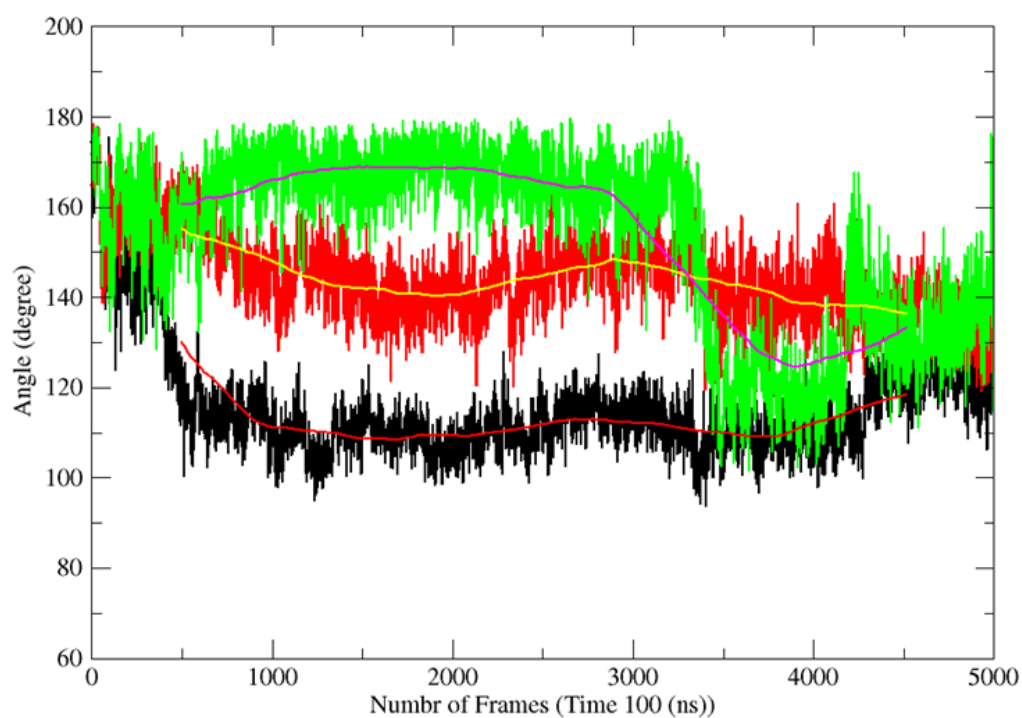
Figure S5.20. The angle of bend in the leading (black), middle (red), and trailing (green) THP strands in the MMP-1/MMP-14•THP complex.

200
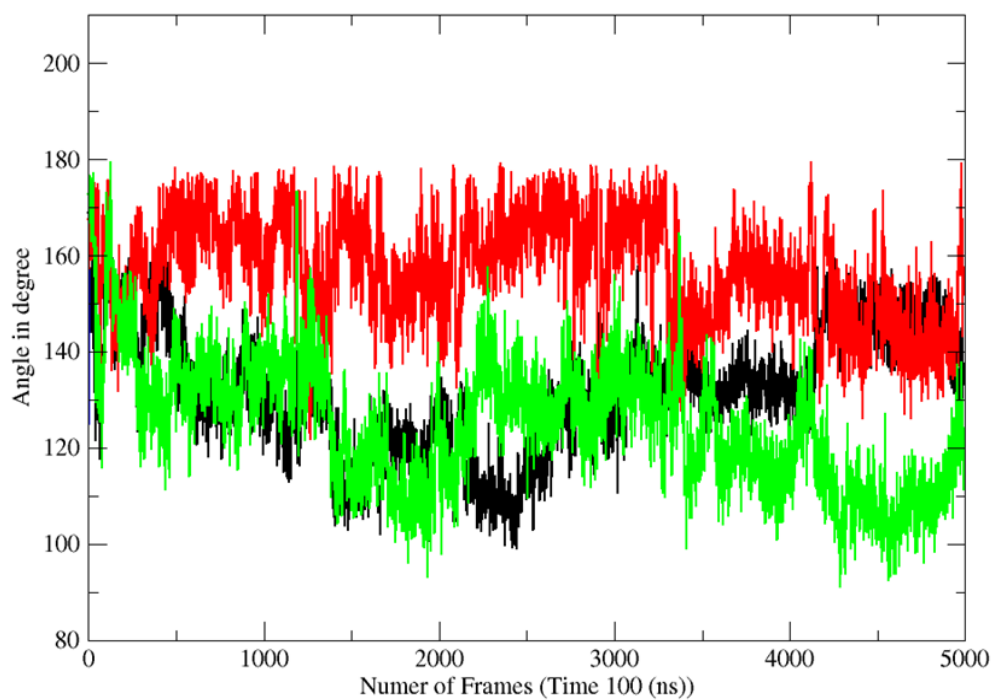
Figure S5.21. The angle of bend in the leading (black), middle (red), and trailing (green) THP strands in the MMP-1/MMP-13•THP complex.
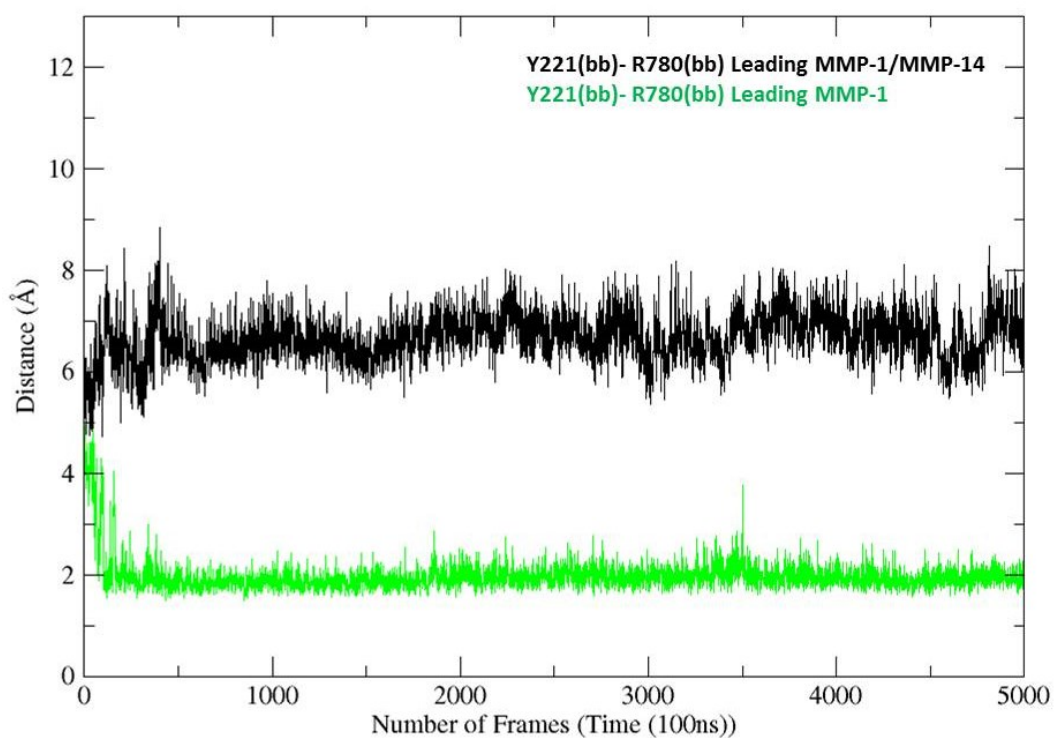
Figure S5.22. The interaction of the leading THP strand with the CAT domain of MMP-1/MMP-14 in comparison to MMP-1. Backbone = bb.
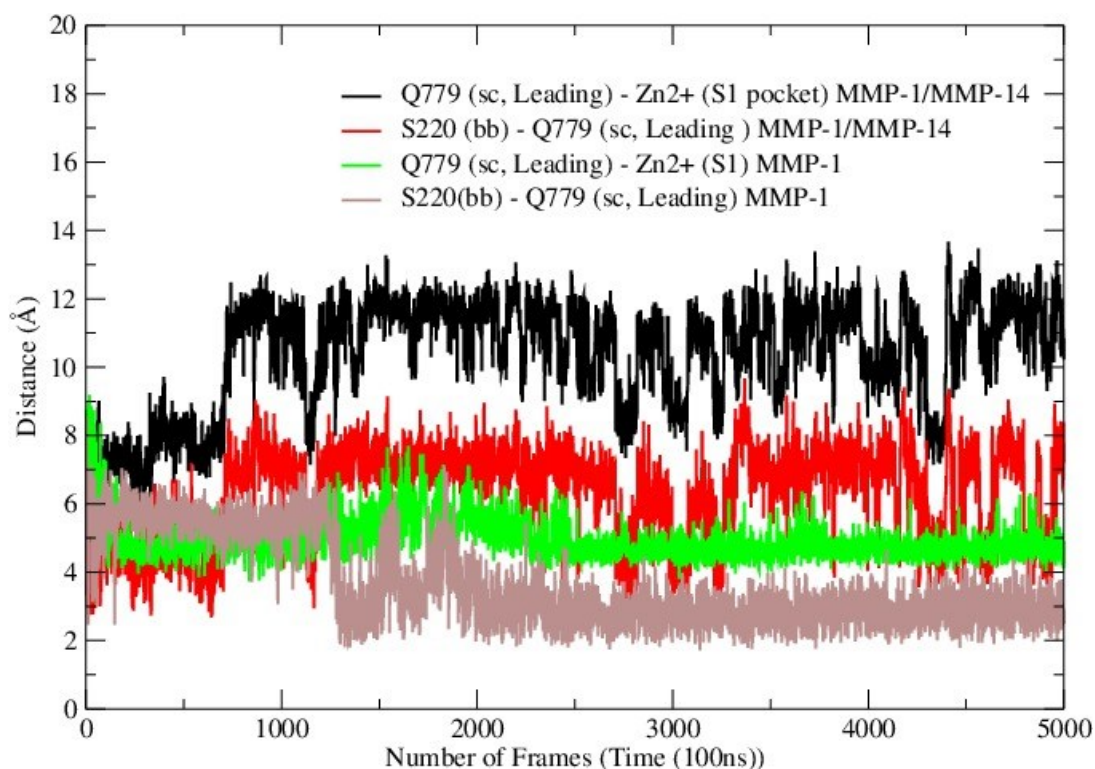
Figure S5.23. The interaction of the leading THP strand with the CAT domain of MMP-1/MMP-14 in comparison to MMP-1. Side chain = sc, backbone = bb.
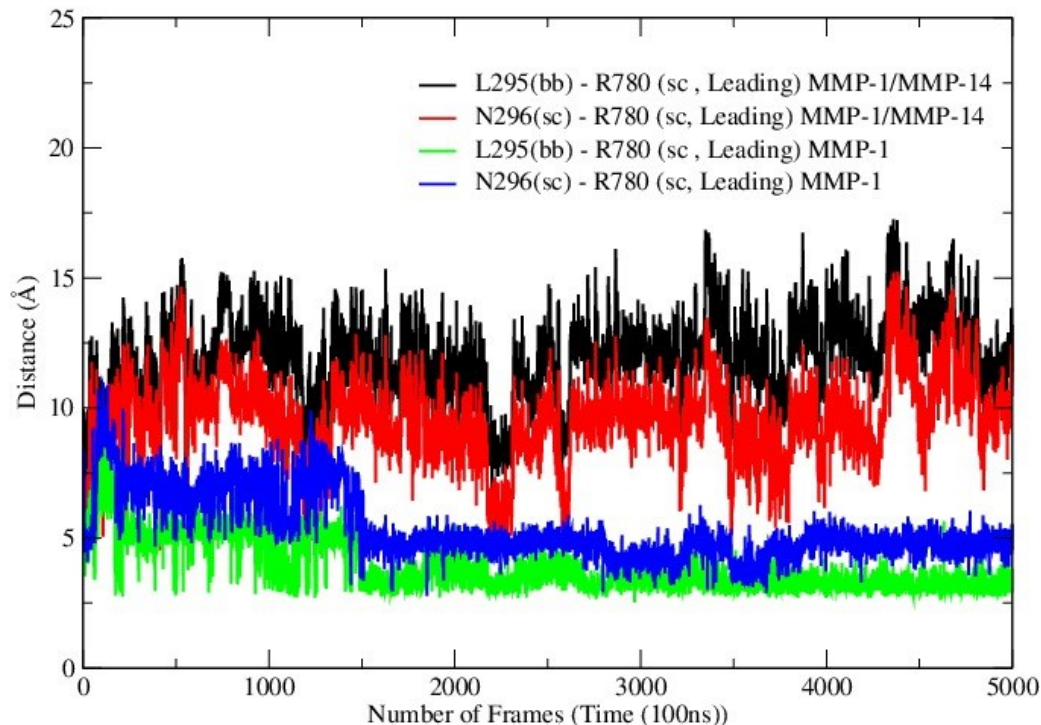
Figure S5.24. The interaction of the leading THP strand with the CAT domain of MMP-1/MMP-14 in comparison to MMP-1. Side chain = sc, backbone = bb.
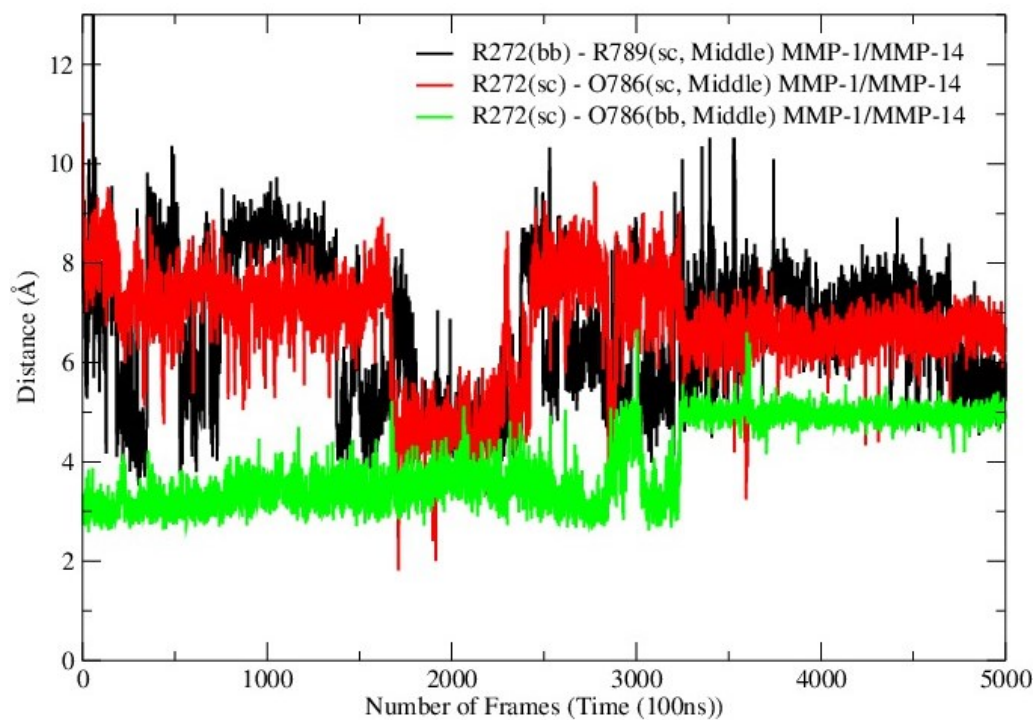
Figure S5.25. Interaction of MMP-1/MMP-14 R272 with the middle THP strand. Side chain = sc, backbone = bb.
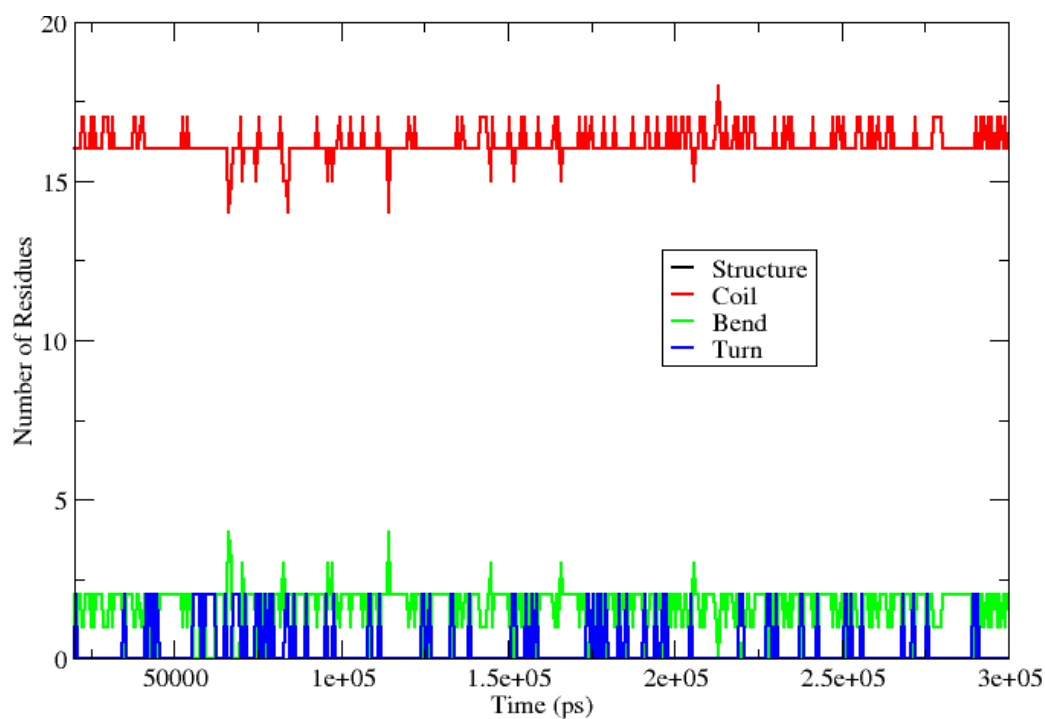
Figure S5.26. Secondary structure prediction of the linker region from the trajectory for MMP-1
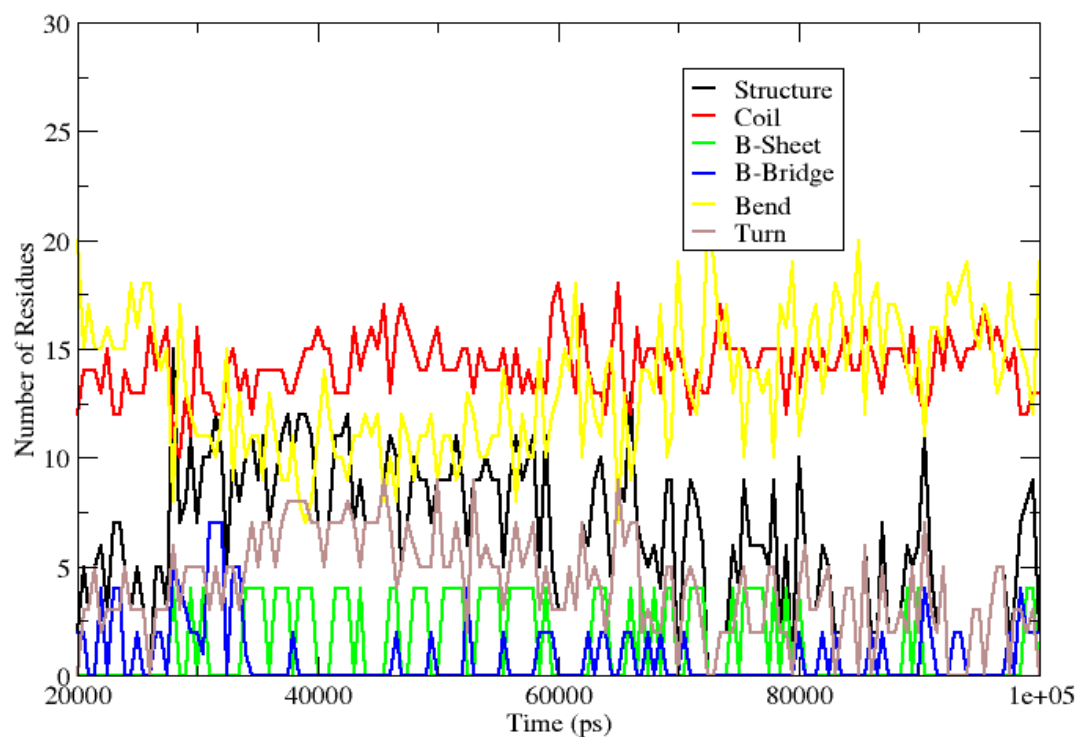
Figure S5.27. Secondary structure assessment of the linker region of MMP-1/MMP-14
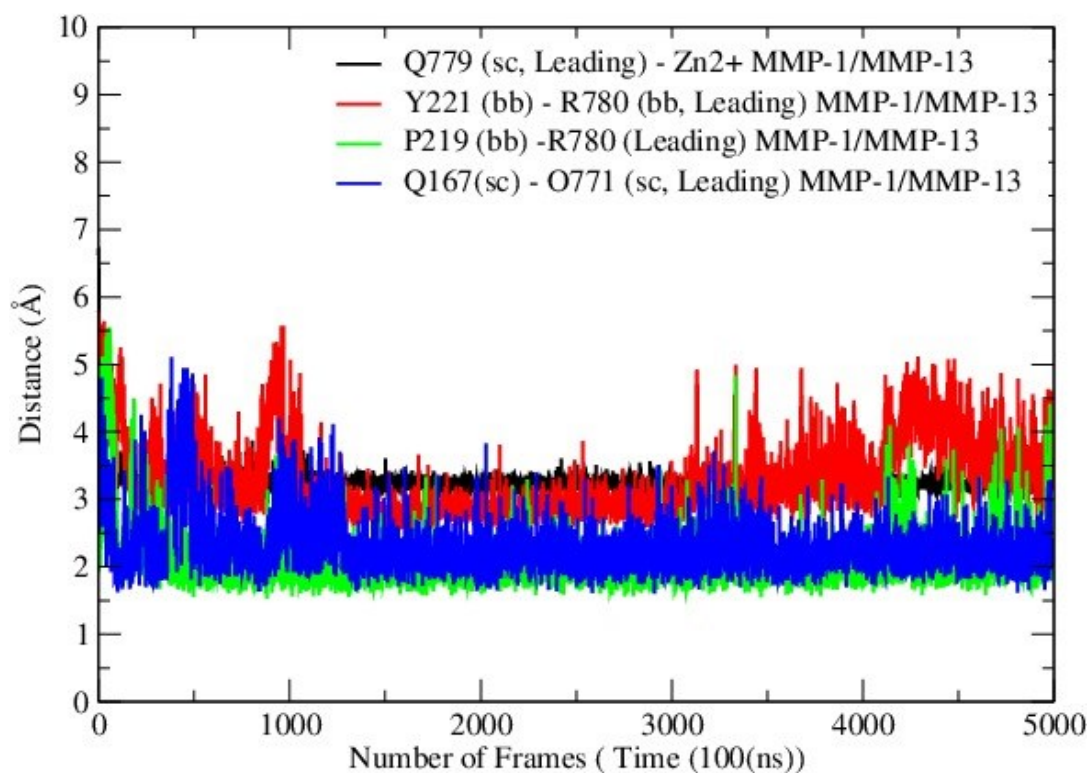
Figure S5.28. The CAT domain interactions of MMP-1/MMP-13 with the THP. Side chain = sc, backbone = bb.
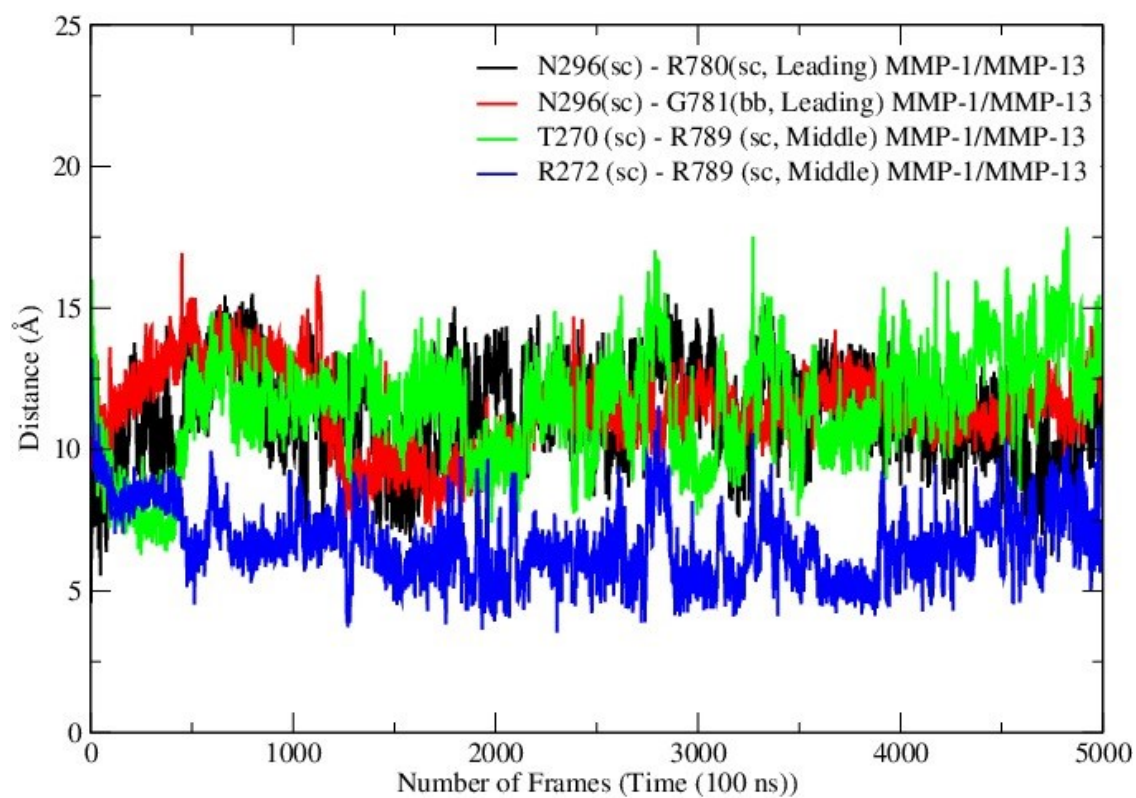
Figure S5.29. MMP-1/MMP-13 HPX domain interactions with the THP. Side chain = sc, backbone = bb.

# CHAPTER SIX   MATRIX METALLOPROTEINASE-1

AND MUTANTS

6.1 INTRODUCTION

The matrix metalloproteinases (MMPs) are key enzymes responsible for modulating the balance of collagen and other extracellular matrix (ECM) components in vertebrates. The catabolism of collagen by MMPs is a prerequisite for normal physiological function of cells in vertebrates [125-128]. MMP-1 has been the subject of a broad range of experimental studies and important conclusions have been drawn about the conformational behaviour of MMP-1 domains in solution. The MMP-1 hemopexin-like (HPX) and catalytic (CAT) domains are connected by a linker and exhibit complex conformational motions in order to bind the substrate and perform the chemical reaction of collagenolysis [131, 135-137]. The existence of equilibrium between the open/extended and closed/collapsed conformation of MMP-1 in solution has been established by experimental methods [131, 137] [156].

The first molecular dynamics (MD) simulation of MMP-1 bound to a collagen-model triple-helical peptide (THP) [19] provided insight into the role of the linker for modulating conformational dynamics between the HPX and CAT domains and revealed the dynamic nature of interactions between the THP, HPX domain, and CAT domain. The study confirmed the closed or "collapsed" state of the MMP-1 X-ray crystallographic structure (PDB 4AUO) [18] characterized by close orientation of the MMP-1 HPX and CAT domains. The MD did not change the MMP-1 conformation to open but allowed the effects of the flexibility on the closed conformation to be explored. Analyses of the dynamics effects including radius of gyration and distances between the centers of masses of the MMP-1 domains were performed [19]. An extensive bioinformatics analysis of 142 MMP

X-ray crystallographic structures revealed structural relationships within the enzyme family and structural arrangements important for inhibitor design [157].

Experimental studies have identified specific residues from the MMP-1 HPX domain which are involved in interactions with the THP[135] [18, 156] . Kinetic studies identified the effects of the mutations on the enzymatic activity of MMP-1 [135] [18.] However, there is little understanding to the atomistic effects of the mutations on MMP-1 structure and the impact on MMP-1 flexibility and binding of substrate. The association of the HPX domain with the leading and middle strand of the THP plays a significant role in properly orienting the scissile peptide bond of the leading strand to the CAT domain [131]. Mutagenesis studies on residues from the MMP-1 HPX domain (**Figure 6.1**) [135] [18] indicated effects on enzyme activity, which suggested a long-range conformational effect of the HPX domain and its influence on the ability of the CAT domain to perform effective catalysis. The mutations in MMP-1 range from a single amino acid substitution (E200A) in the CAT domain to a triple amino acid substitution (F289A/Y290A/P291A) in the HPX domain.

The prior MMP-1 mutagenesis studies provided a background to examine the atomistic effects of MMP-1 mutations. MD simulations are applied herein to understand how MMP-1 mutations perturbed the local structure and also to investigate the long-range effects on MMP-1 structure, dynamics, and interactions with a THP. MD simulations have been successfully applied previously to study the effect of mutations on key structural determinants in different enzymes [16, 17, 19, 101, 158]. In order to understand their influence on MMP-1 structure-function relationships, MD simulations were performed on the seven previously described mutants [18, 135]: E200A, F301Y, F289A/Y290A/P291A, I271A/R272A, L338A/H339A, R272A, and L295S.

Figure 6.1. (A) The 3D structure of inactive human MMP-1 (E200A mutant) complexed with a THP (PDB code: 4AUO[159]) drawn using UCSF Chimera [118]. MMP-1 consists

of the CAT domain, inter-domain linker, and HPX domain displayed in silhouette round ribbon representation. The THP leading, middle, and trailing strands are shown in tube representation in cyan, green, and red, respectively. The Zn^{2+} and Ca^{2+} ions bound to the enzyme are shown in spherical representation in ice blue and green, respectively. The mutations are shown in orange on the MMP-1•THP structure. The snapshot of the active site of the MMP-1•THP was obtained from MD simulation [19]. The two solvent molecules are coordinated to the Zn^{2+} along with three His residues with average distance maintained to the X-ray crystallographic distance using harmonic restraints in the simulation.
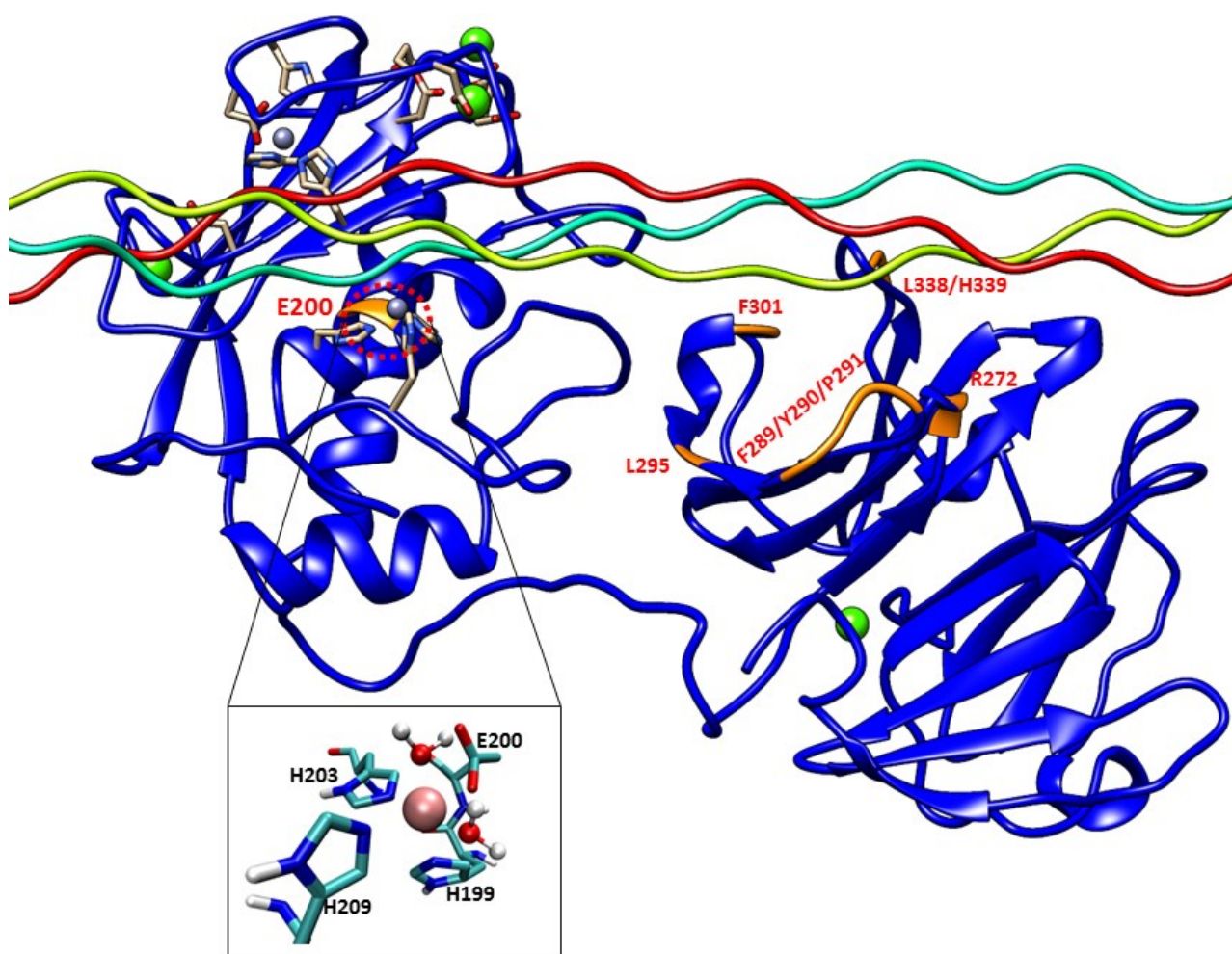
6.2 RESULTS AND DISCUSSION

6.2.1 Overall Conformational Flexibility of MMP-1•THP with Mutated Residues

The Root Mean Square Deviation (RMSD) as a function of time was used to access the structural stability of the simulations. The RMSDs of the mutant forms along with wild-type (WT) MMP-1•THP showed similar trends in that the structures reached equilibration before 20 ns (**Figure 6.2**). The average value of RMSD ranged from 5.3 Å (F289A/Y290A/P291A triple mutant) to 4.5 Å (F301Y) (**Table S6.1**). The flexibilities of individual residues in the MMP-1•THP complex were assessed by using Root Mean Square Fluctuation (RMSF) analysis (**Figure S6.1, Tables S6.1 and S6.2**). The basal level fluctuation for WT MMP-1•THP was considered to be 1.4 Å (based upon the mean value), confirmed by a distribution analysis of the RMSFs (Figure S2). 43% of residues showed fluctuations greater than 1.4 Å and 57% of the residues showed fluctuations less than 1.4 Å (**Table S6.1**). The RMSF plot of the I271A/R272A and F301Y showed a greater number of residues with fluctuations larger than 1.4 Å in contrast to WT, 60 and 54%, respectively (**Table S6.1**). The triple mutant F289A/Y290A/P291A showed the lowest number of residues (29%) with fluctuations > 1.4 Å, followed by L338A/H339A (31%) (**Table S6.1**). The different mutations induced small changes in the number of hydrogen bonds in comparison to the WT MMP-1•THP (**Table S6.3**, **Figure S6.3**).

The average RMSF value of the linker region in all of the mutants except I271A/R272A showed a slight decrease with respect to WT (**Figures 6.3 and S6.4A, Table S6.4**). The change in linker flexibility (decrease or increase) could influence the communication between the HPX and CAT domains and the conformational dynamics of the CAT domain and hence contribute to altered enzyme activity. The largest reduction in linker flexibility was seen in the F289A/Y290A/P291A mutant, with average RMSF value of 1.4 Å in contrast to 2.2 Å for WT MMP-1•THP (**Figure 6.3**, Table **S6.4**).

Experimental studies have identified specific residues within MMP-1 HPX domain blades I and II (**Figure S6.4B**) that interact with the leading and middle strands of the THP [18, 136, 156]. The study performed by Zhao et al. on MT1-MMP also identified blade I and II residues that interact with the THP [160]. Our recent MD study of MMP-1•THP stressed the importance of HPX domain blade I and blade II residues for interactions with the THP [19]. The majority of the experimental HPX domain mutations are located in blade I, apart from the L338A/H339A mutation which is in blade II [136] [18]. The blade I residues constitute the S_{10}' exosite of MMP-1, and mutations in blade I reduced enzymatic activity [135] [136] [18]. The interactions of HPX domain blade I and II residues with the THP also play a significant role in guiding the CAT domain for effective catalysis of the leading strand of the THP [131, 136, 137, 156]. RMSF analysis of blade I residue mutants showed relatively higher flexibility of this region for I271A/R272A (**Figures S6.4C** and **Table S6.4**). The increase in the flexibility of blade I in this mutant with respect to the WT MMP-1•THP could potentially influence THP binding and hence the enzyme activity. Indeed, the I271A/R272A mutation reduced MMP-1 collagenolytic activity [135]. The residues constituting the mutation in blade II showed slightly lower RMSF in comparison to WT MMP-1•THP (**Figure S6.4D**). The mutations of residues in blade II (L338A/H339A) resulted in an increase in the enzyme activity [18]. Alteration (decrease or increase) in the flexibility of the linker region would influence the communication between the HPX and CAT domains, binding of the THP, and thus potentially enzyme activity. The specific atomistic mechanism of this effect would need, however, further studies. In addition, strong correlation between the flexibility of residues in blades I and II and the catalytic activity was not extracted from the current data.

Mutations can influence not only the local structure, but also regions which are distant from the mutation sites [161]. In order to explore these effects we analysed the distance between the THP scissile bond and the catalytic Zn^{2+} in all mutants and the WT (**Table**

S6.3, Figure 6.4). Importantly, in all mutant simulations we found that the above distance is larger than in the WT, which indicates further that the mutations (all of which but one are in the HPX domain) effect interactions between the scissile bond from the leading chain of the THP and the Zn^{2+} from the MMP-1 CAT domain.

The Radius of Gyration (Rg) of the mutants and WT MMP-1•THP showed similar profiles of structural compression as a function of simulation time (**Figure 6.5**, **Table S6.5**), with L295A showing the smallest Rg and R272A the largest. The compressions of the structures during simulations of the mutants are consistent with the MMP-1•THP simulation studies [19] and was consistent with the "closed" or "collapsed" form of MMP-1 observed in the 4AUO X-ray crystallographic structure [19]. The small differences in the averaged values of the Rg and in the distances between centres of mass between both domains (**Figure 6.6**, **Table S6.5**) indicate subtle but distinct effects of the mutations on the MMP-1•THP structure and flexibility.

6.2.2 Conformational and Dynamical Effects of Individual Mutations

The E200A mutation is utilized to greatly suppress the enzymatic activity of MMP-1. The X-ray crystallographic structure of MMP-1•THP (4AUO) incorporated the E200A mutation. In the WT simulations [19], E200 is in close vicinity to the catalytic Zn^{2+} with an average distance of 2.5 Å (**Figures 6.1** and **6.7A**). The carboxylate group of the side chain of E200 interacts with Q779 from the THP leading strand and also makes a hydrogen bond with the N-H group from the backbone of A165 with an average distance of ~3.0 Å (**Figure S6.5A**).
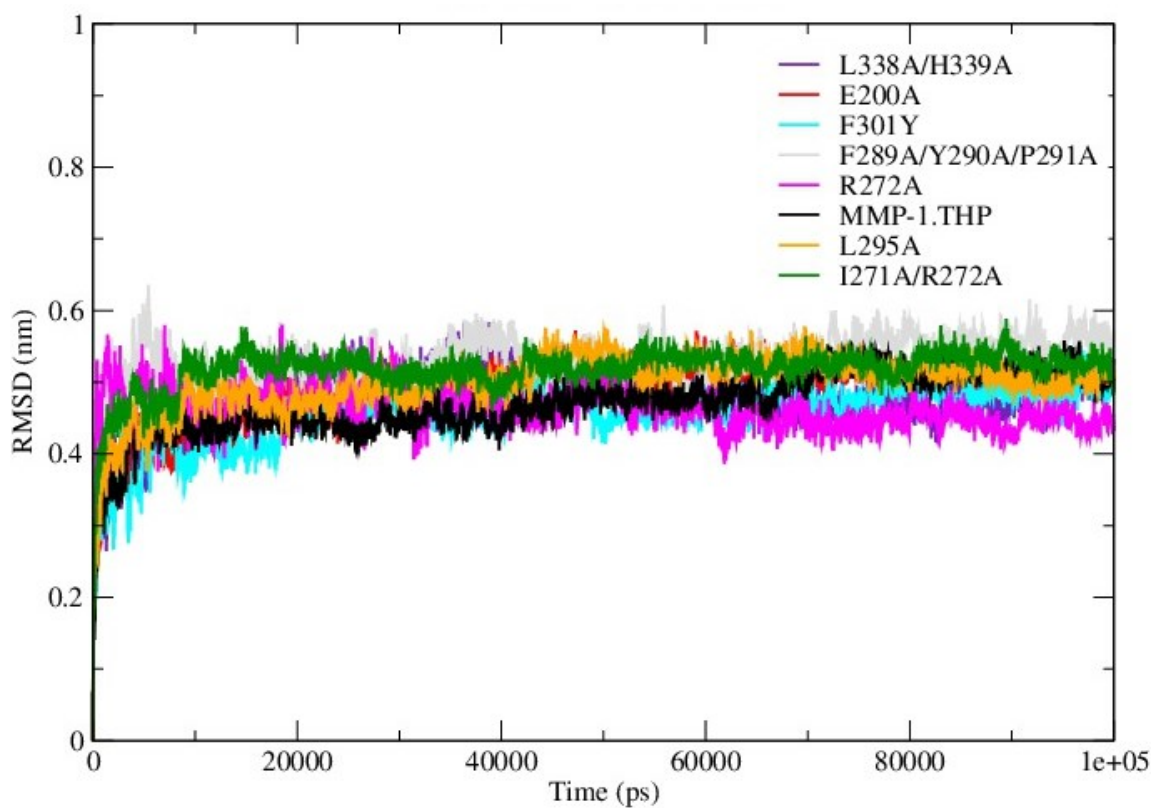
Figure 6.2. RMSD of all Cα atoms of MMP-1•THP complex in comparison to the mutants.
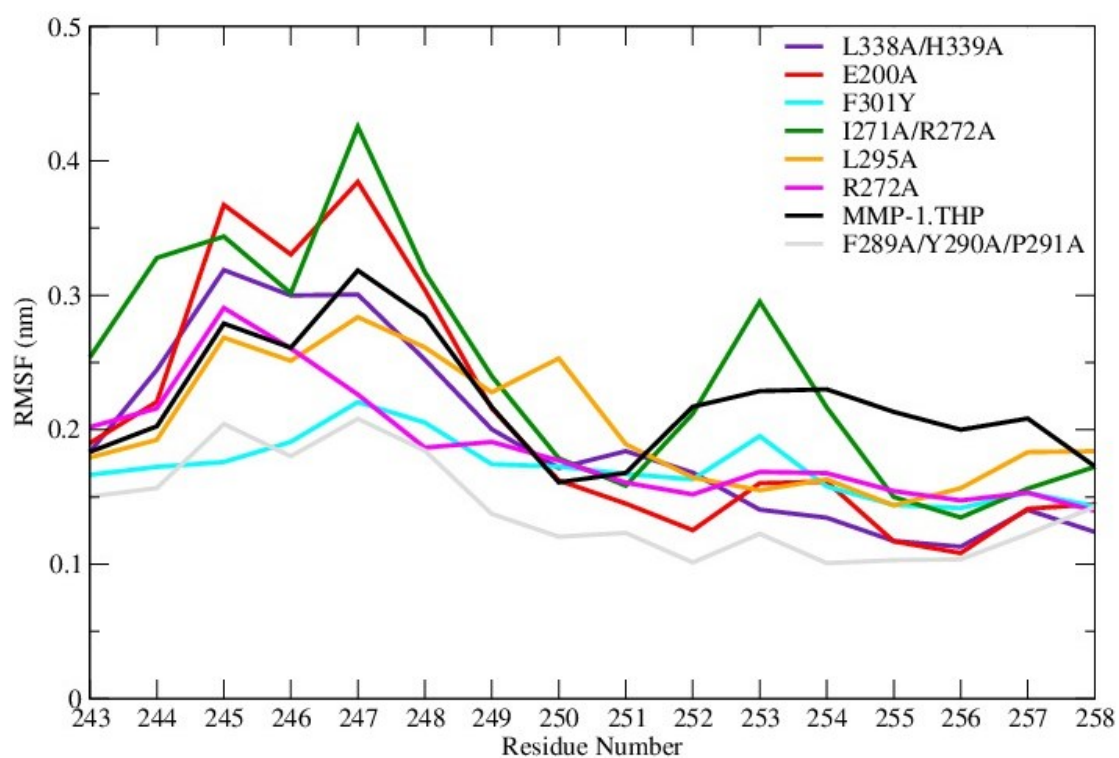
Figure 6.3. RMSF analysis of linker residues of WT MMP-1•THP and mutants.
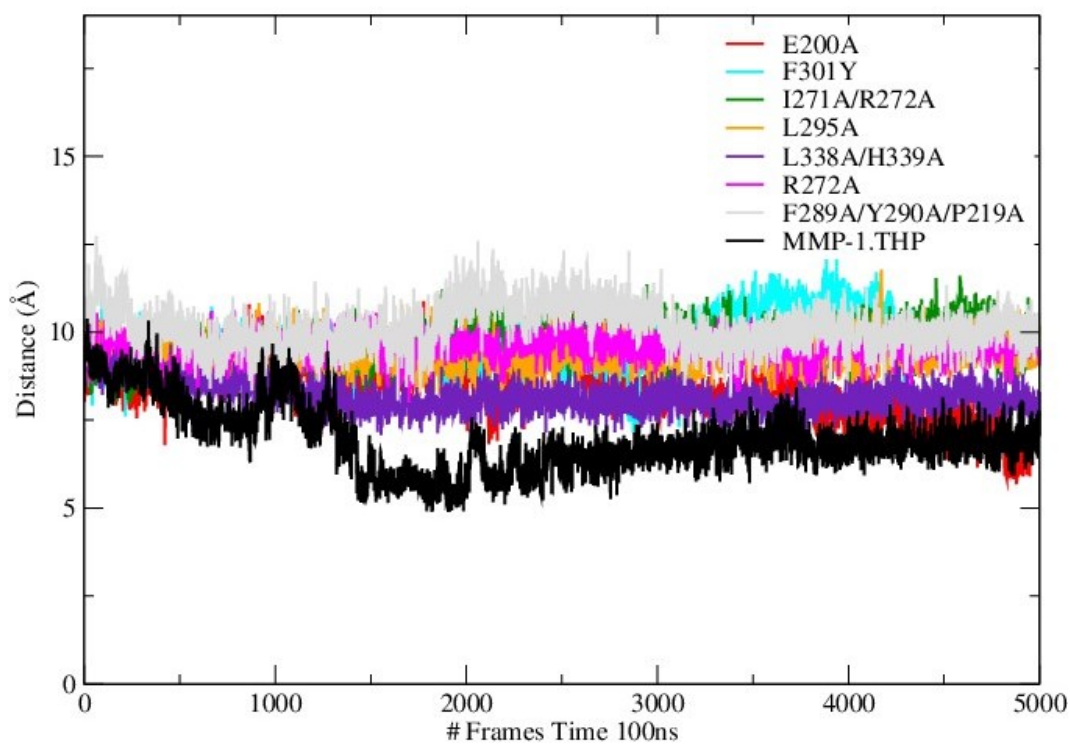
Figure 6.4. Time evolution of the distance between the scissile bond G775-L776 and Zn^{2+} in WT MMP-1•THP and the mutants.
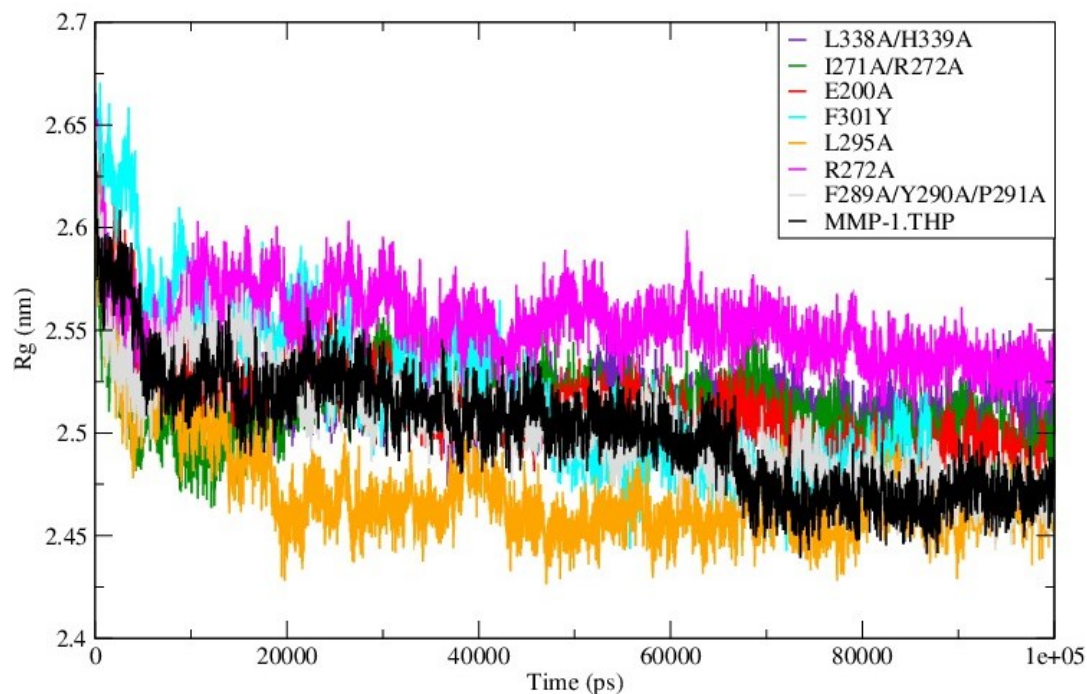
Figure 6.5. The radius of gyration of the mutants and WT MMP-1•THP.
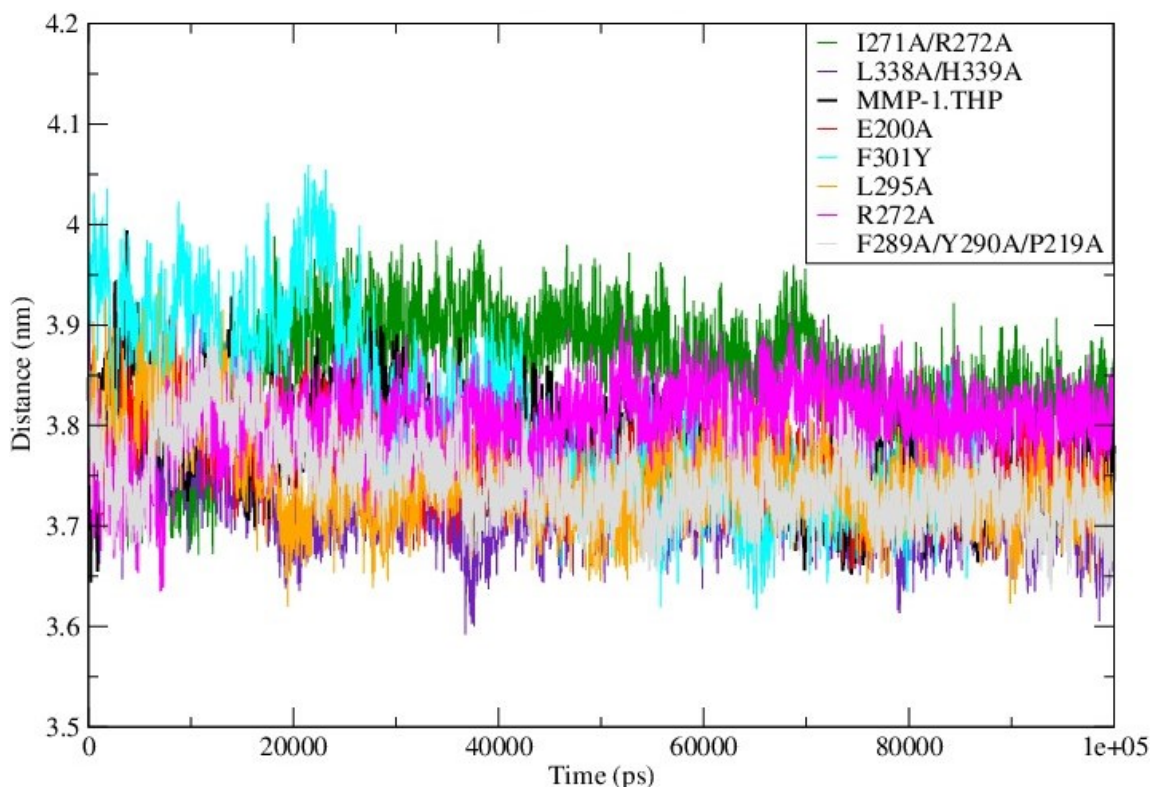
Figure 6.6. The distance between the center of mass of the CAT and HPX domains comparing the mutant forms and WT MMP-1•THP.

In the E200A mutant there were three water molecules which coordinated the active site Zn^{2+} with an average distance of 2.1 Å (**Figures S6.5B, S6.6A**). Two solvent molecules coordinated with the catalytic Zn^{2+} simultaneously made hydrogen bonds with the E200 residue. The interactions between Zn^{2+} and the water molecules were also confirmed by radial distribution analysis (**Figures S6.5C, S6.6B**). The involvement of the water is in agreement with Quantum Mechanics and Molecule Mechanics (QM/MM) studies of the reaction mechanism of MMP-2 [133, 147, 149]. The side chain of Q779 (THP leading strand) accessed the active site in a similar manner as in the WT MMP-1•THP [19]. The side chain of A200 in E200A moved away from the catalytic Zn^{2+} with an average distance of ~7.3 Å with respect to 2.2 Å in MMP-1•THP. The backbone of A200 no longer made hydrogen bonds with the backbone of A165 (**Figure S6.6C**). The RMSF profile of the E200A mutant did not show any significant difference with respect to MMP-1•THP (**Figure S6.1**). However, the loop region between β4-β5 of the CAT domain along with the

215

N-terminal region showed increased fluctuations (**Figure S6.7A**). The residues of the linker region (250-259) and β18-β19 (380-390) showed reduced fluctuations compared with MMP-1•THP (**Figure S6.7B-C, Table S6.2**).

The HPX domain of the E200A mutant showed lower RMSDs with respect to the HPX domain of WT MMP-1•THP (**Figure S6.8**). Understanding of the correlated character of atomistic motions in proteins is vital since it relates their structure to function. This insight might be difficult to obtain experimentally, but can be straightforwardly extracted from the MD trajectories [162]. The dynamics cross correlation analysis of the E200A mutant (**Figure 6.8**) showed correlated motion between the strands of the THP as was observed for MMP-1•THP. However, there was no negative correlated motion seen in the THP strand as was observed for WT MMP-1•THP. An important difference seen in the E200A mutant was that the THP strands showed very low positive correlated motion towards the CAT domain. There was positive correlated motion seen between the THP and the HPX domain. The extents of both positive and negative correlated motion were reduced in the E200A mutation. Thus, although commonly used to provide a locally reduced enzyme activity, the E200A mutation has long range effects.

The F301Y mutation site is located between the loop region of D α-helix and β10 of the HPX domain blade I (**Figures 6.1, 6.7B, and S6.9**), part of an MMP-1 exosite for collagen binding. In the MD simulation of the WT MMP-1•THP, F301 exhibited hydrophobic interactions with the side chains of I782 and L785 from the THP middle strand, with average distances of 4.3 and 4.1 Å, respectively (**Figure S6.10**). This result is inconsistent with the X-ray crystallographic structure but supports results from NMR spectroscopic experiments [136] [18, 131]. The hydroxyl group of Y301 in the F301Y mutant forms hydrogen bonds with the side chains of R285 (2.5 Å) and Q335 (3.1 Å) of the HPX domain. The backbone of L795 also forms a hydrogen bond with the side chain of Y301,

with an average distance of 3.5 Å. I782 and L785 from the THP show weaker hydrophobic interactions in the F301Y mutant compared with WT MMP-1•THP. The RMSF analysis of F301Y showed increased fluctuations of residues from blade III between β15-D-β16 (**Figure S6.11A**). The linker region of the F301Y mutant showed reduced fluctuations in comparison to the linker region of WT MMP-1•THP (**Figure S6.11B**). The residues 294-310 from blade I that belong to the loop region between β9-D-β10, in the vicinity of the F301Y mutation also showed increased flexibility (**Figure S6.11C**). This region is in very close contact with the THP leading strand and makes important binding interactions with the THP as experimentally demonstrated by Arnold and co-workers [136]. The RMSD of the CAT and HPX domains of the F301Y mutant showed lower structural deviation in comparison to MMP-1•THP (**Figure S6.12**). The F301Y mutant in the study performed by Arnold and co-workers [136] had an Rg value (33.6 Å) which was 18% higher compared with WT MMP-1 (Rg = 28.5 Å). These results suggest disruption of the CAT/HPX domain interface in the F301Y mutant due to destabilization of collagen interactions mediated by blade I residues of the HPX domain [136]. The differences between the Rg of the F301A mutant in comparison to WT MMP-1•THP in our study are smaller (due to the length of the simulation) but still indicated a similar trend.

The dynamic cross correlation analysis (DCCA) of F301Y showed relatively strong anti-correlation between the CAT and HPX domain residues in comparison to MMP-1•THP and slightly reduced positive correlation overall (**Figure 6.8**). This is an indication of the sensitive effect of the F301Y mutation on the interactions with the THP and motions of the HPX domain. The F301Y mutant exhibited a ~90% reduction in collagenase activity [18]. The F301Y mutation also showed a 30-fold decrease in THP binding by MMP-1 [136], which was proposed to be due to the absence of interaction of the F301 side chain with the THP molecule rather than large conformational changes in the HPX domain.

The L295S mutation is located in the β9 region of the HPX domain (**Figure 6.7C**). In the MD of WT MMP-1•THP the backbone of L295 forms a hydrogen bond with the side chain of R780 of the THP leading strand with an average distance of 3.5 Å. The residues in the vicinity of L295, such as N296 and E294, also have interactions with the THP leading strand [19]. The side chain of S295 in L295S formed new interactions with the backbone of P256 of the linker region with an average distance of 4.1 Å (**Figure S6.13**). The hydrogen bond with R780 from the THP leading strand in WTMMP-1•THP is not present in L295S (**Figure S6.13**).

The CAT domain of L295S showed slightly increased fluctuations with respect to WT MMP-1•THP (**Figure S6.14A-B**), while the HPX domain exhibited reduced fluctuations in the region encompassing β17-β18 (**Figure S6.14C**). The linker region residues joining the HPX domain also showed slightly reduced flexibility in comparison to WT MMP-1•THP (**Figure S6.14D**). The DCCA analysis of L295S revealed some reduction in both positive and negative motion with respect to MMP-1•THP. The triple-helix strands lost the correlated motion with the HPX domain and showed correlated motion towards the CAT domain (**Figure 6.8**) which indicated potential influence on THP binding by the HPX domain in this mutant. The L295S mutant showed approximately 60% reduction in collagenase activity in comparison to MMP-1 [18].

The R272A and I271A/R272A mutations are located in the blade I region (S_{10}' exosite) of HPX domain (**Figure 6.7D**). R272 has extensive interactions with the THP middle strand in MMP-1•THP [18] [19]. The side chain and the backbone of R272 make hydrogen bonds with the backbone and side chain of O786 and R789, respectively. The aliphatic region of the R272 side chain is involved in formation of hydrophobic interactions with the side chains of R789 and L785 of the THP middle stand [18] [19]. The backbone of I271 forms hydrogen bonds with the backbone of E274 and T269, and the side chain of I271

participates in hydrophobic interactions with the side chain of V321 of blade II [18, 19]. In the R272A mutant, the side chain of A272 made hydrophobic interactions with the side chains of L785 and Q788 of the THP middle strand with average distances of 4.1 and 4.2 Å, respectively (**Figure S6.15A**), however there are no hydrogen bonds with O786 and R789 (**Figure S6.15B**). The RMSF of the residues in the vicinity of R272A (271-276; the loop between β6-β7 of blade I) showed an increased fluctuation in comparison to WT MMP-1•THP (**Figure S6.16A**). These residues interact with the THP [136]. The residues harboring the active site in the CAT domain (210-215; loop between the B-C helix) showed slightly increased fluctuations in contrast to WT MMP-1•THP (**Figure S6.16B**). R272A showed a slight increase in its average Rg (25.5 Å) in comparison to WT MMP-1•THP (24.8 Å).

The I271A/R272A mutant (**Figure 6.7D**) showed an increase in the RMSF of blade I residues ranging from 275-285 (β7-β8) and 295-310 (loop region β9-β10) (**Figure S6.17**) and also increased flexibility seen in residues 345-355 of blade II of the HPX domain (**Table S6.2**). The extent of anti-correlated motions in I271/R272 is similar to WT MMP-1•THP and new anti-correlated motion emerged (**Figure 6.8**). The HPX domain showed anti-correlated motions between the CAT domain and strands of the THP. The linker region exhibited anti-correlated motion towards the HPX domain. New correlated motions between the HPX domain and the CAT domain appeared. The R272A mutant showed a similar profile to dynamic cross correlated motions of I271/R272A, but with smaller magnitudes (**Figure 6.8**).

R272A and I271A/R272A mutations have the most significant effect on the enzyme activity of MMP-1. The I271A/R272A mutant has less than 10% of the collagenase activity of WT MMP-1[135] [18]. Analysis of individual kinetic parameters for R272A and I271A/R272A hydrolysis of a fluorogenic THP (fTHP-17) revealed that K_M increased

and k_{cat} decreased compared with WT MMP-1 [18]. Since both K_M and k_{cat} changed, the R272A and I271A/R272A mutations affected more than just substrate binding. It was suggested that both substrate binding and coupled motions for catalysis were altered by these mutations [135]. The present MD study has identified the long range effects of R272A and I271A/R272A mutations within MMP-1•THP, which was primarily new anti-correlated motions (**Figure 6.8**).

The F289A/Y290A/P291A triple mutant (**Figure 6.7E**) was next considered. The interactions in the X-ray crystallographic structure of residues Y290 and F289 were stable in the MD simulation of the WT MMP-1•THP [19]. The triple mutant showed the very lowest RMSF overall as compared to other mutants in relation to WT MMP-1•THP (**Table S6.1**). The cross correlation analysis showed very low anti-correlated motion observed in F289A/Y290A/P291A (**Figure 6.8**). There was also a significant reduction in the positive correlation motions with respect to WT MMP-1•THP. The THP strands showed positive correlated motion towards the residues of the CAT domain. F289A/Y290A/P291A had ~30% reduction in collagenase activity compared with WT MMP-1 [18].

The last mutant considered was L338A/H339A. L338 and H339 are located on the loop between β13 and β14 of the HPX domain (blade II) and are on average 17 Å away from the THP (**Figure 6.7F**). The X-ray crystallographic structure and MD simulations of the WT do not reveal significant interactions of these residues with the THP. The average RMSF of this mutant was slightly lower than the WT (**Table S6.1**). The DCCA of L338A/H339A (**Figure 6.8**) showed significantly lower anti-correlated motions with respect to WT MMP-1•THP. Residues 100-120 of the CAT domain showed limited negative correlation against the HPX domain β-sheets (residues 310-380). Residues 175-185 of the CAT domain also showed reduced negative correlation towards residues 330-350 of the HPX domain. However, the L338A/H339A mutation also results in an anti-

correlation between the *C*-terminal region of the CAT domain and the *N*-terminal region of the HPX domain that is not observed in the WT or other mutant enzymes (**Figure 6.8**). The anti-correlation may be a shift of an anti-correlation in the WT enzyme (noted as 6 in **Figure 6.8**) further towards the *C*-terminal region of the CAT domain. A unique small correlation was observed between the *C*-terminal regions of the CAT and HPX domains in the L338A/H339A mutant enzyme (**Figure 6.8**). The combined L338A/H339A mutation increased the collagenase activity of MMP-1 by approximately 10% [18] .The change in the anti-correlated motions resulting from the L338A/H339A mutation may play a factor in increased collagenase activity. The changes in the anti-correlated motions can be complex and not unidirectional. In the case of L338A/H339A, a decrease in previously observed anti-correlations and the appearance of a unique anti-correlation are accompanied by increased enzyme activity, while in F289A/Y290A/P291A a virtual complete lack of anti-correlations (as well as decreased correlated motions) are related to decreased enzyme activity. Further studies, including simulations at longer time scales, would be necessary for more detailed analysis of these effects.

It is important to note that the X-ray crystallographic structure of MMP-1•THP [18] does not represent the productive enzyme•substrate (E•S) complex. In contrast, NMR studies have reported a productive complex [156]. Nevertheless, the X-ray crystallographic structure of MMP-1•THP is still the only available structure that presents the atomistic interactions between the enzyme and substrate (the NMR-derived structure is based on docking the THP). Both the X-ray crystallographic and NMR structures of MMP-1•THP are in agreement about most of the THP interactions with the HPX domain and the closed conformation of MMP-1. The structures differ in the position of the THP in the CAT domain, where the NMR-derived structure represents a productive complex, while the X-ray crystallographic structure does not. Most of the mutations are located in the HPX domain, therefore the MD simulations provide relevant insight on the influence of these

mutations on the MMP-1•THP structure and dynamics. In the absence of experimental structures of the mutants (with the exception of E200A), the present study also provided an atomistic view of short- and long-range effects and their impact on the correlated motions of enzyme-substrate complex structures. The effects of the mutations on the different X-ray crystallographic and NMR structures would be the subject of further comparative MD studies.

6.3 CONCLUSIONS

Experimental studies show that a variety of mutations in the HPX and CAT domains of MMP-1•THP have strong effects on the enzyme activity. The present computational studies provide atomistic explanation of these effects and reveal that the mutations have not only local structural effects but also long-range impact on the protein structure, dynamics, and the interactions with a triple-helical substrate. The mutations change the flexibility around the local site but also influence distant regions from the HPX domain, the linker region, and the CAT domain. In addition, the mutations modulate the intensity and the nature of the correlated and anti-correlated motions. An increased collagenase activity in L338A/H339A mutant corresponded to the appearance of a unique anti-correlated motion and decreased correlated motions, while decreased collagenase activity in the other mutations corresponded both to increased and decreased anti-correlated motions. Our studies provide important structural and dynamic information which correlates and helps to explain the experimentally measured effects of the mutations. With the exception of the E200A mutant, there are no X-ray crystallographic structures of MMP-1•THP with mutated residues, and the present study provides this missing structural information and asserts that the mutations have delicate, distinct, and specific effects on the structure, interactions, and dynamics in MMP-1•THP. The magnitudes of the changes in the local interactions and dynamics are in agreement with the effects of mutations in other proteins [16, 17, 19, 158] [84].
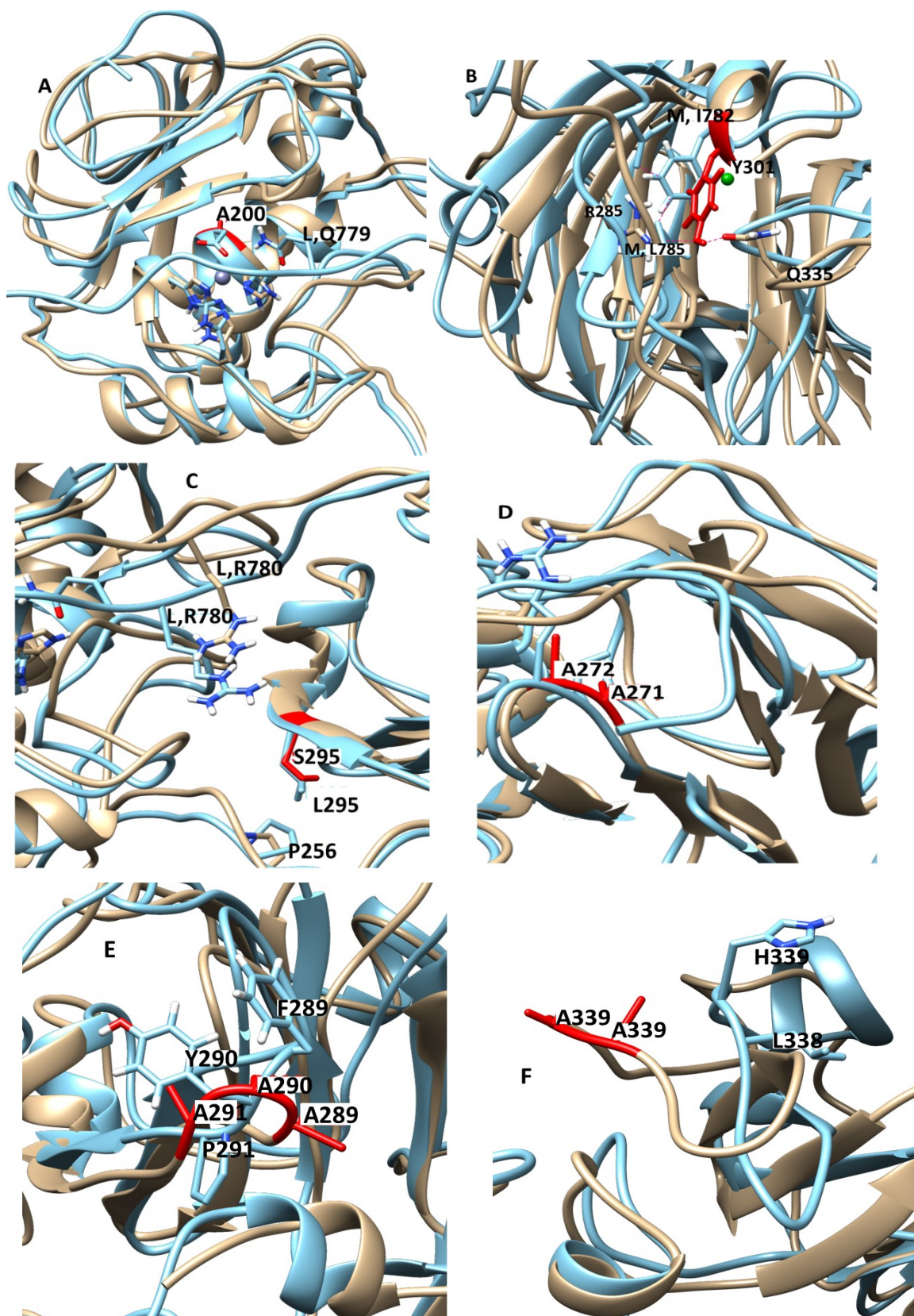
Figure 6.7. Locations of the mutants in the structure of MMP-1•THP. (A) E200A (1.21 Å),
(B) F301Y (1.24 Å), (C) L295S (1.30 Å), (D) I271A/R272A (1.3 Å), (E)

F289A/Y290A/P291A (1.23 Å), and (F) L338A/H339A (1.9 Å). The most populated

cluster of MMP-1•THP (cyan) is superimposed on the most populated cluster structure

from the different mutants (brown). The RMSD between the mutant structure and the WT

MMP-1•THP are indicated in parentheses.
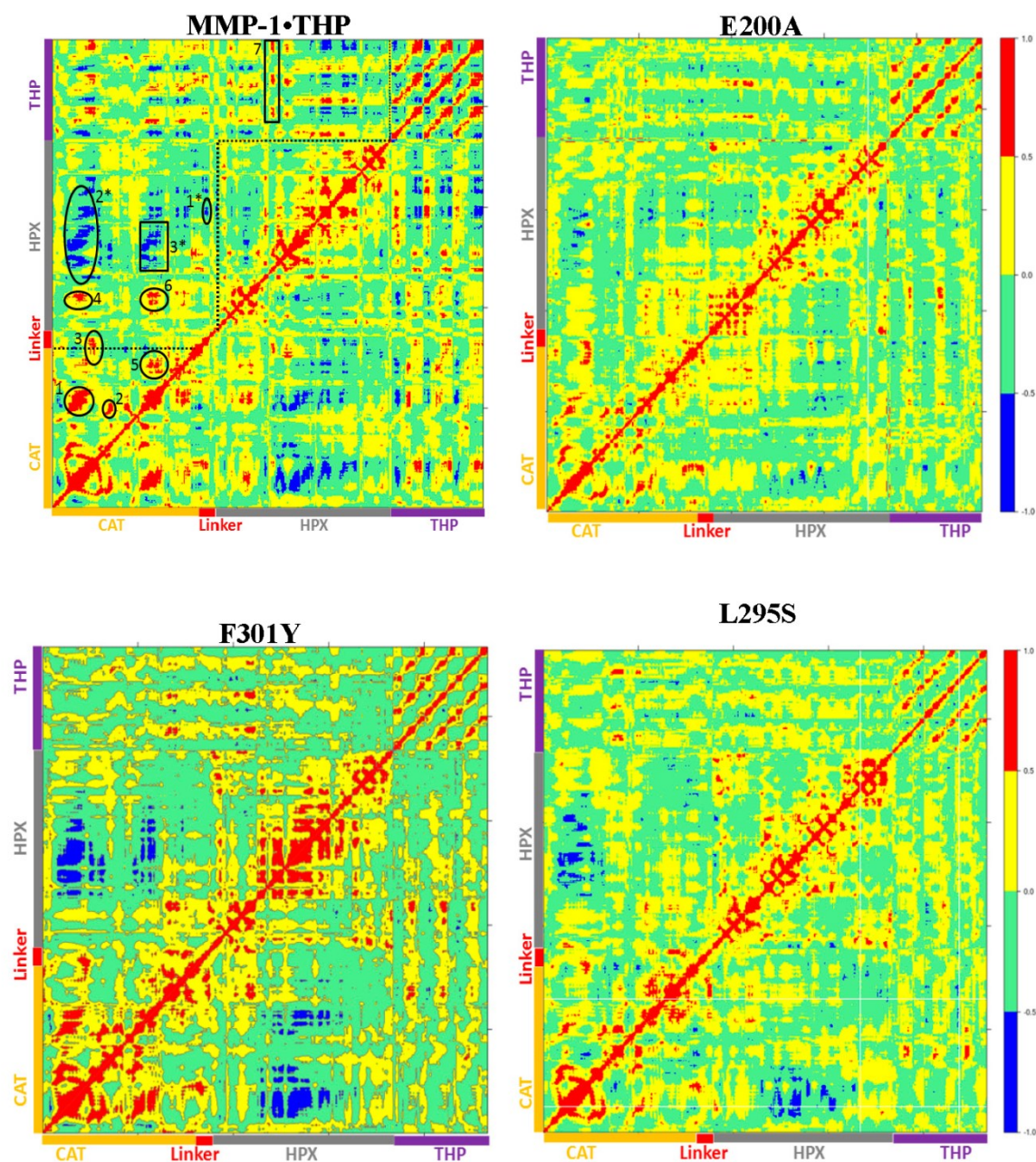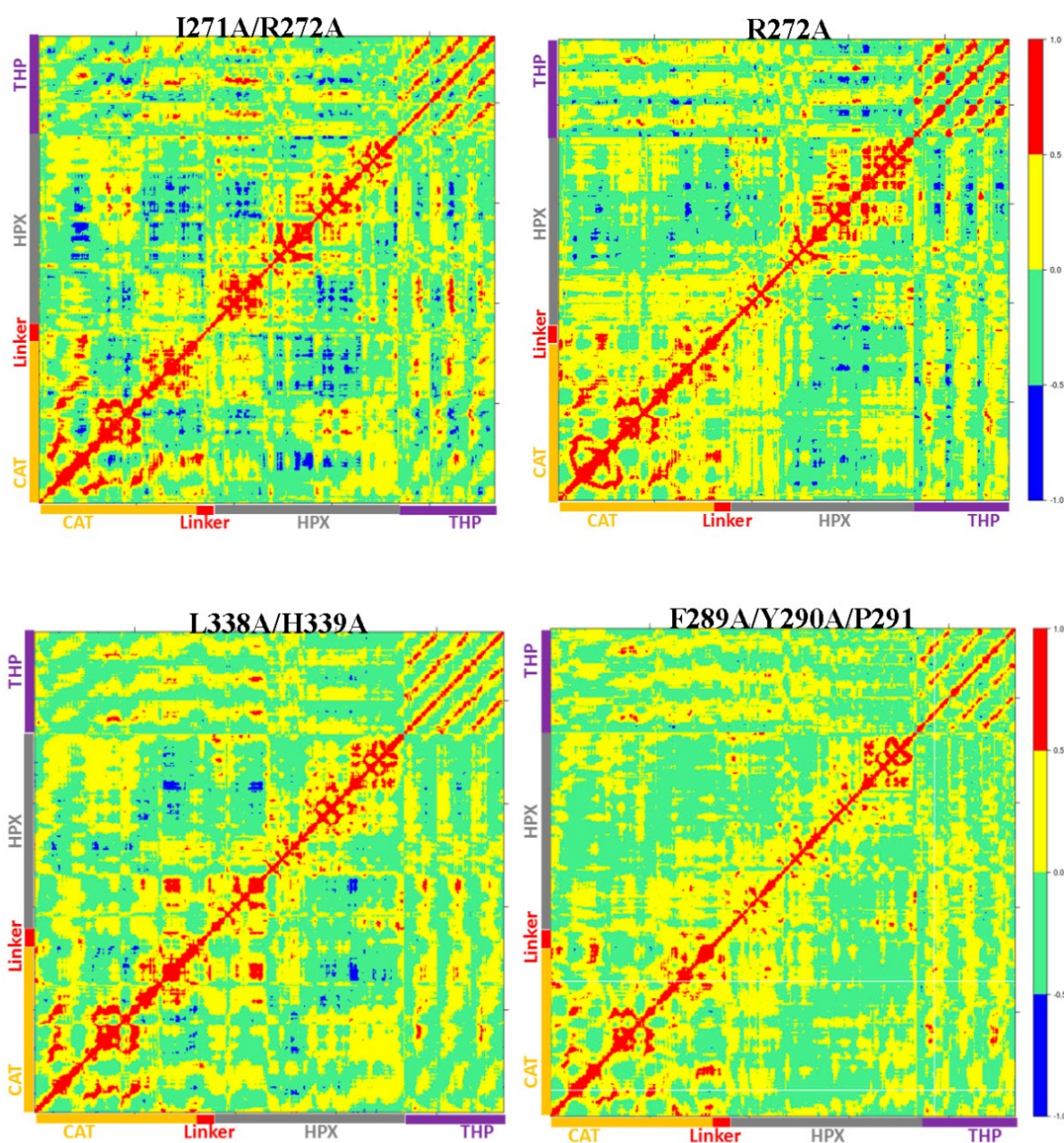
Figure 6.8. DCCA of MMP-1•THP and mutants. The scale of correlated motion ranges from +1 to -1 and represents positive (red) and negative (blue) motions of Cα atoms. The analysis was performed using Bio3D package in R For WT MMP-1•THP, the areas labeled 1-7 indicate positive correlated motions, while the areas labeled 1*-3* indicate negative correlated motions.

6.4 SUPPORTING INFORMATION

Table S6.1. Average RMSD and RMSF of wild type and mutant MMP-1•THP.

MMP-1	Average RMSD	RMSD Standard Deviation	RMSF > 1.4 (Å) (%)	RMSF < 1.4 (Å) (%)	Location of Mutation
WT	4.7	0.44	43	57	
F289A/Y290A/P291A	5.3	0.35	29	71	β8-β9 (HPX)
R272A	4.6	0.29	43	57	β6-β7 (HPX)
L338A/H339A	4.7	0.41	31	69	β13-β14 (HPX)
F301Y	4.5	0.42	54	46	D-β10 (HPX)
L295S	4.9	0.39	47	53	β9 (HPX)
I271A/R272A	5.2	0.92	60	40	β6-β7 (HPX)
E200A	4.8	0.44	38	62	B helix (CAT)

Table S6.2. RMSFs of the structural elements in MMP-1 mutants compared with WT MMP-1•THP. An increase or decrease is based on an average change in RMSF of > 0.3 Å in at least 50% of the residues.

Orange = CAT domain, Green = HPX domain, Red = Linker region.

Mutant	Increased		Decreased	
	Residue span	Structural location	Residue span	Structural location
F289A/Y290A/ P291A			100-128, 146-158, 225-237, 245-258, 319-338, 382-391	β1-A, β3-β4, Linker, E-β11-β12-β13, β17-β19
R272A	81-92, 210-215, 271-280, 350-355, 430-435	N-terminus, B-C, β6-β7, F-β15, G-β23	250-258, 148-158, 180-189	Linker, β3-β4, β5-B
I271/R272A	81-90, 275-285, 295-310, 345-355	N-terminus, β7-β8, β9-D, F-β15		
L338A/H339A	168-172, 288-292	β4-β5, β8-β9	100-109, 147-157, 252-263, 377-391	β1-A, β4-β5, Linker, β17-β19
F301Y	81-90, 168-173, 187-192, 296-309, 364-368	N-terminus, β4-β5, β5-B, β9-D-β10, β15-β16	250-262	Linker
L295S	81-92, 167-172, 206-223, 252-258	N-terminus, β4-β5, B-C, Linker	375-390	β17-β18
E200A	168-173, 83-87	β4-β5, N-terminus	250-259, 380-390	Linker, β18-β19

227

Table S6.3. Average number of hydrogen bonds and average distance between the scissile bond and catalytic Zn^{2+}.

Enzyme	Average number of H-bonds	Average distance between the scissile bond and catalytic Zn^{2+} (Å)
WT	296	6.9
F289A/Y290A/P291A	295	10.1
R272A	293	9.5
L338A/H339A	301	8.1
F301Y	300	9.7
L295S	299	9.3
I271A/R272A	300	9.6
E200A	301	8.4

Table S6.4. The average RMSF value of the linker, blade I and blade II regions, and HPX and CAT domains of mutants in comparison to MMP-1•THP. μ = mean RMSF and σ = standard deviation.

MMP-1	μ linker	σ linker	μ blade I	σ blade I	μ blade II	σ blade II	μ HPX domain	σ HPX domain	μ CAT domain	σ CAT domain
WT	2.2	0.4	1.0	0.5	1.3	0.4	1.3	0.5	1.1	0.3
F289A/ Y290A/ P291A	1.4	0.3	0.9	0.9	1.1	0.4	1.2	0.4	0.9	0.2
R272A	1.8	0.4	1.1	0.5	1.3	0.5	1.4	0.6	1.1	0.2
L338A/ H339A	1.9	0.6	1.1	0.6	1.1	0.4	1.2	0.5	1.1	0.3
F301Y	1.7	0.2	1.1	0.3	1.5	0.3	1.4	0.4	1.3	0.3
L295S	2.0	0.4	1.1	0.3	1.2	0.4	1.2	0.4	1.3	0.5
I271A/ R272A	2.4	0.8	1.3	0.4	1.6	0.6	1.5	0.5	1.2	0.4
E200A	2.0	0.9	1.0	0.4	1.2	0.4	1.2	0.4	1.1	0.3

Table S6.5. Average value of radius of gyration and centre of mass for MMP-1•THP and mutants.

MMP-1	Radius of Gyration (Å)	Center of Mass (Å)
WT	24.8	37.7
F289A/Y290A/P291A	25.0	37.5
R272A	25.5	38.0
L338A/H339A	25.2	37.2
F301Y	25.2	38.1
L295S	24.7	37.5
I271A/R272A	25.1	38.5
E200A	25.1	37.6
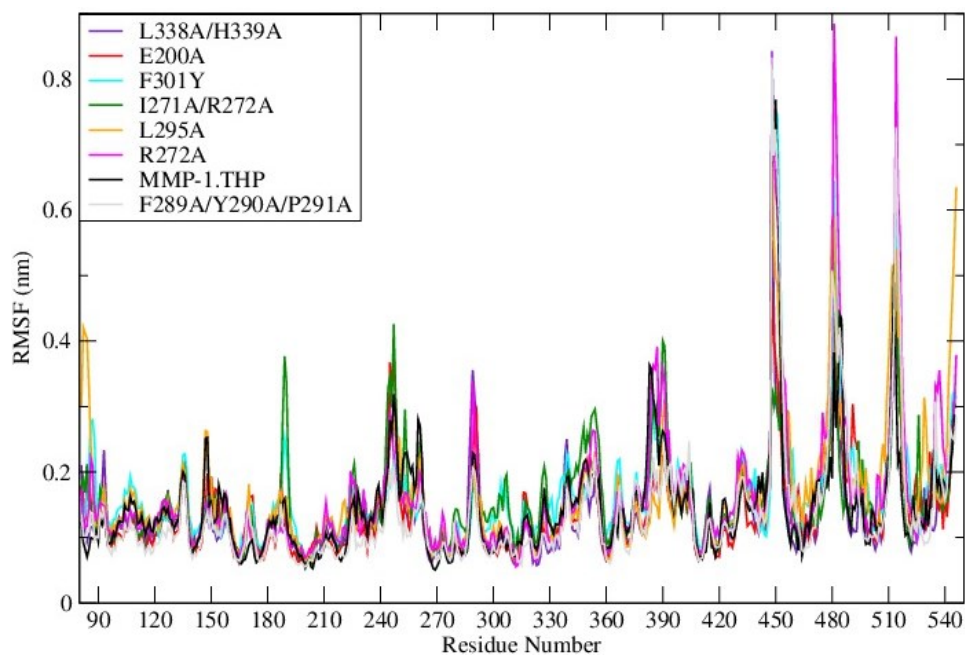
Figure S6.1. RMSF profiles of the mutants in comparison to the WT MMP-1•THP.
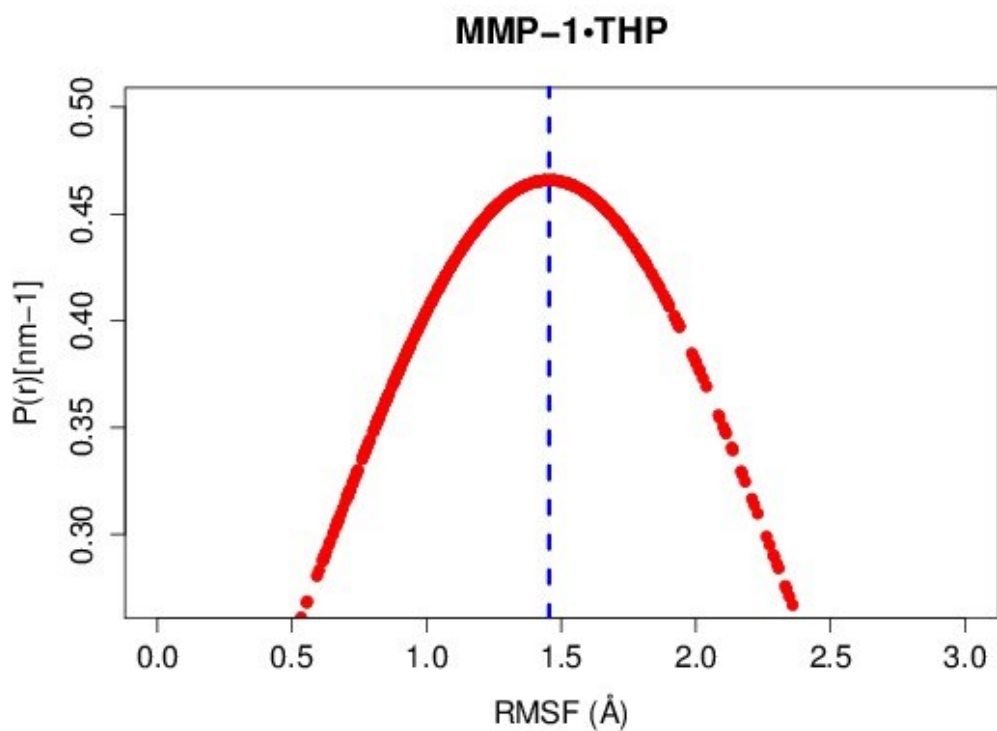
Figure S6.2. Distribution of the RMSF of MMP-1•THP. The blue line indicates the mean value of the distribution (1.4 Å).
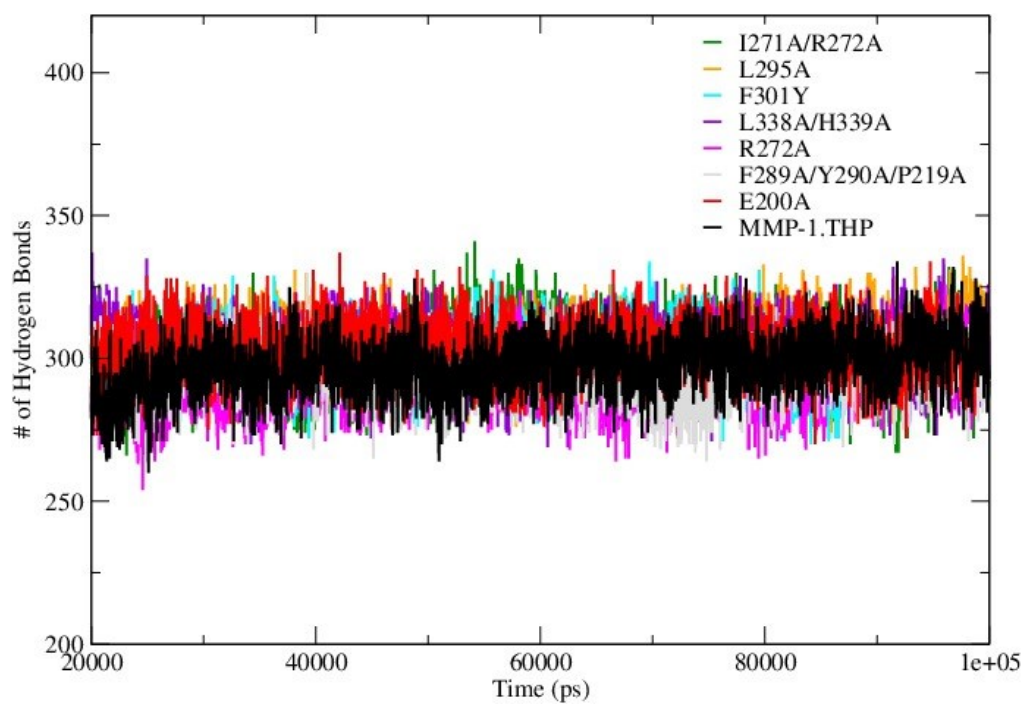
Figure S6.3. Time evolution of the hydrogen bonds in MMP-1•THP and the mutants.
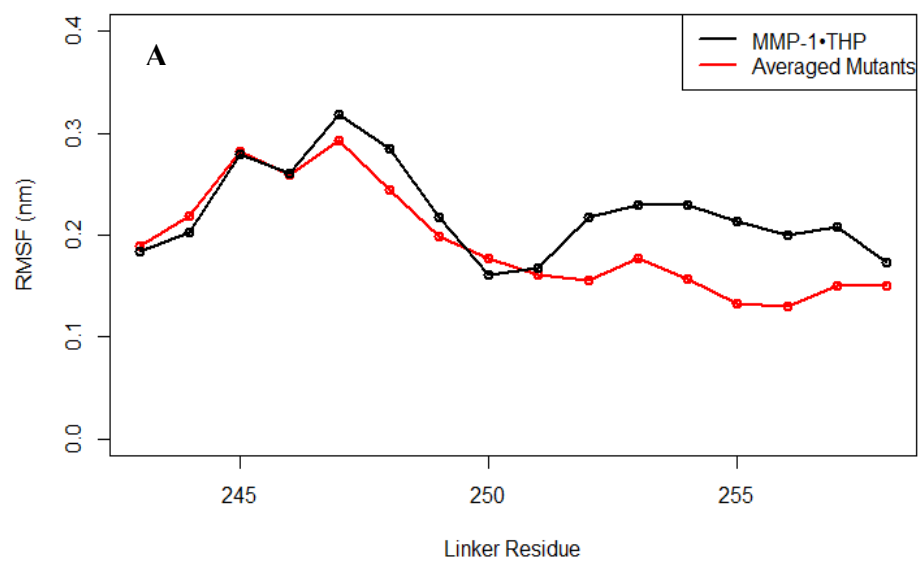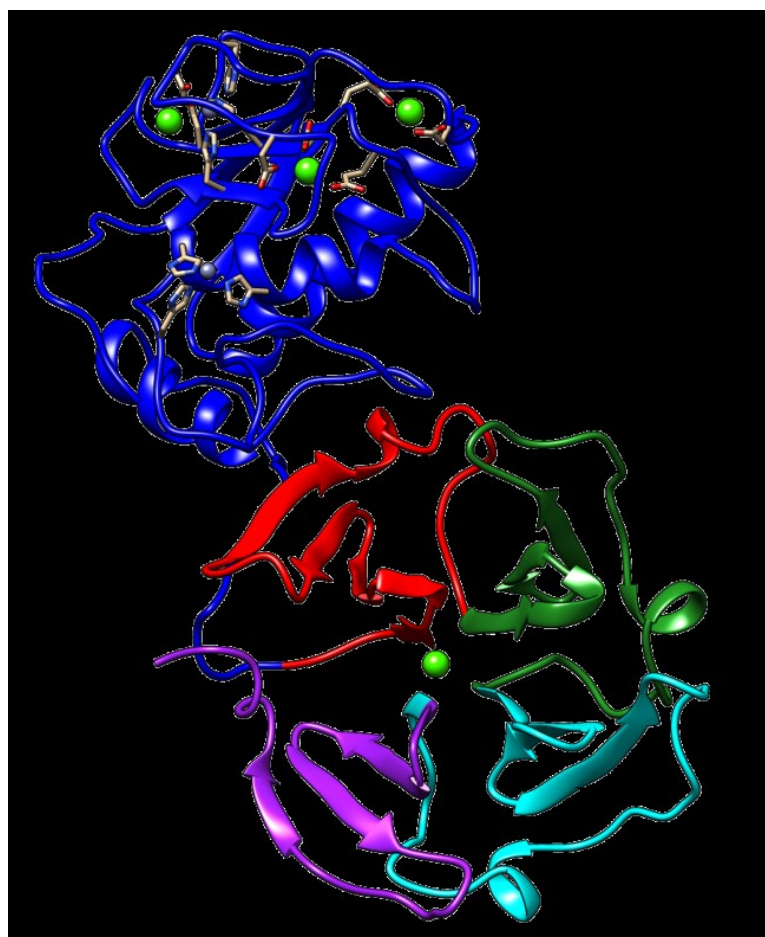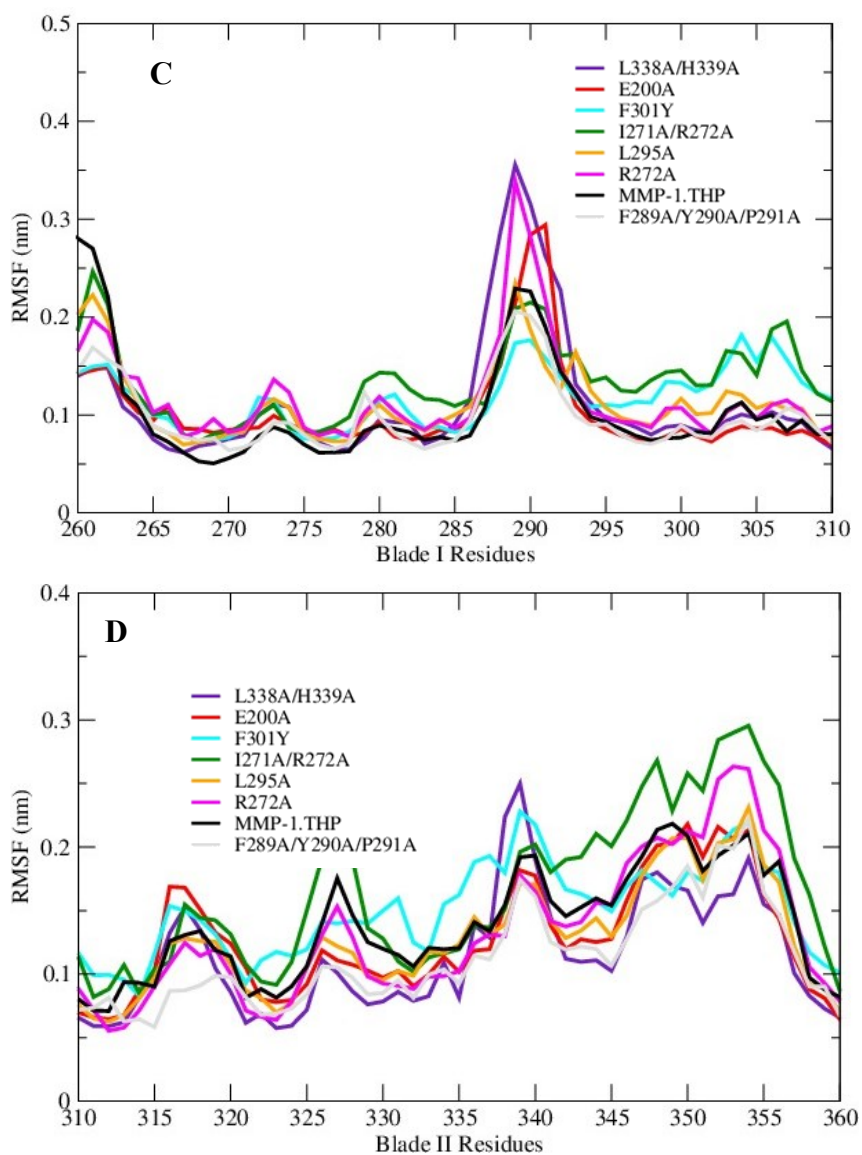
Figure S6.4. RMSF analysis of the linker, blade I and blade II: (A) Averaged RMSF analysis of the linker for all mutants. (B) The numbering of blades of the HPX domain. (C) RMSF of blade I in all mutants and the WT. (D) RMSF analysis of blade II in all mutants in comparison to MMP-1•THP.
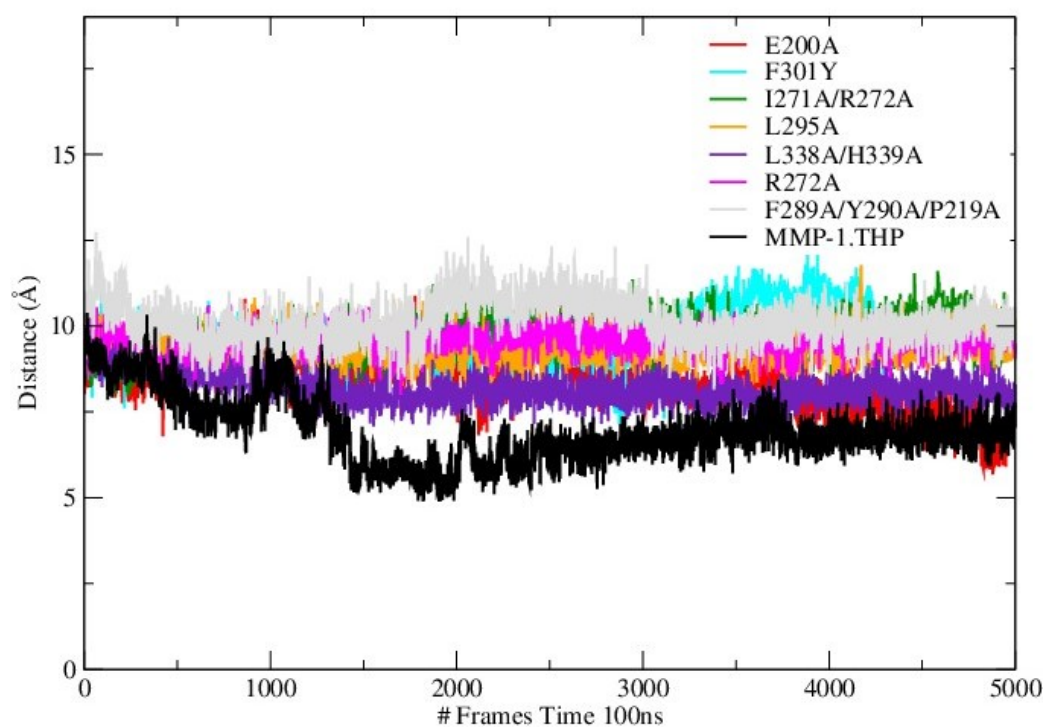
Figure S6.5. Time evolution of the distance between the scissile bond G775-L776 and Zn^{2+} in WT MMP-1•THP and the mutants.
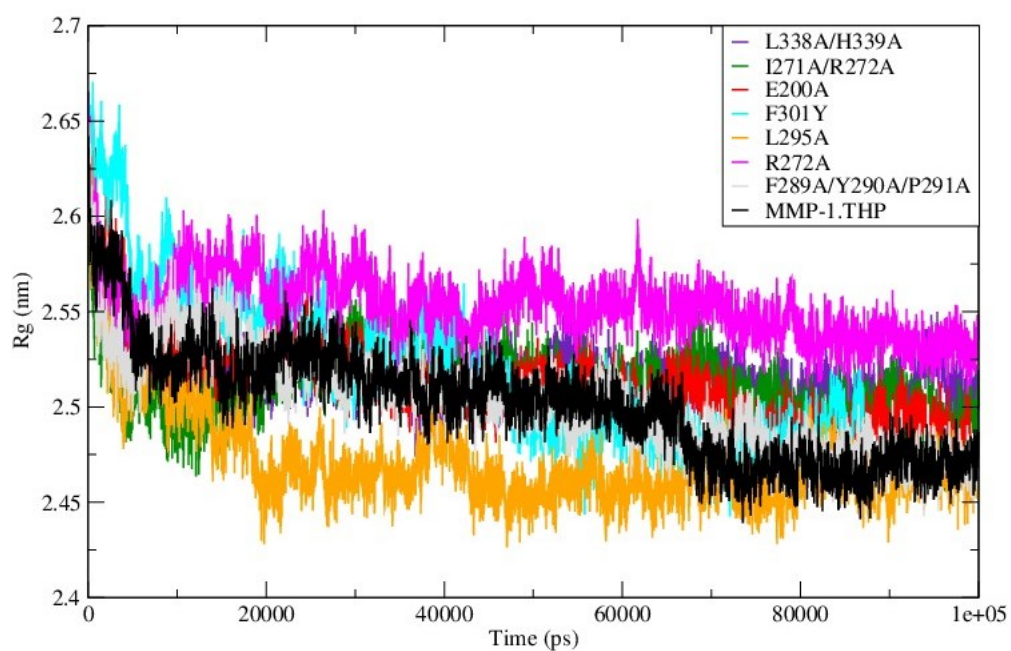
Figure S6.6. The radius of gyration of the mutants and WT MMP-1•THP.
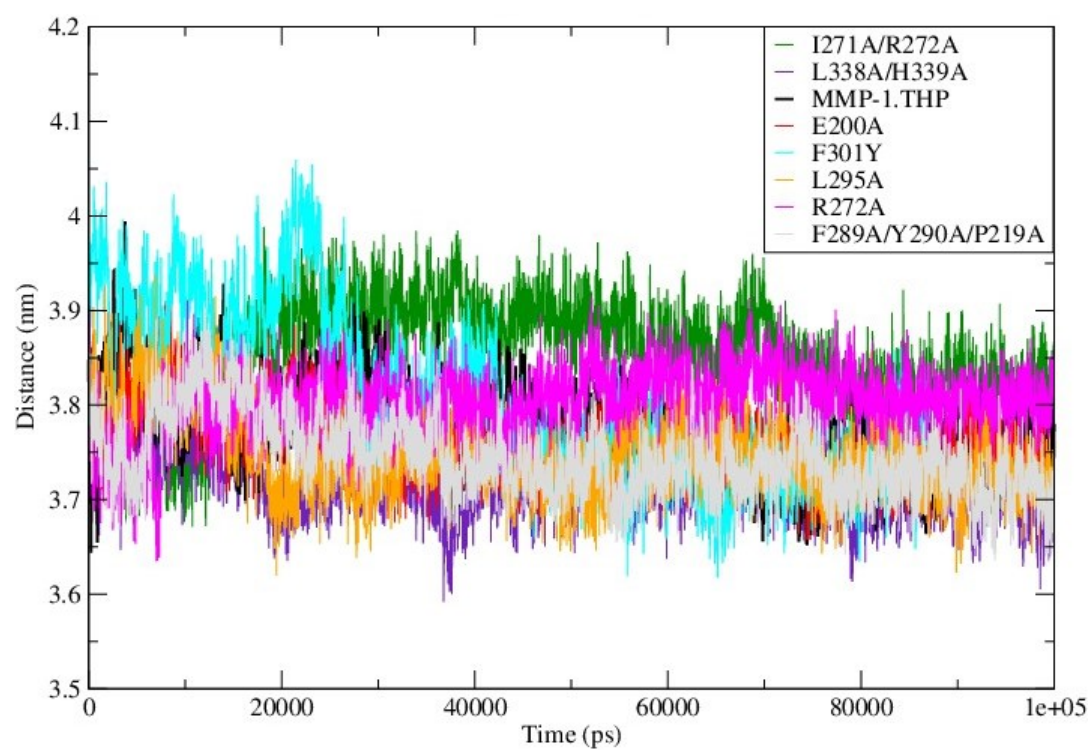
Figure S6.7. The distance between the center of mass of the CAT and HPX domains comparing the mutant forms and WT MMP-1•THP.
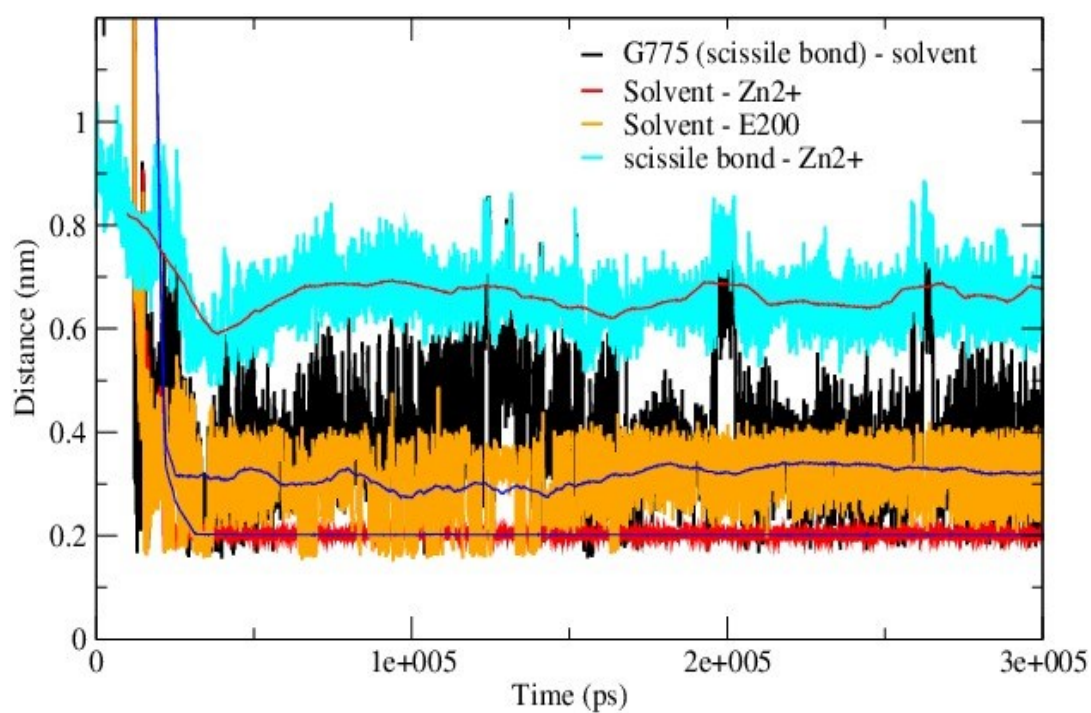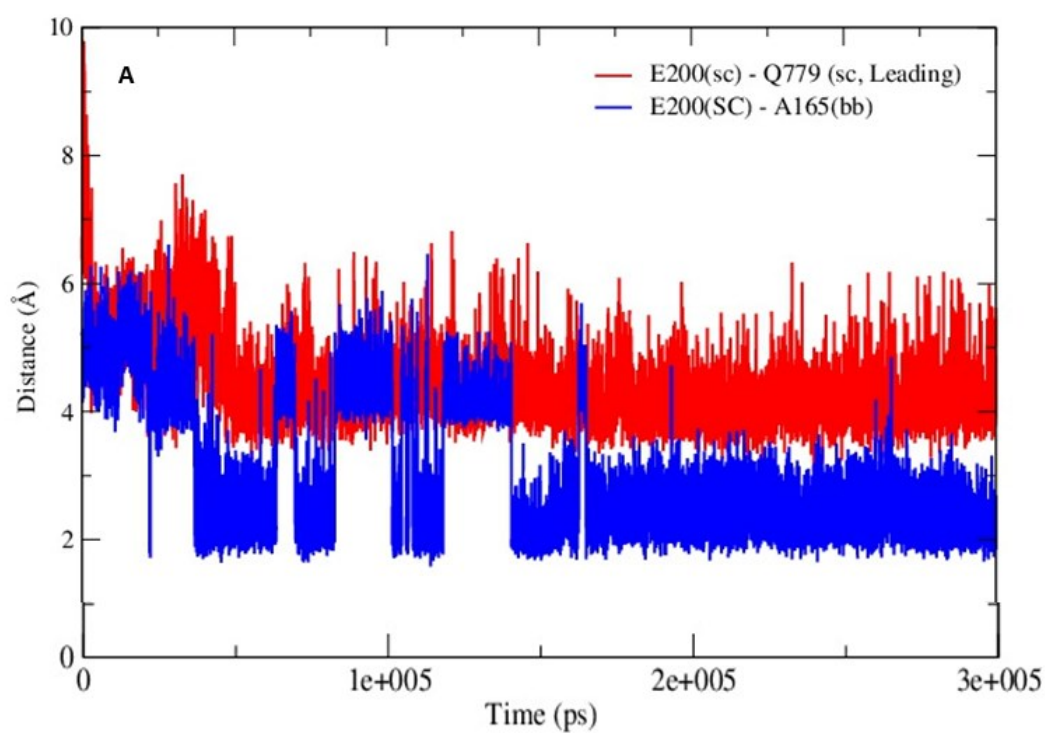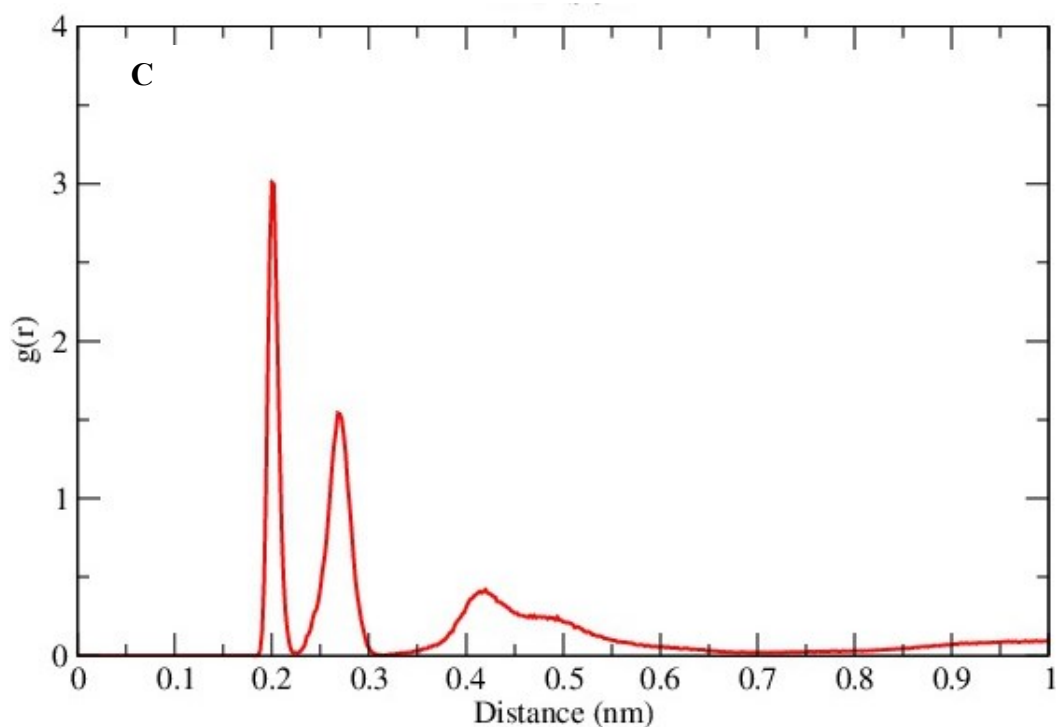
Figure S6.8. Interactions of E200 of MMP-1•THP. (A) The E200 residue interactions with the THP leading strand and forms hydrogen bonds with the backbone of A165. (B) The solvent molecule interactions with E200 and the catalytic Zn^{2+} in the active site. The distance between the centre of mass of the side chain and the backbone were used in this analysis. (c) The radial distribution functions g(r) of the catalytic Zn^{2+} showing the distribution of solvent molecules for 300ns trajectory in MMP-1•THP.

Zn(II) - SOL
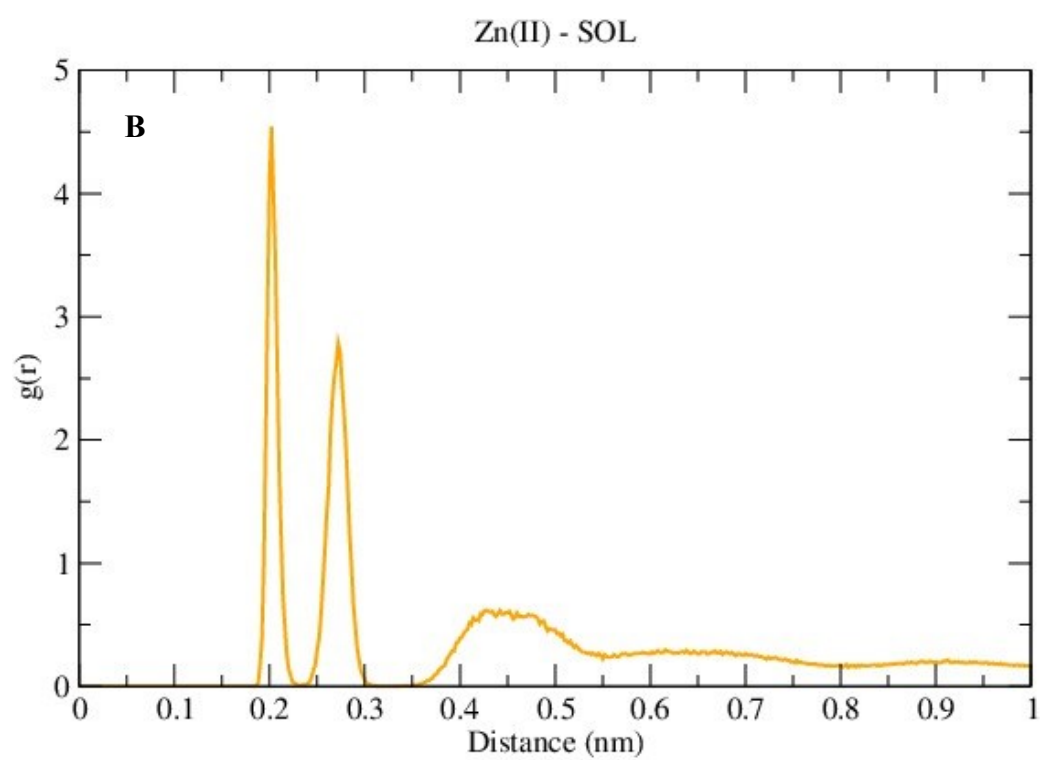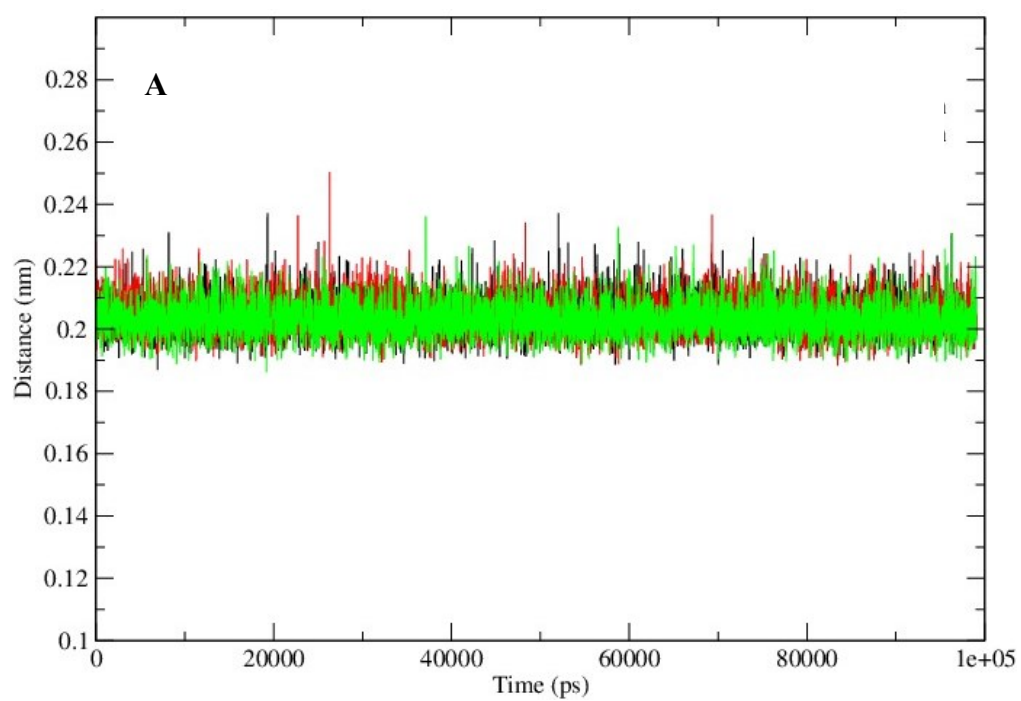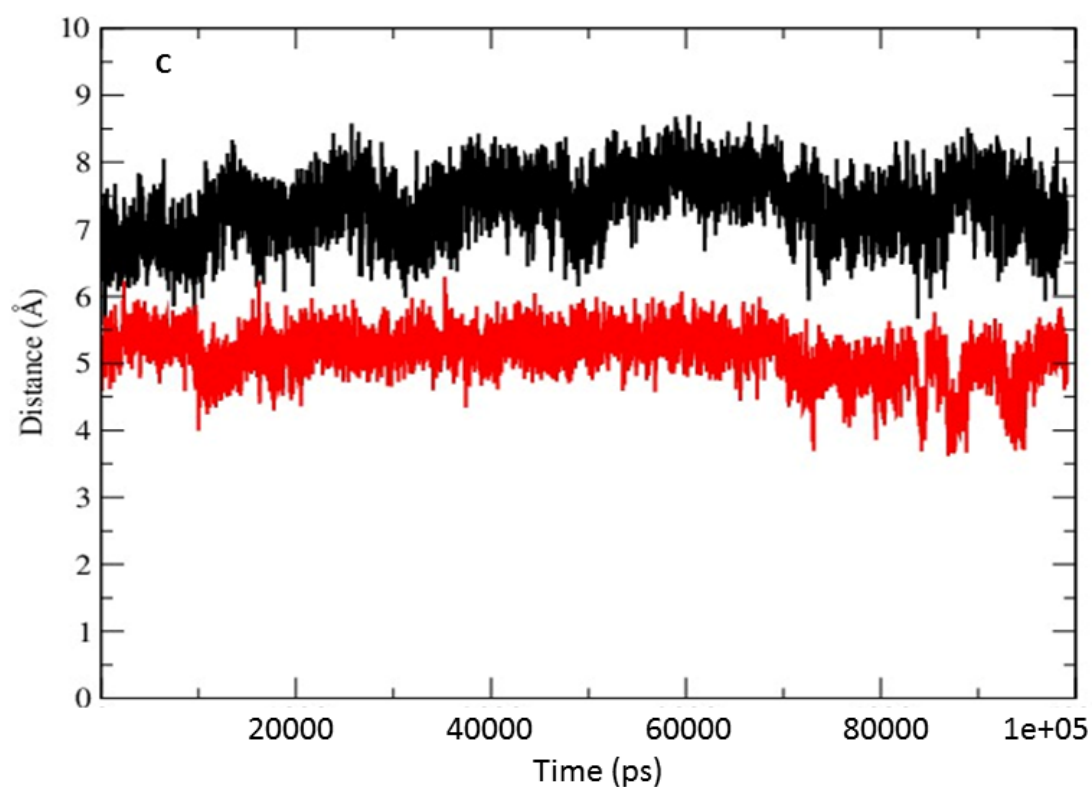
Figure S6.9. Interactions of the E200A mutant. (A) The interactions of the solvent molecules with the catalytic Zn^{2+} in E200A. (B) The radial distribution functions g(r) of the catalytic Zn^{2+} E200A mutant showing the distribution of solvent molecules for 100ns trajectory. (C) The distance between the side chain and backbone of the A200 with respect to the catalytic Zn^{2+} (black) and A165 backbone (red).
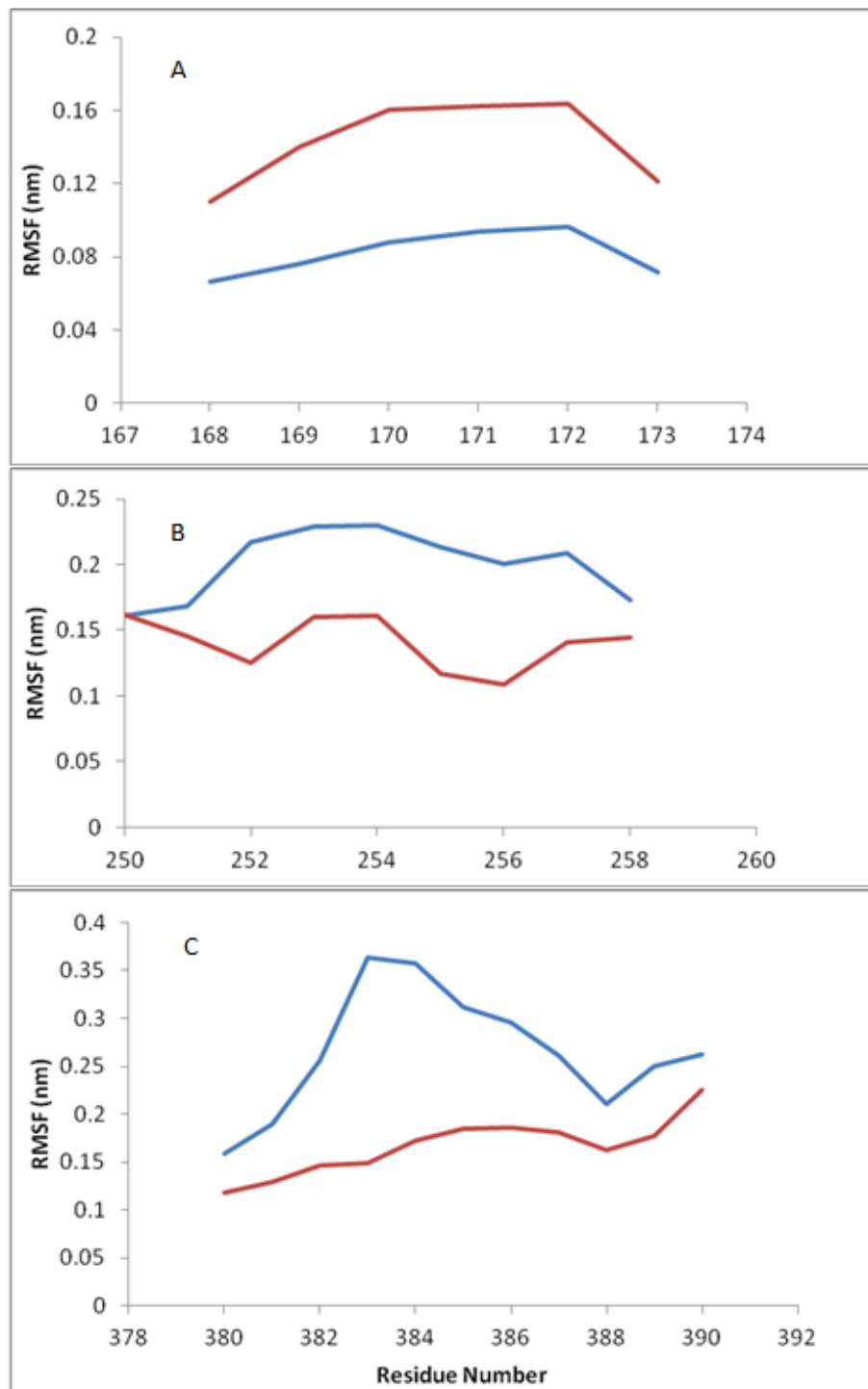
Figure S6.10. The RMSF of regions of E200A (red) in comparison to WT MMP-1•THP (blue). (A) Residues 167-174 belonging to the loop region of β4-β5 of the CAT domain. Residues on this loop coordinate to the structural Zn^{2+}, particularly D175. (B) Residues 250-260 belong to the linker region. (C) Residues spanning 378-392, part of blade III of the HPX domain and belonging to β18-β19.
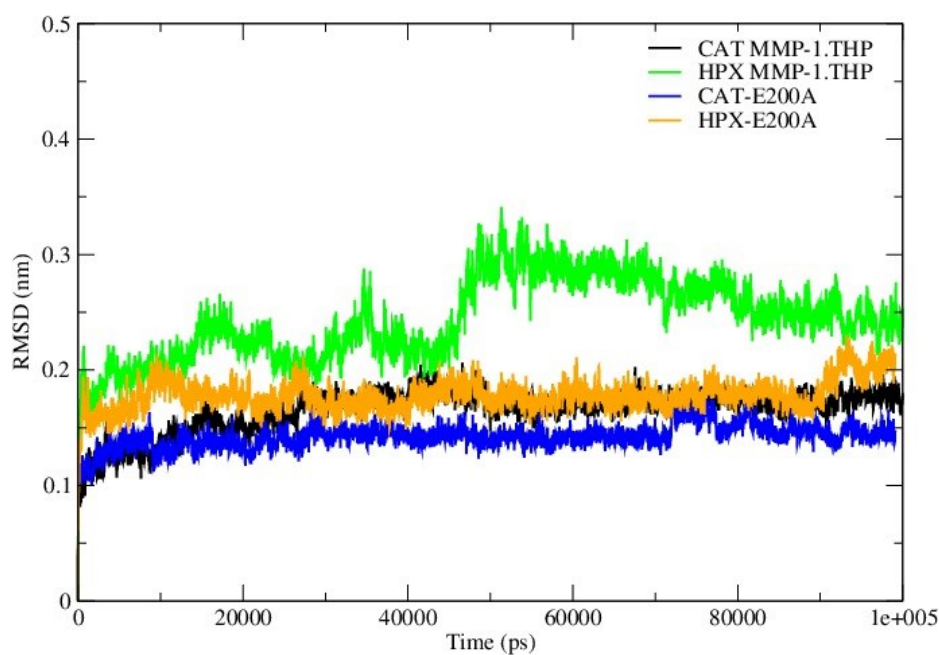
Figure S6.11. The RMSD of the CAT domain, HPX domain, and linker region of E200A

in comparison to the WT MMP-1•THP domains using Cα atoms.
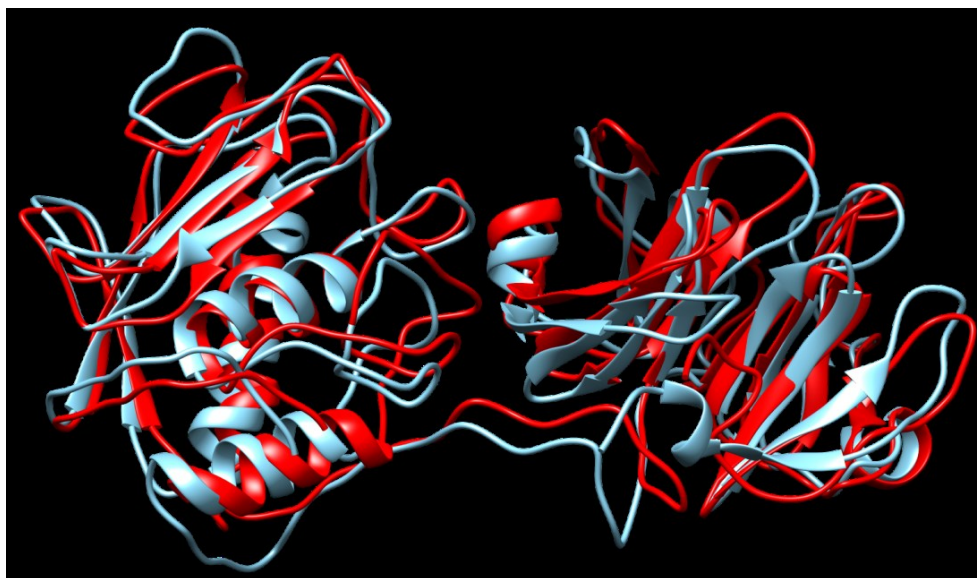
Figure S6.12. The superposition of most populated cluster of F301Y mutant (red) and WT

MMP-1•THP (cyan). There is a clear change in the conformation of blade I in the F301Y

mutant compared with WT MMP-1•THP.
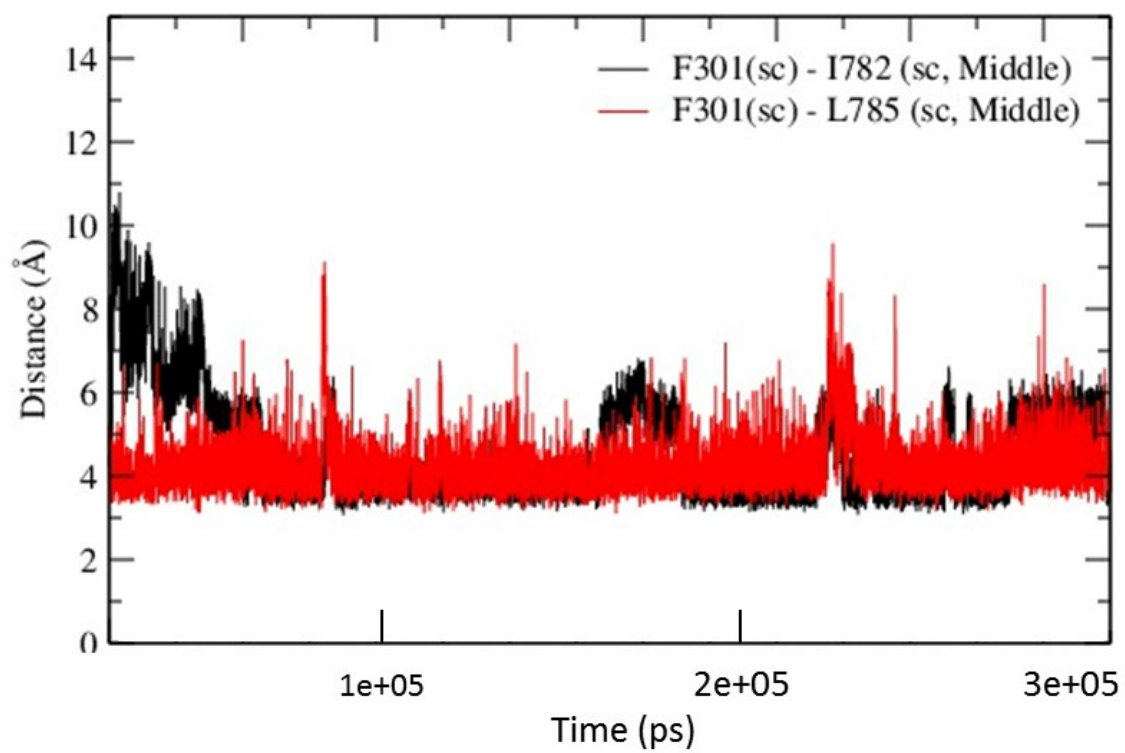
Figure S6.13. Interactions of HPX domain F301 with the THP middle strand in WT MMP-1•THP.
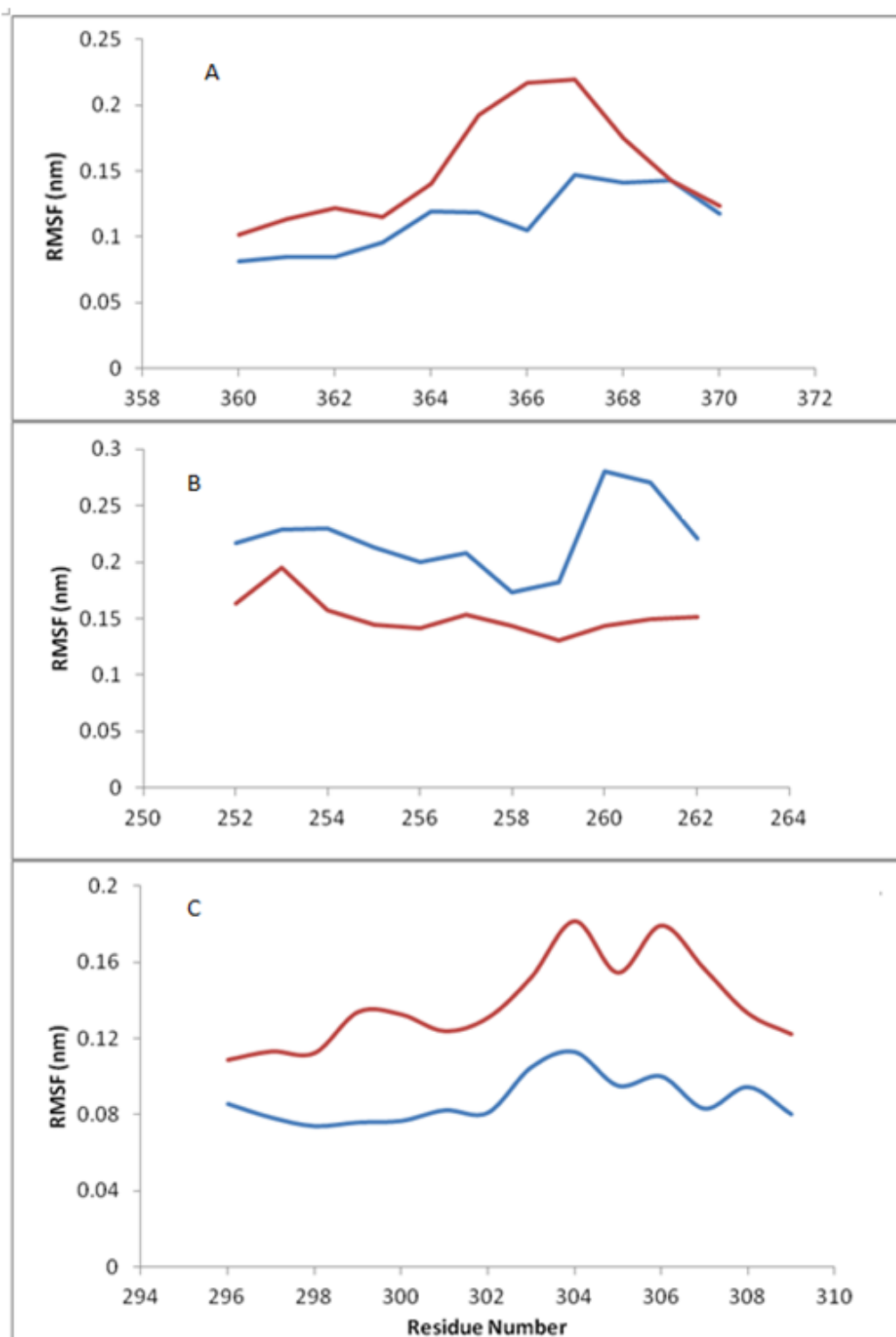
Figure S6.14. The RMSF analysis of the F301Y mutant (red) and WT MMP-1•THP (blue).
(A) The residues spanning 360-372 belong to blade III of the HPX domain (β15-β16). (B)
Residues 250-264 belong to the linker region. (C) Residues 294-310 belong from blade I of
the HPX domain.
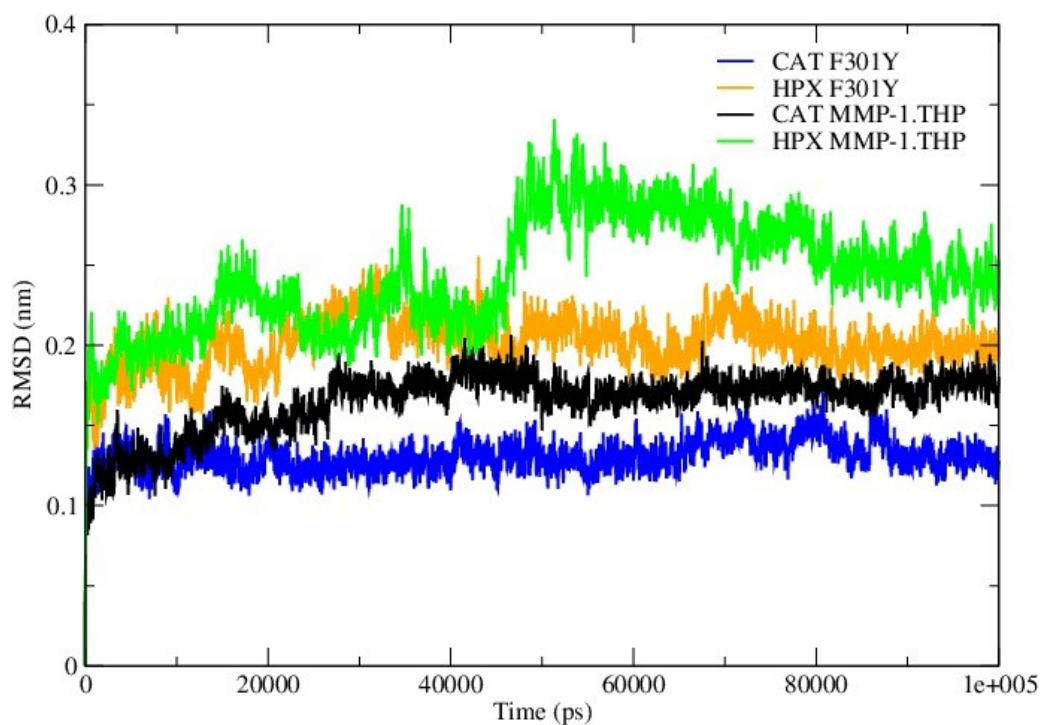
Figure S6.15. The RMSD of individual domains of the F301Y mutant in comparison to WT MMP-1•THP.
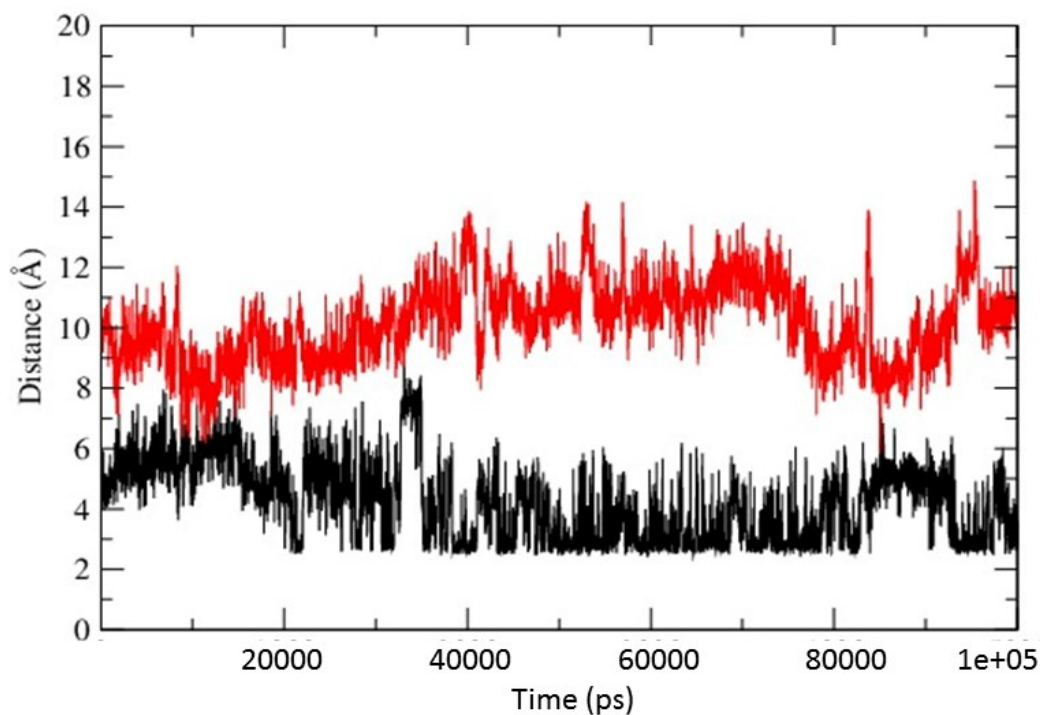
Figure S6.16. Interactions of the L295S mutant. The side chain of S295 in L295S forms interactions with the backbone of P256 of the linker (black) but not with the side chain of R780 (red) of the THP leading strand.
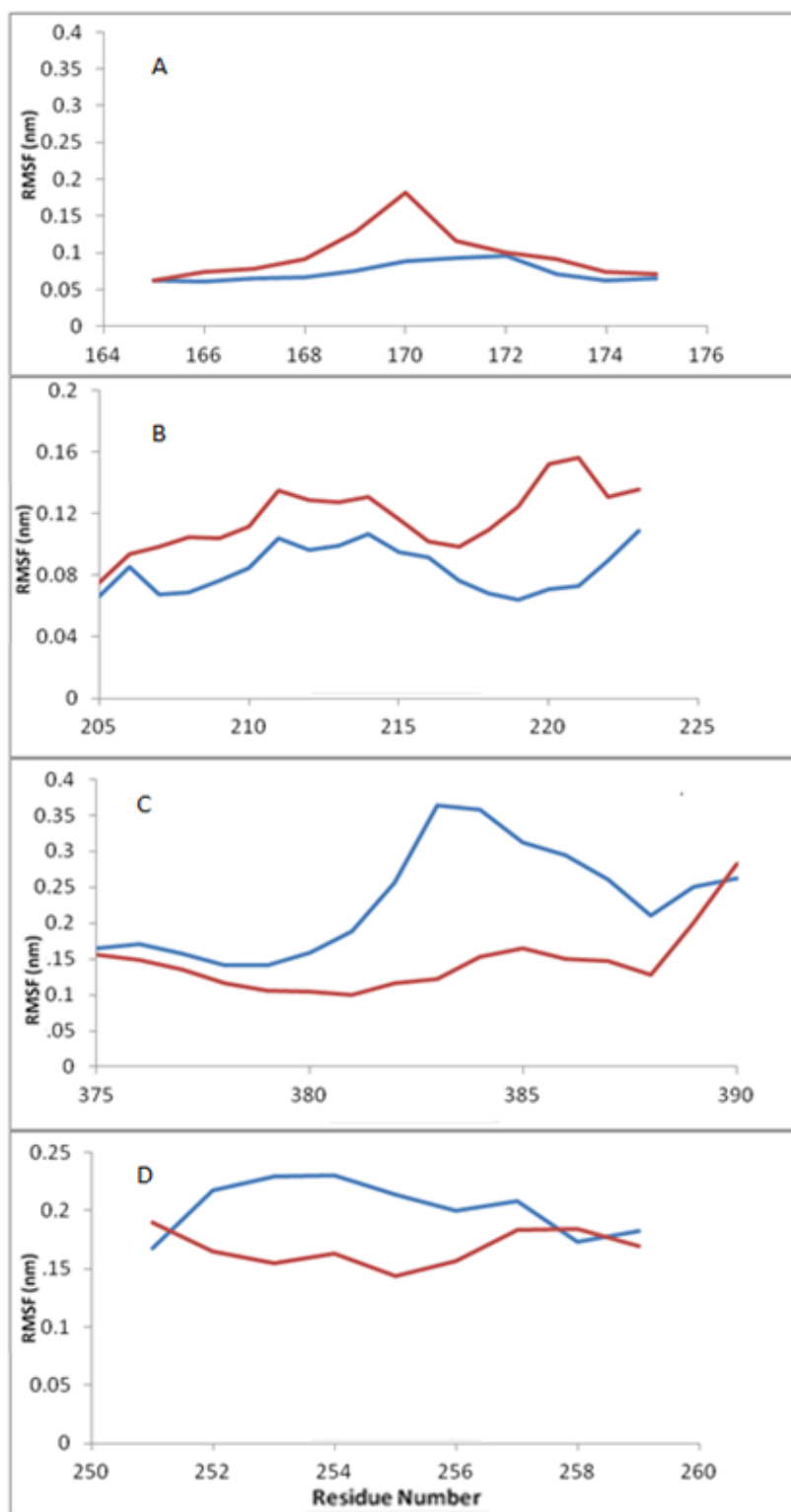
Figure S6.17. RMSF analysis of the L295S mutant (red) in comparison to WT MMP-1•THP (blue). (A) Residues 164-176 belong to β4-β5 of the CAT domain. (B) Residues 205-225 belong to loop region between α-helix B and C of the CAT domain. (C) Residues 375-390 belong to blade III of the HPX domain. (D) residues 250-260 are part of the linker region.

CHAPTER SEVEN   NON HEME 2-OXOGLUTRATE DEPENDENT ENZYMES

7.1. INTRODUCTION

There are an estimated 80 non-heme Fe(II) and 2-oxoglutrate (2OG) dependent dioxygenases found in animals and plants which catalyse an array of biochemical reactions ranging from hydroxylation, desaturation, ring formation and expansion, halogenation, demethylation etc [163-165]. The non-heme Fe(II) and 2-oxoglutrate (2OG) dependent dioxygenases bind molecular oxygen to the Fe (II) centre and results in the production of high-valent iron (IV)-oxo reactive intermediates followed by the oxidation of the carbon hydrogen (C-H) bond [164, 166]. The detailed description of reaction mechanics is shown in **Figure 7.1** [20, 167-169] [170]. The non heme Fe (II) is typically coordinated in a mono-dentate fashion by His-His-Asp catalytic triad while the 2OG binds to the metal centre in a bidentate fashion [168] (**Figure 7.2**). The sixth coordination site is normally coordinated by a water molecule in the crystal structure, however this is displaced by molecular oxygen to bind after the sequential binding by the substrate and cosubstrate (2OG) activating the enzyme to pursue catalysis [167, 168].

7.1.1 2OG dependent dioxygenases that act on Histone proteins

The post translational modifications such as methylation, acetylation, and phosphorylation etc. of histone proteins are the essential instruments of chromatin remodelling which controls important physiological process in the eukaryotes [171]. The control of the methylation status of the lysine residues of histone proteins by histone demethylases enzymes plays an important in stabilizing the structure and function of chromatin [172]. Overexpression or mutations in the histone lysine demethylases (KDM) have been linked to various forms of cancer and have become an attractive drug target to the pharmaceutical industry. There are two families of histone lysine demethylases which either depend on FAD (Flavin Adenine Dinucleotide) or 2OG (2-oxoglutrate) to remove the methylation

mask from the histone proteins [166]. The 2OG dependent histone lysine demethylases specifically act on the N^{ε}-methylated lysine residues of the histone proteins and demethylates the methylated lysine residues, after hydroxylation and release of the hemiaminal intermediate the CH_2OH group decomposes to release formaldehyde leaving the demethylated lysine [173].

PHF8 and KIAA1718 (**Figure 7.2A, B**) (also known as JHDM1D) enzymes are Fe(II)-and 2OG-dependent demethylases belong to the Jumonji C (JmjC) proteins which in turn are part of the cupin superfamily of metalloenzymes [166]. The crystal structure of PHF8 and KIAA1718 reveals the presence of a conserved β-barrel fold (cupin or jelly-roll core fold) in the JmjC domain of both enzymes which is conserved throughout the cupin superfamily [174]. The mutation of F279S in the catalytic JmjC domain of PHF8 results in X-linked mental retardation and also been found in individuals with autism spectrum disorders [175]. The knockout of PHF8 and KIAA1718 gene in zebrafish results in brain defects [176]. Both these enzymes contain a plant homeodomain (PHD), a binding domain and the JmjC catalytic domain [166]. The PHD domain binds to the Lys4-trimethylated histone 3 (H3K4me3) and the JmjC domain demethylase H3K9me2 or H3K27me2 (di-methylated histone residue). The presence of the H3K4me3 on the same histone peptide enhances the catalytic activity of PHF8 JmjC domain on H3K9me2 by 12-fold, however the presence of the H3K4me3 on the same histone peptide diminished the enzyme activity of KIAA1718 JmjC domain on H3K9me2 [174]. The differences in the sequence and conformation of the linker region was proposed to explain the difference in substrate specificity for both enzymes [174]. In particular the linker region of the PHF8 enzyme exhibits a bent conformation allowing both the PHD domain and the JmjC domain to interact with their substrates. In KIAA1718, the linker adopts an extended conformation that prevents the interaction of the H3K9me2 with the JmjC domain when its PHD domain interacts with the H3K4me3 [174]. Similar functionality of the linker region was proposed in MMP-1

(Matrix metalloproteinases 1) enzyme [131, 135, 137] and has been tested by us computationally [19].
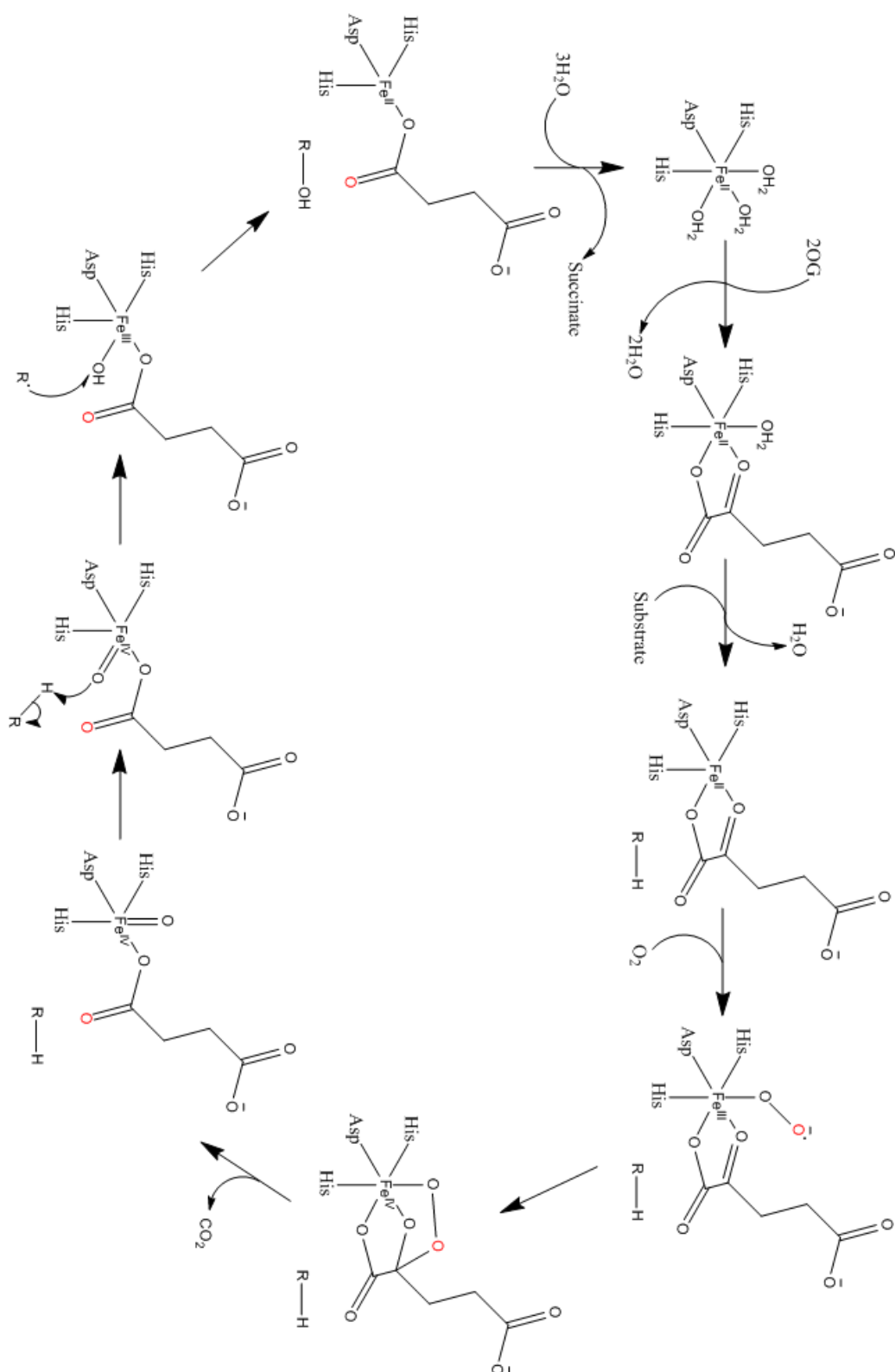
Figure 7.1 The reaction mechanism of non heme 2OG dependent dioxygenases [20, 167-170, 177].

E

F
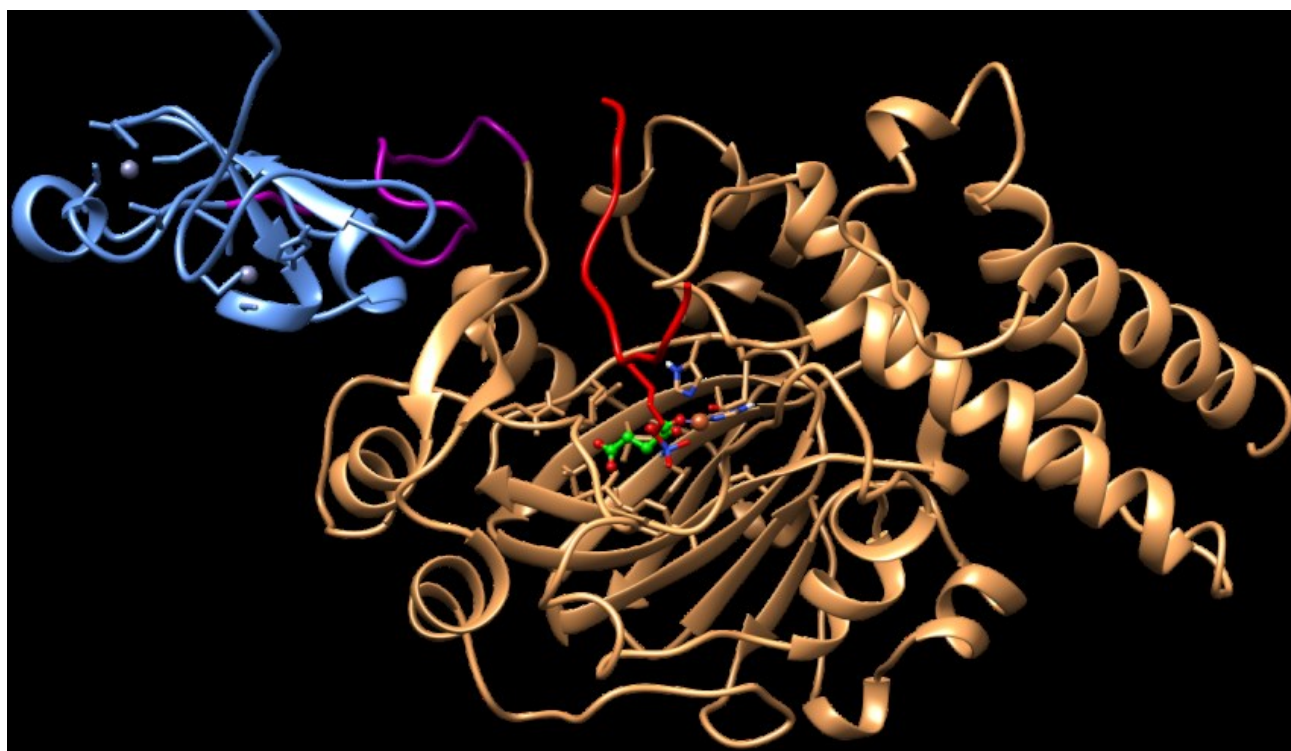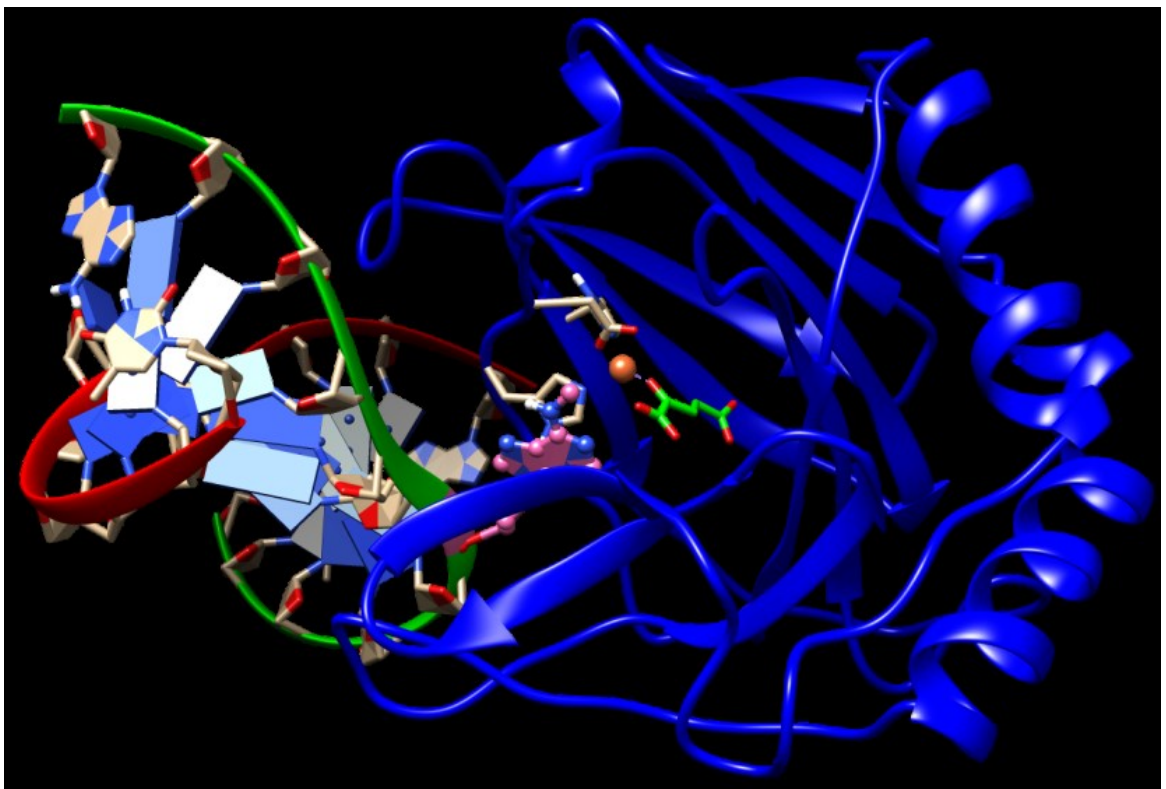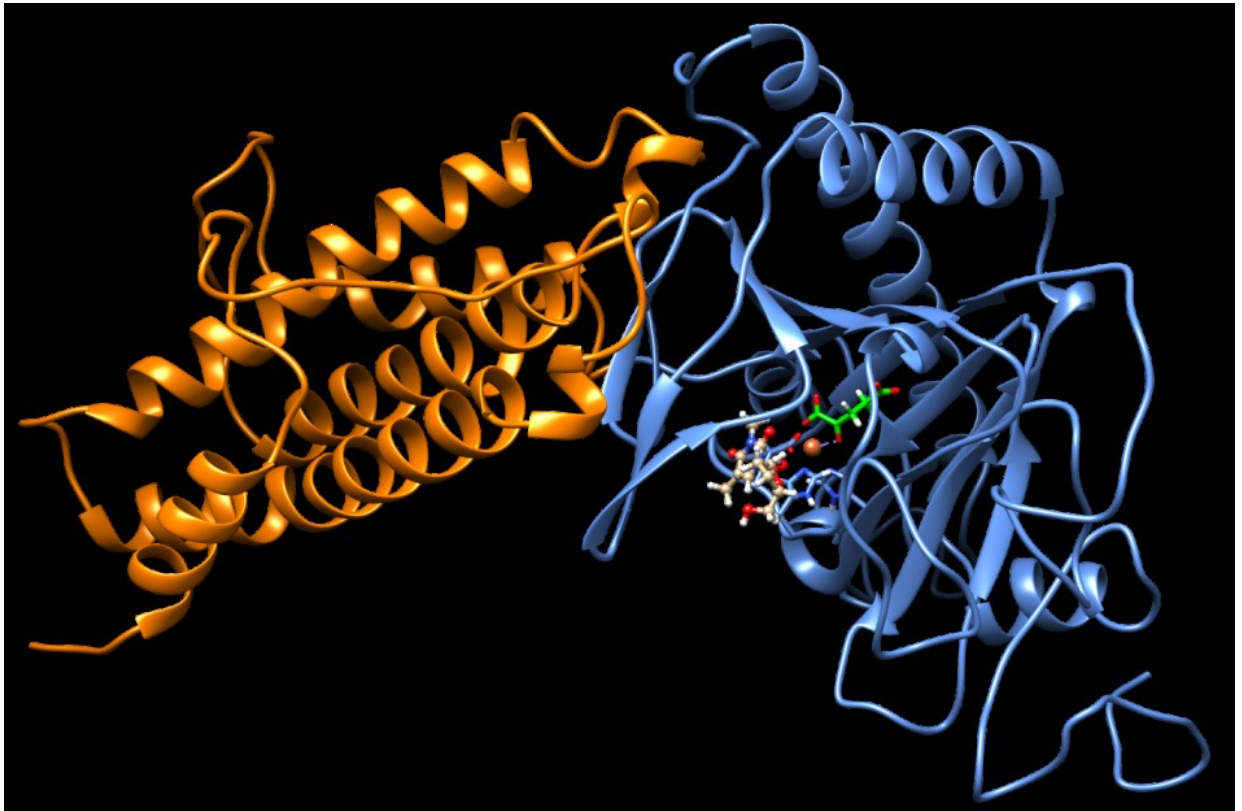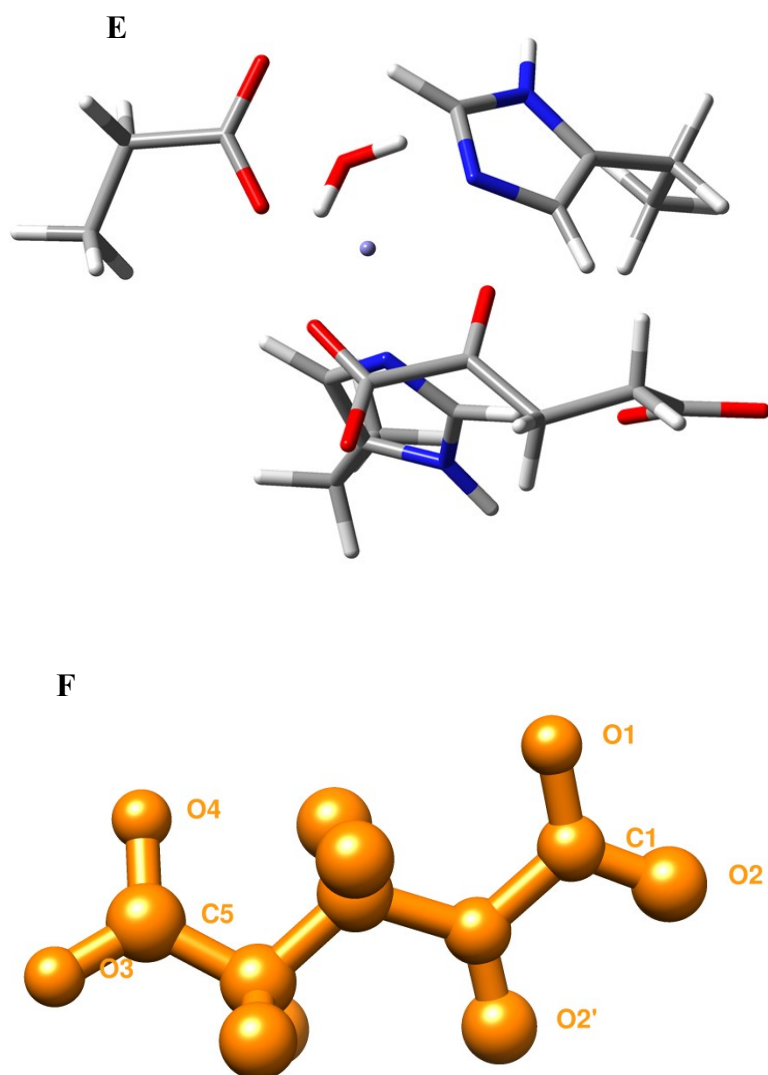
Figure 7.2 Energy minimized X-ray crystal structure of PHF8[174], KIAA1718[174], FTO[178, 179] and AlkB[180] enzymes using UCSF Chimera[118]. (A) PHF8 enzymes consist of a PHD domain and Jumonji domain coloured in cyan and kaki colours respectively. The histone peptide is represented in red colour, 2OG is shown in green liquorice representation and non heme Fe(II) in orange sphere representation. The linker region connecting the PHD and Jumonji domain is coloured in purple representation. (B) KIAA1718 also consists of a PHD and Jumonji domain represented in light blue and

orange colours. The linker region and the histone peptide are shown in purple and red colours respectively. (C) FTO enzyme consist of N terminal and C terminal domains shown here in light blue and orange colours. The 3DT substrate is show in ball and stick representation near the non heme active site. (D) AlkB enzyme complex with double stranded DNA, where the methylated base 6MA is shown in ball and stick representation in pink colour. (E) The geometry optimized active site of FTO non heme iron containing enzymes using Gaussion09 [12]. No negative frequencies were found in the structure indicating local minima on the potential energy landscape. (F) The 2OG atom numbering used in this study.

7.1.2 2OG dependent dioxygenases that act on DNA or RNA

The methylated DNA or RNA bases due to the endogenous and exogenous alkylating agents can be a source of cellular toxicity and mutagenesis and thus require urgent removal for the regulation of the normal metabolism of cellular environment [181]. AlKB and FTO (fat mass and obesity-associated enzymes) (**Figure 7.2C, D**) enzymes belong to superfamily of Fe(II)-and 2OG-dependent dioxygenases which carry out direct dealkylation to remove the methyl group from the base without breaking the sugar phosphate backbone. The crystal structure of AlkB and FTO reveals the presence of a conserved β-barrel fold; however the FTO contains an extra loop in the substrate entry site which inhibits the binding of the double stranded DNA molecule near the active site. Both the AlkB and FTO enzymes have unique ability to recognize different methylated DNA/RNA bases which forms the basis of their unique biological role in cellular environment [163-165, 182]. The crystal structure of human PHF8 [174], KIAA1718 [174], FTO[178] [179] and AlkB [174] reveals the important atomistic details of the enzyme and the ligand binding and possible mechanism, but shows no information about how conformational flexibility and dynamics influence the protein structure and its impact

on the electronic structure of the active site. In order to study the conformational dynamics and its impact on electronic structure we performed long range MD, QM and QM/MM studies.

7.2 METHODS AND PARAMETRIZATION

7.2.1 System Preparation

Histone Lysine Demethylases PHF8 and KIAA1718: The crystal structure of the PHF8 (PDB: 3KV4) [174] in complex with substrate histone (H3) and cofactor analogue N-oxalylglycine was used an initial structure. The missing linker region (residue 65-79) between the PHD and the Jumonji domain was modelled using Modeller software [139]. The crystal structure of KIAA1718 (PDB: 3KV6 chain D used here) complexed with 2-oxoglutarate and oxygen molecule was used as an initial structural for the further structural preparation for this computational study. The oxygen molecule was removed from the crystal structure. The substrate histone H3 was modelled into the active site in the Jumonji domain by superimposing the PHF8 crystal structure with KIAA11718 using Maestro (Schrodinger LLC, New York). The H3K9me2 of histone was placed in vicinity of the metal centre however the H3K4me3 ligand made no interactions with the PHD domain due to extended nature of the PHD in KIAA1718 in comparison to PHF8. In order to solve this issue we modelled the parts of the H3K9me2 and H3K4me2 of histone H3 separately into the structure of KIAA1718 using UCSF Chimera [118]. To understand how the linkers are crucial for the enzyme activity of both PHF8 and KIAA1718 we modelled in the liker of PHF8 into KIAA1718 and vice versa. The modelling was performed using Modeller [139]. The linker of PHF8 consists of residues 65-79 and KIAA1718 consists of residues 97-114. The hybrid structures obtained were then subjected to molecular dynamics simulations.

DNA base Demethylases FTO and AlkB: The crystal structure of FTO (PDB: 4IDZ [178, 179]) was used as the starting structure for building the final structure to be used in this

computation study. The missing residues were added using modeller tool. The final structure of FTO was obtained by modelling the substrate 3-methylthymidine (3-meT) from 3LFM [178] to 4IDZ [179] crystal structure by aligning two structures using Maestro (Schrodinger LLC, New York). This was followed by replace of the Nickle (Ni^{2+}) with the Fe(II) and N-Oxalylglycine with 2-oxoglutarate using GaussView 5.0 [87]. The crystal structure of the AlKB enzyme (PDB: 4NID[180]) in complex with 2-oxoglutarate and double stranded DNA containing N6-methyladenine (m6A) was used in this computational study. The Mn^{2+} was replaced with Fe(II) using GaussView 5.0 [87] in AlkB to make it a productive complex and was used for this computation study. The AlkB enzyme was also modelled with the base only (6MA) instead of double stranded DNA to make comparison with the FTO single base simulation and to see how the DNA effect the conformational dynamics of the AlkB enzyme. The amber parameters for the N6-methyladenine were developed by using the CIF file (PDB: 6MA) using AM1-BCC charge model available in Antechamber and prepgen from AmberTools15 [7]. The wild type structure of AlkB (PDB:3BIE [183]) with substrate bases N1-Methyladenine was also used in this simulation studies. The parameters for the ligand were developed using antechamber [7].

The SwissPDBViewer was used for adding missing atoms and selecting one from the alternative side chain orientations [86]. The protonation states of ionisable sidechains of the proteins were assessed using H++ server [184]. The Histidine residues coordinating with the metal centre were assigned protonation states based upon the visual inspection of their local environment. The cofactor analogue N-Oxalylglycine was modelled to 2-oxoglutarate by replacing the nitrogen with carbon atom using GaussView 5.0 [87] and was used as a cofactor in the computational study. The hydrogen atoms were added to 2-oxoglutarate and 3-methylthymidine by using the reduce programme [185]. The amber parameters for the 2-oxoglutarate and 3-methylthymidine were developed using the general Amber force field (GAFF) [60] using Antechamber. The atomic charges of the cofactors

were calculated based on the electrostatic potential from single point HF/6-31G* calculations using Gaussian09[12]. The restrained electrostatic potential (RESP) [186] method was used for charge fitting procedure. The substrate histone H3 consisted of di and tri methylated lysine residues denoted as H3K9me2 and H3K4me3 respectively in the crystal structure [174]. The amber parameters for the Di-methylated lysine (M2L) and Tri-methylated lysine (M3L) used in this study were obtained from [187].

7.2.2 Metal Centre Parameter Builder

The amber parameters for the active site containing Iron (Fe(II) high spin S=2, M=5, ground state [20, 67, 167, 168, 188, 189] and the coordinating ligands (2-oxoglutarate (bidentate), Histidine, Aspartic acid and water as monodentate) were prepared using the Metal Centre Parameter Builder (MCPB) using MCPB.py [190]. The metal centre parameters were derived based upon the bonded and electrostatic model approach in which the coordinating ligands are connected to metal through covalent bonds. The bond and the angle force constants were derived using the Seminario Method and the point charge parameters for the electrostatic potential were obtained using the ChgModB method (details in S.I with all the parameters and the charges). The Pabis et al have applied the MCPB tools in the past for the description of the mononuclear non heme iron centre and iron-sulfur Rieske cluster. The molecular dynamics simulations run using these parameters have reproduced the geometry of metal-ligand complex successfully [191]. The description of the Zinc ion and its coordinating ligands in the PHD domain were described using Zinc AMBER force field (ZAFF) [192].

7.2.3 MD Simulation

The productive molecular dynamics simulations were performed using GPU version of PMEMD [193] engine integrated with Amber 15[7]. The FF14SB[7] force field was in all the simulation performed and Leap module was used to add missing hydrogen atoms and

counter ions for neutralisation of the protein system. All the systems were immersed into a truncated octahedral box with TIP3P [194] water molecules such that no protein atom was within 10 Å of any box edge. The periodic boundary conditions were employed in all the simulations. Long-range electrostatic interactions have been calculated using the particle mesh Ewald (PME) method [95] with a direct space and vdW cut-off of 8 Å. The various systems were subjected to energy minimization using first steepest descent (5000 steps) followed by conjugate gradient (5000 steps). The solute molecules were restrained using restrained potential of 100 kcal mol^{-1} Å^2 and only solvent and ions were allowed to minimize. This was followed by full minimization of entire system with both steepest descent (5000 steps) and conjugate gradient (5000 steps) to relax the system prior to productive simulation. All the energy minimization, heating and equilibration were performed with CPU version of PMEMED. The systems were them subjected to controlled heating from 0 to 300K at constant volume using Langevin thermostat with a collision frequency of 1 ps^{-1} using a canonical ensemble (NVT) [195] MD simulation for 400 ps. The solute molecules were restrained using harmonic potential of 10 kcal mol^{-1} Å^2 during the heating process. The SHAKE algorithm [196] was used to constrain bonds involving hydrogen. This was followed by equilibration at 300K in an NPT ensemble [94] for 1 ns without restrains on the solute molecules and the Berendsen barostat was used to maintain the pressure at 1 bar [93]. The SHAKE algorithm was used to constrain the bonds of all hydrogen atoms, a time step of 2 fs and the SPFP precision model were used for all MD runs. A production MD run with explicit solvent for continuous 1μs was performed in a NPT ensemble with a target pressure of 1 bar and a pressure coupling constant of 2 ps. The frames from the productive run were save every 10 ps. The trajectories were analysed using CPPTRAJ[197], VMD[97], UCSF Chimera[118] and R (Bio3D[98]). The Root Mean Square Deviation (RMSD) of C α atoms of the protein with respect to minimized crystal structure, Root Mean Square Fluctuations (RMSF), Radius of gyration (Rg),

Electrostatic interactions, hydrogen bonding, Solvent Accessible Surface Area (SASA), and cluster analysis were performed. The Bio3D package [98] in R was used to produce PCA and domain cross correlation.

7.2.4 QM calculations

The snapshots of structure were obtained from the minimized crystal structures of the various systems. The GaussView 5.0 [87] was used to set up QM calculations and Gaussion09[12] code was used to run all QM calculations. In all the calculations, Iron (Fe(II) high spin S=2, M=5, ground state [20, 67, 167, 168, 188, 189] and the coordinating ligands (2-oxoglutarate (bidentate), Histidine, Aspartic acid and water as monodentate) was used. The Histidine and Aspartic acid residues were truncated at the Cβ position and hydrogen atom were added to saturate the bonds. Geometry optimization, frequency calculations and single point calculations were performed with Density Functional Theory (DFT) using unrestricted UBP86 functional [198] with 10 % exact HF (Hartree Fork) exchange [199] with 6-311G* basis set of the Fe and its coordinating atoms (Oxygen and Nitrogen) from the ligand and rest of the atoms we employed 6-31G*. The stationary points obtained were checked using frequency calculations and are available in S.I. The calculations were also performed using unrestricted B3LYP functional where Fe was treated with LANL2DZ [200] basis set and effective core potential (ECP) and rest of the atoms with 6-31G* basis set. A conductor-like polarizable continuum model (CPCM) with ε=4.3 (diethyl ether as solvent) was used in the QM calculations to mimic the hydrophobic active site in the protein [201]. The orbital analysis and the spin densities were obtained from the Gaussion09[12] and Multiwfn [202] .

7.2.5 QM/MM calculations

The snapshots for the QM/MM calculations were obtained randomly in no particular order from the MD trajectory performed on all the systems. The residues of all the enzymes including the water molecules which are within 35 Å of Iron (II) (expect for AlkB where

whole protein and DNA were used and water up to 35 Å) were involved in the QM/MM optimization. These snapshots were first subjected to energy minimization for 10,000 steps by using both steepest descend (5000) and conjugate gradient (5000) algorithms using amber14 [7]. The active site residues were restrained with restrained potential of 100 kcal mol^{-1} Å^2 in the energy minimization in order to maintain the geometry of the active site. The energy minimized snapshots of all the enzymes were prepared using the Schlegel's toolkit TAO for ONIOM calculation in Gaussion09 [12, 203]. The residues which are within 20Å of Iron (II) including water molecules were allowed to move freely during geometry optimization and rest of the system was frozen during geometry optimization in ONIOM. The QM/MM system (**Figure 7.3** QM/MM of PHF8) was prepared using the GaussView 5.0 [87] and all the calculations were run using Gaussion09 code [12]. The Amino acids were assigned with the standard bonded and non-bonded terms available from the ff99SB force field [204] available in Gaussion09 [12]. The mechanical embedding scheme was used in the geometry optimization however we have also used electronic embedding scheme for the few snapshots. The non-bonded van der Walls parameters for the Iron were obtained from the Li *et al* 2013. The QM region in the QM/MM calculation is consistent with the QM calculation performed above and link atoms were used to saturate the dangling bond in the QM/MM calculation. The Harmonic vibrational frequency calculations were performed at (UBP86/GEN: AMBER) lever of theory and the basis set used here are consistent with the QM calculations performed. The stationary points obtained were then subject to single point energy calculation in order to obtain the orbitals and spin density information.
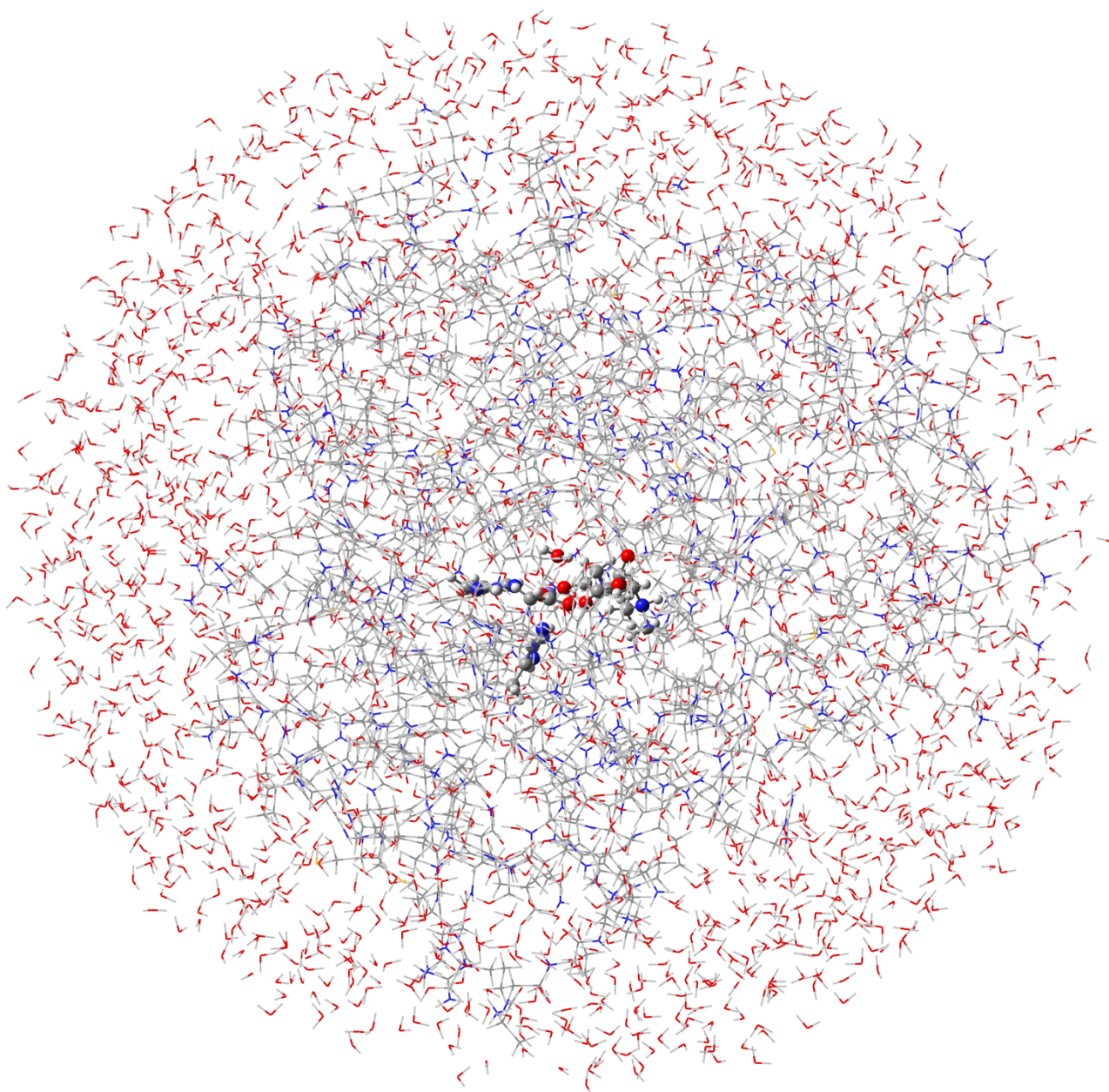
Figure 7.3 The QM/MM systems of PHF8. The residues and water molecules shown in the line representation are part of the MM region and the residues and solvent molecule shown in ball and stick model are part of the QM region. The link atom (Hydrogen) approach is used to saturate the dangling bond between the QM and MM region. The residue within 20 Å of the Fe (II) are free to move during the optimization. The structure is obtained using GaussView5 [87].

7.3 RESULTS AND DISCUSSION

7.3.1 Conformational and Essential Dynamics of PHF8

PHF8; The overall structural stability of PHF8 enzyme along with its PHD, Linker and Jumonji domain were assessed by performing RMSD analysis of Cα atom with respect to energy minimized X-ray crystallographic structure for 1 μs trajectory. PHF8 enzyme showed equilibration around 100 ns with average RMSD value of 3.3 Å and rest of the analysis was performed on the equilibrated trajectory (**Figure S7.1**). The RMSD data suggest the PHD domain (average rmsd value 2.7 Å) showed increase structural deviation in comparison to the Jumoji domain (average rmsd value 1.8 Å). The low RMSD and structural deviation of the Jumonji domain shows its importance for accommodating the active site and provide stability to the active site. The Jumonji domain harbours the jelly like fold which is the common feature for most 2OG dioxygenases enzymes. The jelly fold is composed of beta sheets which run in parallel and antiparallel fashion relative to each other [164, 166]. The RMSD profile of jelly fold domain showed very low structural deviation (average RMSD value of 1.3Å) indicating it underwent limited deviation from its crystal structure which also supports the fact that this region provides stability to non heme active site. The RMSD profile of histone peptide also showed structural deviation with average value of 3.4 Å for 1 μ trajectory. The most significant effect on RMSD value (3.3 Å) was seen in the linker region of PHF8 enzyme which connects PHD and Jumonji domain. The RMSD data shows that the linker region undergoes significantly increases conformational changes in comparison to rest of the protein. These results are in agreement with the crystal structure in which the residue of the linker region was missing and overall the region was assigned as a '*disordered*' region by the crystal structure authors [174]. The RMSF analysis of PHF8 (**Figure S7.2**) enzymes shows the increased flexibility of both the PHD and the linker region in comparison to Jumojoi domain; however the linker region residues showed the highest peak in the plot. The N and C terminal region of protein

showed high peaks which are consistent with the literature [19, 158]. The residues 216 to 230 also showed increased flexibility in Jumonji domain and belong to the loop region which is near the C terminal region of the protein and quite far away from the Jelly fold of PHF8 which harbours active site. The radius of gyration of PHF8 along with PHD, Linker, Jumonji and histone is shown in **Figure S7.3**. The major conformation motions of PHF8 enzyme were identified by performing PCA analysis. The first few eigenvectors describes mainly describes the major conformational motion in biomolecular systems such as proteins. This technique has been applied successfully to find patterns in high dimensional data [205]. The **figure S7.4** describes shows the proportion of variance of first 20 eigenvectors against its eigenvalue rank. The three eigenvectors describes up to 47 % of overall variance in the PHF8 enzyme. The first eigenvector mainly shows the overall motion of PHD and linker region of PHF8 (**Figure S7.5** and **S7.6**). The second and third principal components mainly describe the motion of linker region and terminal regions of the proteins. The domain cross correlation analysis of the PHF8 enzyme is represented in **figure S7.7**. The linker region of the PHF8 enzyme showed region of anticorrelated motion against the Jumoji domain. The histone peptide showed positive correlated motion against both the PHD and Jumonji domain indicating its important interactions with the both domains. The M3L and its neighbouring residues showed positive correlation against the PHD domain especially the histone M3L binding region. The M2L showed positive correlation towards the active site region of the Jumonji domain. The residue of the jelly roll motif showed positive correlation against each other as well.

7.3.1.1 Interactions of 2OG in PHF8

The carboxylate group of the 2OG (**Figure 7.2F**) makes electrostatic interactions with side chain of K264 residue with average distance of 3.5 Å throughout the simulation (**Figure S7.8**). The N189, Y257 and T244 residues stabilized the carboxylate side chain of 2OG by

making hydrogen bonds. The hydrophobic part of the 2OG is then stabilized by interaction with the side chain of I191 and L236. The MD simulations showed that the side chain of S238 make hydrogen bonds with both the K264 and N189 residues and stabilizes these residues to partake interactions with the carboxylate group of the 2OG.

7.3.1.2 Histone Interactions in PHF8

The non-coordinating oxygen (O1) of the 2OG make made hydrogen bonds with the side chain of the M2L with average distance of 3.0Å (**Figure S7.9**) and occupancy of 93 %. The tri-methylated lysine residue (M3L) makes extensive hydrophobic interactions with the side chain of Y7, Y14 and W29 residue (hydrophobic cage) of the PHD domain with average distance of 4.5 Å, 4.7Å and 4.4Å respectively (**Figure S7.10**). The backbone of M20 makes hydrogen bonds with the backbone carbonyl group of the M3L residue. The histone peptide is stabilized by both the PHD and the Jumonji in the PHF8.

7.3.1.3 QM/MM and MD atomistic analysis of PHF8 enzyme

The comparison of the MD, QM and QM/MM distance are shown in **table S7.1** and **S7.2**. The QM/MM calculations of five coordinated iron centre are displayed in **table S7.3**. The C5 carboxylate group of the 2OG is stabilized by K264 side chain with average distance of 3.5 Å in both MD and QM/MM optimization of all the snapshots. This is the common feature of both the PHF8 and KIAA1718 enzyme in which the lysine residue is involved in stabilization of C5 region of 2OG and indicates its importance in the drug design. The O3 and O4 atoms of the C5 carboxylate region makes hydrogen bonds with the hydroxyl group of the Y257 and T244 with average distance of 3.36 Å and 3.50 Å respectively in QM/MM minimization. The same distances are slightly higher for hydrogen bonding in MD simulation where these distance are 3.96 Å and 4.31 Å respectively. The M2L base (NZ) atom was on average 2.99 Å away from the O1 of 2OG during QM/MM

minimization and made stable hydrogen bonds. The MD trajectory of the same distance also showed an average value of 3.5 Å. The side chain of the AP1 (Aspartic acid coordinating to Iron (II), OD2 atom) was stabilized by hydrogen bonds with the N333 residue. In snapshot 300ns the bulky methyl group of the M2L base come close the iron and pushes N333 furthered away from AP1 and OD2 of AP1 is coordinated by water molecule using hydrogen bonds. In snapshot 700ns, the M2L make hydrogen bonds with the N33 side chain and N333 stabilises the OD2 oxygen of the AP1. In snapshot 900ns, the M2L made hydrogen bond with the side chain of N333 and N333 made hydrogen bond with the AP1 (**Figure 7.4**).
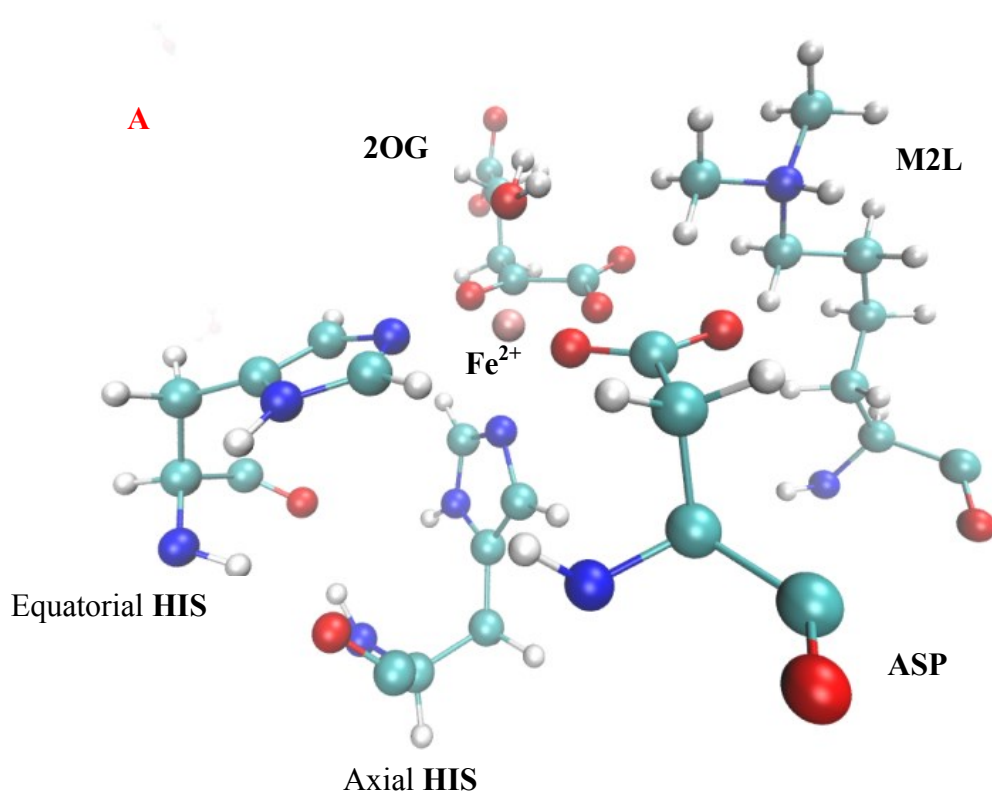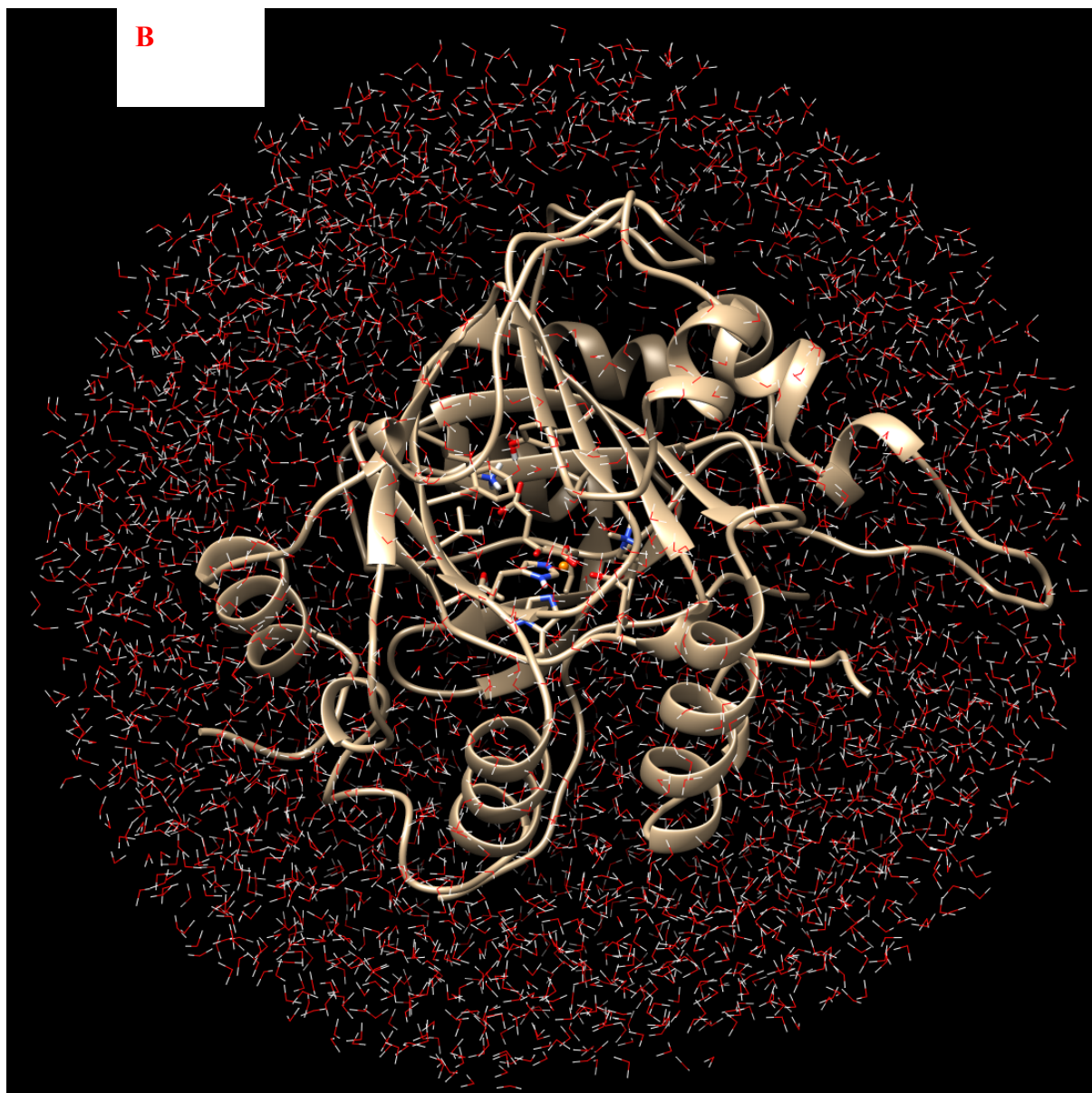
Figure 7.4 The geometry optimized 900ns QM/MM snapshot of the PHF8 enzyme. (A) The active site of the PHF8 enzyme, (B) The full structure of PHF8 enzyme used for the QM/MM optimization.

7.3.2 Conformational dynamics and Essential Dynamics of KIAA1718

The overall structural stability of KIA1718 enzyme along with its PHD, Linker and Jumonji domain were assessed by performing RMSD analysis of Cα atom with respect to energy minimized X-ray crystallographic structure for 1 μs trajectory. KIAA1718 enzyme

showed equilibration around 200 ns with the average RMSD value of 2.1 Å (**Figure S7.11**). The radius of gyration of KIAA1718 and its domain also show equilibration around 200ns as well (**Figure S12**). The PHD and Jumonji domains showed an average RMSD value of 2.5 Å and 1.5 Å respectively for the equilibrated trajectory. The PHD domain makes no interactions with the histone peptide as it is in the 'extended' conformation which might explain relatively increased RMSD profile. The crystal structure describes the distance of hydrophobic cage (potential histone binding site in KIAA1718) in PHD domain to the non heme active site in the Jumonji domain to be ~ 37 Å. The centre of mass of hydrophobic core residue (P32-P37) to active site in the simulation revealed an average distance of 46.2 Å, indicating even more extended conformation of PHD domain as function of simulation time. The centre of mass of the PHD to Jumonji domain also showed an average distance of 40.1 Å during simulation furtherer indicating the extended conformation of the PHD domain. The PHD domain distance to histone peptide remained stable (average distance 29.0 Å) (**Figure S7.13**). The Jumonji domain harbours the jelly-roll motif which is the common feature for most 2OG dioxygenases enzymes. The jelly fold is composed of eight β-strands which run in parallel and antiparallel fashion relative to each other to form a squashed β-sheet barrel type structure [164, 166]. The RMSD profile of jelly fold domain showed very low structural deviation (average RMSD value of 0.6Å) indicating it underwent very limited deviation from its crystal structure which also supports the fact that this region provides stability to non heme active site also act as a place holder for substrate catalysis. The RMSD profile of histone peptide show higher structural deviation to rest of the domains of the KIA1718. Interesting the linker region of KIAA1718 showed very low RMSD with average value of 0.9 Å as predicted by the crystal structure paper which describes the linker region to be in the 'ordered' conformation [174]. The low RMSD value of linker region in context of KIAA1718 means that the linker region plays a very limited role in influencing

conformational dynamics of PHD and Jumonji domain for substrate recognition and catalysis respectively. The RMSF analysis of KIAA1718 enzymes shows increased flexibility of the PHD in comparison to the Jumonji domain (**Figure S7.14**). The PHD domain is responsible for substrate binding and specificity. The lack of interactions of substrate due to extended nature of PHD domain correlates with increased flexibility of residues comprising PHD domain. The linker region residues showed significantly low flexibility and as suggested by crystal structure and have limited influence on the conformational dynamics of both PHD and KIAA1718 domains. The residues 255 to 265 also showed increased flexibility in Jumonji domain and belong to the loop region which is near the C terminal region of the protein and quite far away from the jelly-roll motif of PHF8 which harbours active site. The major conformation motions of the KIAA1718 enzyme were identified by performing PCA analysis (**Figure S7.15**). The scree plot of the first 20 eigenvectors shows that the first two eigenvectors or principal components account for more than half (56.1 %) of overall variance in the KIAA1718. The visual inspection of the first PC shows the movement of PHD domain with respect to the Jumonji domain. The linker region showed no movement with respect to any of the domain and is quite rigid in terms of its conformational dynamics. The individual residue contribution towards the first PC were also computed and showed again the major contribution from PHD domain (residue 33-97) of KIAA1718. The dynamic cross correlation analysis shows that the residues belonging to the jelly roll motif (140-205, 325-340) showed positive correlation against the residue belonging to (140 -205) of the Jumonji domain. The residues of the linker region showed correlated motion towards the residues (160-180) of the Jumonji domain. The residue in vicinity of linker also showed positive correlation against the Jumonji domain residues. The histone peptide residue also showed positive correlation towards the Jumonji domain residues, however showed negative correlation towards the

PHD domain. The residues of the PHD domain showed the region of large anti-correlated motion towards the Jumonji domain.

7.3.2.1 Interactions of 2OG in KIAA1718

The 2OG is located within the jelly roll motif and its non-coordinating C5 carboxylate group penetrates deep in the jelly roll barrel. The C5 of the 2OG makes electrostatic interactions with side chain of K299 residue with average distance of 3.3 Å throughout the simulation (**Figure S7.17**). The Lysine residue in PHF8 was also involved in stabilizing the C5 carboxylate group of 2OG with electrostatic interactions. This is a unique feature of JmJC subfamily however in other families such as FTO, AlkB normally Arginine is involved in the electrostatic interactions with the C5 group. The side chain of Y292 and T279 also forms the part of cofactor binding site and stabilizes the C5 group of the 2OG by making hydrogen bonds with 97 % and 25.6 % occupancy throughout the simulation. The methylene carbon atoms of the 2OG maintained the extended conformation throughout the simulation and were stabilized by cluster of hydrophobic residues I195 and V325 with average distance of 4.7 Å and 4.0 Å respectively. The non-coordinating oxygen atom (O1) of 2OG made hydrogen bonds (occupancy of 92 %, average distance of 3.0 Å) with the side chain of di-methylated lysine residue of histone peptide.

7.3.2.3 QM/MM and MD atomistic analysis of KIAA1718 enzyme

The comparison of the MD, QM and QM/MM distance are shown in **table S7.4** and **S7.5**. The side chain of residue K299 is involved in electrostatic interactions with the non-coordinating C5 of 2OG both in MD simulation and QM/MM snapshots with average distance of 3.35 Å and 3.27 Å respectively. The hydroxyl group of the Y292 residue made consistent hydrogen bonds with the O3 oxygen of the 2OG during the MD minimization and in all the QM/MM snapshots. However the side chain of T279 residue was on average

4.84 Å (MD simulation) away from the O4 of 2OG and only made hydrogen bonds in the 600 ns and 700ns snapshot and in rest of the snapshots, the side of T279 moved slightly away from the 2OG. The N224 side chain was on average 4.22 Å away from the O4 of 2OG during QM/MM minimizations. The M2L base (NZ) atom was on average 3.12 Å away from the O1 of 2OG during QM/MM minimization and made stable hydrogen bonds. The MD trajectory of the same distance also showed an average value of 3.12 Å. In the 600n snapshot, the methylated base of M2L come close to the iron centre, the O1 of 2OG make no hydrogen bonds with the NZ of the base, however the side chain of N368 make hydrogen bonds with the NZ of the M2L substrate. There is also disruption of the AP1 OD2 hydrogen bond with the side chain of N368. These results are similar to the PHF8 enzyme. In the 700ns snapshot, the side chain of the AP1 (OD2) made water-bridge mediated hydrogen bonds with the side chain of the N368 residue. The sidechain N368 play an important role either it stabilizes the M2L when it gets close to the iron centre otherwise it is involved in stabilizing the OD2 of AP1. In the 800ns snapshot the AP1 mediated water bridge hydrogen bond with the N368 residue as well.

7.3.3 Conformational dynamics and Essential Dynamics of FTO

FTO enzyme showed equilibration around 100 ns with average RMSD value of 4.2Å (**Figure S7.18**). The N and C domain of FTO showed average RMSD value of 4.5 Å and 1.7 Å respectively. The RMSD profile of N domain of the FTO showed significantly larger structural deviation in respect to the C domain and indicates overall extend of greater conformational dynamics attributed by N domain of FTO. The major contribution of the overall RMSD of FTO is attributed by N domain. The N domain of the FTO enzyme have greater number of loop region which are known to experience greater structural deviation in respect to the alpha helix or beta sheets [17, 101] . The RMSD profile of residues connecting N to C domains showed average value of 0.7 Å. The jelly roll motif which is

the common feature of DNA/RNA and histone demethylases [164, 166] showed a stable RMSD profile with average value of 0.8 Å. The RMSD of jelly roll motif however showed a slight increase around 900ns and then stabilized for rest of the trajectory. The characteristic L1 loop [178] which is the unique feature of FTO enzyme showed stable RMSD value till 400ns however showed higher conformational dynamics after 500 ns and then stabilized for rest of the trajectory. This loop is involved in substrate recognition. The radius of gyration of FTO and its N and C domains also show equilibration around 100ns with average value of 27.0 Å, 19.2Å and 17.3 Å respectively (**Figure S7.19**). The centre of mass distance between the N and C domain of the FTO enzyme showed an average value of 36.8 Å for 1 μs trajectory (**Figure S7.20**). There was slight increase in the distance between two domains after 400ns during the simulation. The RMSF analysis of FTO is represented in **figure S7.21**. The residues ranging from 160 to 197 show a high flexibility in the N domain region. These residues belong to the loop region and are far away from the active site of FTO enzyme. The residues ranging from 209 to 224 showed increased flexibility. The L1 loop belongs to this region and is important for the FTO. The high flexibility of loop region indicates its vital role in FTO. The residue ranging from 250 to 260 which belong to the loop region show increased flexibility. The residue of the C domain ranging from 346 to 355 which are part of loop region show increased flexibly. It could be concluded from the RMSF analysis that the N domain residues show overall increased conformational flexibility in respect to C domain residues.

The principal components analysis of FTO shows that the first three eigenvector represent 50 % of overall variance in the FTO (**Figure S7.22**). The analysis of residues which contributes to the PC1 and PC2 are represented in **figure S7.23**. The major contribution was made by residues in the N terminal residues of the FTO in both PC1 and PC2. The residues ranging from 160 to 197 show a high flexibility in PC2. The residue belonging to the L1 loop also showed increased contribution in both PC1 and PC2 plots. The visual

inspection of the PC1 (**Figure S7.24**) indicates similar results as shown by residue loading

contribution for each PC. Overall the PCA analysis reveals that the major motions are

experienced in the N domain of the FTO enzyme and the loop region of the N domain

residues plays an important role in overall dynamics of the protein. The jelly roll motif

remain in the intact position and show limited motion which is consistent with the idea that

it harbours the active site of the protein and provide stable environment for catalysis. The

motion of the L1 domain in PCA is an important feature extracted and is consistent with

the RMSD and RMSF analysis.
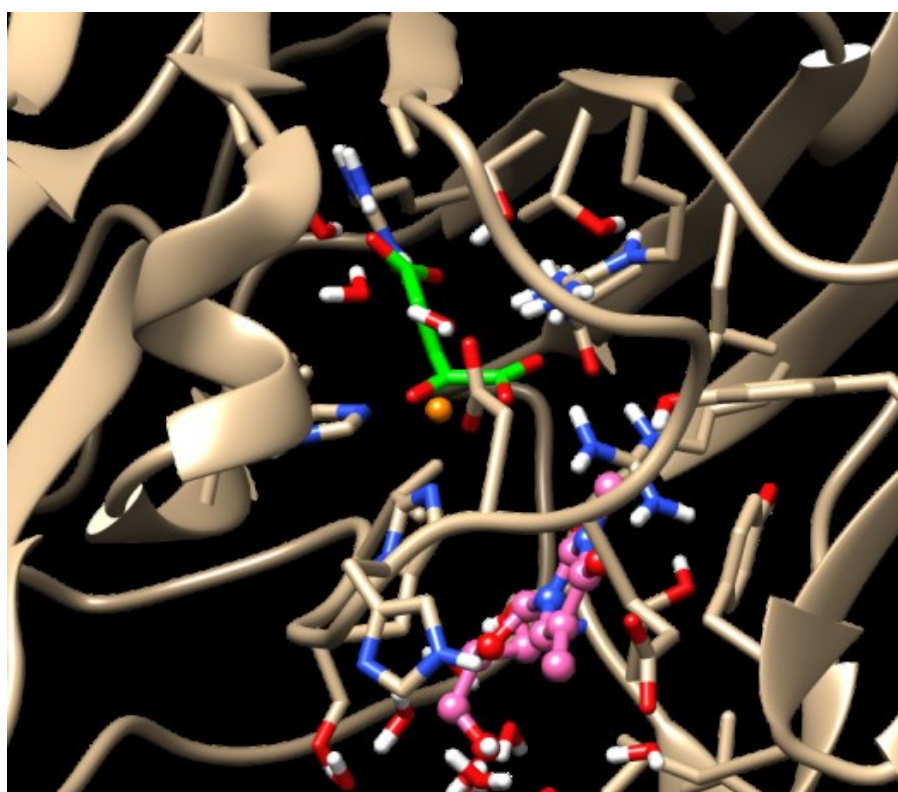
Figure 7.5 The QM/MM optimized 300ns snapshot of the FTO enzyme.

7.3.3.1 QM/MM and MD atomistic analysis of FTO enzyme

The comparison of the MD, QM and QM/MM distance are shown in **table S7.6** and **S7.7**.

The residue R316 stabilized the C5 carboxylate group of the 2OG with electrostatic

interaction through the trajectory during MD simulation average distance of 3.97 Å.

However during the QM/MM minimization of the minimized crystal structure shows the distance to increase to 4.4 Å. However the side chain of R316 (NH2) is involved in making hydrogen bonds with O3 oxygen of the C5 carboxylate group with distance of 1.9Å and angle of 151.9 Å. The QM/MM minimization (**Figure 7.5**) of various snapshots (200, 300, 700 and 800ns) showed consistent distance of ~3.94 between the side chain of R316 and C5 carboxylate group of 2OG. This results show the R316 residue stabilizes the carboxylate group of 2oG by salt bridges rather than hydrogen bonding. The hydroxyl group of Y295 residue made hydrogen bonds (average distance of all snapshots ~2.72 Å) with O3 of C5 carboxylate group of 2OG in 200-800ns snapshots. However the QM/MM minimized structure of the same distance showed distance of 5.13 Å indicating absence of hydrogen bonds. This indicates that the side chain of Y295 moves towards the 2OG cofactor and stabilizes its C5 carboxylate group. The side chain of S318 is involved in hydrogen bonding with the O4 oxygen of C5 group with distance of 2.56Å and angle of 178.10 degree in the minimized QM/MM structure. The similar trend was seen in rest of the snapshots as well with average distance of 2.55 Å. The side chain of N205 makes hydrogen bonds with both the O4 and O1 of 2OG with distance and angle of 1.78 Å, 178.71° and 2.81Å, 123.53° respectively in the minimized QM/MM crystal structure. However the distance in rest of the snapshot gradually increased to 3.74 Å and 3.80 Å respectively. The average distance of all the snapshots was computed to be 3.52 Å and 3.42 Å respectively. These results were consistent with the average distance of 4.5Å and 3.11 Å during the MD simulation of FTO. The side chain of R322 in minimized QM/MM crystal structure made electrostatic interaction with the C1 carboxylate group of the 2OG and simultaneously involved in stabilization of the non-coordinating oxygen (OD2) of the AP1 (Asp coordinated to Iron) with distance of 3.92Å and 3.14Å, 94.9° respectively. The NH2 group of the R322 also was involved in the hydrogen with the O1 oxygen of 2OG C1carboxylate group. The R96 residue was involved in hydrogen bonding with the O1

oxygen of 2OG and side chain of 3DT substrate with distance of 2.98 Å and 2.75Å respectively for minimized QM/MM crystal structure. The rest of the snapshot also showed similar profile apart from 800ns snapshot where there significant increase in distance to 5.10Å and 3.80Å respectively. Overall the side chain of R96 residue play an important role in stabilizing both the substrate and cofactor in FTO enzyme.

7.3.4 Conformational dynamics and Essential Dynamics of AlKB

The overall structural stability of the AlkB enzyme along with double stranded DNA were assessed by performing RMSD analysis of Cα and backbone phosphorus atoms with respect to the minimized crystal structure. The AlkB enzyme showed equilibration around 50 ns with average value of 4.0 Å (**Figure S7.25**). The backbone phosphorus atoms of DNA showed average value RMSD of 4.0 Å for entire length of the simulation. The RMSD of Cα only atoms of the protein showed the significantly lower structural (average value 1.4 Å) in contrast to the DNA molecule, it could be conclude that the overall increase in the RMSD profile of AlkB protein due to the DNA molecule. The jelly roll motif showed low and stable RMSD with average value of 0.8 Å through the simulation. These results are consistent with the above simulation of 2OG dependent proteins. The significantly stable structure of jelly roll motif is an important feature of all the 2OG enzymes [164, 166]. The active site loop consisting of residues from 133 to 140 belong to β6-β7 showed average RMSD value of 1.8Å for 1 μs simulation. The RMSD profile of active site loop showed a stable profile. The residues harbouring the active site are vital for catalysis therefor any unwanted perturbation of the loop might influence the catalytic activity of the enzyme. The radius of gyration of AlkB (Cα+P) showed an average value of 19.5Å and was stable throughout the simulation (**Figure S7.26**). The radius of gyration of protein and DNA molecule also showed a stable profile with average value of 15.8Å and 14.4Å for 1μs trajectory. The Rg of the active site loop of the AlkB protein also showed a stable profile with average value of 6.5Å (**Figure S7.27**). Centre of mass distance of the

protein from DNA indicates the relative movement of the two molecules in respect to each other. The centre of mass distance between the protein and DNA molecule during the simulation maintained and average value of 28.2 Å in contrast to its crystal distance of ~ 24.0 Å (**Figure S7.28**). The increase in the distance during simulation indicates the more open structure of the AlkB enzyme in contrast to its crystal structure. The RMSF analysis (**Figure S7.29**) of AlkB enzymes reveals region of increased flexibility. The residue ranging from 81-90 which belong to loop region between β4 and B alpha helix showed increased flexibility. The residue of the active site loop (133-140) belonging to β6-β7 showed increased flexibility and very important according to the crystal structure authors. The active site loop harbours D133 which make is part of the active site and residues in this loop make important interactions with the methylated DNA bases. These residues are significantly far away from the active site. The residue (160-170) of β9 and loop connecting β8-β9 showed increased flexibility and are located in close vicinity to DNA molecule. The residue 179 to 185 belonging to the C alpha helix and β11 also showed increased flexibility, β11 harbours the active site H187 residue. The backbone atoms of the terminal ends of the DNA molecule showed increased fluctuations which is consistent with the N and C terminal region of the protein.

The principal components analysis of the AlKB enzyme which consistent of Cα and P atoms from protein and DNA molecules respectively showed major motions experienced by enzyme during 1μs simulation. (**Figure S7.30**). The scree plot of the eigenvectors showed that first four eigenvectors represented more than half of the total variance in the data set. The projection of PC1 on the atomic coordinates of the showed major motions in the AlkB enzyme which were mainly restricted to N and C terminal region of the protein. The active site loop showed motion towards the C alpha helix and β11. The loop region belongs to β4 and B alpha helix also experienced motions which were away from the

active site. Mainly the loop region of the AlkB enzymes showed motions. The jelly roll motif showed stability in the PCA analysis. The major motion experienced DNA during PCA analysis was mainly found on the terminal regions as shown by individual DNA bases loading of the PC1 and PC2 projection on the structure. The methylated DNA bases show it motion showed rather limited motion (**Figure S7.31**). The domain cross correlation analysis of AlkB is represented in **figure S7.32**. The residues 17-25 belong to β1 show correlation against residues 175-180 of β10. The β10 belong to the jelly roll motif. The residues harbouring to β2 (59-64) show correlation towards residues 115-122 of β5 and both these beta sheets are part of jelly roll motif. The residues of the active site loop (133-139; β6-β7) show correlation towards 187-195 (loop and β11). The β11 is part of jelly roll motif and harbours active site residues which are involved in coordination with Fe(II). The residues152-160 part of β8 shows correlation towards 187-195. In essence it could be said that the jelly roll motif show correlation against other. Interesting the residues (53-58) belonging to loop between A helix and β2 show correlation against middle of the DNA molecule in particular with 6MA.

7.3.4.1 QM/MM and MD analysis of AlkB

The comparison of the MD, QM and QM/MM distance are shown in **table S7.8** and **S7.9**. The residue R204 stabilized the C5 carboxylate group of the 2OG with electrostatic interaction through the trajectory during MD simulation average distance of 3.97 Å. However during the QM/MM minimization of the minimized crystal structure shows the distance to increase to 4.06 Å. The QM/MM minimization of various snapshots (200, 400 and 500ns) showed consistent distance of ~3.9 between the side chain of R204 and C5 carboxylate group of 2OG. This results show the R204 residue stabilizes the carboxylate group of 2oG by salt bridges rather than hydrogen bonding. The hydroxyl group of Y122 residue made hydrogen bonds (average distance of all snapshots ~2.70 Å) with O3 of C5

carboxylate group of 2OG in all the QM/MM snapshots (**Figure 7.6**). The side chain of N206 is involved in hydrogen bonding with the O4 oxygen of C5 group with an average distance of 3.737Å in the QM/MM structures. The R210 side chain in 200ns snap shot stabilized the OD2 of AP1.
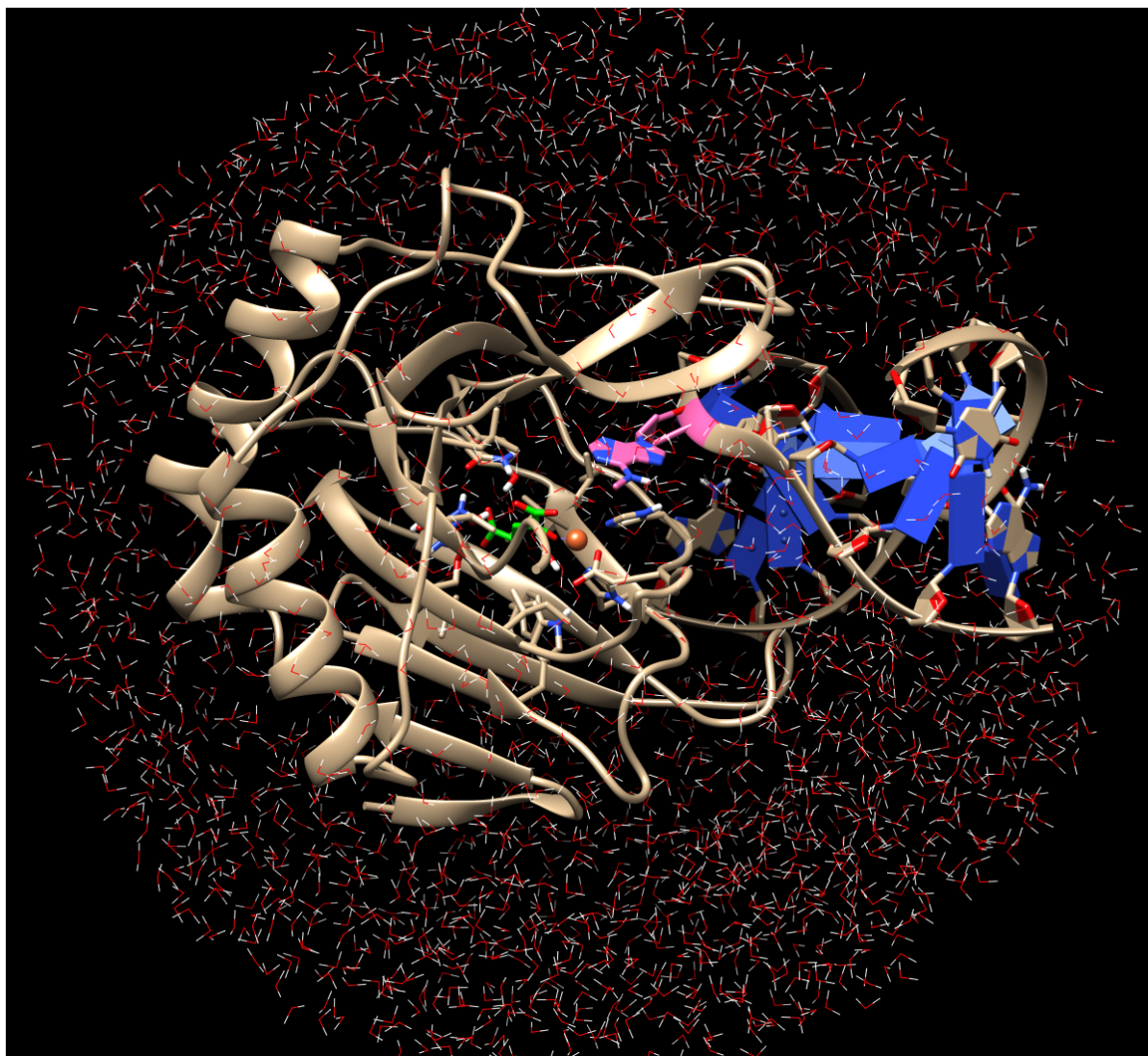
Figure 7.6 The QM/MM optimized structure of the AlkB enzyme.

7.4 CONCLUSION

The non-heme Fe (II) active site parameters developed by MCPB reproduced the active site very well in the MD simulations for 1μs. The linker region is important in the histone lysine demethylase enzymes and the exchange of the linker region between these enzymes has an impact on the overall dynamics. The PHD domain binds to the tri-methylated lysine residue and lack of histone binding to the PHD domain increases its fluctuations. The relative free energy of binding indicates the stronger influence of vdW interactions form the PHD domain to histone binding. The jelly roll motif is stable in all 2OG dependent enzymes and showed extend of the positive correlated motions. The QM /MM ONIOM model maintained the overall architecture of the non-heme Fe (II) active site in contrast to the QM cluster model. The effect of conformational flexibility was most importantly seen in the orientation of the H3K9me2, non-coordinating carboxylic side chain of 2OG and OD2 of the Aspartic acid. The stabilization of the 2OG by second spheres residues in all fours enzyme are consistent in both the MD and QM/MM optimizations. The MD and QM/MM simulations of the non-heme enzymes foster our understanding of the properties of the activated iron-oxygen complex. This provides us with the valuable information to design novel iron based catalyst which are both inexpensive and environmentally friendly.
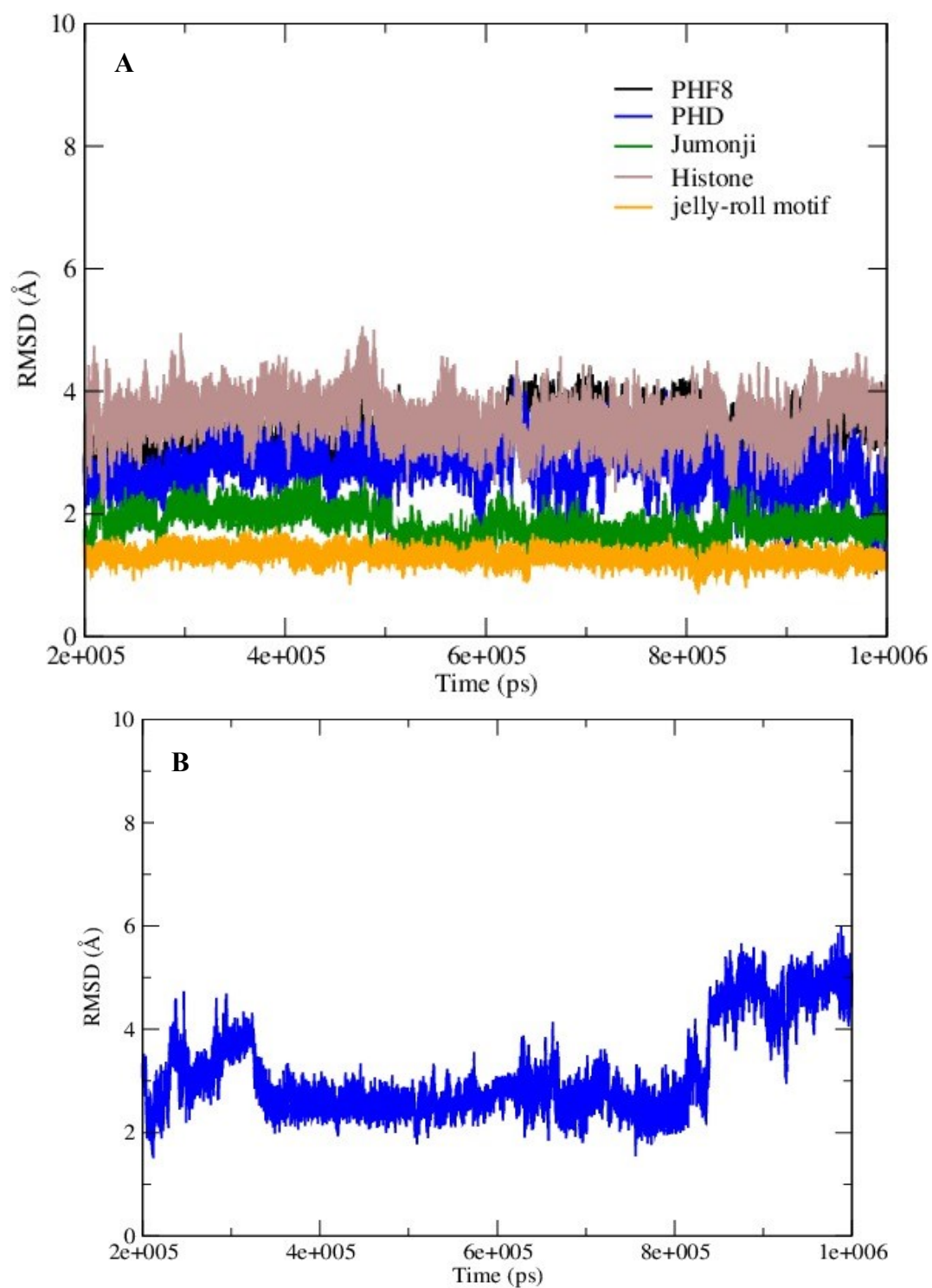
Figure S7.1 The RMSD profile of the PHF8 enzyme along with its individual domains (PHD, Linker, Jumonji) and Histone substrate peptide. (A) RMSD of the PHD, Jumonji, Histone and Jelly roll motif, (B) The RMSD profile of the linker region. The rms command

in ccpptraj was used for the analysis. The trajectory was saved every 10 ps. The equilibrated trajectory from 200 to 1000 ns was used for the analysis.
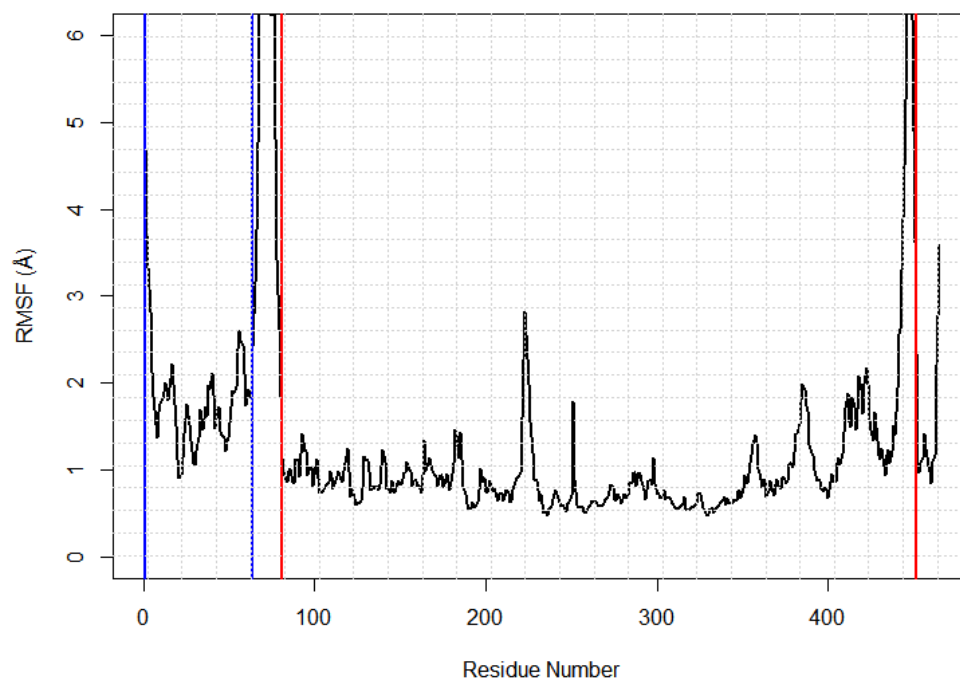
Figure S7.2 The RMSF profile of the PHF8 protein using Cα using cpptraj. The PHD domain of the protein is in blue column followed by linker region sandwiched between the PHD and Jumoji domain. The analysis was carried out on the equilibrate trajectory.
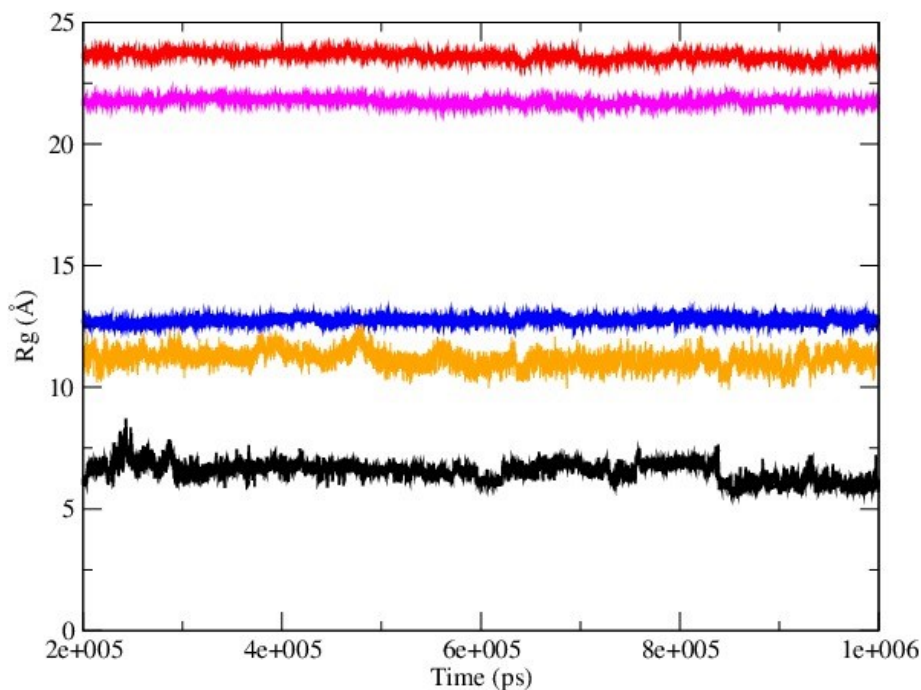
Figure S7.3 The Radius of gyration of the PHF8 and its domains using Cα using cpptraj.
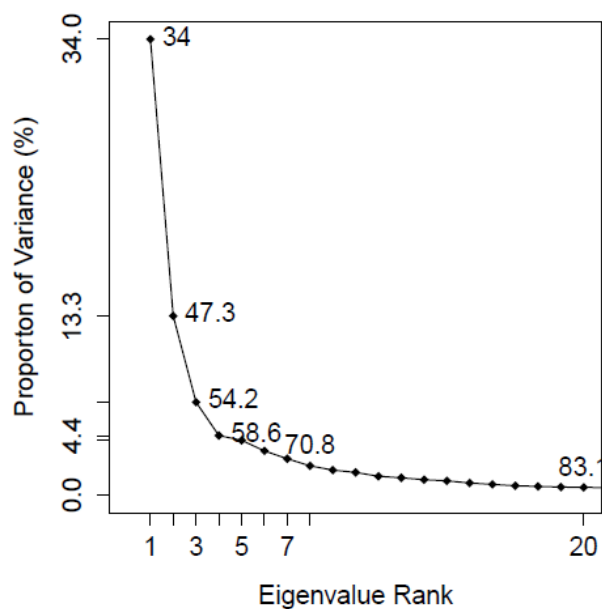
Figure S7.4 The plot of first 20 eigenvectors of PHF8 done using Cα atoms using Bio3d package in R
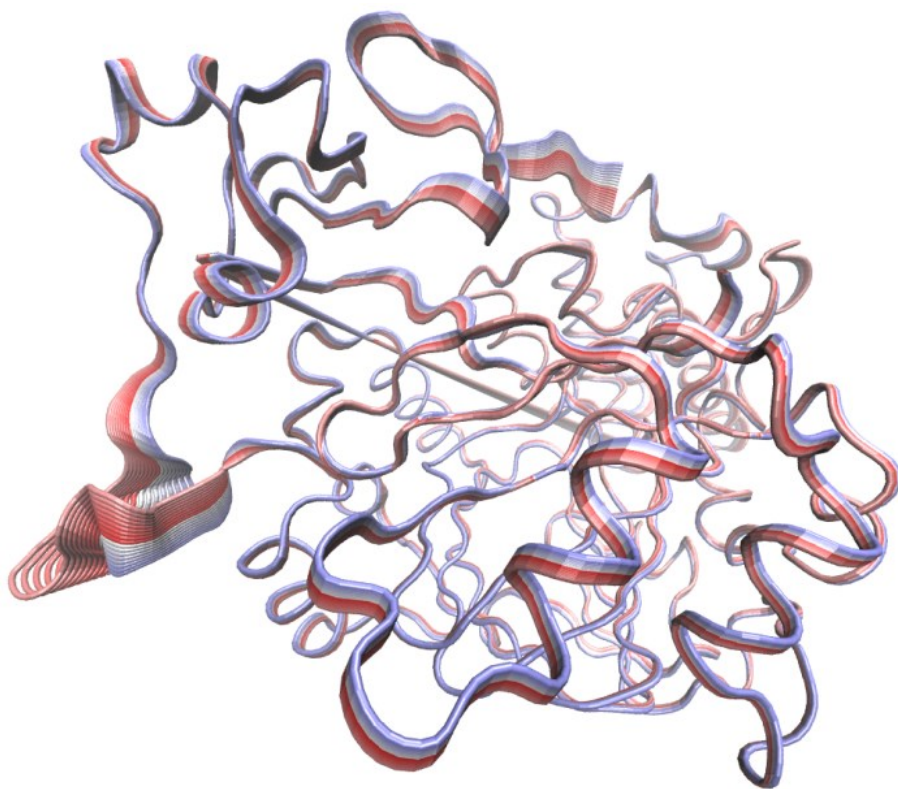
Figure S7.5 The projection of first eigenvector showing the major motion in the linker and the PHD domain of PHF8. The greater extend of atomic fluctuations of PHD and linker. The colour scale from blue to red represents low to high atomic displacements.
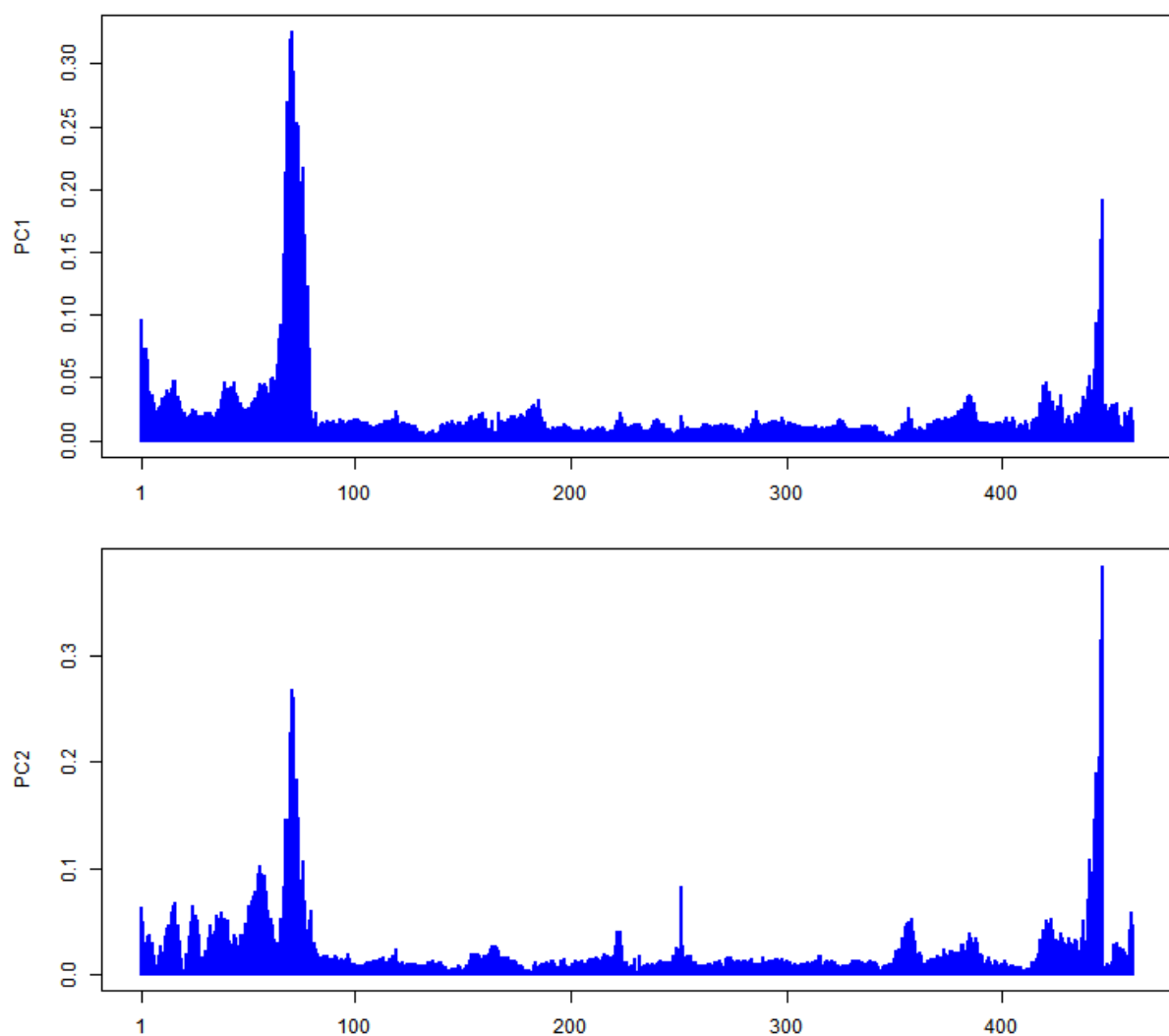
Figure S7.6 The residue wise loading of individual residues contributions (Å) to first two

eigenvectors. The residue 1-65 belong to PHD domain and the linker residues are 65-79
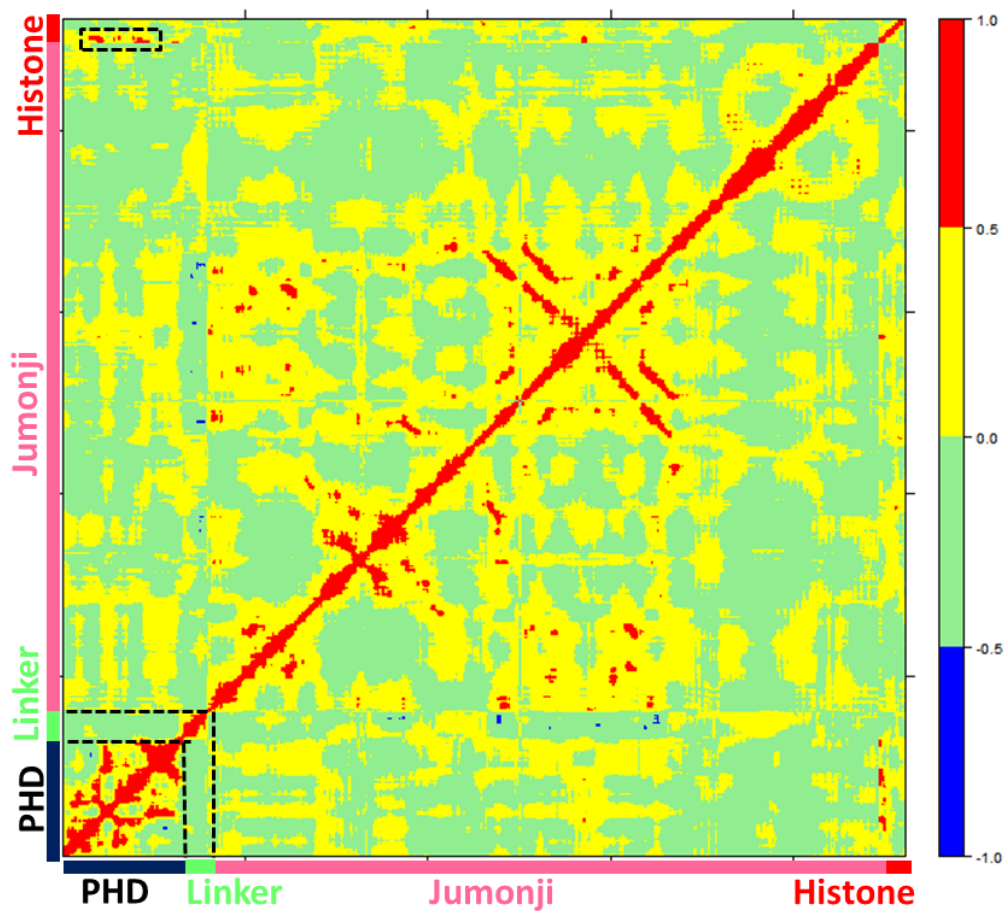
Figure S7.7 The domain cross correlation of the PHF8 enzyme using Bio3D in R package.
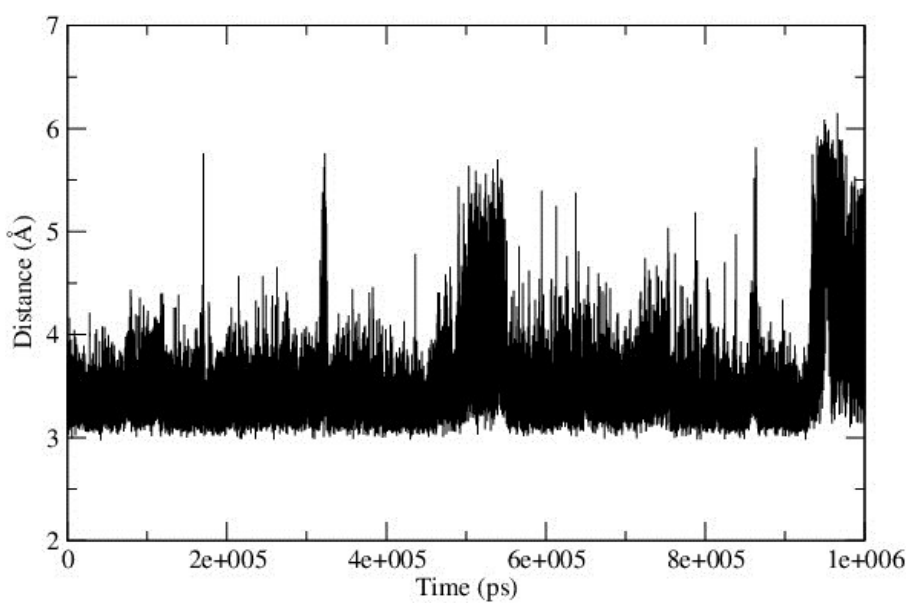
Figure S7.8 The interaction of the side chain of the K264 with the carboxylate group of

2OG.
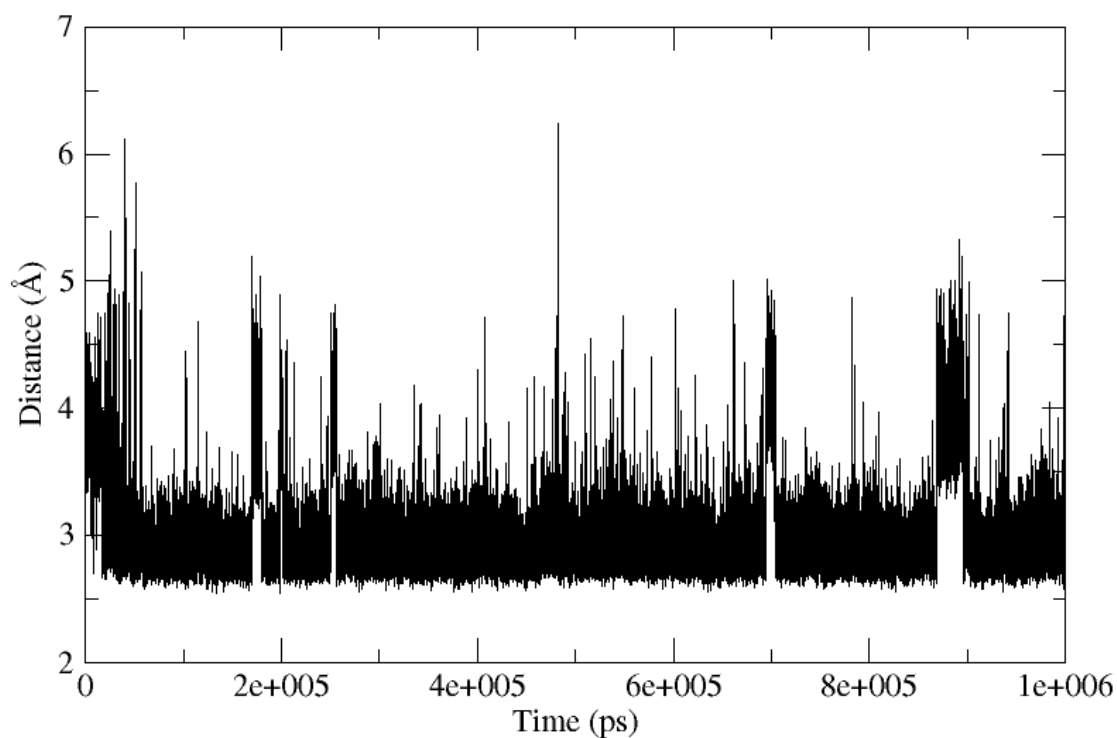
Figure S7.9 The interactions of the substrate M2L with the 2OG in the PHF8 enzyme.
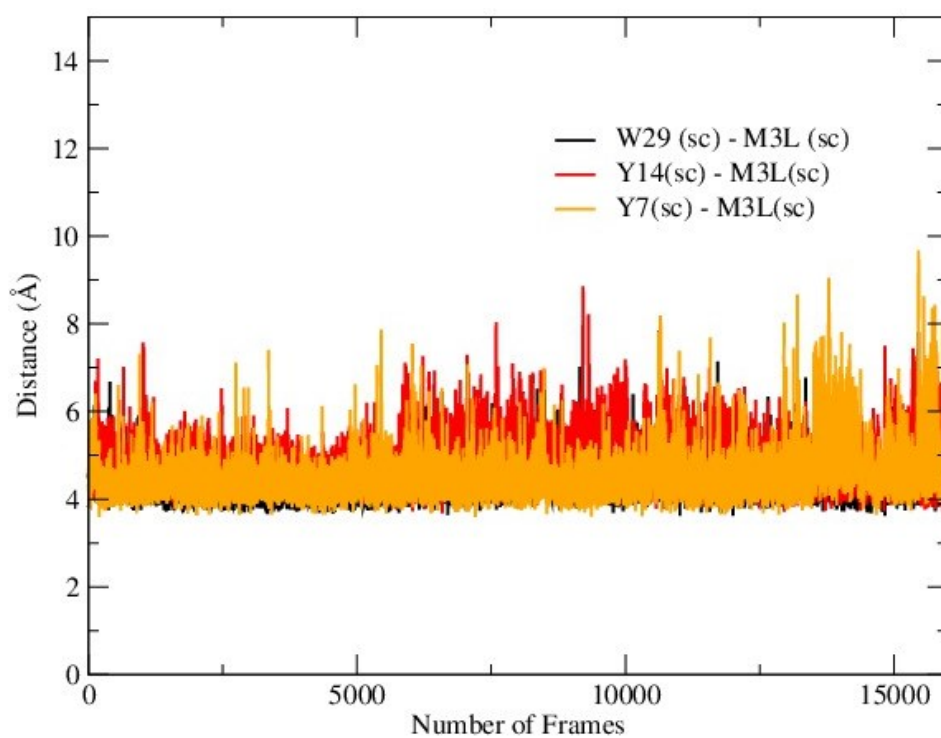
Figure S7.10 The interactions of the substrate M3L with the PHD domain in the PHF8

enzyme. Note here numbers of frames are equivalent to one microsecond trajectory.

Table S7.1 **PHF8** QM/MM analysis of distance to the non heme Iron (II) of the ligands

Name	Equatorial His	Axial His	ASP	2OG (C1)	2OG (C2)	H2O	di-methylated
QM Cluster minimized	**2.12**	**2.16**	**2.05**	**2.12**	**2.12**	**2.17**	**5.15**
Minimized	2.13	2.17	2.04	2.11	2.26	2.24	4.60
300ns	**2.13**	**2.13**	**2.00**	**2.17**	**2.15**	**2.29**	**4.28**
600ns	2.15	2.14	2.03	2.15	2.15	2.21	4.53
700ns	2.16	2.13	2.12	2.08	2.22	2.21	4.21
800ns	2.13	2.16	2.06	2.17	2.21	2.23	4.45
800ns (EE)	**2.14**	**2.16**	**2.09**	**2.09**	**2.15**	**2.20**	**4.50**
900ns	2.13	2.15	2.09	2.09	2.21	2.17	4.14
1000ns	2.11	2.22	1.98	2.19	2.17	2.19	4.57
QM/MM (Mec)(μ*)	2.135	2.16	2.05	2.13	2.19	2.21	4.41
QM/MM (Mec)(σ*)	0.0151	0.029	0.0479	0.040	0.0410	0.036	0.175
MD μ**	2.30	2.15	2.022	1.98	2.11		4.64
MD (σ*)	0.0912	0.0740	0.0651	0.060	0.116		0.66

EE Electrostatic embedding, rest is all mechanical embedding schemes.

μ* here is only performed for the analysis which has been done using Mechanical embedding scheme under ONIOM calculations.

μ** here is the MD simulation averaged over 1μs trajectory.

σ* here is the standard deviation of the QM/MM minimization.

C5: here is the non-coordinating carboxylate group of the 2OG cofactor in FTO.

O1: The non-coordinating oxygen of C1 of 2OG

Table S7.2 **PHF8** QM/MM analysis of PHF8 of the second sphere interactions with active site.

Name	N189 (sc) O4 2OG	Y257 (sc) – O3 2OG	T244 (sc) – O4 2OG	K264 (sc) –C5 2OG	M2L (sc) – O1 2OG	N333 – O2 AP1
Minimized	**2.93**	**4.41**	**2.63**	**3.56**	**2.64**	**3.02**
300ns	**4.39**	**2.64**	**2.66**	**3.50**	**2.62**	**5.76**
600ns	4.04	2.72	4.33	3.19	2.63	3.53
700ns	3.17	4.32	2.59	3.48	4.07	3.36
800ns	5.11	2.71	4.41	3.19	2.64	3.73
800ns (EE)	5.31	2.68	4.49	3.20	2.78	3.62
900ns	4.76	2.69	4.29	3.16	3.93	3.31
1000ns	2.91	4.73	2.61	5.33	2.60	3.75
QM/MM (Mec)(μ^*)	**4.07**	**3.36**	**3.50**	**3.58**	**2.99**	**3.76**
QM/MM (Mec)(σ^*)	**0.974**	**0.9382**	**0.941**	**0.727**	**0.627**	**0.843**
MD μ^{**}	**5.35**	**3.96**	**4.312**	**3.505**	**3.505**	**3.52**
MD (σ^*)	**1.197**	**1.187**	**1.241**	**0.386**	**0.3441**	**0.344**

Table S7.3 PHF8 QM/MM analysis of distance to the non heme Iron (II) of the ligand with 5-coordinated

| Name	Equatorial His	Axial His	ASP	2OG (C1)	2OG (C2)	di-methylated
Minimized	2.15	2.17	2.02	2.06	2.40	4.51
Minimized (EE)	**2.09**	**2.09**	**1.97**	**2.04**	**2.24**	**4.95**
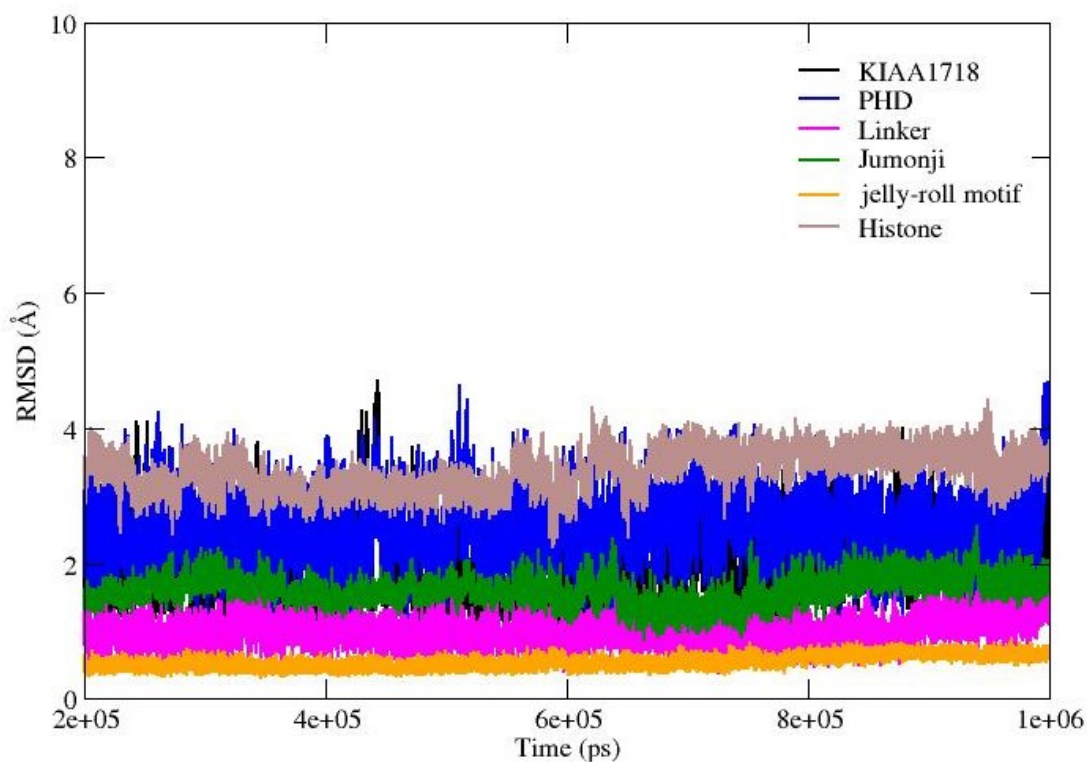
Figure S7.11 The RMSD profile of the KIAA1718 enzyme along with its individual domains (PHD, Linker and Jumonji), Histone substrate peptide and jelly-roll motif. The rms command in ccpptraj was used for the analysis. The trajectory was saved every 10 ps. The equilibrated trajectory from 200 to 1000 ns was used for the analysis.

Figure S7.12 The Radius of gyration of the KIAA1718 and its domains using Cα using cpptraj.
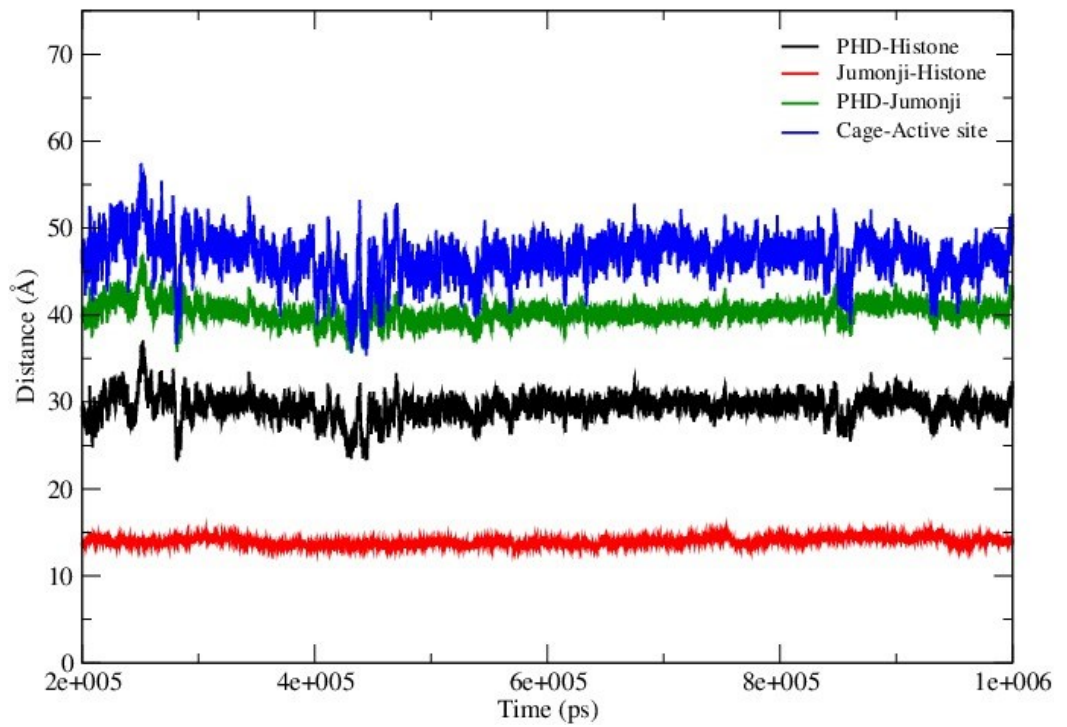
Figure S7.13 The centre of mass of the KIAA1718 domains for the equilibrated trajectory.
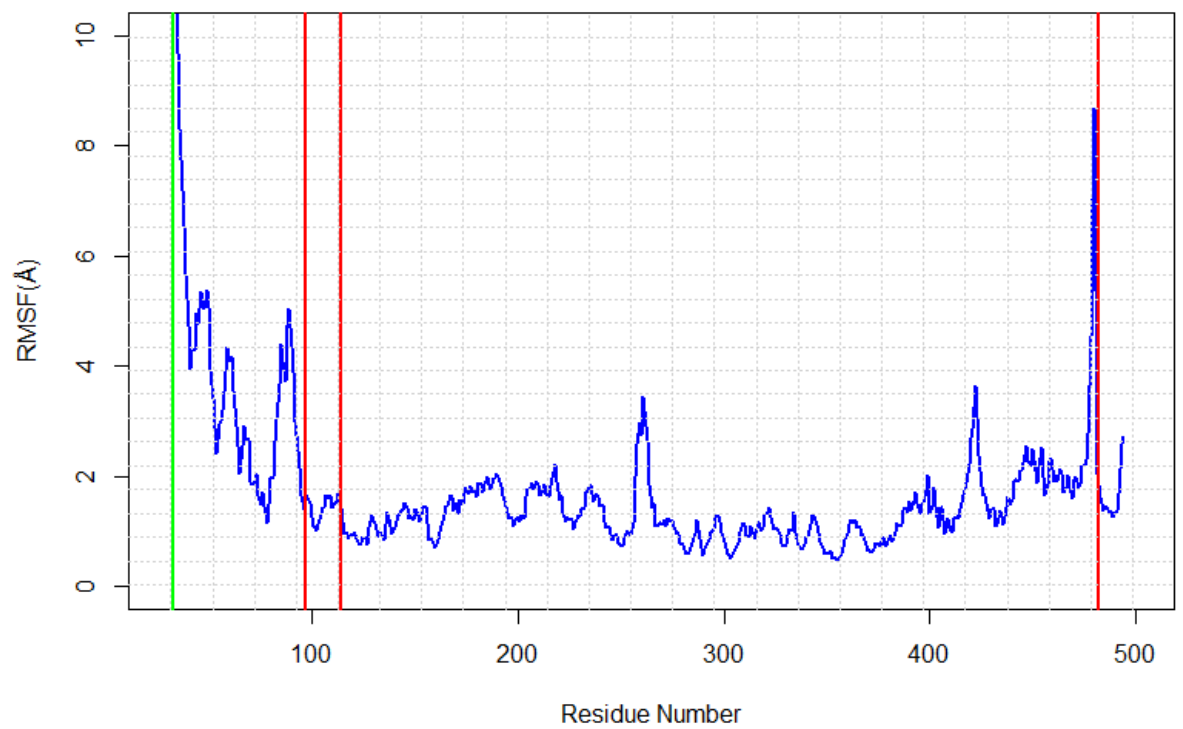
Figure S7.14 The RMSF profile of the KIAA1718 protein using Cα using cpptraj.
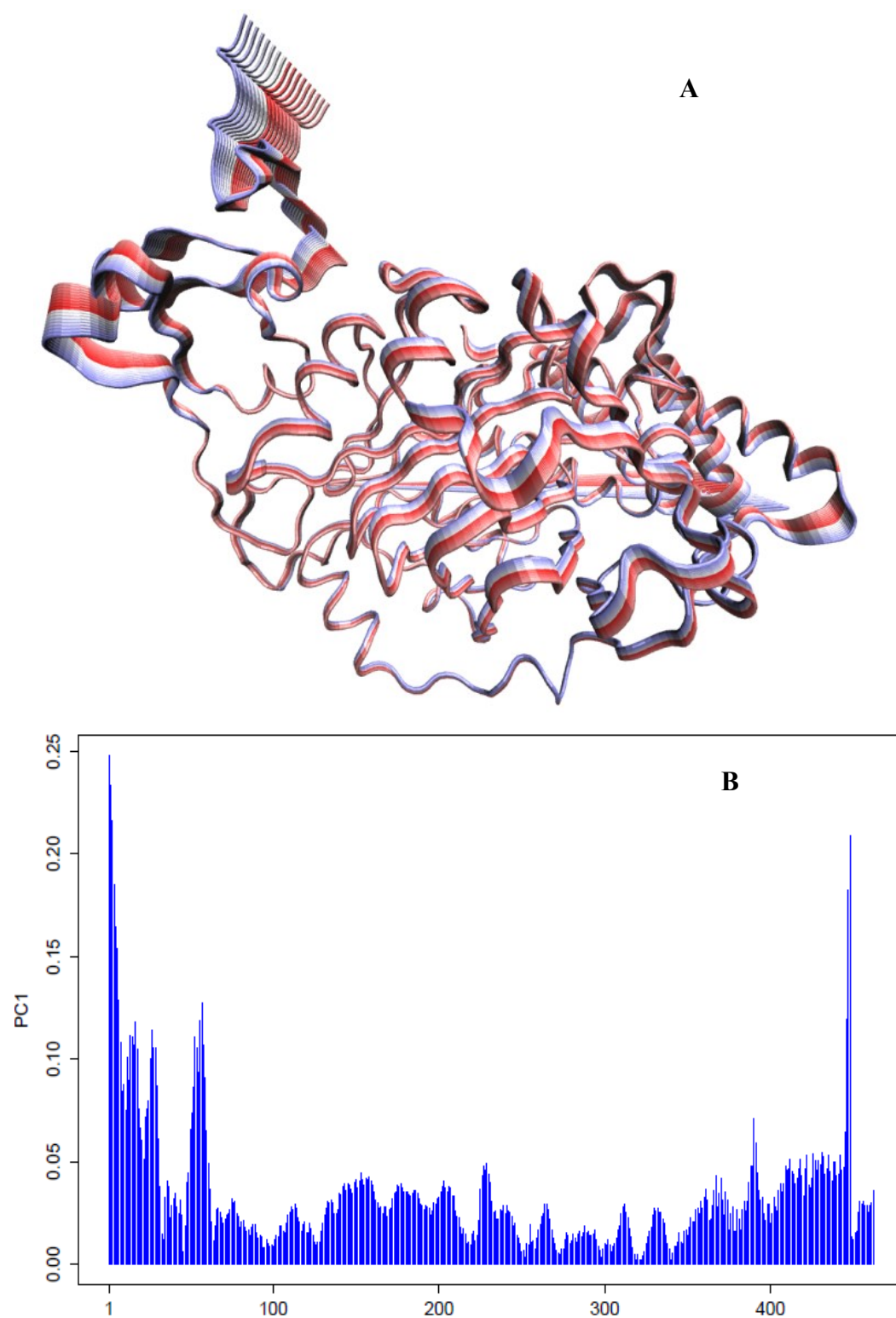
Figure S7.15 (A) The projection of first eigenvector showing the major motion in the PHD domain with respect to the Jumonji domain. (B) The linker residue show very limited

motion. The residue wise loading of the individual residues contribution (Å) of the first PC in the KIAA1718.
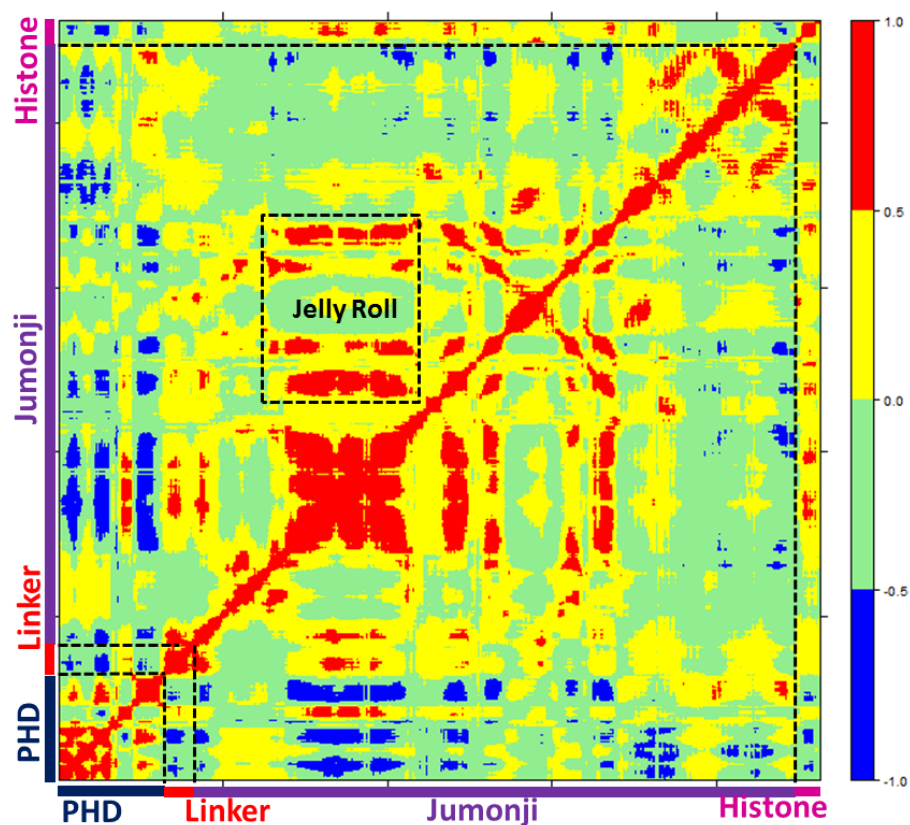
Figure S7.16 The correlated and anti-correlated motion of KIAA1718 performed using Cα atoms.
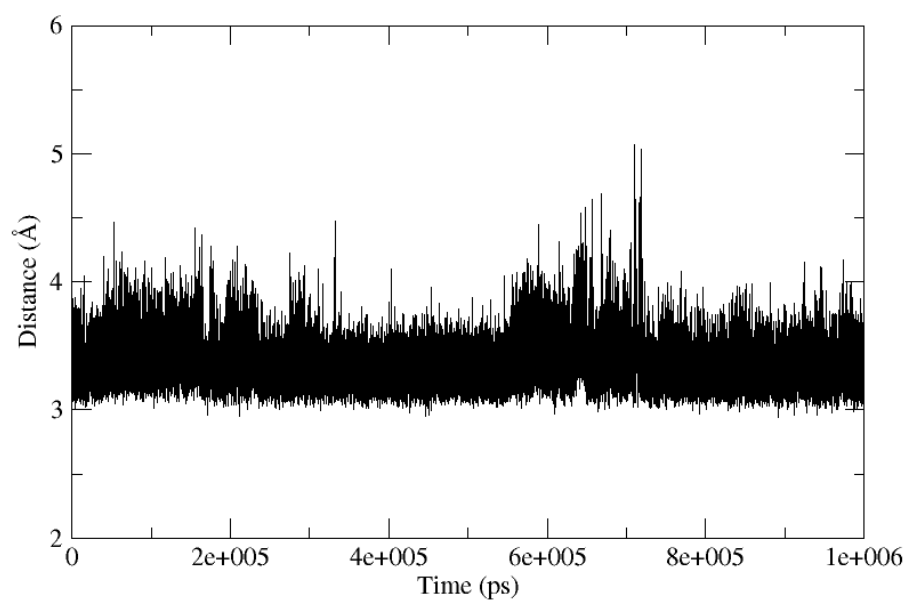
Figure S7.17 The electrostatic interactions of K299 with C5 of 2OG.

Table S7.4 KIAA1718 QM/MM analysis of distance to the non heme Iron (II) of the ligands

Name	Equatorial His	Axial His	ASP	2OG (C1)	2OG (C2)	H2O	di-methylated
QM Cluster minimized	**2.17**	**2.14**	**2.08**	**2.11**	**2.11**	**2.17**	**5.15**
600ns	2.15	2.18	2.05	2.09	2.31	2.19	3.77
700ns	2.15	2.13	2.11	2.13	2.24	2.19	4.51
700ns(EE)	**2.15**	**2.13**	**2.13**	**2.07**	**2.21**	**2.18**	**4.63**
800ns	2.13	2.14	1.98	2.25	2.11	2.20	4.30
900ns	2.12	2.16	2.00	2.19	2.19	2.20	4.26
1000ns	2.17	2.13	1.98	2.09	2.25	2.20	3.76
QM/MM (Mec)(μ*)	**2.14**	**2.14**	**2.04**	**2.13**	**2.21**	**2.19**	4.20
QM/MM (Mec)(σ*)	**0.0176**	**0.0207**	**0.066**	**0.070**	**0.067**	**0.0081**	0.366
MD μ**	**2.29**	**2.15**	**2.03**	**1.927**	**2.13**		**4.582**
MD (σ*)	**0.0904**	**0.075**	**0.065**	**0.060**	**0.115**		**0.818**

Table S7.5 KIA1718 QM/MM analysis of KIA1718 of the second sphere interactions with active site.

Name	N224 (sc) O4 2OG	Y292 (sc) – O3 2OG	T279 (sc) – O4 2OG	K299 (sc) –C5 2OG	M2L (sc) – O1 2OG	N368 – O2 AP1
600ns	4.08	2.63	2.60	3.39	4.00	3.72
700ns	4.03	2.65	2.62	3.31	2.66	3.59
700ns (EE)	3.83	2.74	2.64	3.34	2.78	3.57
800ns	4.58	2.61	3.57	3.13	2.65	4.19
900ns	4.60	2.58	4.03	3.25	2.66	4.68
1000ns	4.23	2.61	4.11	3.24	3.97	4.88
QM/MM (Mec)(μ*)	**4.22**	**2.64**	**3.26**	**3.27**	**3.12**	**4.10**
QM/MM (Mec)(σ*)	**0.310**	**0.056**	**0.726**	**0.091**	**0.672**	**0.572**
MD μ**	**5.45**	**3.61**	**4.84**	**3.35**	**3.058**	**4.50**
MD (σ*)	**0.742**	**1.018**	**0.892**	**0.160**	**0.430**	**0.568**
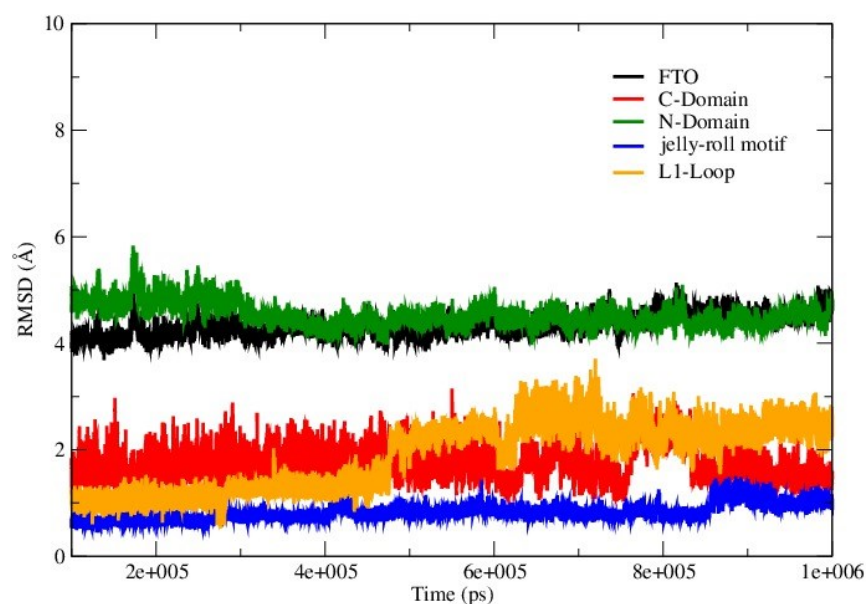
Figure S7.18 The RMSD profile of the FTO enzyme along with its individual domains (N, and C domain), L1 loop and jelly-roll motif. The rms command in ccpptraj was used for

the analysis. The trajectory was saved every 10 ps. The equilibrated trajectory from 100 to 1000 ns was used for the analysis.
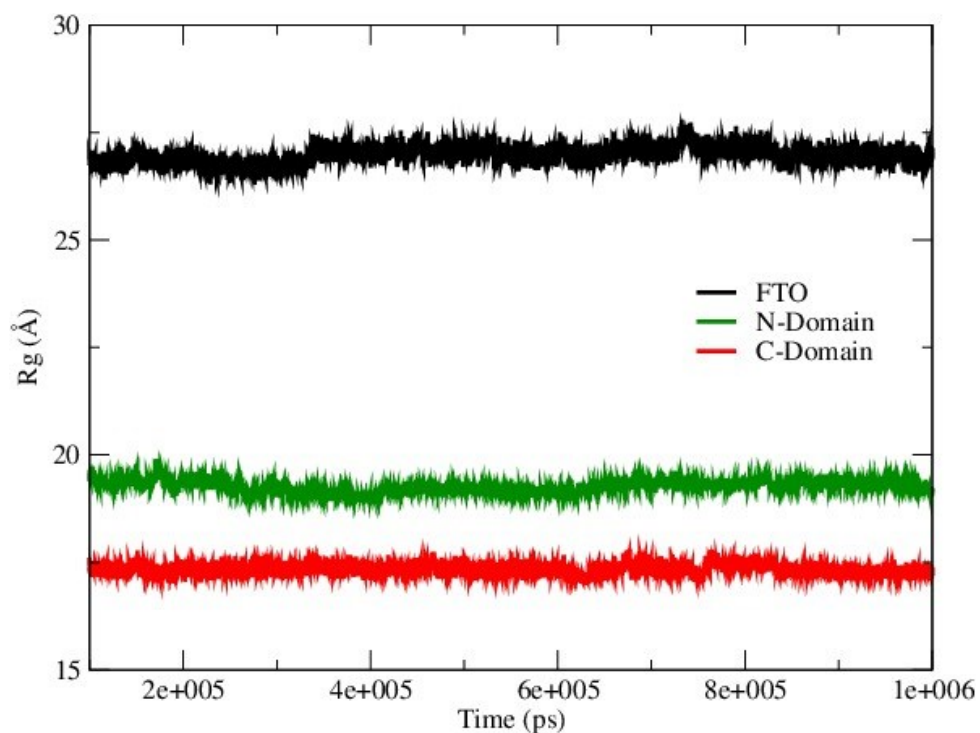
Figure S7.19 The Radius of gyration of the FTO and its domains using Cα using cpptraj
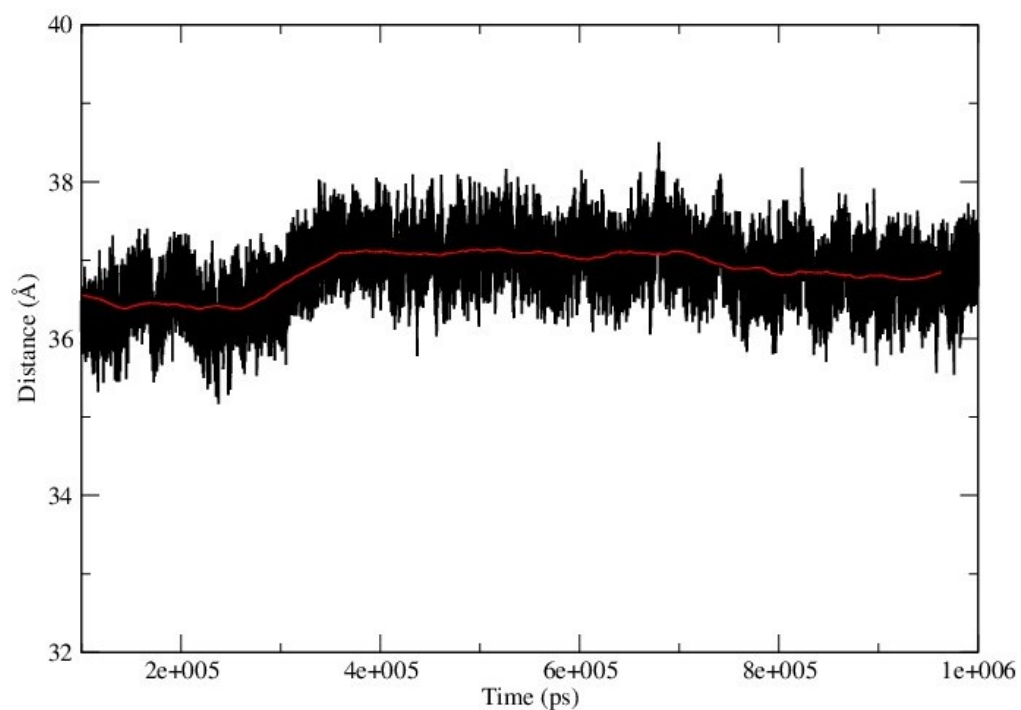
Figure S7.20 The centre of mass between N and C domains using Cα using cpptraj for equilibrated trajectory of FTO.
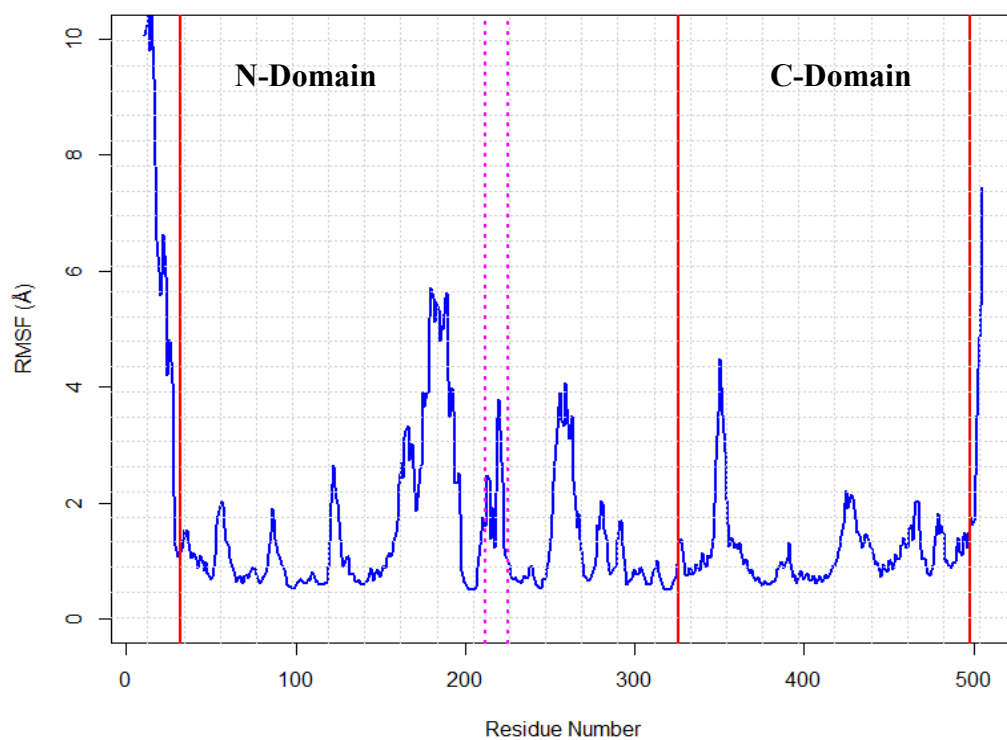
Figure S7.21 The RMSF analysis of the Cα of FTO enzyme. The residue marked with dotted magenta lines belong to the L1 loop.
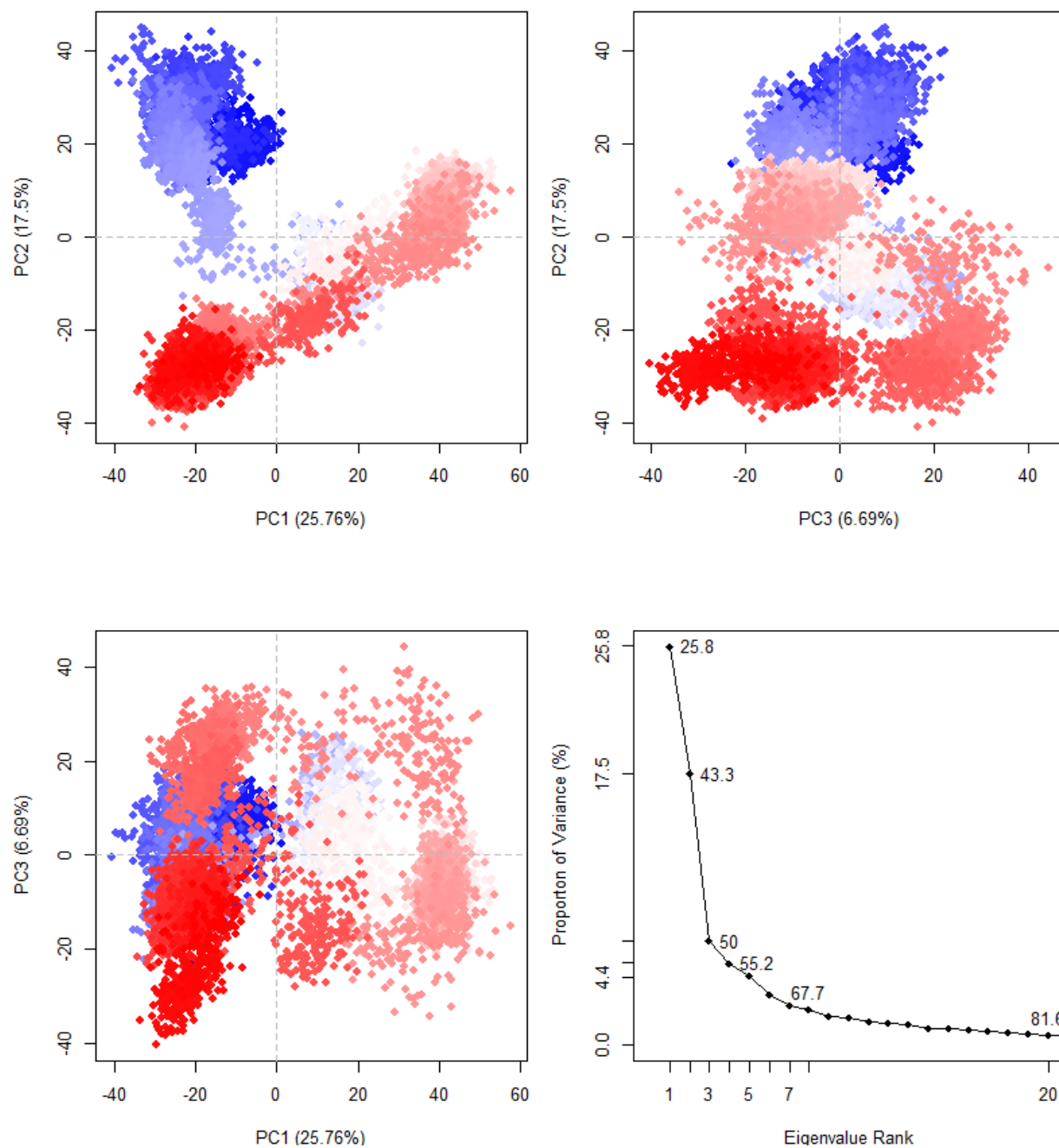
Figure S7.22 The projection of first three principal components for FTO equilibrated trajectory for 100 ns to 1000ns. The Scree plot of the first 20 eigenvectors showing variances. The first three eigenvector representing half of the overall variance in the data set.
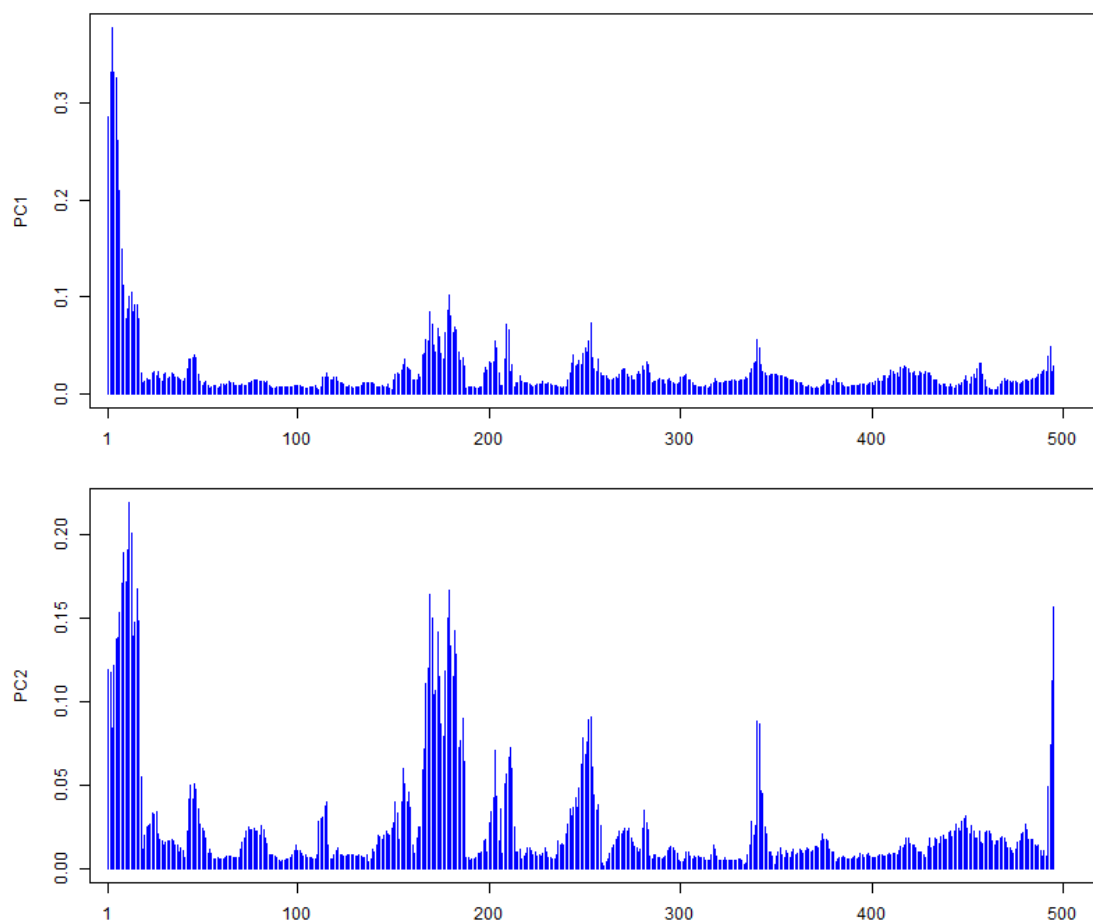
Figure S7.23 The contribution of individual residues towards the PC1 and PC2 in the FTO
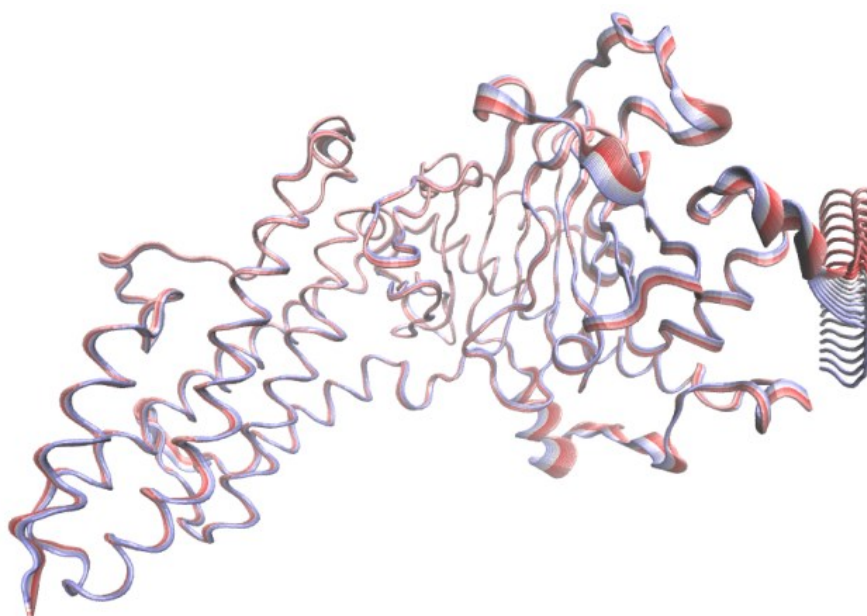
Figure S7.24 The PC1 projection on the FTO. The major motions are seen in the N terminal motion and the L1 loop region of the protein. The residues ranging from 160 to 197 of the N domain of the FTO.

Table S7.6 FTO QM/MM analysis of distance to the non heme Iron (II) of the ligands

Name	Equatorial His	Axial His	ASP	2OG (C1)	2OG (C2)	H2O	Base (methyl)
Minimized	2.17	2.13	2.11	2.11	2.23	2.20	4.67
QM Cluster minimized	2.15	2.13	2.08	2.08	2.15	2.21	5.51
* Minimized (EE)	2.18	2.12	2.13	2.09	2.16	2.16	4.71
Minimized (Lanl2dz-ECP)	2.25	2.20	2.15	2.08	2.32	2.24	4.66
* Minimized (EE) (Lanl2dz-ECP)	2.20	2.16	2.20	2.07	2.40	2.20	4.36
200ns	2.12	2.13	2.13	2.15	2.13	2.17	4.65
300ns	2.12	2.16	2.21	2.05	2.19	2.22	5.01
700ns	2.09	2.17	2.15	2.13	2.15	2.20	5.34
800ns	2.14	2.12	2.14	2.14	2.16	2.16	5.82
QM/MM (Mec)(μ*)	2.13	2.14	2.15	2.12	2.17	2.19	5.1
QM/MM (Mec)(σ*)	0.029	0.022	0.038	0.039	0.038	0.024	0.492
MD μ**	2.18	2.14	1.98	2.15	2.33		4.50
MD (σ*)	0.085	0.078	0.075	0.102	0.142		0.30

EE Electrostatic embedding, rest is all mechanical embedding schemes.

μ* here is only performed for the analysis which has been done using Mechanical embedding scheme under ONIOM calculations.

μ** here is the MD simulation averaged over 1μs trajectory.

σ* here is the standard deviation of the QM/MM minimization.

C5: here is the non-coordinating carboxylate group of the 2OG cofactor in FTO.

O1: The non-coordinating oxygen of C1 of 2OG

Table S7.7 Table FTO QM/MM analysis of FTO of the second sphere interactions with active site

Name	R316 (sc) – C5	Y295 (sc) – C5	S318(sc) –C5	N205 (sc) – C5	N205 (sc) – O1	R96 (sc) – O1	R96 (sc) – 3dt	R322 (sc) – AP1	R322 (sc) – O1
Minimized	4.43	5.31	2.56	2.80	3.47	2.98	2.75	3.14	2.71
* Minimized (EE)	4.38	5.16	2.62	2.86	3.34	2.99	2.78	3.46	2.76
Minimized (Lanl2dz-ECP)	4.44	5.34	2.56	2.80	3.59	2.90	2.81	2.77	2.71
* Minimized (EE) (Lanl2dz-ECP)	4.39	5.31	2.61	2.85	3.43	3.01	2.89	2.87	2.77
200ns	3.96	2.72	2.55	3.72	2.99	2.91	2.75	3.43	3.19
300ns	3.96	2.73	2.56	3.60	2.89	2.76	2.82	2.76	3.85
700ns	3.93	2.68	2.55	3.76	3.96	2.95	2.80	2.74	3.08
800ns	3.94	2.76	2.54	3.74	3.80	5.10	3.80	3.84	3.30
QM/MM (Mec)(μ*)	4.04	3.24	2.55	3.52	3.422	3.34	2.98	3.18	3.22
QM/MM (Mec)(σ*)	0.216	1.157	0.0083	0.409	0.475	0.987	0.457	0.466	0.413
MD μ**	3.97	3.71	3.25	4.55	3.11	2.94	2.89	2.87	4.40
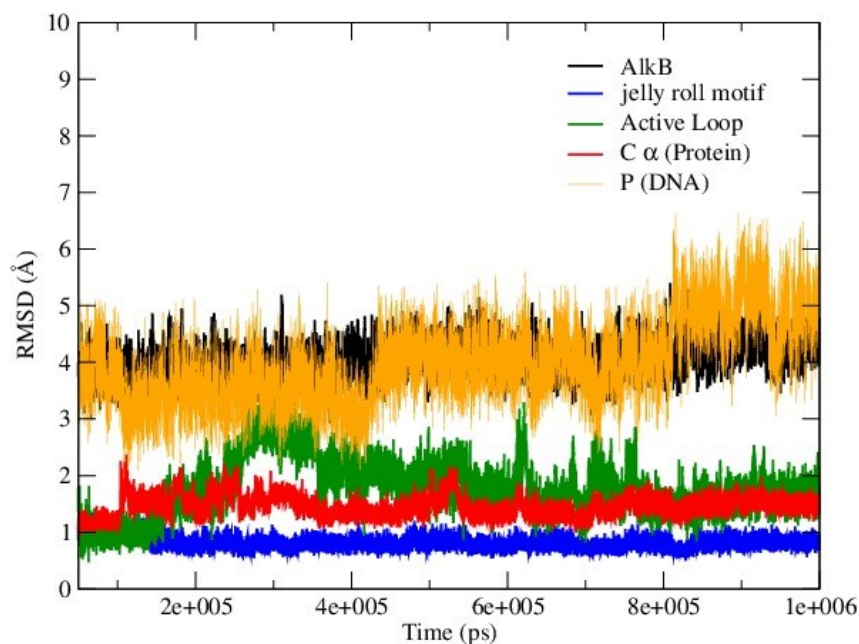
Figure S7.25 The RMSD profile of the AlkB enzyme along with the active loop, jelly-roll motif and DNA. The rms command in ccpptraj was used for the analysis. The Cα and backbone phosphorus atoms were used for the rmsd analysis. The overall rmsd of full complex AlkB was performed using both Cα and P atoms of protein and DNA respectively. The trajectory was saved every 10 ps. The equilibrated trajectory from 50 to 1000 ns was used for the analysis.
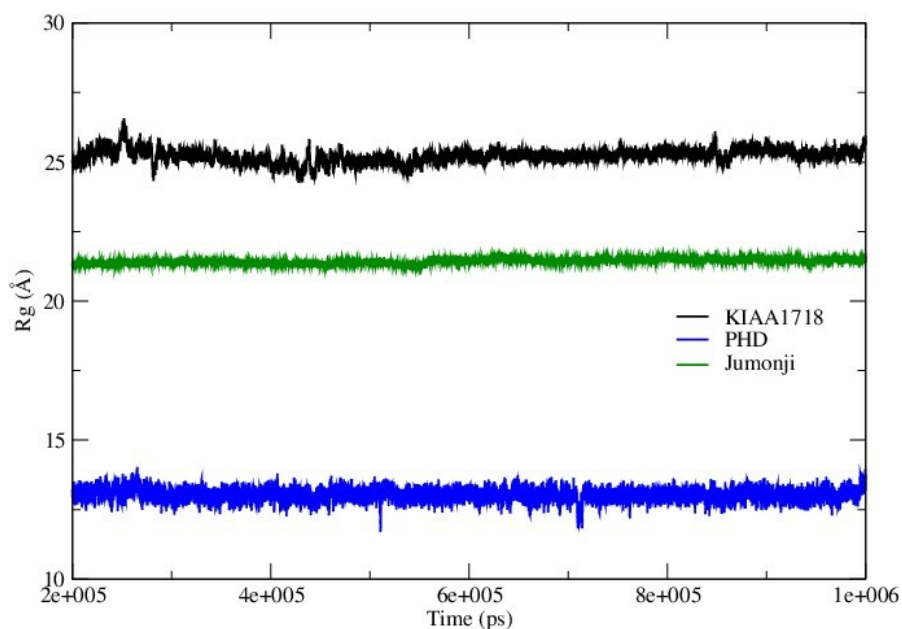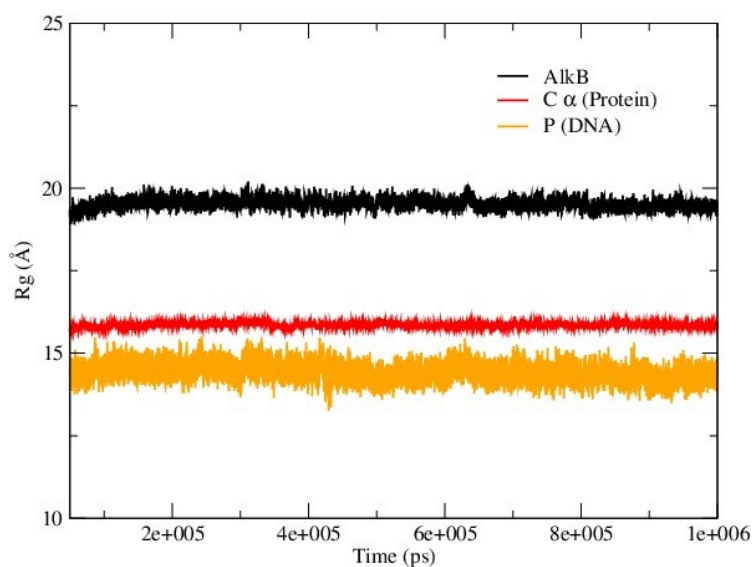

Figure S7.26 The radius of gyration of the AlkB protein. The Cα and P atoms were used for the analysis. The equilibrated trajectory was used.
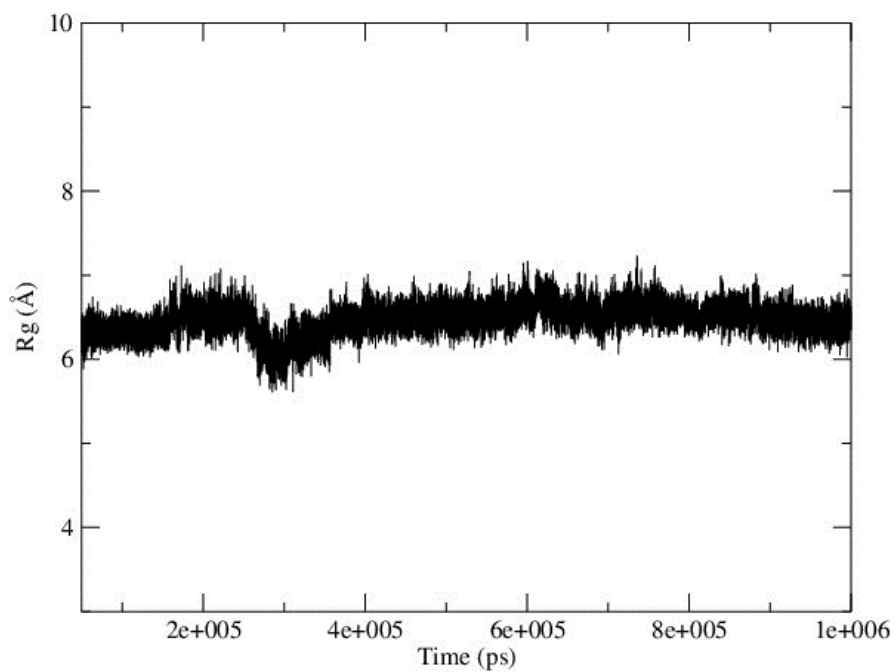
Figure S7.27 The radius of gyration of the active site loop of the AlkB protein using Cα of the protein.
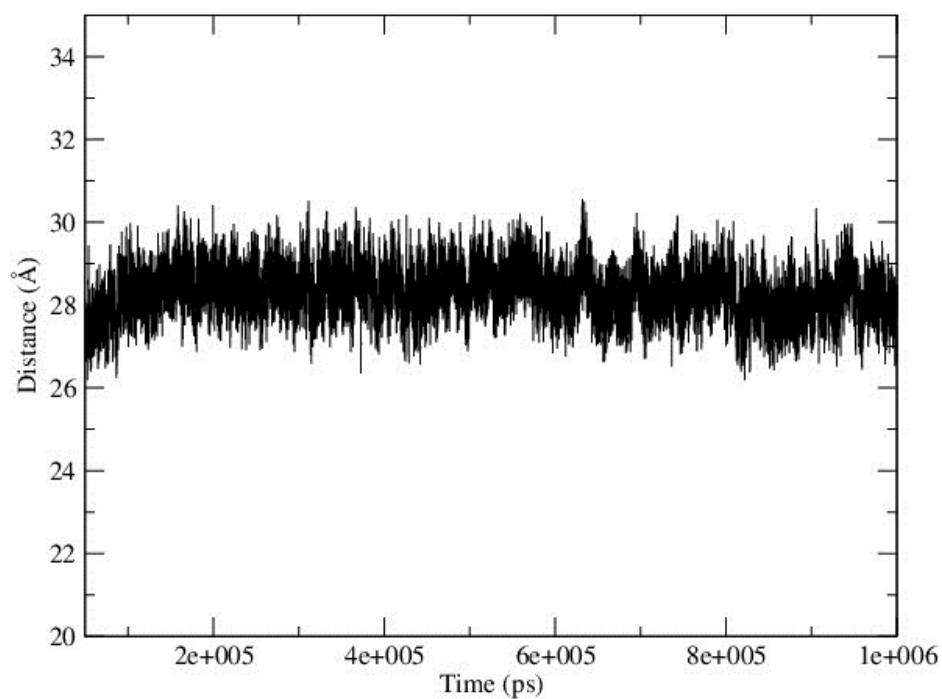
Figure S7.28 The centre of mass of AlkB for protein and DNA molecule.
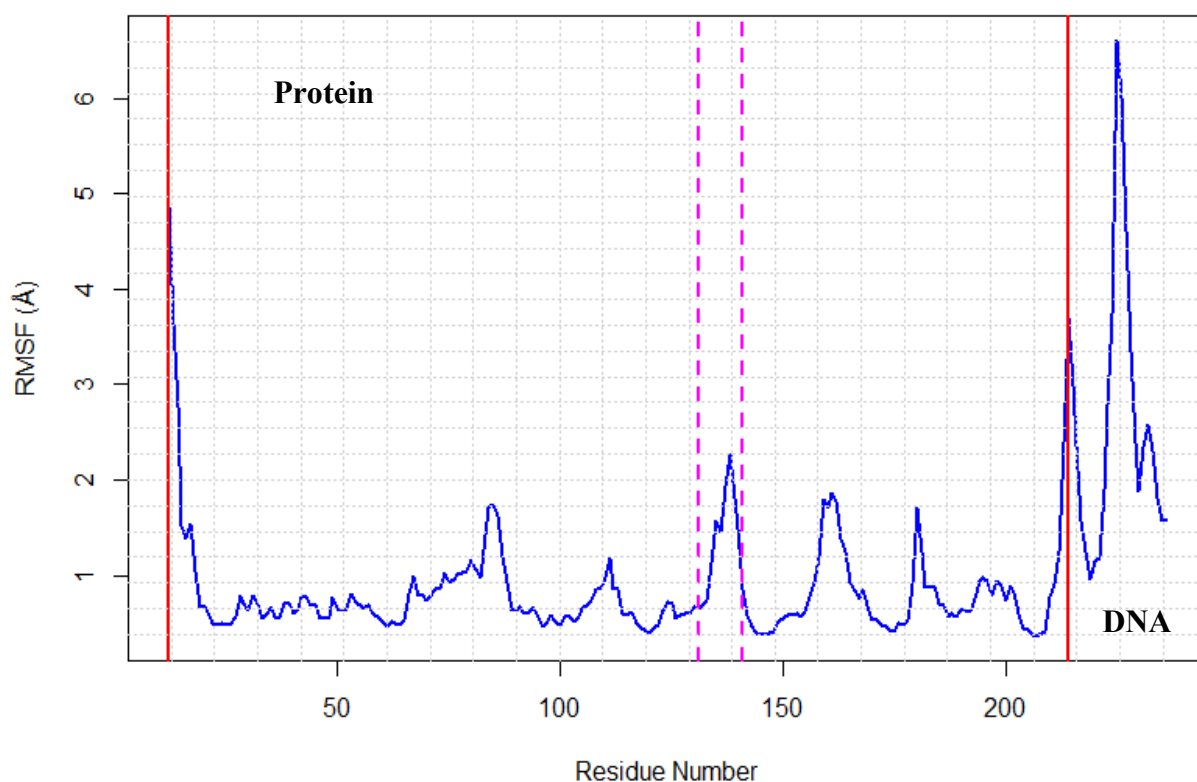
Figure S7.29 The RMSF of the protein and DNA molecule using Cα and P atoms for AlkB for equilibrated trajectory from 50 to 1000ns. The crystal structure residues start from 12-214 for protein and are marked by red lines. The double stranded DNA molecule starts from 2-13. The dotted pink line indicates the active site loop (133-140) in the crystal structure.
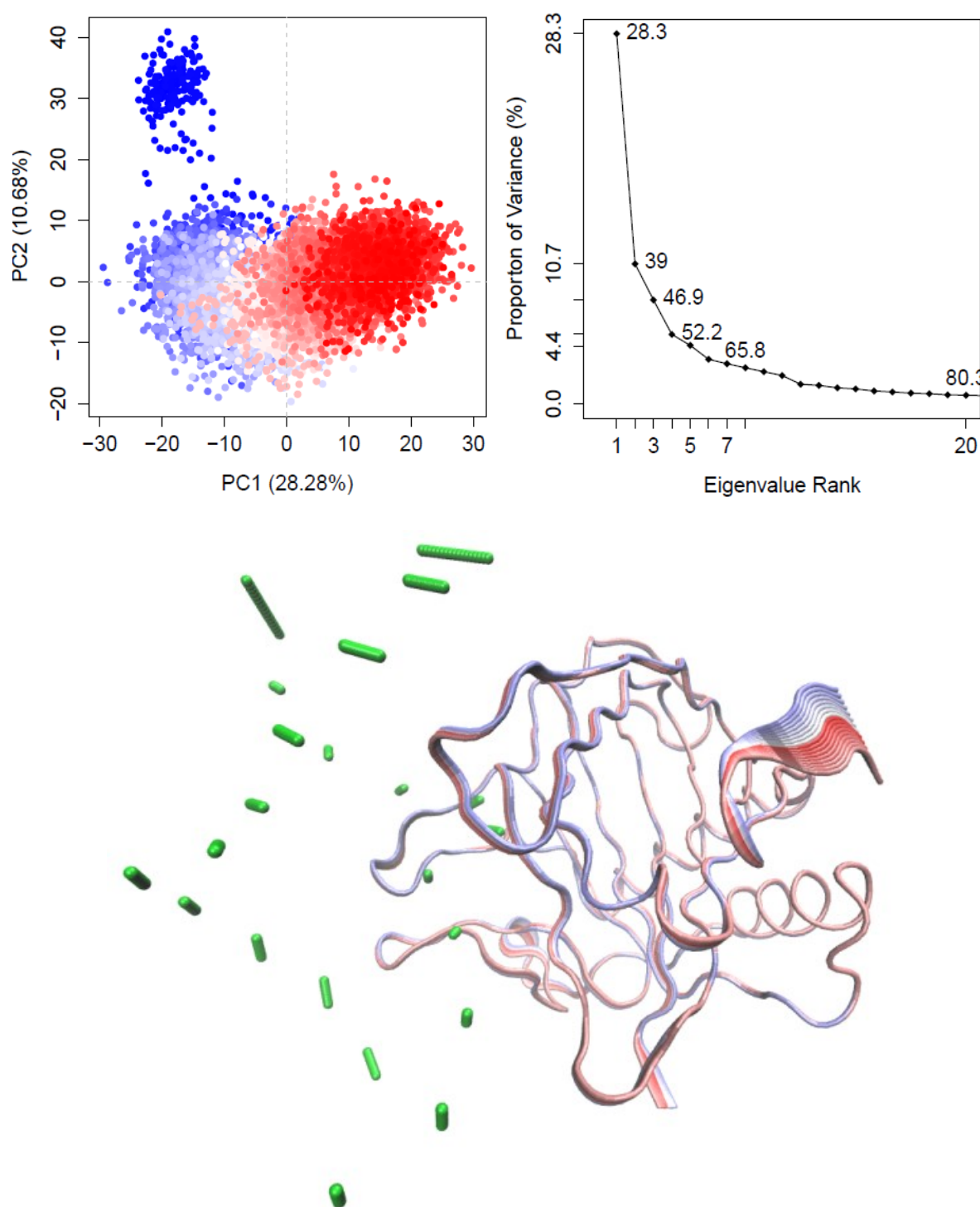
Figure S7.30 The PCA analysis of the AlkB enzyme (PC1 vsPC2, Scree plot and projection of PC1 on to enzyme indicating the major motions in the protein by removing the noise from translational and rotational data from the MD trajectory).
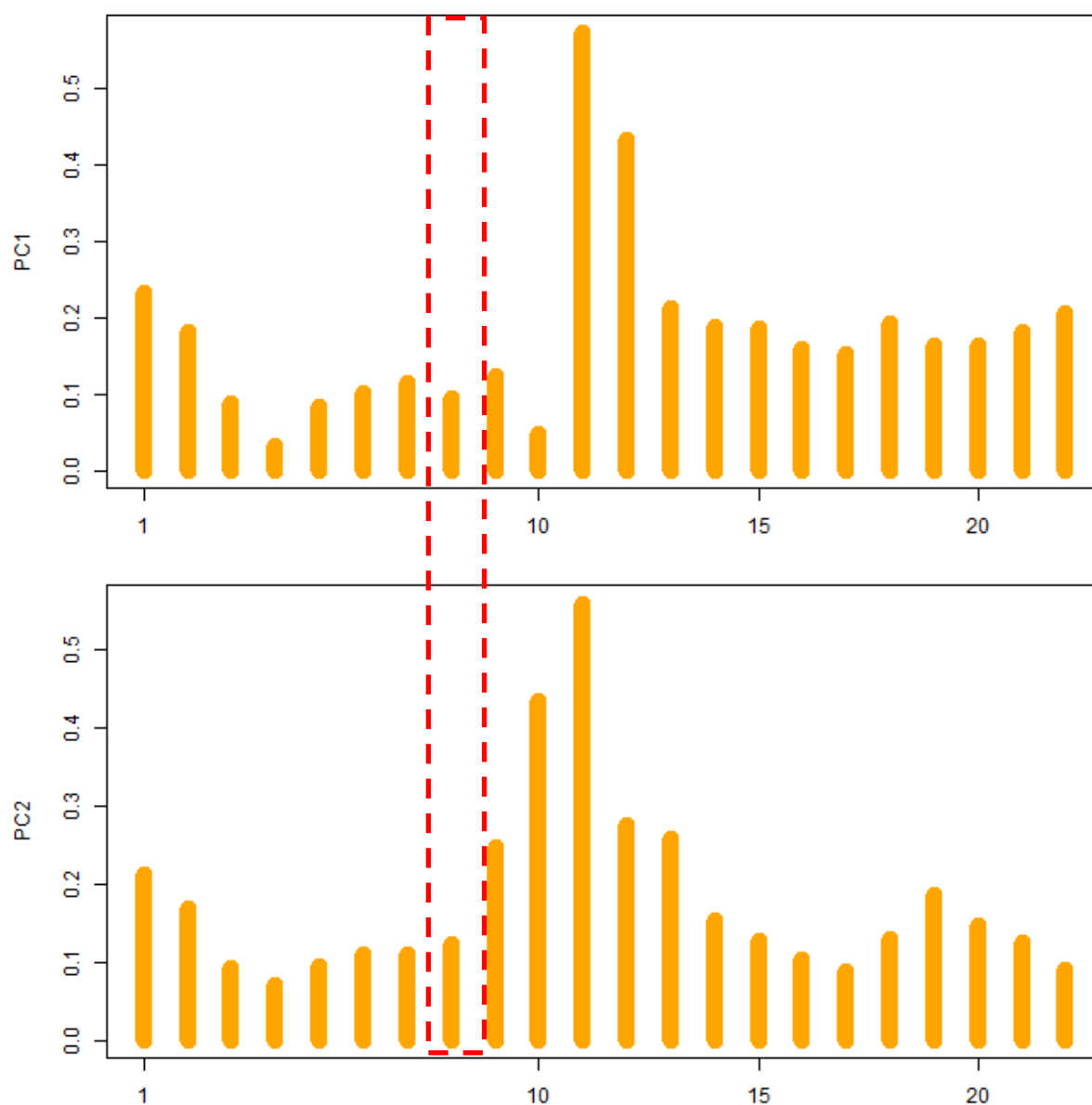
Figure S7.31 The contribution of individual DNA base towards the PC1 and PC2 in the AlkB. The region marked in the red dotted line is methylated DNA baes 6MA.
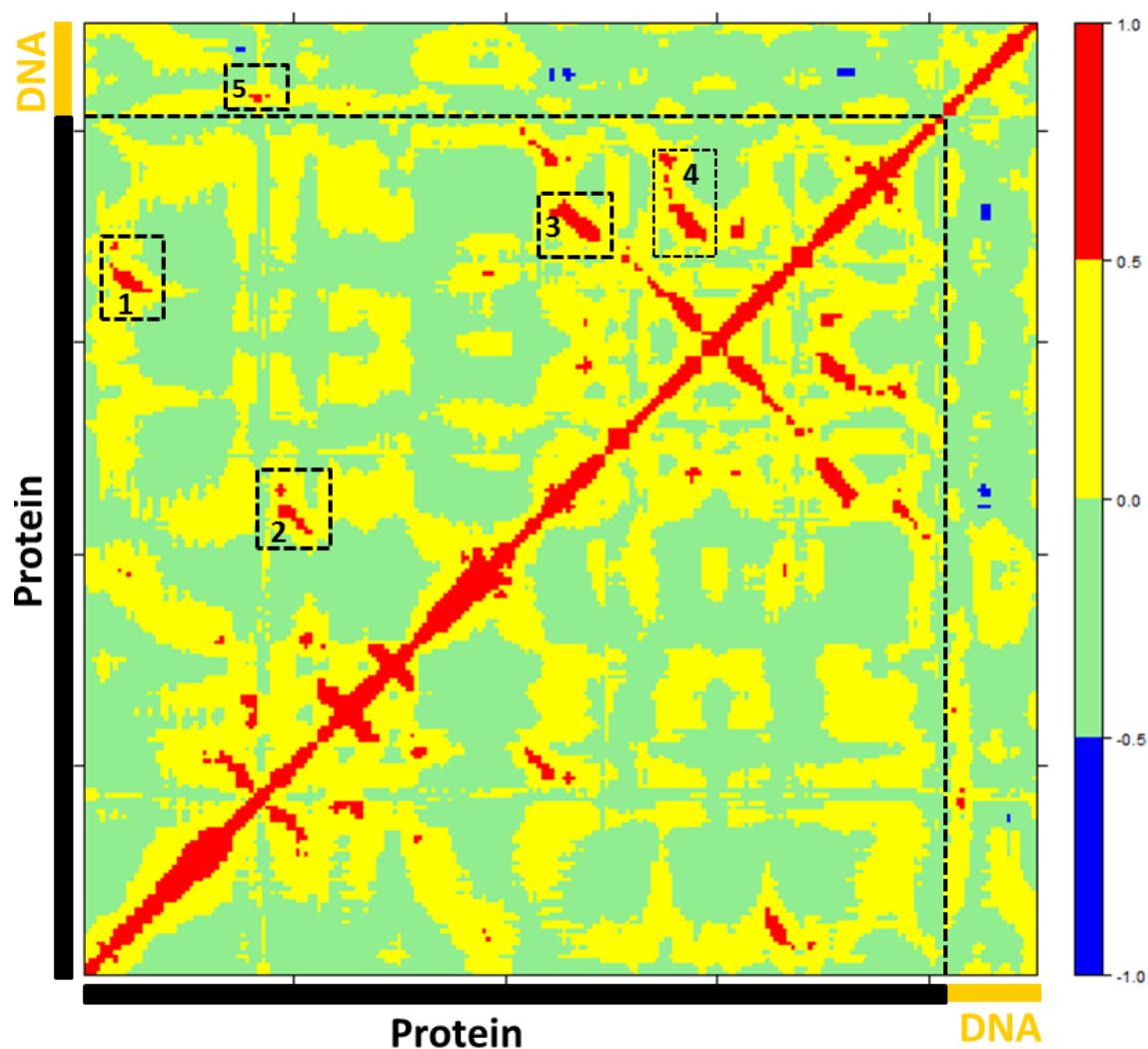
Figure S7.32 The DCCM of the AlkB enzyme along with protein and DNA molecule. (1) residues 17-15 show correlation against 175-180, (2) residues (59-64) show correlation against 115-122, (3) residues 130-137 show correlation against 187-195, (4) residue 152-160 show correlation against 187-195 residues, (5) residue 53-58 show correlation against 6MA residue.

Table S7.8 AlKB QM/MM analysis of distance to the non heme Iron (II) of the ligands

Name	Equatorial His	Axial His	ASP	2OG (C1)	2OG (C2)	H2O	Base (methyl)
QM Cluster minimized	2.13	2.14	2.07	2.08	2.13	2.20	7.31
Minimized (3BIE 1MA)	2.12	2.11	2.05	2.01	2.36	2.12	4.88
200ns (6MA)	2.10	2.13	2.06	2.09	2.27	2.29	7.21
Minimized	2.16	2.18	2.34	2.04	2.29	2.20	5.66
200ns	2.18	2.15	2.15	2.06	2.27	2.23	7.25
400ns	2.19	2.17	2.11	2.04	2.27	2.22	6.69
500ns	2.18	2.18	2.13	2.02	2.27	2.23	8.50
QM/MM (Mec)(μ*)	2.18	2.17	2.18	2.04	2.27	2.22	7.025
QM/MM (Mec)(σ*)	0.0125	0.0141	0.106	0.0163	0.01	0.0141	1.183
MD μ**	2.30	2.13	2.10	2.00	2.19		7.44
MD (σ*)	0.113	0.079	0.109	0.063	0.0967		0.787

Table S7.9 AlkB QM/MM analysis of AlKB (PDB: 4nid, 3bie) of the second sphere interactions with active site.

Name	R204 (sc) – C5	Y122 (sc) – C5-(O3)	N206 (sc)- C5-(O4)	N120 (sc) O1	T208 (sc) – O1	W167 (sc) – AP1 (OD2)	R210(sc) - AP1 (OD2
Minimized (3BIE)	**4.05**	**2.62**	**3.06**	**5.16**	**3.44**	**5.49**	**2.85**
200ns – (6MA-base only)	**4.01**	**2.59**	**3.25**	**3.10**	**4.00**	**6.78**	**5.53**
Minimized	3.96	2.65	2.89	2.87	2.90	2.90	5.67
200ns	3.99	2.69	3.52	2.90	4.08	5.19	2.87
400ns	3.96	2.74	3.58	2.86	3.95	3.97	3.87
500ns	3.99	2.71	3.52	2.82	3.50	5.45	7.17
QM/MM (Mec)(μ*)	**3.97**	**2.70**	**3.37**	**2.86**	**3.60**	**4.37**	**4.89**
QM/MM (Mec)(σ*)	**0.0173**	**0.037**	**0.326**	**0.033**	**0.533**	**1.177**	**1.908**
MD μ**	**4.06**	**3.02**	**3.47**	**2.98**	**4.18**	**6.54**	**7.48**
MD (σ*)	**0.159**	**0.820**	**0.350**	**0.215**	**0.781**	**1.483**	**2.703**

Non Heme Iron (II) parameters using MCPB

MASS

Y1 14.01   0.530   sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y2 16.00   0.434   carboxyl and phosphate group oxygen
Y3 14.01   0.530   sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y4 16.00   0.434   Oxygen with one connected atom
Y5 16.00   0.434   Oxygen with one connected atom
Y6 16.00   0.000   oxygen in TIP3P water
M1 55.85   Fe ion

BOND

Y2-M1 51.3 2.0031   Created by Seminario method using MCPB.py
Y3-M1 33.2 2.1816   Created by Seminario method using MCPB.py
M1-Y6 68.6 2.0608   Created by Seminario method using MCPB.py
Y5-M1 0.0 2.3463   Created by Seminario method using MCPB.py
Y4-M1 15.7 2.1592   Created by Seminario method using MCPB.py
Y1-M1 41.0 2.1534   Created by Seminario method using MCPB.py
Y1-CV 410.0 1.394   JCC,7,(1986),230; HIS
CR-Y1 488.0 1.335   JCC,7,(1986),230; HIS
c -Y5 648.0 1.2140   SOURCE1   36820.0165
CO-Y2 656.0 1.2500
c -Y4 648.0 1.2140   SOURCE1   36820.0165
CR-Y3 488.0 1.335   JCC,7,(1986),230; HIS
Y3-CV 410.0 1.394   JCC,7,(1986),230; HIS
Y6-HO 553.0 0.9572   TIP3P water amber

ANGL

c -Y4-M1 65.16   118.50   Created by Seminario method using MCPB.py
CR-Y1-M1 55.83   129.66   Created by Seminario method using MCPB.py
Y1-M1-Y5 30.49   90.79   Created by Seminario method using MCPB.py
Y1-M1-Y4 42.10   92.32   Created by Seminario method using MCPB.py
Y1-M1-Y2 37.90   92.87   Created by Seminario method using MCPB.py
Y4-M1-Y6 51.18   75.44   Created by Seminario method using MCPB.py
M1-Y6-HO 26.52   101.02   Created by Seminario method using MCPB.py
Y4-M1-Y5 53.55   71.54   Created by Seminario method using MCPB.py
CR-Y3-M1 56.84   121.84   Created by Seminario method using MCPB.py
Y3-M1-Y5 33.79   82.29   Created by Seminario method using MCPB.py
Y2-M1-Y5 32.34   175.71   Created by Seminario method using MCPB.py
Y3-M1-Y6 37.98   98.74   Created by Seminario method using MCPB.py
Y3-M1-Y4 40.98   153.62   Created by Seminario method using MCPB.py
Y2-M1-Y4 49.01   106.06   Created by Seminario method using MCPB.py
M1-Y1-CV 54.04   122.63   Created by Seminario method using MCPB.py
Y2-M1-Y6 36.36   92.47   Created by Seminario method using MCPB.py
CO-Y2-M1 47.81   140.11   Created by Seminario method using MCPB.py
Y5-M1-Y6 36.95   83.50   Created by Seminario method using MCPB.py
Y1-M1-Y3 36.61   91.36   Created by Seminario method using MCPB.py
M1-Y3-CV 57.22   127.77   Created by Seminario method using MCPB.py
Y2-M1-Y3 33.77   99.83   Created by Seminario method using MCPB.py

c -Y5-M1   57.20   111.53   Created by Seminario method using MCPB.py
Y1-M1-Y6   35.15   167.60   Created by Seminario method using MCPB.py
Y1-CV-H4   50.0   120.00   AA his
Y1-CR-H5   50.0   120.00   AA his
CR-Y3-CV   70.0   117.00   AA his
2C-CO-Y2   70.0   117.00
c3-c -Y5   68.03   123.11   SOURCE3   267 3.0977
NA-CR-Y3   70.0   120.00   AA his
CC-CV-Y1   70.0   120.00   AA his
Y3-CR-H5   50.0   120.00   AA his
CR-Y1-CV   70.0   117.00   AA his
Y3-CV-H4   50.0   120.00   AA his
Y4-c -o   78.17   130.38   SOURCE4   429 1.0315
O2-CO-Y2   80.0   126.00
NA-CR-Y1   70.0   120.00   AA his
Y4-c -c   67.16   120.99   SOURCE4   233 2.0333
CC-CV-Y3   70.0   120.00   AA his
c -c -Y5   67.16   120.99   SOURCE4   233 2.0333
HO-Y6-HO   100.   104.52   TIP3P water amber

DIHE
Y1-M1-Y5-c   3   0.00   0.0   3.
Y4-M1-Y1-CV   3   0.00   0.0   3.
c3-c -Y5-M1   3   0.00   0.0   3.
Y3-M1-Y4-c   3   0.00   0.0   3.
c -Y5-M1-Y6   3   0.00   0.0   3.
CR-Y1-M1-Y3   3   0.00   0.0   3.
NA-CR-Y3-M1   3   0.00   0.0   3.
M1-Y1-CR-H5   3   0.00   0.0   3.
CR-Y1-M1-Y6   3   0.00   0.0   3.
CR-Y1-M1-Y5   3   0.00   0.0   3.
CR-Y3-M1-Y4   3   0.00   0.0   3.
M1-Y3-CV-H4   3   0.00   0.0   3.
Y4-M1-Y3-CV   3   0.00   0.0   3.
M1-Y4-c -o   3   0.00   0.0   3.
2C-CO-Y2-M1   3   0.00   0.0   3.
c -Y4-M1-Y6   3   0.00   0.0   3.
Y5-M1-Y3-CV   3   0.00   0.0   3.
NA-CR-Y1-M1   3   0.00   0.0   3.
Y3-M1-Y6-HO   3   0.00   0.0   3.
Y4-M1-Y6-HO   3   0.00   0.0   3.
CR-Y1-M1-Y2   3   0.00   0.0   3.
Y1-M1-Y3-CR   3   0.00   0.0   3.
Y5-M1-Y1-CV   3   0.00   0.0   3.
O2-CO-Y2-M1   3   0.00   0.0   3.
Y2-CO-2C-HC   1   0.000   0.0   2.
Y2-M1-Y6-HO   3   0.00   0.0   3.
Y1-M1-Y4-c   3   0.00   0.0   3.
Y1-M1-Y6-HO   3   0.00   0.0   3.
c -Y5-M1-Y4   3   0.00   0.0   3.
Y3-M1-Y1-CV   3   0.00   0.0   3.

```
CR-Y3-M1-Y6    3    0.00        0.0            3.
Y3-M1-Y5-c     3    0.00        0.0            3.
c -c -Y5-M1     3    0.00        0.0            3.
CO-Y2-M1-Y6    3    0.00        0.0            3.
CX-2C-CO-Y2    1    0.031       180.0         -4.
CX-2C-CO-Y2    1    0.000       0.0           -3.
CX-2C-CO-Y2    1    0.769       180.0         -2.
CX-2C-CO-Y2    1    0.000       0.0            1.
Y6-M1-Y3-CV    3    0.00        0.0            3.
M1-Y1-CV-H4    3    0.00        0.0            3.
M1-Y4-c -c      3    0.00        0.0            3.
Y6-M1-Y1-CV    3    0.00        0.0            3.
Y2-M1-Y3-CV    3    0.00        0.0            3.
Y2-M1-Y3-CR    3    0.00        0.0            3.
Y1-M1-Y3-CV    3    0.00        0.0            3.
c -Y4-M1-Y5     3    0.00        0.0            3.
CC-CV-Y1-M1    3    0.00        0.0            3.
CO-Y2-M1-Y5    3    0.00        0.0            3.
CR-Y1-M1-Y4    3    0.00        0.0            3.
Y1-M1-Y2-CO    3    0.00        0.0            3.
Y5-M1-Y6-HO    3    0.00        0.0            3.
CR-Y3-M1-Y5    3    0.00        0.0            3.
Y2-M1-Y1-CV    3    0.00        0.0            3.
hc-c3-c -Y5     1    0.80        0.0           -1.      Junmei et al, 1999
hc-c3-c -Y5     1    0.08        180.0          3.      Junmei et al, 1999
CC-CV-Y3-M1    3    0.00        0.0            3.
M1-Y3-CR-H5    3    0.00        0.0            3.
Y2-M1-Y4-c     3    0.00        0.0            3.
CO-Y2-M1-Y4    3    0.00        0.0            3.
CO-Y2-M1-Y3    3    0.00        0.0            3.
Y2-M1-Y5-c     3    0.00        0.0            3.
X -CR-Y1-X     2    10.00       180.0          2.      JCC,7,(1986),230
X -CR-Y3-X     2    10.00       180.0          2.      JCC,7,(1986),230
X -CV-Y3-X     2    4.80        180.0          2.      JCC,7,(1986),230
X -CV-Y1-X     2    4.80        180.0          2.      JCC,7,(1986),230


IMPR
X -X -c -Y4          10.5        180.         2.        JCC,7,(1986),230
X -O2-CO-Y2         10.5        180.         2.
X -Y4-c -o           1.1         180.         2.        JCC,7,(1986),230
X -X -c -Y5          10.5        180.         2.        JCC,7,(1986),230


NONB
  Y1      1.8240 0.1700         OPLS
  Y2      1.6612 0.2100         OPLS
  Y3      1.8240 0.1700         OPLS
  Y4      1.6612 0.2100         OPLS
  Y5      1.6612 0.2100         OPLS
  Y6      1.7683 0.1520         TIP3P water model
  M1      1.409  0.0172100000   IOD set from Li et al. JCTC, 2013, 9, 2733
```

CHAPTER EIGHT   CONCLUSION

The MD reveal that the conformational flexibility and dynamics influences the structure-functions relationships of enzymes [1]. Conformational flexibility is of key importance for understanding enzymes interactions with its ligands, understanding the effects of mutations and can be implemented in protein engineering and drug design [206] [207]. In TPST-2 enzyme, the MD studies described the key effects of protein flexibility on the dimerization of TPST-2 and the dynamic nature of the interactions between the apoenzyme, the cofactor PAPS and the substrate peptide C4P5Y3. The study also suggested the more open conformation of the substrate peptide and its stabilization by hydrophobic residues located in vicinity of the active site [16, 158]. The MD of the NirE enzyme provides valuable insight into the enzyme structure-function relationships at atomistic level. The study discriminated between the two candidate arginine residues for the initial proton abstraction step indicating that R*149 is the likely proton abstractor. The binding of the substrate UP2 contributes to the stabilization of the structure of the full enzyme complex [17]. MD simulations of the X-ray crystallographic structure of MMP-1 [18] revealed that conformational changes exercise an important role and influence the interactions between the enzyme HPX and CAT domains and revealed the existence of extensive correlated motions [19]. The interactions between the THP chains and the enzyme also have a flexible nature. Importantly, the study confirm prior NMR experimental results [131] that the X-ray crystallographic structure does not represent the functional enzyme-substrate complex. The MD simulations with exchanged linker regions from MMP-14 and MMP-13 reveal the important role of the linker region to influence overall enzyme flexibility, the pattern of the correlated motions between the HPX and CAT domains, and the interactions between the triple-helical substrate and the enzyme [19]. The mutations influence the flexibility not only of the nearby residues around the site, but also have effect on the different closer or distant parts of the enzyme molecule which was seen in TPST-2, NirE

309

and MMP-1 enzymes [17, 19, 101, 158]. The MD studies on the non-heme Fe(II) and 2-oxoglutrate (2OG) dependent dioxygenases provided curtail knowledge about the conformational effect of the linker region in PHF8 and KIAA1718 enzyme which was missing in the crystal structure [174]. The N domain of the FTO enzyme and the L1 loop region showed increased conformational flexibility and dynamics. The MD study also described the nature of the important second sphere interactions which stabilize the Fe (II) active site in KIA1718, PHF8, FTO and AlKB enzyme. The orientation of the histone peptide and DNA bases during MD simulations provided valuable insight into the enzyme catalysis. The QM/MM optimized structure of the reactant complex showed the effect of the conformational flexibility on the active site and its impact on the catalysis.

The results obtained from the MD and QM/MM simulations provide important information regarding the flexibility of the important residues which are involved in catalysis. The geometric and electronic information obtained from the stationary points of the catalytic cycle are vital for the design of new inhibitor and biocatalysts in chemical biology [1, 2] and biotechnology industry [208]. The combination of directed evolution techniques and MD simulations has already produced enzymes with enhanced activity and increased thermal stability [205]. The effect of the mutations in or around the active site in the enzymes further provide atomistic insight into why certain mutants have high or low enzyme activity. The mutation of the residue also provides vital information for the task of improving the enzyme activity towards a broad range of substrates. The combinations of computational and experimental techniques have contributed to both the industrial and scientific community, helping to better understand biological catalysis.

CHAPTER NINE FUTURE WORK

8.1 METADYNAMICS AND FREE ENERGY CALCULATIONS

Aim: (1) To explore the large conformational changes which occur upon binding of doubly and triply methylated histone peptides to PHF8 and KIA using enhancing sampling methods followed by characterization of the free energy of binding of the histone peptide using free energy perturbation methods. (2) Identification of the differences in histone peptide binding in both PHF8 and KIAA1718 enzymes.

Methods: In order to achieve this knowledge we will perform metadynamics[209], free energy perturbation and umbrella sampling simulations [210]. The classical atomistic Molecular Dynamics (MD) simulations will play an important role in the study of the conformational dynamics in the biomolecular systems however this requires a lot of computational time in order to sample the transitions separated by high energy barriers ("rare events"), which are expected to take place on long-timescale; these include processes such as histone binding and unbinding[209]. To overcome this issue, diverse approaches aimed at accelerating conformational transition in the framework of MD simulations have been developed. One promising technique is metadynamics, in this sampling is accelerated by adding an external potential bias as a sum of repulsive gaussians, which act on few degrees of freedom, named collective variables (CVs). The external potential added allows the ligand to escape from local minima (the bound state) and prevents the ligand from revisiting already sampled minima[5]. Metadynamics is a powerful approach as it not only allows us to characterized free-energy minima, but also to detects the metastable minima and transition states[6], which is useful for kinetic studies. To verify the accuracy of the free-energy surface rebuild along well-defined CV with metadynamics, other MD-derived methodology will be applied. In details, free energy calculations [210] will be also carried out using the thermodynamic end states methods and

umbrella sampling, to accurately predict the binding of the histone to the PHF8 and KIAA1718 enzymes.

8.2 PLAN OF THE STUDY

I) Structure selection; The crystal structure of the PHF8 (PDB: 3KV4) [174] in complex with the substrate histone (H3) and the cofactor analogue N-oxalylglycine will be used as an initial structure. The missing linker region (residue 65-79) between the PHD and the Jumonji domain will be modelled using the Modeller software [139]. The crystal structure of KIAA1718 (PDB: 3KV6) [174] complexed with 2-oxoglutarate (2OG) and dioxygen will also be used as an initial structure.

II) Metadynamics MD simulations of both PHF8 and KIAA1718 enzymes will be performed using the GPU-accelerated version of Desmond software. The enzymes will be simulated in order to capture the major conformational changes which would happen during the histone peptide binding events. A free-energy surface of the binding process will be reconstructed along specific CVs able to distinguish the bound and unbound state of histone peptide to PH8 and KIAA1718 enzymes. Identified stationary points will be used as reference structures to perform additional simulations aimed at an obtaining an accurate energetics of this process.

III) Free energy calculations which will be performed in Desmond software to study the energy changes related to histone binding and unbinding, the potential of the mean force will be calculated using umbrella sampling for the histone binding and unbinding events. The free energy analysis will provide valuable information on the metastable states of the histone peptide during its binding and unbinding to the PH8 and KIAA1718 enzymes.

8.3 EXPECTED RESULTS

The outcome of the project will provide us with the following information; (1) binding free energies of the histone peptide to the KIAA1718 and PHF8 enzymes; (2) Observations of large scale conformational motions during the metadynamics dynamics simulations. These

results will be useful in understanding the large conformational motion experienced by these enzymes along with the binding energies of histone peptides. The results obtained from this project will aid the pharmaceutical industry in rational based drug design.

CHAPTER TEN   REFERENCES

1.   Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.
2.   Karplus, M. and G.A. Petsko, *Molecular dynamics simulations in biology.* Nature, 1990. **347**(6294): p. 631-639.
3.   Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
4.   Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Computer Physics Communications, 1995. **91**(1): p. 43-56.
5.   Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
6.   Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* Computer Physics Communications, 1995. **91**(1): p. 1-41.
7.   Case, D., et al., *Amber 2015 (University of California, San Francisco).* There is no corresponding record for this reference, 2015.
8.   Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme.* J Mol Biol, 1976. **103**(2): p. 227-49.
9.   Chung, L.W., et al., *The ONIOM method and its applications.* Chemical reviews, 2015. **115**(12): p. 5678-5796.
10.   Maseras, F. and K. Morokuma, *IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states.* Journal of Computational Chemistry, 1995. **16**(9): p. 1170-1179.
11.   Matsubara, T., et al., *Application of the new "integrated MO+ MM"(IMOMM) method to the organometallic reaction Pt (PR3) 2+ H2 (R= H, Me, t-Bu, and Ph).* The Journal of Physical Chemistry, 1996. **100**(7): p. 2573-2580.
12.   Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09.* Gaussian, Inc, 2009. **Revision D.01**(Wallingford CT).
13.   Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives1.* Journal of Molecular Structure: THEOCHEM, 1999. **461–462**: p. 1-21.
14.   Niehrs, C. and W.B. Huttner, *Purification and characterization of tyrosylprotein sulfotransferase.* The EMBO journal, 1990. **9**(1): p. 35.
15.   Teramoto, T., et al., *Crystal structure of human tyrosylprotein sulfotransferase-2 reveals the mechanism of protein tyrosine sulfation reaction.* Nat. Commun., 2013. **4**: p. 1572.
16.   Singh, W., et al., *Conformational flexibility influences structure-function relationships in tyrosyl protein sulfotransferase-2.* RSC Advances, 2016. **6**(14): p. 11344-11352.

17.   Singh, W., et al., *Conformational Dynamics, Ligand Binding and Effects of Mutations in NirE an S-Adenosyl-L-Methionine Dependent Methyltransferase.* Sci Rep, 2016. **6**: p. 20107.

18.   Manika, S.W., et al., *Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12461-6.

19.   Singh, W., et al., *Importance of Linker Region in Matrix Metalloproteinase-1 Domain Interactions.* RSC Advances, 2016.

20.   Holm, R.H. and E.I. Solomon, *Introduction: Bioinorganic Enzymology II.* Chemical reviews, 2014. **114**(8): p. 4039-4040.

21.   Mulholland, A.J., *Computational enzymology: modelling the mechanisms of biological catalysts.* Biochemical Society Transactions, 2008. **36**(1): p. 22-26.

22.   Rask-Andersen, M., M.S. Almén, and H.B. Schiöth, *Trends in the exploitation of novel drug targets.* Nature reviews Drug discovery, 2011. **10**(8): p. 579-590.

23.   Imming, P., C. Sinning, and A. Meyer, *Drugs, their targets and the nature and number of drug targets.* Nature reviews Drug discovery, 2006. **5**(10): p. 821-834.

24.   Adrio, J.L. and A.L. Demain, *Microbial enzymes: tools for biotechnological processes.* Biomolecules, 2014. **4**(1): p. 117-139.

25.   Lonsdale, R., K.E. Ranaghan, and A.J. Mulholland, *Computational enzymology.* Chemical Communications, 2010. **46**(14): p. 2354-2372.

26.   van der Kamp, M.W. and A.J. Mulholland, *Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology.* Biochemistry, 2013. **52**(16): p. 2708-2728.

27.   Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Computational enzymatic catalysis–clarifying enzymatic mechanisms with the help of computers.* Physical Chemistry Chemical Physics, 2012. **14**(36): p. 12431-12441.

28.   Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-615.

29.   Siegbahn, P.E. and F. Himo, *The quantum chemical cluster approach for modeling enzyme reactions.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. **1**(3): p. 323-336.

30.   Blomberg, M.R., et al., *Quantum chemical studies of mechanisms for metalloenzymes.* Chemical reviews, 2014. **114**(7): p. 3601-3658.

31.   Siegbahn, P.E. and T. Borowski, *Modeling enzymatic reactions involving transition metals.* Accounts of chemical research, 2006. **39**(10): p. 729-738.

32.   Field, M.J., P.A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations.* Journal of Computational Chemistry, 1990. **11**(6): p. 700-733.

33.   Duarte, F., et al., *Recent advances in QM/MM free energy calculations using reference potentials.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2015. **1850**(5): p. 954-965.

34.   van der Kamp, M.W., et al., *Testing high-level QM/MM methods for modeling enzyme reactions: acetyl-CoA deprotonation in citrate synthase.* The Journal of Physical Chemistry B, 2010. **114**(34): p. 11303-11314.

35.   Soriano, A., et al., *Catalysis in glycine N-methyltransferase: Testing the electrostatic stabilization and compression hypothesis.* Biochemistry, 2006. **45**(50): p. 14917-14925.

36.   Christov, C.Z., et al., *Conformational Effects on the pro-S Hydrogen Abstraction Reaction in Cyclooxygenase-1: An Integrated QM/MM and MD Study.* Biophysical journal, 2013. **104**(5): p. L5-L7.

37.   Lodola, A., et al., *Identification of productive inhibitor binding orientation in fatty acid amide hydrolase (FAAH) by QM/MM mechanistic modelling.* Chemical Communications, 2008(2): p. 214-216.

38. Acevedo, O. and W.L. Jorgensen, *Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions.* Accounts of chemical research, 2009. **43**(1): p. 142-151.

39. Schatz, G.C. and M.A. Ratner, *Quantum mechanics in chemistry.* 1993: Courier Corporation.

40. Born, M., *The statistical interpretation of quantum mechanics.* Nobel Lecture, 1954. **11**: p. 1942-1962.

41. Koch, W. and M.C. Holthausen, *A chemist's guide to density functional theory.* 2015: John Wiley & Sons.

42. Lewars, E., *The concept of the potential energy surface.* Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, 2003: p. 9-41.

43. Shapiro, M. and P. Brumer, *Principles of the quantum control of molecular processes.* Principles of the Quantum Control of Molecular Processes, by Moshe Shapiro, Paul Brumer, pp. 250. ISBN 0-471-24184-9. Wiley-VCH, February 2003., 2003: p. 250.

44. Echenique, P. and J.L. Alonso, *A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry.* Molecular Physics, 2007. **105**(23-24): p. 3057-3098.

45. McQuarrie, D.A., *Quantum chemistry.* 2008: University Science Books.

46. Costagliola, S., et al., *Tyrosine sulfation is required for agonist recognition by glycoprotein hormone receptors.* The EMBO journal, 2002. **21**(4): p. 504-513.

47. Holme, T.A., *Review of Pioneers of Quantum Chemistry.* Journal of Chemical Education, 2014. **91**(6): p. 773-775.

48. Møller, C. and M.S. Plesset, *Note on an approximation treatment for many-electron systems.* Physical Review, 1934. **46**(7): p. 618.

49. Friesner, R.A., *Ab initio quantum chemistry: methodology and applications.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6648-6653.

50. Bartlett, R.J. and M. Musiał, *Coupled-cluster theory in quantum chemistry.* Reviews of Modern Physics, 2007. **79**(1): p. 291.

51. Levine, I.N., *Quantum chemistry.* Vol. 6. 2009: Pearson Prentice Hall Upper Saddle River, NJ.

52. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review, 1964. **136**(3B): p. B864-B871.

53. Parr, R.G., *Density functional theory of atoms and molecules,* in *Horizons of Quantum Chemistry.* 1980, Springer. p. 5-15.

54. Eschrig, H., *The fundamentals of density functional theory.* Vol. 32. 1996: Springer.

55. Cohen, A.J., P. Mori-Sánchez, and W. Yang, *Challenges for density functional theory.* Chemical Reviews, 2011. **112**(1): p. 289-320.

56. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.

57. Schuler, L.D., X. Daura, and W.F. Van Gunsteren, *An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase.* Journal of Computational Chemistry, 2001. **22**(11): p. 1205-1218.

58. MacKerell Jr, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins†.* The journal of physical chemistry B, 1998. **102**(18): p. 3586-3616.

59. Kaminski, G.A., et al., *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides.* The Journal of Physical Chemistry B, 2001. **105**(28): p. 6474-6487.

60. Wang, J., et al., *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-74.

61. Vanommeslaeghe, K., E.P. Raman, and A.D. MacKerell Jr, *Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges.* Journal of chemical information and modeling, 2012. **52**(12): p. 3155-3168.

62. Verlet, L., *Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical review, 1967. **159**(1): p. 98.

63. Van Gunsteren, W. and H. Berendsen, *A leap-frog algorithm for stochastic dynamics.* Molecular Simulation, 1988. **1**(3): p. 173-185.

64. Beeman, D., *Some multistep methods for use in molecular dynamics calculations.* Journal of Computational Physics, 1976. **20**(2): p. 130-139.

65. Hansson, T., C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations.* Current opinion in structural biology, 2002. **12**(2): p. 190-196.

66. Senn, H.M. and W. Thiel, *QM/MM methods for biomolecular systems.* Angewandte Chemie International Edition, 2009. **48**(7): p. 1198-1229.

67. Lundberg, M. and K. Morokuma, *Protein environment facilitates O2 binding in non-heme iron enzyme. An insight from ONIOM calculations on isopenicillin N synthase (IPNS).* The Journal of Physical Chemistry B, 2007. **111**(31): p. 9380-9389.

68. Vreven, T., et al., *Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints.* J Comput Chem, 2003. **24**(6): p. 760-9.

69. Hu, H. and W. Yang, *Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes.* Journal of Molecular Structure: THEOCHEM, 2009. **898**(1): p. 17-30.

70. Friesner, R.A. and V. Guallar, *Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis.* Annu. Rev. Phys. Chem., 2005. **56**: p. 389-427.

71. Niehrs, C., J. Stinchcombe, and W. Huttner, *Two membrane-bound forms of tyrosylprotein sulfotransferase as revealed by phase partitioning in Triton X-114.* European journal of cell biology, 1992. **58**(1): p. 35-43.

72. Huttner, W.B., *Tyrosine sulfation and the secretory pathway.* Annu Rev Physiol, 1988. **50**: p. 363-76.

73. Moore, K.L., *The biology and enzymology of protein tyrosine O-sulfation.* J Biol Chem, 2003. **278**(27): p. 24243-6.

74. Kehoe, J.W. and C.R. Bertozzi, *Tyrosine sulfation: a modulator of extracellular protein-protein interactions.* Chem Biol, 2000. **7**(3): p. R57-61.

75. Zhu, J.Z., et al., *Tyrosine sulfation influences the chemokine binding selectivity of peptides derived from chemokine receptor CCR3.* Biochemistry, 2011. **50**(9): p. 1524-1534.

76. Stone, M.J., et al., *Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins.* New biotechnology, 2009. **25**(5): p. 299-317.

77. Choe, H. and M. Farzan, *Chapter 7. Tyrosine sulfation of HIV-1 coreceptors and other chemokine receptors.* Methods Enzymol, 2009. **461**: p. 147-70.

78. Nishimura, Y., T. Wakita, and H. Shimizu, *Tyrosine sulfation of the amino terminus of PSGL-1 is critical for enterovirus 71 infection.* PLoS Pathog, 2010. **6**(11): p. e1001174.

79. Kraemer, S., et al., *Hetero-oligomerization of chemokine receptors: diversity and relevance for function.* Current medicinal chemistry, 2013. **20**(20): p. 2524-2536.

80. Zhu, L., Q. Zhao, and B. Wu, *Structure-based studies of chemokine receptors.* Current opinion in structural biology, 2013. **23**(4): p. 539-546.

81. Beisswanger, R., et al., *Existence of distinct tyrosylprotein sulfotransferase genes: molecular characterization of tyrosylprotein sulfotransferase-2.* Proceedings of the National Academy of Sciences, 1998. **95**(19): p. 11134-11139.

82. Ouyang, Y.-B. and K.L. Moore, *Molecular cloning and expression of human and mouse tyrosylprotein sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in Caenorhabditis elegans.* Journal of Biological Chemistry, 1998. **273**(38): p. 24770-24774.

83. Teramoto, T., et al., *Crystal structure of human tyrosylprotein sulfotransferase-2 reveals the mechanism of protein tyrosine sulfation reaction.* Nature communications, 2013. **4**: p. 1572.

84. Karabencheva-Christova, T.G., et al., *Conformational effects on the circular dichroism of Human Carbonic Anhydrase II: a multilevel computational study.* PLoS One, 2013. **8**(2): p. e56874.

85. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.

86. Guex, N. and M.C. Peitsch, *SWISS‐MODEL and the Swiss‐Pdb Viewer: an environment for comparative protein modeling.* electrophoresis, 1997. **18**(15): p. 2714-2723.

87. Dennington, R.K., T.; Millam, *GaussView, Version 5.* J. Semichem Inc, 2009. **Shawnee Mission KS**.

88. Vriend, G., *WHAT IF: a molecular modeling and drug design program.* Journal of molecular graphics, 1990. **8**(1): p. 52-56.

89. Schuttelkopf, A.W. and D.M. van Aalten, *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 8): p. 1355-63.

90. Gesteira, T.F., et al., *Insights into the N-sulfation mechanism: molecular dynamics simulations of the N-sulfotransferase domain of ndst1 and mutants.* PloS one, 2013. **8**(8): p. e70880.

91. Fletcher, R. and M.J. Powell, *A rapidly convergent descent method for minimization.* The Computer Journal, 1963. **6**(2): p. 163-168.

92. Hestenes, M.R. and E. Stiefel, *Methods of conjugate gradients for solving linear systems.* 1952.

93. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and Hermans, J. , *Interaction models for water in relation to protein hydration.* Intermolecular forces(ed. B. Pullman), 1981: p. 331—342.

94. McDonald, I.R., *NpT-ensemble Monte Carlo calculations for binary liquid mixtures.* Molecular Physics, 1972. **23**(1): p. 41-58.

95. Ewald, P.P., *Die Berechnung optischer und elektrostatischer Gitterpotentiale.* Annalen der Physik, 1921. **369**(3): p. 253-287.

96. Hess, B., *P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation.* Journal of Chemical Theory and Computation, 2007. **4**(1): p. 116-122.

97. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

98. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures.* Bioinformatics, 2006. **22**(21): p. 2695-2696.

99. Kakuta, Y., et al., *Crystal structure of estrogen sulphotransferase.* Nat Struct Biol, 1997. **4**(11): p. 904-8.

100. Kasahara, K., I. Fukuda, and H. Nakamura, *A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer–DNA Complex.* 2014.

101. Singh, W., et al., *Conformational Flexibility Influences Structure-Function Relationships in TyrosylProtein Sulfotransferase - 2.* RSC Advances, 2016.

102. Warispreet Singh, T.G.K.-C., Gary W. Black, Jon Ainsley, Lynn Dover & Christo Z. Christov, *Conformational Dynamics, Ligand Binding and Effects of Mutations in NirE an S-Adenosyl-L-Methionine Dependent Methyltransferase.* Sci. Rep., 2015.

103. Karabencheva, T.G., et al., *How does conformational flexibility influence key structural features involved in activation of anaplastic lymphoma kinase?* Mol Biosyst, 2014. **10**(6): p. 1490-5.

104. Kruschel, D. and B. Zagrovic, *Conformational averaging in structural biology: issues, challenges and computational solutions.* Molecular BioSystems, 2009. **5**(12): p. 1606-1616.

105. Kormos, B.L., A.M. Baranger, and D.L. Beveridge, *Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A-RNA complex.* J Am Chem Soc, 2006. **128**(28): p. 8992-3.

106. Rüdiger, W., *Chlorophyll metabolism: From outer space down to the molecular level.* Phytochemistry, 1997. **46**(7): p. 1151-1167.

107. Battersby, A.R., *Tetrapyrroles: the pigments of life.* Natural Product Reports, 2000. **17**(6): p. 507-526.

108. Layer, G., et al., *Structure and function of enzymes in heme biosynthesis.* Protein Sci, 2010. **19**(6): p. 1137-61.

109. Bali, S., et al., *Recent advances in the biosynthesis of modified tetrapyrroles: the discovery of an alternative pathway for the formation of heme and heme d 1.* Cell Mol Life Sci, 2014. **71**(15): p. 2837-63.

110. Zajicek, R.S., et al., *d(1) haem biogenesis - assessing the roles of three nir gene products.* Febs j, 2009. **276**(21): p. 6399-411.

111. Baker, S.C., et al., *Cytochrome cd1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to beta-propeller folds.* J Mol Biol, 1997. **269**(3): p. 440-55.

112. Fulop, V., et al., *The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd1.* Cell, 1995. **81**(3): p. 369-77.

113. Carlson, C.A. and J.L. Ingraham, *Comparison of denitrification by Pseudomonas stutzeri, Pseudomonas aeruginosa, and Paracoccus denitrificans.* Applied and Environmental Microbiology, 1983. **45**(4): p. 1247-1253.

114. Zumft, W.G., *Cell biology and molecular basis of denitrification.* Microbiol Mol Biol Rev, 1997. **61**(4): p. 533-616.

115. Hentzer, M., et al., *Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors.* Embo j, 2003. **22**(15): p. 3803-15.

116. Yap-Bondoc, F., et al., *C-methylation occurs during the biosynthesis of heme d1.* J Biol Chem, 1990. **265**(23): p. 13498-500.

117. Storbeck, S., et al., *Crystal structure of the heme d1 biosynthesis enzyme NirE in complex with its substrate reveals new insights into the catalytic mechanism of S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferases.* J Biol Chem, 2011. **286**(30): p. 26754-67.

118. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

119. Blanche, F., et al., *Purification, characterization, and molecular cloning of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from Methanobacterium ivanovii.* J Bacteriol, 1991. **173**(15): p. 4637-45.

120. Blanche, F., et al., *Purification and characterization of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from Pseudomonas denitrificans.* J Bacteriol, 1989. **171**(8): p. 4222-31.

121. DeLano, W.L., *The PyMOL molecular graphics system.* 2002.

122. Ihaka, R. and R. Gentleman, *R: a language for data analysis and graphics.* Journal of computational and graphical statistics, 1996. **5**(3): p. 299-314.

123. Lange, O.F. and H. Grubmüller, *Generalized correlation for biomolecular dynamics.* Proteins: Structure, Function, and Bioinformatics, 2006. **62**(4): p. 1053-1061.

124. Kadler, K.E., et al., *Collagens at a glance.* Journal of Cell Science, 2007. **120**(12): p. 1955-1958.

125. Egeblad, M. and Z. Werb, *New functions for the matrix metalloproteinases in cancer progression.* Nat Rev Cancer, 2002. **2**(3): p. 161-74.

126. Song, F., et al., *Matrix metalloproteinase dependent and independent collagen degradation.* Front Biosci, 2006. **11**: p. 3100-20.

127. Fingleton, B., *Matrix metalloproteinases as valid clinical targets.* Curr Pharm Des, 2007. **13**(3): p. 333-46.

128. Amar, S. and G.B. Fields, *Potential clinical implications of recent matrix metalloproteinase inhibitor design strategies.* Expert Rev Proteomics, 2015. **12**(5): p. 445-7.

129. Page-McCaw, A., A.J. Ewald, and Z. Werb, *Matrix metalloproteinases and the regulation of tissue remodelling.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 221-233.

130. Fields, G.B., *Interstitial collagen catabolism.* J Biol Chem, 2013. **288**(13): p. 8785-93.

131. Bertini, I., et al., *Structural Basis for Matrix Metalloproteinase 1-Catalyzed Collagenolysis.* Journal of the American Chemical Society, 2012. **134**(4): p. 2100-2110.

132. Diaz, N., D. Suarez, and H. Valdes, *From the X-ray compact structure to the elongated form of the full-length MMP-2 enzyme in solution: a molecular dynamics study.* J Am Chem Soc, 2008. **130**(43): p. 14070-1.

133. Diaz, N. and D. Suarez, *Alternative interdomain configurations of the full-length MMP-2 enzyme explored by molecular dynamics simulations.* J Phys Chem B, 2012. **116**(9): p. 2677-86.

134. Lu, K.G. and C.M. Stultz, *Insight into the degradation of type-I collagen fibrils by MMP-8.* J Mol Biol, 2013. **425**(10): p. 1815-25.

135. Lauer-Fields, J.L., et al., *Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity.* J Biol Chem, 2009. **284**(36): p. 24017-24.

136. Arnold, L.H., et al., *The interface between catalytic and hemopexin domains in matrix metalloproteinase-1 conceals a collagen binding exosite.* J Biol Chem, 2011. **286**(52): p. 45073-82.

137. Cerofolini, L., et al., *Examination of matrix metalloproteinase-1 in solution: a preference for the pre-collagenolysis state.* J Biol Chem, 2013. **288**(42): p. 30659-71.

138. Stivala, A., et al., *Automatic generation of protein structure cartoons with Pro-origami.* Bioinformatics, 2011. **27**(23): p. 3315-6.

139. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models.* Methods Enzymol, 2003. **374**: p. 461-91.

140. Hestenes, M. and E. Stiefel, *Methods of Conjugate Gradients for Solving Linear Systems.* J.Res.Nat.Bur.Stand 1952. **49**(6): p. 409-436.

141. Berendsen, H.J.C., et al., *Interaction Models for Water in Relation to Protein Hydration*, in *Intermolecular Forces*, B. Pullman, Editor. 1981, Springer Netherlands. p. 331-342.

142. Chelvanayagam, G., et al., *A combinatorial distance-constraint approach to predicting protein tertiary models from known secondary structure.* Fold Des, 1998. **3**(3): p. 149-60.

143. Diaz, N. and D. Suarez, *Extensive simulations of the full-length matrix metalloproteinase-2 enzyme in a prereactive complex with a collagen triple-helical peptide.* Biochemistry, 2015. **54**(5): p. 1243-58.

144. Azhagiya Singam, E.R., V. Rajapandian, and V. Subramanian, *Molecular dynamics simulation study on the interaction of collagen-like peptides with gelatinase-A (MMP-2).* Biopolymers, 2014. **101**(7): p. 779-94.

145. Suarez, E., N. Diaz, and D. Suarez, *Entropic control of the relative stability of triple-helical collagen peptide models.* J Phys Chem B, 2008. **112**(47): p. 15248-55.

146. Klein, T.E. and C.C. Huang, *Computational investigations of structural changes resulting from point mutations in a collagen-like peptide.* Biopolymers, 1999. **49**(2): p. 167-83.

147. Diaz, N. and D. Suarez, *Peptide hydrolysis catalyzed by matrix metalloproteinase 2: a computational study.* J Phys Chem B, 2008. **112**(28): p. 8412-24.

148. Diaz, N., D. Suarez, and H. Valdes, *Unraveling the molecular structure of the catalytic domain of matrix metalloproteinase-2 in complex with a triple-helical peptide by means of molecular dynamics simulations.* Biochemistry, 2013. **52**(47): p. 8556-69.

149. Diaz, N., D. Suarez, and E. Suarez, *Kinetic and binding effects in peptide substrate selectivity of matrix metalloproteinase-2: Molecular dynamics and QM/MM calculations.* Proteins, 2010. **78**(1): p. 1-11.

150. Maciej J. Stawikowski, G.B.F., *Matrix Metalloproteinases Biology*. 2015, John Wiley & Sons: New York. p. 1-22.

151. Hirose, T., et al., *Structure-function relationship of human neutrophil collagenase: identification of regions responsible for substrate specificity and general proteinase activity.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2569-73.

152. Murphy, G., et al., *The role of the C-terminal domain in collagenase and stromelysin specificity.* J Biol Chem, 1992. **267**(14): p. 9612-8.

153. Chung, L., et al., *Identification of the (183)RWTNNFREY(191) region as a critical segment of matrix metalloproteinase 1 for the expression of collagenolytic activity.* J Biol Chem, 2000. **275**(38): p. 29610-7.

154. Tsukada, H. and T. Pourmotabbed, *Unexpected crucial role of residue 272 in substrate specificity of fibroblast collagenase.* J Biol Chem, 2002. **277**(30): p. 27378-84.

155. Fasciglione, G.F., et al., *The collagenolytic action of MMP-1 is regulated by the interaction between the catalytic domain and the hinge region.* J Biol Inorg Chem, 2012. **17**(4): p. 663-72.

156. Bertini, I., et al., *Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1).* Journal of Biological Chemistry, 2009. **284**(19): p. 12821-12828.

157. Nicolotti, O., et al., *Screening of Matrix Metalloproteinases Available from the Protein Data Bank: Insights into Biological Functions, Domain Organization, and Zinc Binding Groups.* Journal of Chemical Information and Modeling, 2007. **47**(6): p. 2439-2448.

158. Singh, W., et al., *Dimerization and ligand binding in tyrosylprotein sulfotransferase-2 are influenced by molecular motions.* RSC Advances, 2016. **6**(22): p. 18542-18548.

159. Manka, S.W., et al., *Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1.* Proceedings of the National Academy of Sciences, 2012. **109**(31): p. 12461-12466.

160. Zhao, Y., et al., *Transient collagen triple helix binding to a key metalloproteinase in invasion and development.* Structure, 2015. **23**(2): p. 257-69.

161. Mangiatordi, G.F., et al., *Challenging AQP4 druggability for NMO-IgG antibody binding using molecular dynamics and molecular interaction fields.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2015. **1848**(7): p. 1462-1471.

162. Hünenberger, P., A. Mark, and W. Van Gunsteren, *Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations.* Journal of molecular biology, 1995. **252**(4): p. 492-503.

163. Hausinger, R.P., *Fe (II)/α-ketoglutarate-dependent hydroxylases and related enzymes.* Critical reviews in biochemistry and molecular biology, 2004. **39**(1): p. 21-68.

164. Schofield, C.J. and Z. Zhang, *Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes.* Current opinion in structural biology, 1999. **9**(6): p. 722-731.

165. Loenarz, C., et al., *PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nε-dimethyl lysine demethylase.* Human molecular genetics, 2010. **19**(2): p. 217-222.

166. Clifton, I.J., et al., *Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins.* Journal of inorganic biochemistry, 2006. **100**(4): p. 644-669.

167. Solomon, E.I., et al., *Geometric and electronic structure/function correlations in non-heme iron enzymes.* Chemical reviews, 2000. **100**(1): p. 235-350.

168. Solomon, E.I., A. Decker, and N. Lehnert, *Non-heme iron enzymes: contrasts to heme catalysis.* Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 3589-3594.

169. Costas, M., et al., *Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates.* Chemical reviews, 2004. **104**(2): p. 939-986.

170. Wong, S.D., et al., *Elucidation of the Fe (IV)= O intermediate in the catalytic cycle of the halogenase SyrB2.* Nature, 2013. **499**(7458): p. 320-323.

171. Tsai, C.-L., Y. Shi, and J.A. Tainer, *How substrate specificity is imposed on a histone demethylase—lessons from KDM2A.* Genes & development, 2014. **28**(16): p. 1735-1738.

172. Klose, R.J., E.M. Kallin, and Y. Zhang, *JmjC-domain-containing proteins and histone demethylation.* Nature Reviews Genetics, 2006. **7**(9): p. 715-727.

173. Tsukada, Y.-i., et al., *Histone demethylation by a family of JmjC domain-containing proteins.* Nature, 2006. **439**(7078): p. 811-816.

174. Horton, J.R., et al., *Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases.* Nat Struct Mol Biol, 2010. **17**(1): p. 38-43.

175. Tsukada, Y.-i., T. Ishitani, and K.I. Nakayama, *KDM7 is a dual demethylase for histone H3 Lys 9 and Lys 27 and functions in brain development.* Genes & development, 2010. **24**(5): p. 432-437.

176. Qi, H.H., et al., *Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development.* Nature, 2010. **466**(7305): p. 503-507.

177. Cho, J., et al., *Structure and reactivity of a mononuclear non-haem iron (III)–peroxo complex.* Nature, 2011. **478**(7370): p. 502-505.

178. Han, Z., et al., *Crystal structure of the FTO protein reveals basis for its substrate specificity.* Nature, 2010. **464**(7292): p. 1205-1209.

179. Aik, W., et al., *Structural basis for inhibition of the fat mass and obesity associated protein (FTO).* Journal of medicinal chemistry, 2013. **56**(9): p. 3680-3688.

180. Zhu, C. and C. Yi, *Switching Demethylation Activities between AlkB Family RNA/DNA Demethylases through Exchange of Active‐Site Residues.* Angewandte Chemie International Edition, 2014. **53**(14): p. 3659-3662.

181. Robertson, K.D., *DNA methylation and human disease.* Nature Reviews Genetics, 2005. **6**(8): p. 597-610.

182. Fedeles, B.I., et al., *The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond.* Journal of Biological Chemistry, 2015. **290**(34): p. 20734-20742.

183. Yang, C.-G., et al., *Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA.* Nature, 2008. **452**(7190): p. 961-965.

184. Anandakrishnan, R., B. Aguilar, and A.V. Onufriev, *H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations.* Nucleic acids research, 2012. **40**(W1): p. W537-W541.

185. Word, J.M., et al., *Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation.* J Mol Biol, 1999. **285**(4): p. 1735-47.

186. Cornell, W.D., et al., *Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation.* Journal of the American Chemical Society, 1993. **115**(21): p. 9620-9631.

187. Papageorgiou, D. *AMBER parameters for modified amino acids*. 2013 Mar 27, 2013; Available from: http://pc164.materials.uoi.gr/dpapageo/amberparams.php.

188. Cortopassi, W.A., et al., *Dioxygen binding in the active site of histone demethylase JMJD2A and the role of the protein environment.* Chemistry–A European Journal, 2015. **21**(52): p. 18983-18992.

189. Wang, B., et al., *Computations Reveal a Rich Mechanistic Variation of Demethylation of N-Methylated DNA/RNA Nucleotides by FTO.* ACS Catalysis, 2015. **5**(12): p. 7077-7090.

190. Li, P. and K.M. Merz, *MCPB.py: A Python Based Metal Center Parameter Builder.* Journal of Chemical Information and Modeling, 2016. **56**(4): p. 599-604.

191. Pabis, A., et al., *Molecular Dynamics Simulation of Nitrobenzene Dioxygenase Using AMBER Force Field.* Journal of Chemical Theory and Computation, 2014. **10**(6): p. 2246-2254.

192. Peters, M.B., et al., *Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF).* Journal of Chemical Theory and Computation, 2010. **6**(9): p. 2935-2947.

193. Götz, A.W., et al., *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born.* Journal of chemical theory and computation, 2012. **8**(5): p. 1542-1555.

194. Price, D.J. and C.L. Brooks III, *A modified TIP3P water potential for simulation with Ewald summation.* The Journal of chemical physics, 2004. **121**(20): p. 10096-10103.

195. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* The Journal of chemical physics, 2007. **126**(1): p. 014101.

196. Ryckaert, J.-P., G. Ciccotti, and H.J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

197. Roe, D.R. and T.E. Cheatham III, *PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data.* Journal of chemical theory and computation, 2013. **9**(7): p. 3084-3095.

198. Soldatova, A.V., M. Ibrahim, and T.G. Spiro, *Electronic structure and ligand vibrations in FeNO, CoNO, and FeOO porphyrin adducts.* Inorganic chemistry, 2013. **52**(13): p. 7478-7486.

199. Jeziorska, M., et al., *Exact Hartree-Fock exchange in one-dimensional metals. II.* Physical Review B, 1990. **41**(18): p. 12473.

200. Couty, M. and M.B. Hall, *Basis sets for transition metals: Optimized outer p functions.* Journal of computational chemistry, 1996. **17**(11): p. 1359-1370.

201. Barone, V. and M. Cossi, *Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model.* The Journal of Physical Chemistry A, 1998. **102**(11): p. 1995-2001.

202. Lu, T. and F. Chen, *Multiwfn: a multifunctional wavefunction analyzer.* Journal of Computational Chemistry, 2012. **33**(5): p. 580-592.

203. Tao, P. and H.B. Schlegel, *A toolkit to assist ONIOM calculations.* Journal of computational chemistry, 2010. **31**(12): p. 2363-2369.

204. Lindorff‑Larsen, K., et al., *Improved side‑chain torsion potentials for the Amber ff99SB protein force field.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(8): p. 1950-1958.

205. Wolf, A. and K.N. Kirschner, *Principal component and clustering analysis on molecular dynamics data of the ribosomal L11· 23S subdomain.* Journal of molecular modeling, 2013. **19**(2): p. 539-549.

206. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery.* BMC biology, 2011. **9**(1): p. 1.

207. Borhani, D.W. and D.E. Shaw, *The future of molecular dynamics simulations in drug discovery.* Journal of computer-aided molecular design, 2012. **26**(1): p. 15-26.

208. Hospital, A., et al., *Molecular dynamics simulations: advances and applications.* Advances and applications in bioinformatics and chemistry: AABC, 2015. **10**: p. 37.

209. Alessandro, L. and L.G. Francesco, *Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science.* Reports on Progress in Physics, 2008. **71**(12): p. 126601.

210. Lybrand, T.P., *Computer simulation of biomolecular systems using molecular dynamics and free energy perturbation methods.* Reviews in Computational Chemistry, Volume 1, 1990: p. 295-320.

Computational Study of the Structure, the Flexibility, and the Electronic Circular Dichroism of Staurosporine – a Powerful Protein Kinase Inhibitor

Tatyana G. Karabencheva-Christova, Warispreet Singh, and Christo Z. Christov

Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom

Reprint requests to C. Z. C.; E-mail: christo.christov@northumbria.ac.uk

This paper is dedicated to the 75th birthday of Prof. Jörg Fleischhauer.

Staurosporine (STU) is a microbial alkaloid which is an universal kinase inhibitor. In order to understand its mechanism of action it is important to explore its structure-properties relationships. In this paper we provide the results of a computational study of the structure, the chiroptical properties, the conformational flexibility of STU as well as the correlation between the electronic circular dichroism (ECD) spectra and the structure of its complex with anaplastic lymphoma kinase.

Key words: Staurosporine; Computational Chemistry; Electronic Circular; Dichroism; Molecular Dynamics; Anaplastic Lymphoma Kinase.

1. Introduction

Staurosporine (STU) is a microbial alkaloid which is a universal kinase inhibitor that has a strong cytotoxic effect on cancer cells [1] and also exhibits antimicrobial, hypotensive activity against yeast and fungi. Moreover, it causes platelet aggregation, inhibition of smooth muscle contraction, and activation of macrophages [2, 3]. STU and its analogues are widespread in nature. The title compound was first isolated from actinomycete *Streptomyces staurosporeus* in 1977 [4] but it also occurs in cyanobacteria and marine invertebrates, including sponges, tunicates, and mollusks [3, 5]. STU belongs to the indolocarbazole family of compounds which is presented in the living organisms only in form of its indolo(2,3-a)carbazoles [6] isomers. STU is a derivative of indolo(2,3-a)pyrrole(3,4-c)carbazole [2, 7]. The biosynthesis of the indolocarbazole ring in STU involves two tryptophan molecules, the sugar moiety from glucose, and methionine [8]. The products of several genes participate in the biosynthetic process [3]. The chemical structure of STU (Fig. 1) can be divided into two units:

i) an aminodesoxy sugar which adopts the boat conformation in the solid state, and is linked to the aglycone through a pair of C–N bonds and

ii) a planar indolocarbazole heterocycle [9]. STU was also chemically synthesized [8] and the structure of STU and its absolute configuration (2′S, 3′R 4′R, 6′R) were determined using X-ray crystallography [10]. Anti-tumour properties of STU have extensively been reported in scientific literature over the past few decades [7, 11, 12]. STU is a highly potent and non-specific inhibitor of various forms of kinases [13], and to date about 337 three dimensional X-ray crystal structures of STU and its analogues exist in the RCSB protein data bank (PDB) (www.rcsb.org) [14]. The crystal structures reveal that STU binds with high affinity to adenosine triphosphate (ATP) binding sites which are usually located between nitrogen and carbon terminal domains of kinases. The indolo(2,3-a)carbazole moiety tends to occupy the hydrophobic adenine-binding pocket while the lactamate group might form hydrogen bonds to the backbone of the hinge segment [3]. The sugar moiety tends to adopt a boat-shaped conformation in the solid state and might participate in hydrophobic interactions as well as in hydrogen bonds within the ribose binding site (Fig. 2) [15]. The high number of hydrogen bonds between the methylamino nitrogen of the glycosidic ring with surrounding residues might also contribute to the binding affinity [3]. Thus STU exhibits binding affinity towards a broad spectrum of

PAPER

Conformational flexibility influences structure–function relationships in tyrosyl protein sulfotransferase-2†

Warispreet Singh,[a] Tatyana G. Karabencheva-Christova,*[a] Gary W. Black,[a] Olivier Sparagano[b] and Christo Z. Christov*[a]

Tyrosine sulfation is a very important posttranslational modification of proteins. It is catalyzed by tyrosylprotein sulfotransferase and recently became increasingly important for biomedicine and pharmacy. An important insight about structure–activity relationships of human tyrosylprotein sulfotransferase has been received by elucidating the crystal structure, but there is still no understanding about how conformational flexibility and dynamics which are fundamental protein properties influence structure–function relationships of the enzyme. In order to provide this missing but crucially important knowledge we performed a comprehensive atomistic molecular dynamics study which revealed that (i) the conformational flexibility influences sensitively key structural determinants and interactions between the enzyme, the substrate and the cofactor; (ii) a more open conformation adopted by the substrate for binding in TPST 2; (iii) the mutations of key residues related with catalysis and binding change alter the enzyme structure and influence important interactions between the enzyme, the cofactor and the substrate.

Introduction

Tyrosine sulfated proteins are the product of post translational modification (PTM), where a sulfate group is covalently added to the hydroxyl group of tyrosine residues of the polypeptide chain.[1] The *trans*-Golgi networks in the cell harbor tyrosylprotein sulfotransferase (TPST) which is an enzyme responsible for the tyrosine sulfation reaction.[1,2] The tyrosine sulfation is very well characterized in eukaryotes and it has been predicted that up to 1% of tyrosine residues in the eukaryotic proteome have the potential to be sulfated.[3] Tyrosine sulfation plays a crucial role in protein–protein interactions in the extracellular environment and recently became increasingly important for biomedicine and as a target for drug design.[4–8] For example, the interaction of HIV's gp120 protein with CCR5 in order to get entry to CD4+ T-lymphocytes requires sulfated tyrosine residues.[9] The enterovirus71 virus responsible for neurological diseases in children depends on the tyrosine sulfation of PSGL-1 on leukocytes in order to gain entry to cells and cause infection.[10] In fact, approximately sixty immune system proteins have been shown to contain tyrosine sulfated residues.[11,12] TPSTs catalyze the transfer of a negatively charged sulfate group from the universal sulfate donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to the hydroxyl group of a tyrosine residue of polypeptide to form a tyrosine O4-sulfate ester and adenosine 3′,5′-diphosphate.[1] In humans, two isoforms TPTST-1, and TPST-2, are encoded by the TPST gene have been identified.[13] The molecular weight of TPST-1 (370 residues) and TPST-2 (377 residues) isoforms are 42.2 and 41.9 kDa respectively.[4] TPST has a type II transmembrane topology *i.e.* a short N terminal cytoplasmic domain, 17 residue transmembrane domain, and a luminal catalytic domain.[14] The enzyme has two N-glycosylation sites, that are four cysteine residues on the luminal oriented side of enzyme.[7] The first crystal structure of core domain of human TPST-2 (designated TPST2ΔC18 encompassing from G43 to L359), complex with PAP and the substrate peptide C4P5Y3 was solved at a resolution of 1.9 Å (ref. 15) (Fig. 1A). C4PY3 contains only one tyrosine sulfate acceptor site and consists of six acidic residues, thus giving the substrate an overall charge of minus six.[15] The catalytic domain of TPTS-2 comprises a single α/β motif with a five-stranded parallel β-sheet, flanked on both sides by α helices.[15] The TPST-2 exists as a homodimer and the two subunits of the dimer are designated as protomer A and B.[15] The crystal structure of human TPST-2 reveals the important atomistic details of the enzyme and the ligand binding and possible mechanism, but shows no information about how conformational flexibility and dynamics influence protein structure, its structural

Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. E-mail: christo.christov@northumbria.ac.uk; tatyana.karabencheva-christova@northumbria.ac.uk; Tel: +44 (0)191 243 7964; +44 (0)191 243 4277

[b]Vice Chancellor's Office, Coventry University, Coventry, CV1 5FB, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5ra25365a

Dimerization and ligand binding in tyrosylprotein sulfotransferase-2 are influenced by molecular motions†

Warispreet Singh,[a] Tatyana G. Karabencheva-Christova,*[a] Olivier Sparagano,[b] Gary W. Black,[a] Petar Y. Petrov[c] and Christo Z. Christov[a]

Tyrosylprotein sulfotransferase-2 catalyses important, but a less explored posttranslational modification of proteins. The crystallographic structure of the full complex, containing the enzyme, PAPS cofactor and the peptide substrate (WTFC) revealed important structural and atomistic details about the overall structure, binding sites and interactions in TPST-2, but is not informative about the conformational flexibility, which is a fundamental protein property. In a recent computational study we analyzed the impact of conformational flexibility on key geometric determinants and interactions in WTFC. However, there are still unexplained effects about how the conformational dynamics influences the formation of the catalytically active dimer form of the enzyme and also how the binding of the cofactor and the substrate to the apoenzyme influences the structure and the pattern of correlated motions in the enzyme. In order to provide the missing knowledge we performed a molecular dynamics study on the dimerization mutant W113A, the apoenzyme (APO), enzyme–cofactor complex (ES), and enzyme–substrate complex (ES) and compared the results to the WTFC. We identified new hydrophobic interactions important for the stabilization of the enzyme dimer and provided structural insight about the role of the key mutation W113A for the dimerization in the absence of experimentally-derived structure. In addition we showed that the binding of the substrate and cofactor to the apoenzyme contributes to the stability of the whole active complex, influences the local interactions in the binding site and importantly, affects the pattern of the correlated motions in the entire molecule.

Introduction

Sulfation of tyrosine residues in proteins is a very important post synthetic modification, which is still explored little. The reaction is catalyzed by an enzyme – Tyrosylprotein sulfotransferase (TPST).[1,2] Tyrosine sulfation is important for cell signaling and is also target for drug design.[3,4] Sulfated tyrosine residues are important for the interaction of HIV's gp120 protein with CCR5 in order to get entry to CD4+ T-lymphocytes.[5] Sulfated PSGL-1 is involved in the infection process of enterovirus 71 virus, which is responsible for neurological diseases in children.[6] Sulfated tyrosine residues are found in approximately sixty immune system proteins.[7,8] The sulfation of protein tyrosine residues is carried out as enzyme catalyzed transfer of

negatively charged sulfate group from the sulfate donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to the hydroxyl group of a tyrosine leading to a tyrosine O4-sulfate ester and adenosine 3′,5′-diphosphate.[2] Two isoforms TPTST-1 and TPST-2, exist in humans.[9] TPST has a type II transmembrane topology *i.e.* a short N-terminal cytoplasmic domain, a 17 residue transmembrane domain, and a luminal catalytic domain.[10] The enzyme has two *N*-glycosylation sites, that are four cysteine residues on the luminal oriented side of enzyme.[4] The crystal structure of the catalytic domain of human TPST-2 (denoted as TPST2ΔC18, containing the sequence from G43 to L359) in complex with PAP (3′-phosphoadenosine 5′-phosphate) and the substrate (the peptide C4P5Y3) was solved at a resolution of 1.9 Å[11] (Fig. 1). The substrate peptide contains only one tyrosine residue which is a sulfate acceptor and is surrounded by six acidic residues, which provide the substrate negative charge of minus six.[11] The catalytic domain of TPTS-2, consists of a α/β motif with a five-stranded parallel β-sheet, surrounded by α helices.[11] The TPST-2 exists as a homodimer, which is the catalytically active state of the enzyme, where two subunits of the dimer are designated as protomer A and protomer B[11] and both participate in the formation of the binding site and the catalytically active complex. The alpha helices α2–α4 play an

[a]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. E-mail: christo.christov@northumbria.ac.uk; tatyana.karabencheva-christova@northumbria.ac.uk; Tel: +44 (0)191 243 7964, +44 (0)191 243 4077

[b]Vice Chancellor's Office, Coventry University, Coventry, CV1 5FB, UK

[c]Department of Organic Chemistry, Faculty of Chemistry and Pharmacy, Sofia University "St Kliment Ohridski", Sofia, Bulgaria

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ra01899h

SCIENTIFIC REPORTS

Conformational Dynamics, Ligand Binding and Effects of Mutations in NirE an S-Adenosyl-L-Methionine Dependent Methyltransferase

Warispreet Singh*, Tatyana G. Karabencheva-Christova*, Gary W. Black, Jon Ainsley, Lynn Dover & Christo Z. Christov

Heme d1, a vital tetrapyrrol involved in the denitrification processes is synthesized from its precursor molecule precorrin-2 in a chemical reaction catalysed by an S-adenosyl-L-methionine (SAM) dependent Methyltransferase (NirE). The NirE enzyme catalyses the transfer of a methyl group from the SAM to uroporphyrinogen III and serves as a novel potential drug target for the pharmaceutical industry. An important insight into the structure-activity relationships of NirE has been revealed by elucidating its crystal structure, but there is still no understanding about how conformational flexibility influences structure, cofactor and substrate binding by the enzyme as well as the structural effects of mutations of residues involved in binding and catalysis. In order to provide this missing but very important information we performed a comprehensive atomistic molecular dynamics study which revealed that i) the binding of the substrate contributes to the stabilization of the structure of the full complex; ii) conformational changes influence the orientation of the pyrrole rings in the substrate, iii) more open conformation of enzyme active site to accommodate the substrate as an outcome of conformational motions; and iv) the mutations of binding and active site residues lead to sensitive structural changes which influence binding and catalysis.

Uroporphyrinogen III (uro'gen III or UP2) acts as a common scaffold for the synthesis of diverse tetrapyrroles such as chlorophylls[1], cobalamins, siroheme, phytochromobilin, heme d1, and coenzyme F430[2,3]. Heme d1 is an iron-containing dioxo-isobacteriochlorin which acts as a cofactor for cytochrome cd1 nitrite reductase enzyme[4]. Cytochrome cd1 nitrite reductase is the only enzyme in which heme d1 is a cofactor and where it functions as a site for nitrite reduction[5] i.e. reduction of nitrite to nitric oxide and water[6,7]. This denitrification process is a respiratory mechanism for many bacteria including human pathogen *Pseudomonas aeruginosa*, and thus represents a potential drug target[8,9]. Direct inhibitors targeting specific enzymatic processes of *P. aeroginonsa* have distinct advantages over general antibiotics; they reduce the selection pressure towards antibiotic resistance in bacterial populations and can be used in combination with antibiotics to increase effectiveness and lower dosage requirements. With the rise of more and more antibiotic resistant bacteria it has become imperative to design new therapies that reduce the likelihood of creating multiple antibiotic resistant bacterial strains[10]. The synthesis of heme d1 proceeds via precorrin-2 which is the product of two methyl group transfers to uro'gen III[11]. The methyl groups are transferred from S-adenosyl-L-methionine (SAM) to uro'gen III by SAM-dependent Uro'gen III MethylTransferase (SUMT) called NirE[4,11]. The crystal structure of *P. aeruginosa* NirE in complex with its substrate uro'gen III and the reaction by-product SAH (S-Adenosyl-L-homocysteine) was solved recently[12] (Figs 1 and 2). The NirE enzyme is a homodimer; each monomer consists of two domains A and B which are connected by a shorter linker region of four residues length. The active site pocket of NirE is located between the two domains of each monomer however residues from both monomers contribute to each active site[12]. NirE is subject to substrate inhibition at high concentrations of uro'gen III and product inhibition at high concentrations of SAH[13,14]. The crystal structure describes the detailed binding of UP2 indicating that it is exposed to solvent and

Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.K-C. (email: tatyana.karabencheva-christova@northumbria.ac.uk) or C.Z.C. (email: christo.christov@northumbria.ac.uk)

Importance of the Linker Region in Matrix Metalloproteinase-1 Domain Interactions

Warispreet Singh[a], **Gregg B. Fields**[b,c,*], **Christo Z. Christov**[a,*], and **Tatyana G. Karabencheva-Christova**[a,*]

[a]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom

[b]Department of Chemistry & Biochemistry, Florida Atlantic University, Jupiter, FL 33458, USA

[c]Department of Chemistry, The Scripps Research Institute/Scripps Florida, Jupiter, FL 33458, USA

Abstract

Collagenolysis is catalyzed by enzymes from the matrix metalloproteinase (MMP) family, where one of the most studied is MMP-1. The X-ray crystallographic structure of MMP-1 complexed with a collagen-model triple-helical peptide (THP) provided important atomistic information, but few details on the effects of the conformational flexibility on catalysis. In addition, the role of the linker region between the catalytic (CAT) and hemopexin-like (HPX) domains was not defined. In order to reveal the dynamics and correlations of MMP-1 comprehensive atomistic molecular dynamics simulations of an MMP-1•THP complex was performed. To examine the role of the linker region for MMP-1 function simulations with linker regions from MT1-MMP/MMP-14 and MMP-13 replacing the MMP-1 linker region were performed. The MD studies were in good agreement with the experimental observation that in the MMP-1•THP X-ray crystallographic structure MMP-1 is in a "closed" conformation. MD revealed that the interactions of the THP with the both the CAT and HPX domains of MMP-1 are dynamic in nature, and the linker region of MMP-1 influences the interactions and dynamics of both the CAT and HPX domains and collagen binding to MMP-1.

Introduction

One of the primary components of the extracellular matrix (ECM) is collagen, which is also the most abundant protein in mammals[1]. There are at least 29 different types of collagens that occur in vertebrates. The collagen molecule consists of three polypeptide strands (α chains) that organize themselves in a ropelike triple-helix conformation, stabilized by inter-

Article

Effects of Mutations on Structure–Function Relationships of Matrix Metalloproteinase-1

Warispreet Singh [1], Gregg B. Fields [2,3,*], Christo Z. Christov [1,*] and Tatyana G. Karabencheva-Christova [1,*]

[1]   Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK; warispreet.singh@northumbria.ac.uk

[2]   Department of Chemistry & Biochemistry, Florida Atlantic University, Jupiter, FL 33458, USA

[3]   Department of Chemistry, The Scripps Research Institute/Scripps Florida, Jupiter, FL 33458, USA

*   Correspondence: fieldsg@fau.edu (G.B.F.); christo.christov@northumbria.ac.uk (C.Z.C.); tatyana.karabencheva-christova@northumbria.ac.uk (T.G.K.-C.); Tel.: +1-561-799-8577 (G.B.F.); +44-191-243-7964 (C.Z.C.); +44-191-243-4277 (T.G.K.-C.)

Abstract: Matrix metalloproteinase-1 (MMP-1) is one of the most widely studied enzymes involved in collagen degradation. Mutations of specific residues in the MMP-1 hemopexin-like (HPX) domain have been shown to modulate activity of the MMP-1 catalytic (CAT) domain. In order to reveal the structural and conformational effects of such mutations, a molecular dynamics (MD) study was performed of in silico mutated residues in the X-ray crystallographic structure of MMP-1 complexed with a collagen-model triple-helical peptide (THP). The results indicate an important role of the mutated residues in MMP-1 interactions with the THP and communication between the CAT and the HPX domains. Each mutation has a distinct impact on the correlated motions in the MMP-1•THP. An increased collagenase activity corresponded to the appearance of a unique anti-correlated motion and decreased correlated motions, while decreased collagenase activity corresponded both to increased and decreased anti-correlated motions.

Keywords: matrix metalloproteinase-1; conformational flexibility; molecular dynamics simulations; mutations; correlated motions

---

1. Introduction

The matrix metalloproteinases (MMPs) are key enzymes responsible for modulating the balance of collagen and other extracellular matrix (ECM) components in vertebrates. The catabolism of collagen by MMPs is a prerequisite for normal physiological function of cells in vertebrates [1–4]. MMP-1 has been the subject of a broad range of experimental studies and important conclusions have been drawn about the conformational behavior of MMP-1 domains in solution. The MMP-1 hemopexin-like (HPX) and catalytic (CAT) domains are connected by a linker and exhibit complex conformational motions in order to bind the substrate and perform the chemical reaction of collagenolysis [5–9]. The existence of an equilibrium between the open/extended and closed/collapsed conformation of MMP-1 in solution has been established by experimental methods [8–10].

The first molecular dynamics (MD) simulation of MMP-1 bound to a collagen-model triple-helical peptide (THP) [11] provided insight into the role of the linker for modulating conformational dynamics between the HPX and CAT domains and revealed the dynamic nature of interactions between the THP, HPX domain, and CAT domain. The study confirmed the closed or collapsed state of the MMP-1 X-ray crystallographic structure (PDB 4AUO) [7] characterized by close orientation of the MMP-1 HPX and CAT domains. The MD did not cause the MMP-1 conformation to open but allowed the