IDENTIFICATION OF PTP1B DYNAMICS USING FREQUENCY RESPONSE ANALYSIS

by

Dilek EREN

B.S., Chemical Engineering, Yeditepe University, 2008

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University
2012

ACKNOWLEDGEMENTS

First and foremost, I would like to thank to my thesis supervisor Assist. Prof. Burak Alakent, who has supported me throughout my thesis with his friendship, and knowledge. He was abundantly helpful and has assisted me with his valuable suggestions in numerous ways. I am especially grateful to him for always being available when I needed the most.

I would like to thank to my thesis committee members Türkan Haliloğlu and Teoman Turgut for their participations in the project and for their constructive comments.

I am grateful to all my friends, especially İrem Şen, Harun Ferit Özbakır, Nihan Ünal, Aybike Nurocak and Hatice Aksoy for always being there for me. Special thanks are due to Kerem Uguz for allowing me to study in his laboratory during my thesis.

My sincere thanks to all instructors and assistants in the Chemical Engineering Department of Boğaziçi University who made significant efforts on my education and future career by their guidance, experience, and knowledge.

My deepest gratitude goes to my family for their unflagging love, ever-lasting encouragement, and support. Finally, I want to thank to Ali Emre Hortaçsu and his family for their encouragement whenever I need, endless love, understanding and their help.

ABSTRACT

IDENTIFICATION OF PTP1B DYNAMICS USING FREQUENCY RESPONSE ANALYSIS

Protein tyrosine phosphatase 1B (PTP1B) plays a key role as negative regulator of insulin and leptin signaling. In human PTP1B, a significant conformation change is seen in the WPD loop from open (inactive) to closed (active) conformation, when a ligand is bound to the enzyme. The aim of this study is to see how functionally important distant portions, such as S-loop, are perturbed by disturbances in the active site, and suggest a method which could be utilized to find communication pathways between distant regions in a protein, using frequency response techniques. Therefore, five different periods of TMD simulation cycles (one closing-opening motion of WPD loop) were applied between WPD loop in open (WPD_{open}) and in closed (WPD_{closed}) conformations of PTP1B. TMD potential force is applied on the WPD loop and the R-loop Using Discrete Fourier Transform to filter out the perturbed frequency components from random fluctuations, it was observed that the effects of this continuous periodic excitation of the active site spread throughout the whole protein, manifesting itself with increase in mobility of distant regions. Increase in fluctuations are seen not only in the vicinity of WPD loop and R- loop, but also at some distant parts of PTP1B, not directly in contact with the active site or WPD loop, i.e. α1', α2', pTyr recognition loop, R-loop, L11, WPD loop, α3, P-loop, Q-loop and α6. Moreover, most of the displacements of the residues are practically in the same direction with those of the crystal structures. A surprising result is also obtained in the functionally important region, S-loop. Although there is no mobility increase in S-loop with respect to equilibrium MD simulations, opening/closing of WPD loop makes the first eigenvector of S-loop displacements at the base frequency of oscillations more aligned with that between the crystal structures. This result shows that allostericity is not always manifested as increase in residue mobility, but also in adjustment of the directionality of residue fluctuations. TMD simulations at higher frequencies of WPD loop motions show that amplitude of oscillations of many of the distant regions, and even the active site, decreases, as that would be expected from a linear system.

ÖZET

FREKANS YANITI METODU İLE PTP1B DİNAMİKLERİNİN TANILANMASI

Protein Tirosin Fosfataz 1B (PTP1B) enzimi insülin ve leptin sinyal ileti sisteminde negatif düzenleyici bir görevi olması sebebi ile akıllı bir ilaç tasarımı hedefidir. Ligand bağlanmasıyla PTP1B yapısındaki WPD halkasında açıktan (inaktif) kapalıya (aktif) bir konformasyonel değişiklik görülmüştür. Bu çalışmanın amacı, aktif bölgeye uygulanan bozan etkenlerin, uzaktaki fonksiyonel önemi olan bölgelerde (örneğin S döngüsü) nasıl değişiklikler meydana getirdiğini görmek, ve frekans yanıtı metodu kullanılarak protein içi sinyal iletiminin anlaşılmasına yardımcı olmaktır. Bu amaçla, WPD döngüsünün açık ($WPD_{açık}$) ve kapalı ($WDP_{kapalı}$) durumları arasında farklı periyotlara (WPD döngüsünün bir kez açılıp kapanma hareketi) sahip beş adet hedefli moleküler dinamik (TMD) simülasyonu gerçekleştirilmiştir. Simülasyonlarda kuvvet sadece WPD ve R döngüsündeki atomlara uygulanmıştır. Ayrık Fourier Dönüşümü metodu yardımı ile proteinin hedef alınan frekanslardaki salınımlarını, rastlantısal salınımlarından filtrelemek mümkün olmuştur. Periyodik olarak aktif bölgenin uyarılması etkisini PTP1B'nin birbirine uzak bölgelerinde kendisini salınımlarında artış olarak göstermiştir. Salınımdaki artış sadece WPD döngüsü ve R döngüsü çevresinde değil, PTP1B'de aktif bölge ve WPD döngüsüyle direkt kontakt halinde olamayan birbirine uzak bölgeler arasında da görülmüştür. Örneğin, α1', α2', pTyr tanıma döngüsü, R döngüsü, döngü 11, WPD döngüsü, α3, P döngüsü, Q döngüsü ve α6'da. Dahası, hareketliliğinde artış görülen kalıntıların büyük bir kısmının hareket yönleri kristal yapılarınki ile aynı yönde olduğu görülmüştür. Diğer bir sonuç ise, PTP1B'nin fonksiyonu açısından önemli olan S döngüsünde görülmüştür. S döngüsünün hareketliliğinde MD simülasyonlarına kıyasla bir artış gözlemlenmezken, WPD döngüsünün açılma ve kapanmasıyla S döngüsünün temel salınım frekansındaki birinci öz vektörünün kristal yapılar yönündeki hareketi temsil ettiği görülmüştür. Bu sonuçlar allosterik etkinin sadece kalıntıların hareketliliğini arttırmadığını, aynı zamanda kalıntıları kristal yapıya göre yönlendirdiği gözlemlenmiştir. WPD döngüsünün daha yüksek frekanslı TMD simülasyonlarında, birçok uzak bölgenin, hatta aktif bölgenin bile salınım genişliğinin doğrusal sistemlerde görüldüğü gibi düştüğü görülmüştür.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a_i	Acceleration of particle i
b_0	Minimum energy bond length
F_i	Force acting on particle i
$G(jw)$	Frequency response
$G(s)$	Transfer function
i	The particle
Im	Imaginary number
K_b	Bond force constant
K_φ	Dihedral force constant
K_θ	Angle force constant
m_i	Mass of particle i
n	Multiplicity of the function
ns	Nanosecond
R_C	Cutoff distance
Re	Real number
r^N	$3N$ atomic coordinates
θ	Angle
V	Potential energy of the system
$\tilde{x}_i$	Reference position of the particle i

α	Alpha
$\mathring{A}$	Angstrom
β	Beta
χ	Chi
$^{\circ}$	Degree
ω	Frequency
δ	Phase shift in CHARMM forcefield
Φ	Phase lag in frequency response
ϕ	Dihedral angle phi
ψ	Dihedral angle psi

LIST OF ACRONYMS / ABBREVIATIONS

aaRS	Aminoacyl tRNA synthetase
Ala	Alanine
Arg	Argenine
Asp	Aspartic Acid
atm	Atmosphere
C	Carbon
CHARMM	Chemistry at Harvard Molecular Mechanics
Cys	Cysteine
C_α	Carbon Alpha
DFT	Discrete Fourier Transform
DSPs	Dual-specifity phosphatases
fs	Femtosecond
GHz	GigaHertz
Gln	Glutamine
Glu	Glutamine Acid
Gly	Glycine
H	Hydrogen
His	Histidine
IR	Insulin receptors
IRSs	Insulin receptor substrates
L	Loop
Leu	Leucine

Lys	Lysine
M	Molarity
MATLAB	Matrix Laboratory
MD	Equilibrium Molecular Dynamics
MD_{closed}	Molecular Dynamics Simulation in closed conformation
MD_{open}	Molecular Dynamics Simulation in open conformation
Met	Methionine
MM	Molecular mechanics
MSF	Mean Square Fluctuation
N	Nitrogen
NAMD	Nanoscale Molecular Dynamics
NMR	Nuclear magnetic resonance
ns	Nanosecond
NtrC	Nitrogen regulatory protein
O	Oxygen
PCA	Principal component analysis
PDB	Protein Data Bank
Phe	Phenylalanine
PI3K	Phosphatidylinositol 3-kinase
PME	Particle-mesh Ewald
ps	Picosecond
PTKs	Protein tyrosine kinases
PTP1B	Protein tyrosine phosphatase 1B
PTPs	Protein tyrosine phosphatases

pTyr	Phosphotyrosine
RMSD	Root-mean square deviation
RMSF	Root Mean Square Fluctuation
RPTPs	Transmembrane receptor PTPs
Ser	Serine
TC-PTP	T-cell PTP1B
Thr	Threonine
TMD	Targeted Molecular Dynamics
tRNA	Transfer Ribonucleic acid
Trp	Trytophan
Tyr	Tyrosine
vdW	Van der Waals
VMD	Visual Molecular Dynamics
WPD_{closed}	WPD loop in closed conformation
WPD_{open}	WPD loop in open conformation

1. INTRODUCTION

Reversible protein tyrosine phosphorylation regulated by balanced actions of protein tyrosine kinases (PTKs), which catalyze the addition of phosphate to tyrosine residue, and phosphatases (PTPs), which remove the phosphate group from phosphotyrosine (pTyr), is an important regulatory mechanism in the signal transduction pathways that control fundamental processes such as cell growth, differentiation, and metabolism. Protein tyrosine phosphatase 1B (PTP1B) was the first isolated member of PTP family. Hence, it becomes a model for studying structure, function, and regulation of protein tyrosine phosphatases (PTPs). PTP1B is responsible for the negative regulation of insulin and leptin signaling. In the past ten years, research on this protein has indicated that deficiency of PTP1B enhances insulin sensitivity resulting in improving glycemic control, and PTP1B null mice are resistant to high fat diet-induced obesity. Therefore, PTP1B is an attractive drug target for the treatment of type II diabetes, insulin resistance, and obesity.

Crystallographic studies show that PTP1B can exist in two different conformations. In the unliganded form, the catalytically important WPD loop is in an "open" conformation, i.e. substrate can easily access the binding pocket. When an inhibitor or a substrate binds to PTP1B, WPD loop moves towards the active site and forms a tight binding pocket with the residues in the active site of PTP1B. WPD loop closure is an essential part of the catalytic mechanism of PTP1B. Understanding the mechanism of WPD loop conformational transition is of great importance, since this may give insight into how this conformational change can be manipulated.

The aim of this thesis is to investigate how PTP1B dynamically responds to conformational changes in the WPD loop and propose communication pathways between the active site, and the functionally important distant portions, such as $\alpha 7$ and S-loop. Equilibrium Molecular Dynamics (MD) simulations, named MD_{open} and MD_{closed} simulations, of PTP1B and Targeted Molecular Dynamics (TMD) simulations, named TMD1 to TMD5, at different frequencies (one closing-opening motion of WPD loop) between the WPD_{open} (WPD loop in open conformation) and WPD_{closed} (WPD loop in closed conformation) conformations of PTP1B are performed, and resulting trajectories are

analyzed using Discrete Fourier Transform (DFT). TMD force is applied only on the atoms of both R-loop (Gly110 to Gly120) and the WPD loop (Thr176 to Ser187) between the initial and target structures of PTP1B, and the rest of the protein, on which TMD force is not directly applied, is expected to be perturbed by this periodic input. This helps to investigate the communication mechanism between the ligand-binding region, and the functionally important distant portions.

Initially, structural analysis and residue mobility analysis of the equilibrium simulations were performed. Then, structural analyses of TMD1 simulation were performed to elucidate the effects of the applied periodic force on WPD loop. Significant interactions that help to stabilize the WPD loop in the open and closed conformations were monitored to understand the WPD loop opening/closing mechanism. In the following section, frequency response analysis of WPD loop fluctuations in TMD1 was performed by DFT on MATLAB. It was seen that in addition to the fundamental frequency (0.2 GHz), the first few harmonics have contributions to WPD loop movement. Subsequently, 80 ns TMD1 simulation is divided into different segments, like 10 ns, 20 ns and 40 ns segments, to check the convergence of residue fluctuations at the base and upper harmonic frequencies. In the following section, it was checked whether the residue displacements at 0.2 GHz are in the same direction with those in the crystal structures. Furthermore, it is seen that residue mobility profile is significantly similar at different frequencies of TMD simulations, however as the frequency of the applied force decreases, the total mobility of PTP1B increases. Last, effects of the forcing function frequencies on all TMD simulations were analyzed at the same Fourier coefficients. A summary of the results are included in Conclusions, and followed by Recommendations for future work.

2. PROTEIN TYROSINE PHOSPHATASE

Phosphorylation is the addition of a phosphate group to one of the amino acid side chains of a protein. Protein tyrosine phosphorylation reaction is a reversible mechanism controlled by balanced and opposing actions of protein tyrosine phosthatases (PTPs) and kinases (PTKs) to catalyze phosphorylation and dephosphorylation reactions, respectively (Figure 2.1) [1].

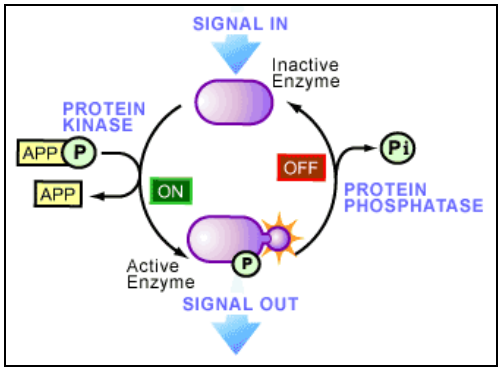

Figure 2.1. Protein kinases and protein phosphatases [1].

Phosphates, which are added to protein, are negatively charged (each phosphate group with two negative charges). Their addition to a protein causes a change in the characteristics of the protein, usually in its conformational structure (Figure 2.2). Due to the dephosphorylation reaction (the removal of the phosphate ion), the protein turns into its original conformation. These two different conformations of PTPs make the protein enzymatically active or inactive. Therefore, it can be said that phosphorylation of the protein acts as a molecular switch [2].

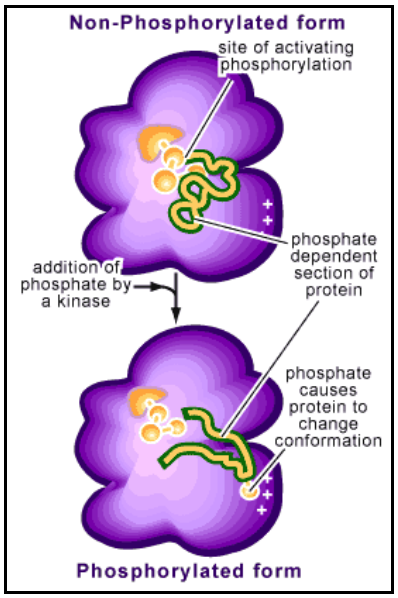

Figure 2.2. Conformational changes of a protein caused by phosphorylation reaction [1].

More than a hundred PTPs are identified in the human genes. In general, PTPs are classified into two main groups: “classical PTPs” which have protein tyrosine specificity and “dual-specifity phosphatases” (DSPs), which also dephosphorylate Ser/Thr residues as well as Tyr residues in proteins [3]. The classical PTPs is also subdivided into two subfamilies; transmembrane receptor PTPs (RPTPs) and non-transmembrane PTPs which are present into the cytoplasm. These two classes of enzymes are characterized by highly conserved catalytic domains consisting of ~250 amino acids with an 11 residue-sequence (I/V)HCXAGXXR(S/T)G, which provides a phosphate binding site (X represents any aminoacid). In that sequence, arginine (R) and cysteine (C) are the catalytically important residues to provide the nucleophilic properties in the catalysis. Diversity within the PTP family arises from the variety of sequence in the noncatalytic domains attached to the -N and -C termini of the catalytic domains [4, 5].

2.1. Protein Tyrosine Phosphatase 1B

Protein tyrosine phosphatase 1B (PTP1B) is the first isolated PTP in the homogeneous form. Therefore, it becomes a model for the studies of PTPs [6].

2.1.1. Biological significance of PTP1B

Reversible protein tyrosine phosphorylation is a significant regulatory mechanism, such as cell growth, proliferation, differentiation, metabolism, and migration, in both eukaryotic and prokaryotic organisms [7]. Abnormalities in this process can cause major diseases, such as type II diabetes and obesity [8].

Researches show that PTP1B is an important drug target for the treatment of diabetes, obesity, and cancer for several reasons: PTP1B is a negative regulator of insulin receptor phosphorylation and signaling; mice lacking PTP1B are resistant to weight gain, have increased sensitivity to insulin, are healthy and fertile, and display no phenotypic abnormalities [9]. In contrast, mice lacking T-cell PTP1B (TC-PTP), sharing 71% amino acid identity and common targets with PTP1B, die after birth due to severe defects in hematopoietic system. These differences between these two proteins are attributed to intrinsic substrate specificity in addition to their different tissue expression patterns and noncatalytic domains [3].

2.1.2. PTP1B and Insulin Signaling

Insulin activates the cells to take up glucose and store it as a glycogen. The role of the PTP1B on the insulin signaling pathway is mediated upstream from phosphatidylinositol 3-kinase (PI3K) and involve dephosphoylation of the insulin receptors (IR) and insulin receptor substrates (IRSs) [10]. Insulin binds to the IR and this cause a conformational change. This conformational change activates the kinase activity of its activation loop. This auto-phosphorylation of the activation loop further causes with other sites to phosphorylate including the IRS1. The phosphorylated IRS1 stimulates the activity of PI3K and phosphoinositide-dependent kinase 1 or 2 which in turn active Akt to promote the glucose uptake and glycogen synthesis (Figure 2.3).

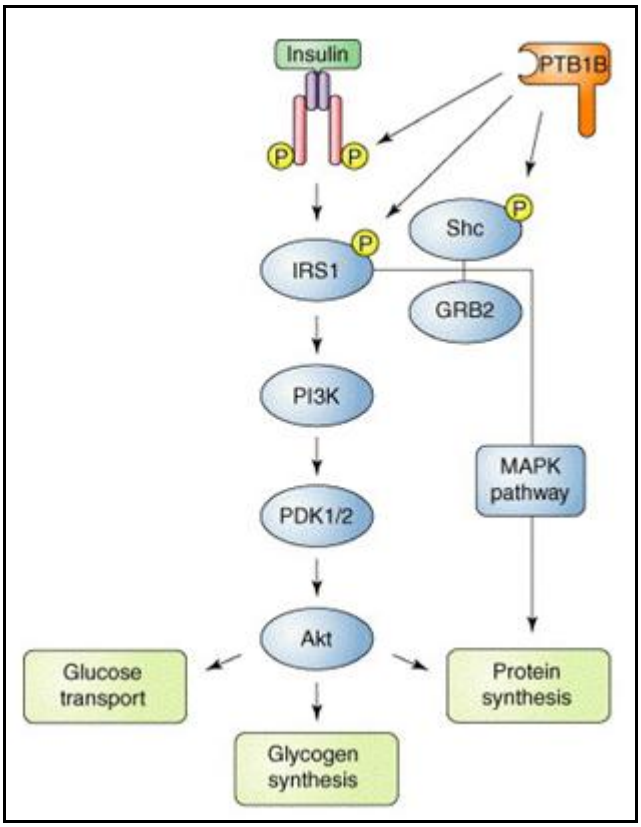

Figure 2.3. Insulin Signaling Pathway [10].

2.1.3. Tertiary Structure of PTP1B

Full length PTP1B is composed of 435 amino acids and a single domain. For biochemical studies, PTP1B structure resolved between residues 1 to 298 is used. Tertiary structure of PTP1B consists of 12-β strands and nine α-helices [11]. Ten of the β-strands (except β5 and β6) form a twisted mixed β-sheet conformation in the center surrounded by α-helices [12]. The last β-strand, β12 is at near the center of the β-sheets region. The center of the β-sheet is covered by α2 helix on one side and α3 and α4 helices on the other (Figure 2.4) [13].

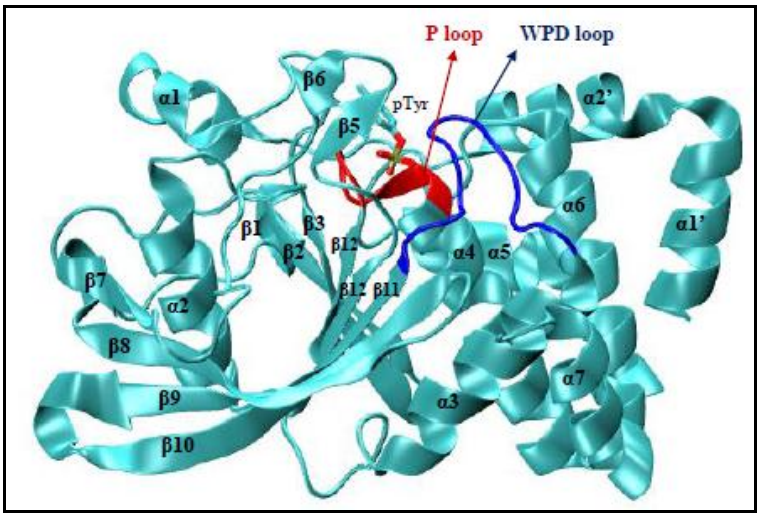

Figure 2.4. Tertiary structure of liganded PTP1B (PDB code:1PTU). WPD and P-loop are shown in blue and red, respectively [14].

The active site of PTP1B is defined with a flexible WPD loop that possesses a catalytically important residue Asp181. Catalytic site or P-loop, which is highly conserved among PTPs, is formed by residues His214 to Ser222 and placed in the deep crevice of PTP1B. P-loop is responsible for providing selectivity for pTyr containing protein substrates with its special sequence motif. Cys215 and Arg221 on the P-loop are significant residues for catalysis [15].

Another conserved loop of PTP1B is the substrate recognition site (residues from Arg43 to Tyr46). Tyr46 on the pTyr loop is responsible for the insertion of the substrate into the catalytic site by interacting with the aromatic ring of the phosphotyrosine in the substrate and determines the depth of the catalytic cleft.

The third conserved loop is WPD loop, which is comprised of the residues between Thr178 to Ser187 (Figure 2.5). The most important and conserved residue on the WPD loop is Asp181 involved in both two steps of catalytic mechanism of PTP1B. It should be noted that WPD loop undergoes a substantial conformational change (from open to closed conformation) with the substrate binding into the active site. Crystallography studies have shown that this conformational change WPD loop of PTP1B is essential for the catalytic mechanism of PTP1B [16]. Open and closed conformations of the WPD loop refers that

the enzyme is in enzymatically inactive and active state, respectively. Ligand free PTP1B alternates in between open and closed conformations of WPD loop [17], and ligand binding affects the residence time of WPD in the closed conformation. Figure 2.6 and Figure 2.7 show the crystal structures of WPD loop in open conformation (PDB IDs: 2F6F), denoted by WPD_{open}, and WPD loop in closed conformation (PDB IDs: 1SUG), denoted by WPD_{closed}, aligned with respect to C_α atoms between Glu2 and Asp298.

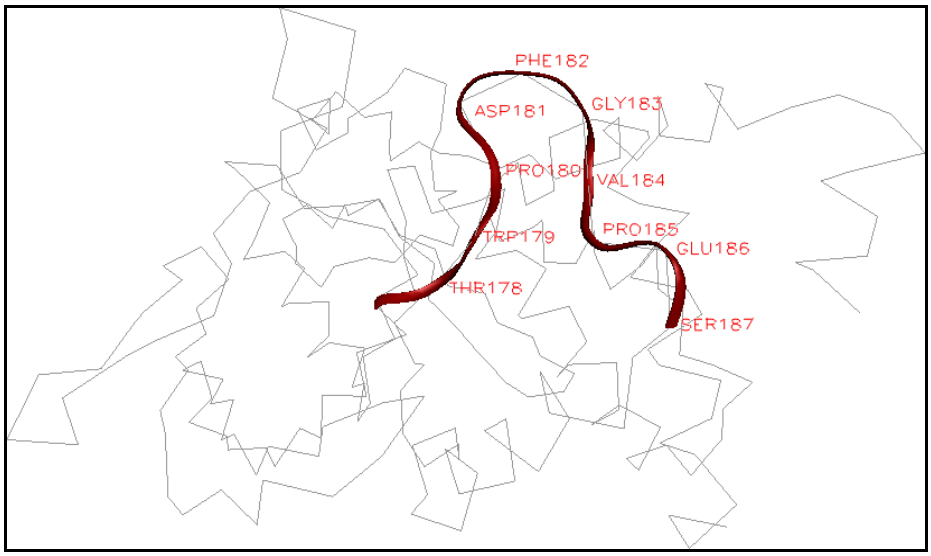

Figure 2.5. WPD loop in PTP1B.

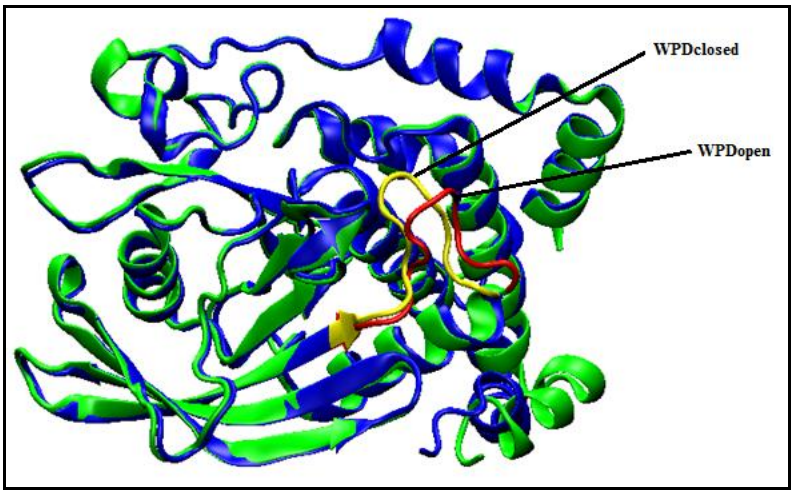

Figure 2.6. WPD_{open} (green) and WPD_{closed} (blue) conformation of PTP1B. WPD loop shown in red and yellow for open and closed conformation, respectively.

Another catalytically important region in PTP1B is the Q-loop (Gln262 to Gln266). Gln262 is essential to hold a water molecule in the active site at the second step of the catalytic mechanism. Other biochemically important regions are α1' (Glu6 to Lys12), α2' (Ser13 to Glu26), loop 11 (L11, Lys150 to Tyr153), α3 helix (Ala189 to Glu200), S-loop (Ser201 to Gly209), α6 helix (Asp265 to Met282) and α7 helix (Ser285 to Asp298) [9, 18].

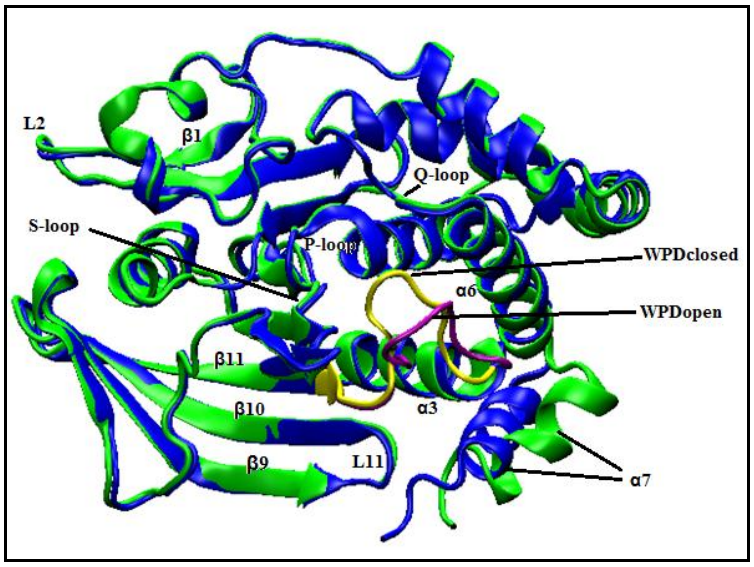

Figure 2.7. Tertiary structure of PTP1B. The α helices and 12 β strands are labeled. WPD_{open} and WPD_{closed} conformations of PTP1B are shown in pink and yellow, respectively.

2.1.4. Catalytic mechanism of PTP1B

Catalytic mechanism of PTP1B proceeds in two-steps. In the first step, the pTyr-containing substrate is engaged to the cysteine residue (Cys215) in the active site with the nucleophilic attack by the sulfur atom of the thiolate side chain of the cysteine amino acid to the substrate phosphate. WPD loop closes over the phenyl ring of pTyr to adopt a closed conformation. This conformational change (closure of WPD) allows the side chain of the conserved acidic residue Asp181 on the WPD loop function as a general acid to transfer a proton to the substrate. This activates the cysteinyl-phosphate intermediate of PTP1B. In the second step, phospho-enzyme intermediate is hydrolyzed by a nucleophilic water molecule, mediated by Gln262 and Asp181, which now functions as a general base to take

a proton away from water in the intermediate. The water molecule forms hydrogen bonds to Gln262 and makes interaction with the carboxylate oxygen of the Asp181. Then, the phosphate is released after the opening of the WPD loop. The closed conformation of the WPD loop in the intermediate has two purposes: the first one is to bring the sidechain of the Asp181 near the catalytic site and to cover the entrance of this site by the help of the side chain of Phe182 (Figure 2.8) [4, 17].

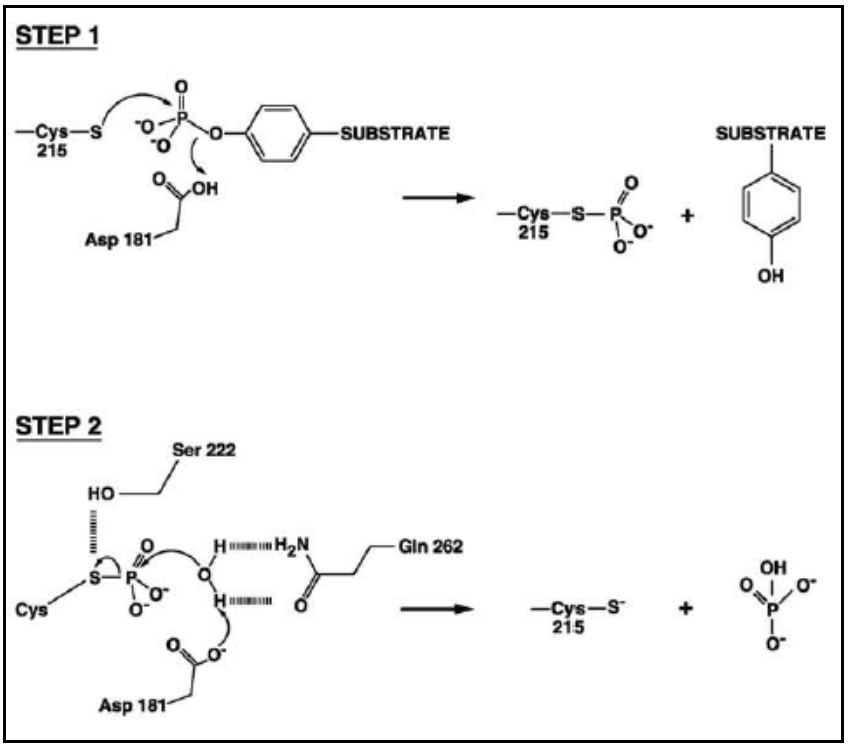

Figure 2.8. Schematic representation of the catalytic mechanism of PTP1B [9].

2.2. Protein Structural Changes upon Ligand Binding and Communication Pathways in Proteins

Various studies have been carried out to understand the propagation of signals between distant sites of the protein. Biological systems are like networks and physical interactions of the appropriate molecule may cause response in the protein system. Allosteric impact is an important key element in elucidating the biologic functions of proteins since allostery can regulate the protein activities as a molecular switch, organize enzymatic and signaling pathways [19, 20].

Understanding the mechanism of the WPD loop dynamics may be significant in facilitating the design of selective inhibitors. Apart from the crystallographic and biochemical studies, detailed computational studies with molecular dynamic simulations are required to have an insight on the WPD loop dynamics. Due to the significant roles of PTPs in signaling pathways, discovery of the inhibitors with suitable selectivity and bioavailability is the principal aim. However, development of such a selective phosphotyrosine (pTyr) mimetics is not an easy task to achieve [9, 16]. Weismann et al. described an allosteric site in PTP1B, ~20 Å away from the catalytic pocket. PTP1B in complex with allosteric inhibitors prevents the closure of the WPD loop, which means deactivation of the enzyme. This allosteric site is not well conserved as the catalytic site of the PTP1B providing these allosteric inhibitors are highly selective for PTP1B (Figure 2.9) [16].

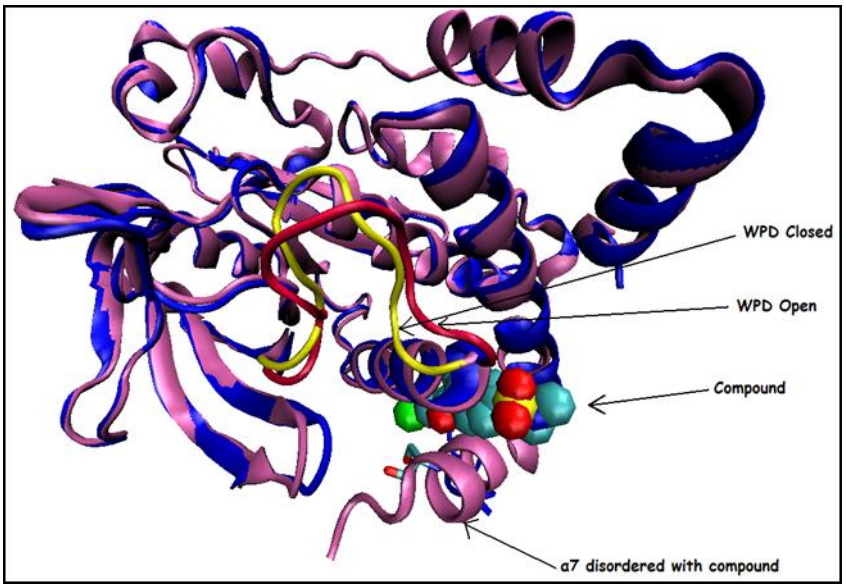

Figure 2.9. Allosteric inhibition of PTP1B. Allosteric inhibitor (shown in vdW spheres) bound to α-6-α-7 region of PTP1B.

For several years, it has been accepted that allostery causes conformational change in the backbone. Therefore, understanding allostery is very important to have an insight about the communication pathways in proteins. Todays, allosteric effects are not only referred as the conformational change at one site, it is also seen as an affecting mechanism without a

backbone conformational change. Current studies show that dynamic fluctuations in residues are also an alternative communication way in proteins [21].

Freire and coworkers proposed that perturbation on binding region with the ligand binding can cause higher fluctuations at the remote regions from the binding site and stabilization of the binding region is accomplished by the loss of the stability at the rest of the protein [22]. There is a thermodynamic relation between function, binding and stability. In Freire et al. [22], a perturbation response analysis was developed by using molecular mechanics (MM) methodology to distinguish residue stabilities of proteins. MM methodology is an approach based on energy minimization altered by inserting a displacement at a selected residue of protein. It is asserted that when a small perturbation occurs in a region of the molecule, there happens a communication between the flexible and rigid parts of the protein and a new global minima was reached by energy minimization [23].

Structural changes upon ligand binding can be also explained by low-frequency normal modes, which describe the pattern of the equilibrium fluctuations. According to the linear response theory, the response of the protein upon ligand binding is highly related with the equilibrium fluctuations in the ligand-free state. Therefore, it was tried to find a linear relationship in between the response of the protein to ligand binding and the equilibrium fluctuations to predict the structural changes. MD simulations for the three protein systems confirm the linear relation between equilibrium fluctuations in the unliganded form and the structural change upon ligand binding [19].

To understand the mechanism of the communication between distant domains of proteins, several studies on different proteins were performed. For instance, Ghosh and coworkers examined the allosteric effect driven by the communication between distal residues for different ligand-bound complexes of MetRS [20]. The aim is to find the communication pathway between the tRNA binding site and the catalytic site in aminoacyl tRNA synthetase (aaRS). MD simulations of different ligand-bound complexes of MetRS were used to identify the residues connected by noncovalent interactions between anticodon and aminoacylation region of MetRS. It is seen that tRNA binding causes higher

flexibility to the MetRS for ligand recognition by long-range communication and allosteric effects can also be at residue level [20, 24].

Protein phosphorylation is one of the other most significant mechanisms in both eukaryotic and prokaryotic cells. Phosphorylation of a protein generally causes conformational changes on the protein and acts as molecular switch by turning the activities on or off. In order to understand how phosphorylation changes the global conformation of a protein rather than local structure, two different systems in which substantial structural changes seen for the phosphorylated and unphosphorylated cystatin and NtrC (nitrogen regulatory protein) [24] were analyzed. The aim of study performed by Latzer and coworkers is to elucidate how phosphorylation causes these changes in protein conformation. Principal component analysis (PCA) was performed in the MD simulations and conformations of the phosphorylated and unphosphorylated forms of the two test proteins were projected onto the first two principal components. By this method, significant residue contacts upon phosphorylation were determined and small conformational changes i.e. residue-residue interactions were determined on the cystatine by using linear response method [24].

In order to elucidate the long-range communication in protein that cause two different conformational changes with the local perturbations (ligand binding), NMR spectroscopy was used in the study of Zhuravleva and coworkers [25]. NMR provides information about the propagations of perturbations that cannot be detected quantitatively in experimental tools, such as change in side-chain dynamics, chemical shifts, and residual dipole couplings. Therefore, free and barstar-bound forms of barnase were examined. There are very few differences between the two structures of barnase and it is aimed to propose an alternative mechanism for the propagation of signal between distant regions of the protein. It is found that binding regions form a network coupled by hydrogen bonds and arranged as a surface providing a mobile interface between relatively rigid domains [25].

3. METHODS

In this section, first computer simulation methods are introduced. Then conventional tools used to analyze molecular trajectories from computer simulations are discussed. In the final part, a summary of frequency response technique, used to identify the protein response to forced perturbations in the loop region in the current study, is given.

3.1. Computer Simulations

Computer simulations have been widely applied to investigate the structural properties of the molecules and microscopic interactions between each other. They serve as a supplement to laboratory experiments. Under some extreme conditions such as very high temperature and pressure, it may be difficult or impossible to carry out experiments. With the use of computer simulations, working with a high-temperature plasma, a nuclear reactor or a planetary core would be perfectly feasible. In addition, investigation of the structure of molecules in detail and their motions are difficult experimentally. However, simulations can extract this type of information easily compared to real experiments. In another sense, computer simulations act as a bridge between theory and experiment. It is possible to test the theory by conducting a computer simulation on the computer with the use of appropriate model. Then, the model is tested by comparing with experimental results. Due to this joining role of the computer simulations in between theory and experiment, this technique is referred as "computer experiments" [26, 27].

3.1.1. Molecular Dynamic Simulations

Molecular dynamics simulation (MD) is one of the significant simulation methods. It is related with the study of biological molecules, which are allowed to interact with each other for a certain period of time based on known physics rules, mathematics and chemistry. Molecular systems include a vast number of particles; the usage of MD simulations facilitates finding the properties of such systems. In summary, MD simulations

give an insight into time dependent behavior of a molecular system and used to investigate the structure, dynamics, function, and relationship between molecular structures [26, 27].

Successful simulations of biomolecular systems can give information about the basic biological processes, fluctuations, and conformational changes of proteins and they can provide new insight to drug design and structure determination [28].

In MD simulations, motion of atoms are simulated as a function of time according to Newton's equation of motion given by

$$F_i = m_i \cdot a_i \tag{3.1}$$

The force can also be expressed as the gradient of the potential energy V,

$$F_i = -\nabla_i V \tag{3.2}$$

Combining these two equations yields,

$$F_i = -\frac{dV}{dr_i} = m_i \cdot \frac{d^2 r_i}{dt^2} \tag{3.3}$$

The force on atom i is denoted by F_i. Calculation of forces acting on the atoms requires the calculation of the gradient of the potential energy $V(r^N)$ to the changes in position as function of time, where $r^N=(r_1, r_2, \ldots r_N)$ represents the $3N$ atomic coordinates.

MD simulations consist of three main steps: minimization, equilibration, and data collection steps. The purpose of energy minimization is to find the conformation of a molecule in which its local energy is minimum. The next step in MD simulations is equilibration. The aim of the equilibration stage is to remove all kinetic energy from the system that is given during the heating process, and to convert this energy into potential energy. This reduces the thermal noise in the structures and potential energies. When the potential energy of the system reaches a constant level, it can be said that the system is equilibrated. Last but not least, in the data collection stage MD trajectories are recorded for the further analysis [29].

3.1.2. CHARMM Force Field

CHARMM is a widely used general and flexible program that can model various macromolecular systems (isolated molecules, molecules in solutions and molecules in crystalline solids) with empirical energy functions. It includes a set of force fields for MD simulations and analysis package related with them. CHARMM can be used to model structures, to minimize the energy of the structures, to perform a normal mode or molecular dynamic simulation and to analyze the structural, equilibrium, and dynamic properties in these calculations [30, 31]. There are several versions of CHARMM force field. In this study CHARMM27, the most recent versions of the force field, is used.

The empirical energy function of CHARMM force field consists of many terms based on intramolecular and intermolecular interaction (Equation 3.4)

$$U_{bonded} = \sum_{bonds} K_b \left(b - b_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{torsions} K_\phi \left(1 + \cos\left(n\phi - \delta\right)\right)^2 + \sum_{impropers} K_\varphi \left(\varphi - \varphi_0\right)^2 + \sum_{Urey-Bradley} K_{UB} \left(r_{1,3} - \varphi_{1,3;0}\right)^2 \quad (3.4)$$

$$U_{nonbonded} = 4\pi D \sum_{electrostatics} \frac{q_i q_j}{r_{ij}} + \sum_{VDW} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \quad (3.5)$$

In Equation 3.4, the first term represents the covalent bond stretching interaction between two atoms in the system, where K_b is the bond force constant and b-b_0 is the distance that atom moves from the minimum energy bond length, b_0. The second term describes the angles between each pair of covalent bonds where K_θ is the angle force constant and θ- θ_0 is the angle from equilibrium between three bonded atoms. The next term represents the dihedral angle potential based on atom pairs separated by three covalent bonds, where K_φ is the dihedral force constant, n is the multiplicity of the function, ϕ is the dihedral angle and δ is the phase shift. This term supplies a better force field near the minimum energy geometry. The final two terms of Equation 3.4 represent the impropers and Urey-Bradley component.

Equation 3.5 represents the nonbonded interactions between atom pairs of *i* and *j*, which corresponds to the van der Waal's forces approximated by Lennard-Jones potential and electrostatic interactions [31, 32].

3.1.3. NAMD

Nanoscale Molecular Dynamics (NAMD) is one of the parallel molecular dynamics program used to develop high-performance simulation of large biological molecular systems in realistic environments. Molecular dynamics program NAMD and its assistant molecular graphics program VMD are used to examine protein structures. NAMD simplifies the access to dynamic information and supplies a productive molecular modeling tool [33].

All atoms in MD simulation experience a model force field accounting for the interaction of that atom with the rest of the protein. The force computations are divided into two parts: bonded force and non-bonded force computations. NAMD uses a common potential energy function that describe stretching, bending, torsional bonded interactions, van der Waal's forces and electrostatic interactions between every nonbonded pair of atoms in the system [30]. NAMD uses a user-specified cutoff distance for long-range interactions and particle-mesh Ewald (PME) method, a fast numerical method, for full electrostatic computations. CHARMM and AMBER force field specifications are used in parameterizations of NAMD. Moreover, NAMD enables the user to apply external forces to the system [30, 33].

NAMD2 is one of the new parallel programs coded for the use of large parallel machines in a scalable manner. It uses a unique decomposition strategy combining the advantages of spatial and force decomposition permitting the program to utilize a large number of processors. The non-bonded force computations of NAMD2 require calculation of pairwise interactions between atoms. A cutoff distance, R_C, is used for the non-bonded interactions. The design of NAMD2 is suitable for the additional new features. NAMD2 uses CHARMM force fields and X-PLOR coordinate and molecular structure files. In addition, NAMD2 uses an adaptive measurement-based periodic load balancing strategy that achieves good performance even for difficult MD simulations [30, 33].

3.2. Conventional Analyses of Computer Simulation Trajectories

3.2.1. Root Mean Square Deviations (RMSD)

Root-mean square deviation (RMSD) is a widely used measure of the average distance between equivalent C_α atoms of two superimposed proteins. Its calculation is quite simply (Equation. 3.6), the square root of the average squared distance between a pair of structures (*x* and *y*) with equivalent atoms (*N*).

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|x_i - y_i\right|^2} \tag{3.6}$$

RMSD calculations were performed on both MD and TMD simulations to measure the mobility of the protein's backbone [34, 35].

3.2.2. Mean Square Fluctuation (MSF)

The mean square fluctuation (MSF) is a measure of deviation of the C_α atoms over time from the averaged structure (Equation 3.7). Different from RMSD, the average is taken over time for MSF. The calculation of MSF is as follows;

$$MSF = \frac{1}{T}\sum_{t_j=1}^{T}\left(x_i\left(t_j\right) - \tilde{x}_i\right)^2 \tag{3.7}$$

where *i* is the particle, *T* is the time which one wants to average, $\tilde{x}_i$ is the reference position of the particle *i* (usually the average conformation). Root mean square fluctuation (RMSF) is the square root of MSF [34, 35].

3.2.3. Principal Component Analysis (PCA)

Studies of collective motions stemmed from atomic interactions in proteins showed that there is a relation between function of the protein and its dynamics [36, 37]. There are many techniques to investigate the relation between collective motion, function, and

folding, like Principal Component Analysis (PCA) [36]. The purpose of PCA is to identify the "essential" dynamics of the system [38]. While, by the help of the MD simulations, it is possible to observe conformational changes in proteins, it is difficult extract functionally important motions from complex protein motions [36, 39]. Using collective coordinates is one of the solutions to this problem. Recent studies have shown that functionally relevant motions are along the direction of the few collective coordinates that have dominating contributions to atomic fluctuations [37]. In order to detect the fluctuations done together with large amplitudes, PCA can be used on the large set of experimental structures.

PCA is related with diagonalization of the covariance matrix of atomic fluctuations to produce collective variables that are sorted according to their contribution to the total mean-square fluctuation (MSF). By the help of the PCA, trajectories onto the collective variables with the largest contributions to the total MSF is projected and significant contributions to the total MSF can be determined. It is important to note that MD simulations have to be prolonged for convergence; therefore, eigenvectors of the first and second half of the MD simulations or between two different simulations has to be compared in order to eliminate the impacts of the intrinsic noise. In general, first few eigenvectors describes the functional motions of the proteins [36].

Algebraically, principal components are concerned with linear combinations of the p random variables $X_1, X_2, \ldots, X_p$. These linear combinations of random variables represents a new coordinate system ($Y_1, Y_2, \ldots, Y_p$) obtained by rotating original system with $X_1, X_2, \ldots, X_p$ as the coordinate axes (Equation 3.8-3.10). The new coordinate axes represent the directions with maximum variability and provide a simpler description of the covariance structure [40]. Principal components only depend on the covariance matrix Σ of random variables.

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \tag{3.8}$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \tag{3.9}$$

...

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p \tag{3.10}$$

3.3. Frequency Response of Dynamic Systems

Recently, number of studies have been done for determining the information flow between residues, utilizing techniques such as computation of the energy correlation matrix, or transfer entropy concept [41-44]. Frequency response techniques may bring an alternative solution to this hot topic. The protein can be perturbed in the form of changing the environmental factors or forcing the protein structure itself to target location by introducing ligand binding, site-specific perturbations, etc. The response of the protein to these perturbations can be measured directly such as RMSD from the initial position, and by frequency domain techniques such as Fourier Transform or indirectly (kinetic parameters).

One way to find the response of a system is to apply a test signal to the input and look at the output to see how it responds. Many test signals are possible, but a simple and useful test signal is the sine wave. This is because the output of a linear system with a sine wave input is also a sine wave, but with different amplitude and phase. By measuring the output amplitude and phase of a system over a range of frequencies of the input sine wave, a particular version of the dynamic response is built-this is called the frequency response (Figure 3.1). Frequency response analysis provides useful insights into stability and performance characteristics of the controlled system [45].

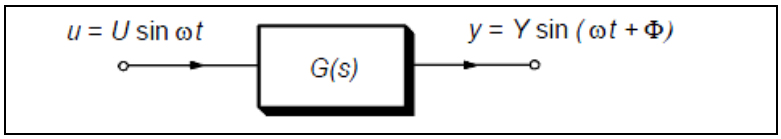

Figure 3.1. Linear transfer function with sine wave input [45].

If a system has transfer function $G(s)$, then the output response at a particular frequency $\omega = 2\pi f$ is given by the gain and phase of the frequency response $G(jw)$ at that frequency ω. For the system shown in Figure 3.1, the input and output signals are

$$u(t) = U \cdot sin(\omega t) \qquad t > 0 \tag{3.11}$$

$$y(t) = Y \cdot sin(\omega t + \Phi) \qquad t >> 0 \tag{3.12}$$

The corresponding gain and phase are given by:

$$\text{Gain at } \omega = |G(j\omega)| = \frac{Y}{U} \tag{3.13}$$

$$\text{Phase at } \omega = G(j\omega) = \Phi \tag{3.14}$$

The amplitude and phase are changed by the system, but the frequency remains the same. This is shown in Figure 3.2.

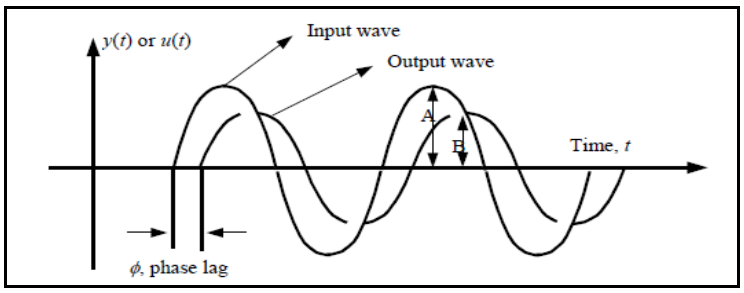

Figure 3.2. Frequency response [45].

The output wave lags behind the input Φ, defined as the *phase lag*. The output amplitude is different from the input, and the amplitude ratio is defined as the ratio of the maximum amplitude of the output over the maximum amplitude of the input.

$$Amplitude\ Ratio(AR) = Y / U \tag{3.15}$$

By measuring the gain and phase over a range of frequencies, the full frequency response of the system can be plotted. There are a number of ways to represent the frequency response of a process. Two of these are *Bode plots* and *Nyquist plots* (Figure 3.3). The Nyquist plot is a diagram of the imaginary part (Im) of the frequency response $G(jw)$, against the real part (Re) of the frequency response $G(jw)$. The Bode diagram shows the plot of the logarithm of the frequency response gain $\log|G(jw)|$ and the phase, $\phi(jw)$ [46].

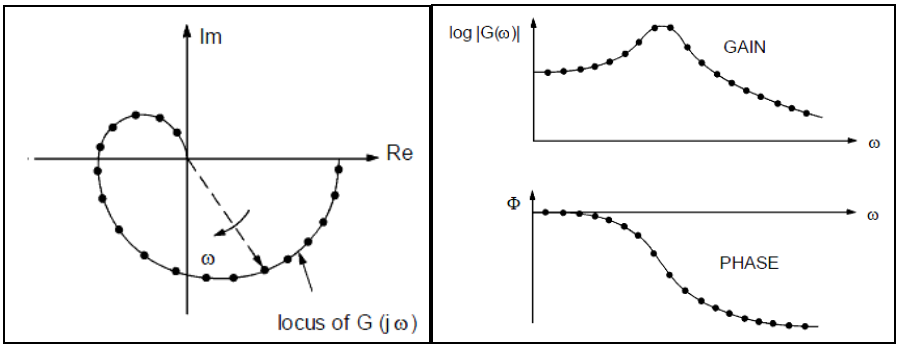

Figure 3.3. (a) Presentation of frequency response data: Nyquist Plot, (b) Presentation of frequency response data: Bode Plot [46].

While examination of communication paths of the protein from the active site to the allosteric regions of the protein, Bode plots of the distant regions of the protein can give information on how information flows (the pathway, the amplification of the signal, phase delay, etc.)

4. CONVENTIONAL ANALYSES OF MD AND TMD SIMULATIONS OF PTP1B

In this study, targeted molecular dynamics (TMD) is used to study the conformation changes in PTP1B. The approach relies on applying directed forces on appropriate segments of protein (perturbation by forcing of WPD loop and R-loop in PTP1B) and recording the resulting changes in the residue coordinates. The initial and final protein conformations for the TMD simulations are obtained from equilibrium MD simulations of PTP1B in open and closed conformations.

4.1. Equilibrium MD simulations of PTP1B in open (WPD_{open}) and closed (WPD_{closed}) conformations

Initial atomic coordinates for the WPD_{open} and WPD_{closed} conformations of PTP1B are obtained from the protein data bank with the PDB IDs: 2F6F and 1SUG, respectively. First of all, some changes were made on the crystal structures of PTP1B. Cys215 is taken to be in thiolate form [47], while Asp181 is taken to be protonated state [48, 49]. In the crystal structure of 2F6F, residues numbered from -3 to 0 are truncated and Phe295 is back mutated to Ser295 using VMD program. In 1SUG crystal structure, Leu299 is removed and Met1 is added to this structure by using 2F6F structure as a template. Adding the Met1 to the crystal structure of 1SUG is to obtain the same residue number for WPD_{open} and WPD_{closed} conformations. In order to estimate the missing sidechain and hydrogen coordinates of the crystal structures, psfgen package of VMD was used.

Crystal waters within 6 Å of PTP1B were kept, while each protein structure was solvated within a layer of 10 Å of TIP3 water in a rectangular box. Then, the system was neutralized by adding 0.5 M sodium and chloride ions by using autoionize package of VMD. The size of the final system (periodic box) for WPD_{open} structure (prepared using the crystal structure with PDB ID: 2F6F) is 79.6 × 80.5 × 69.1 Å and contained 4830 protein atoms and 12005 water molecules. For WPD_{closed} structure (prepared using the crystal structure with PDB ID: 1SUG), the size of the periodic box is 85.2 × 74.4 × 63.2 Å

and contained 4830 protein atoms and 10767 water molecules. Periodic boundary conditions on the system, and the long-range interactions were performed with the particle-mesh-Ewald method [50]. Non-bonded cutoff parameter of the force field was taken as 12 Å.

Energy of the whole system was minimized using steepest descent algorithm. Minimization took a total of 3000 steps. Heavy atoms (non-hydrogen atoms) were constrained during the minimization, and these constraints were relaxed gradually at every 250 step. The minimized conformations of WPD_{open} and WPD_{closed} structures are expected to have small RMSDs to their crystal structures. The average RMSD of C_α atoms for WPD_{open} and WPD_{closed} conformations with respect to their crystal structures are 0.56 Å and 0.47 Å, respectively.

All MD simulations were carried out using the NAMD2 [33] program with the CHARMM27 [51] forcefield at an integration time step of 1 fs. MD simulations, initialized from the minimized structures of WPD_{open} and WPD_{closed} conformations, are named MD_{open} and MD_{closed}, respectively. Both MD_{open} and MD_{closed} simulations were carried out at constant pressure of 1 atm using Langevin piston pressure control. Temperature was gradually increased from 50 K to 300 K with 50 K increments at the beginning of the simulations (~50 ps), and the rest of the simulations were carried out at a constant temperature of 300 K using Langevin temperature control. The lengths of the two MD simulations were 40 ns. A summary of equilibrium MD simulations are given in Table 4.1.

Table 4.1. A list of equilibrium MD simulations performed and analyzed in the current study.

Name of MD simulations	Starting Crystal Structure	Time
MD_{open}	2F6F	40 ns
MD_{closed}	1SUG	40 ns

4.1.1. Structural Analysis of MD_{open} simulation

RMSDs of C_α atoms of the whole protein and WPD loop in MD simulation structures were determined with respect to the crystal structure (PDB IDs: 2F6F). RMSD of all C_α atoms in MD_{open} simulation atoms levels off at ~1.5-2.0 Å away from the crystal structure in 40 ns simulation (Figure 4.1a), showing the stability of the MD simulation. RMSD of the WPD loop trace (C_α atoms) with respect to its conformation in open and closed crystal structures are maintained at ~1.7 Å and 4 Å, respectively (Figure 4.1b) showing that WPD loop maintains its open conformation and there is no trend of WPD loop closure.

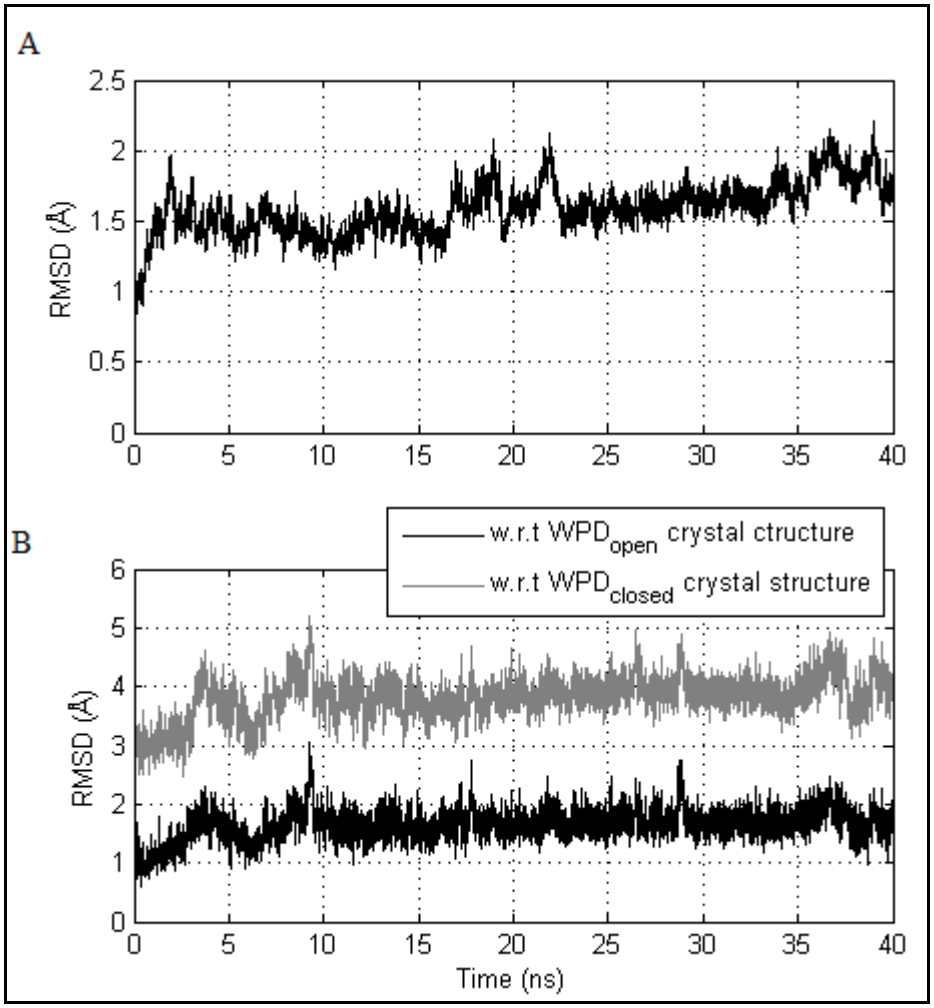

Figure 4.1. (a) RMSDs of the C_α atoms of the whole protein, and (b) RMSDs of the C_α atoms of the WPD loop with respect to WPD_{open} (black) and WPD_{closed} (grey) crystal structures.

In order to see whether major conformational changes took place in the MD simulation, the average conformation of the MD_{open} simulation was compared with the crystal structure in Figure 4.2. RMSD of C_α atoms of the whole protein is ~1.2 Å, and the only significant difference is seen in the highly flexible α7 helix.

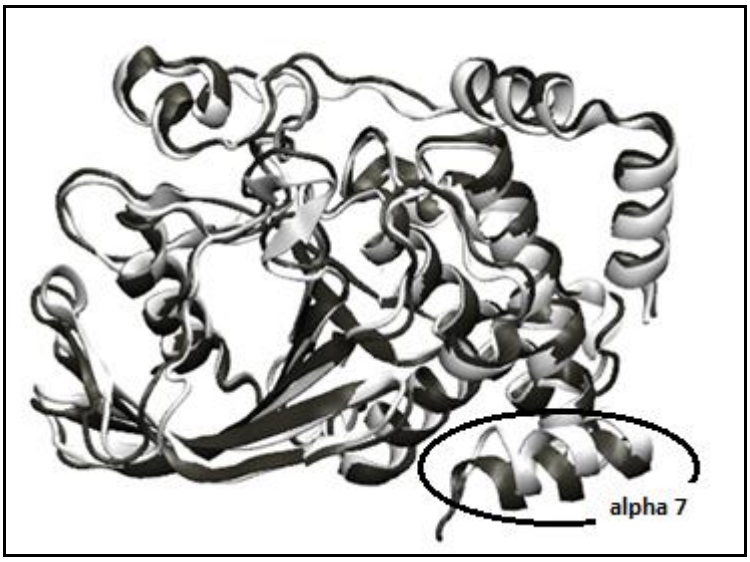

Figure 4.2. The cartoon representation of the average conformation of the MD_{open} simulation (black) and the crystal structure (white, PDB IDs: 2F6F) aligned with respect to their C_α atoms.

4.1.2. Residue Mobility in MD_{open} simulation

To measure the residue mobility in MD_{open} simulation, residue based RMSF of the simulation structures based on C_α atoms were calculated. The residue-based average RMSF of the MD_{open} simulation is 0.81 Å. Highly mobile regions in MD_{open} simulation are loop regions of N- and C- terminus of α1, L2, S-loop and α7 (Figure 4.3). The correlation value between the C_α atom displacements of the MD_{open} simulation and scaled B-factors of the C_α atoms (gray lines in Figure 4.3), as obtained from the crystal structure, is 0.69, confirming the reliability of the dynamic information obtained from the MD_{open} simulation.

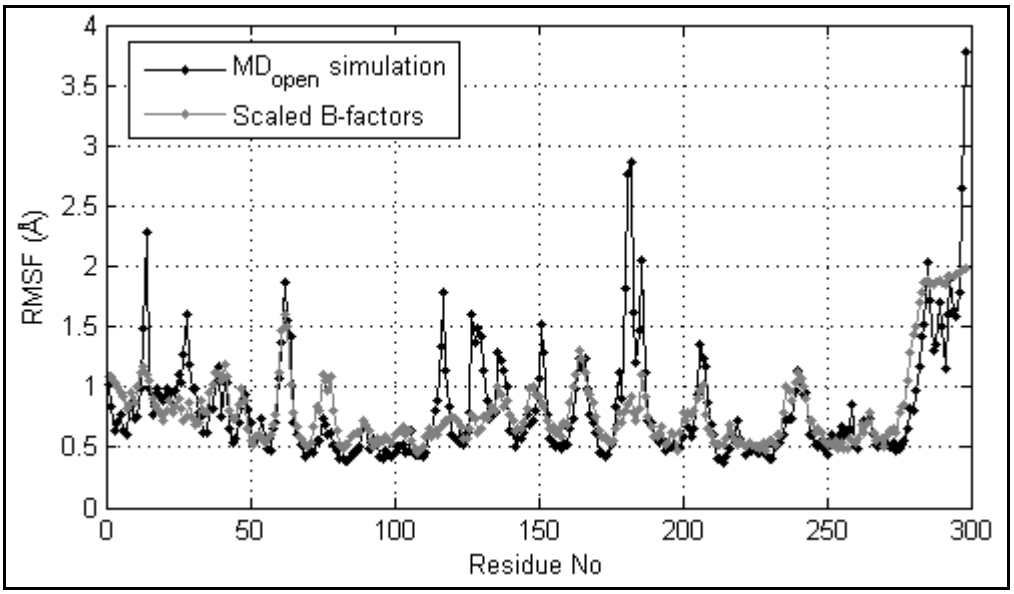

Figure 4.3. RMSF of the residues during the 40 ns MD_{open} simulation.

4.1.3. Structural Analysis of MD_{closed} simulation

The same analysis was also performed for MD_{closed} simulation. In order to have an insight about the overall structural changes in MD_{closed} simulation, the RMSDs of C_α atoms of the whole protein and WPD loop were calculated in MD simulation structures after aligning all C_α atoms of PTP1B with respect to the closed form crystal structure (PDB IDs: 1SUG). RMSD of all C_α atoms in MD_{closed} simulation stabilizes at ~1.5-2.0 Å (similar to that in MD_{open} simulation) away from its structure in 40 ns simulation (Figure 4.4a). As expected, MD_{closed} simulation stabilizes during the 40 ns time simulation. Figure 4.4b shows the RMSD of the C_α atoms of WPD loop with respect to the crystal structures of open and closed conformations. Distance of the WPD loop trace with respect to its conformation in closed and open crystal structures is maintained at ~1 Å and ~3 Å, respectively, indicating that WPD loop does not have a tendency of opening during the course of the MD_{closed} simulation. Furthermore, RMSD of the WPD loop with respect to its starting conformation is smaller in MD_{closed} simulation (~1 Å), as opposed to MD_{open} simulation (~1.7 Å).

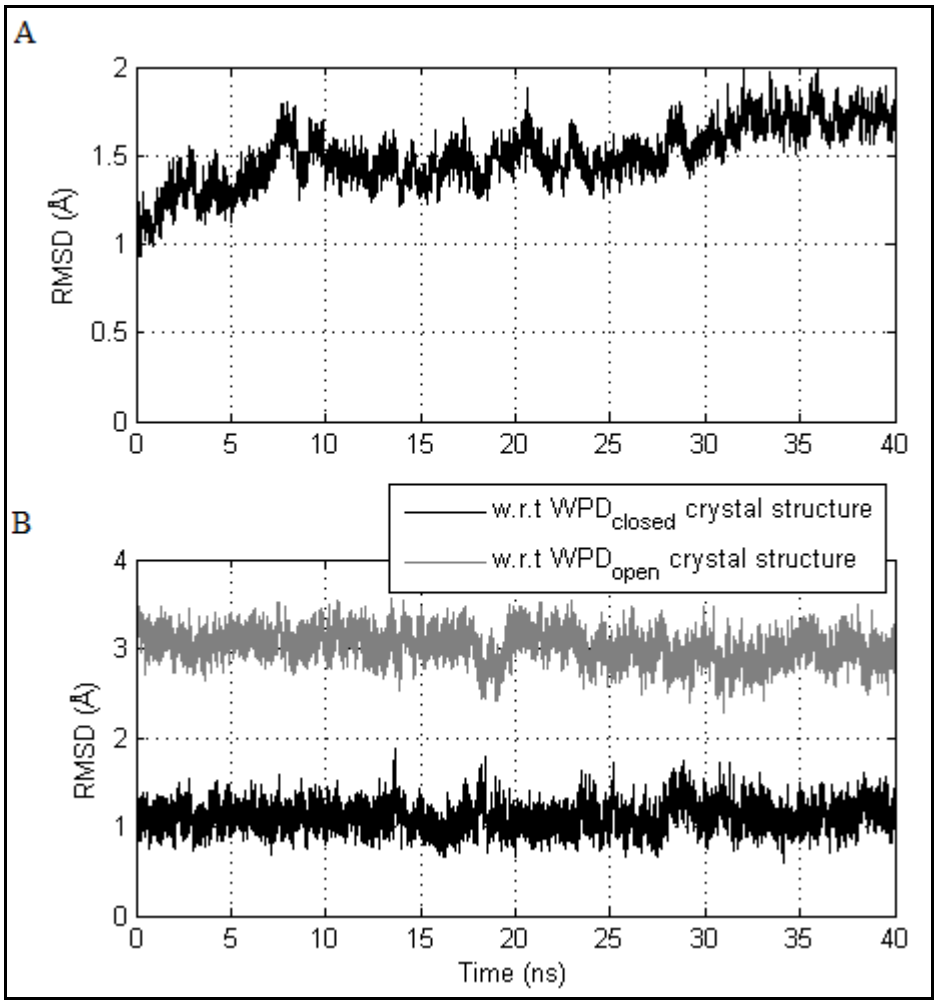

Figure 4.4. (a) RMSDs of the C_α atoms of the whole protein, and (b) RMSDs of the Cα atoms of the WPD loop with respect to WPD_{open} (grey) and WPD_{closed} (black) crystal structures.

Figure 4.5 shows the superposition of the average structure of MD_{closed} simulation trajectory (black) to its starting crystal structure (PDB IDs: 1SUG, white) with respect to the C_α atoms. The RMSD between the two structures is 1.16 Å, a moderately low value. Furthermore, visual examination shows no significant difference between the structures.

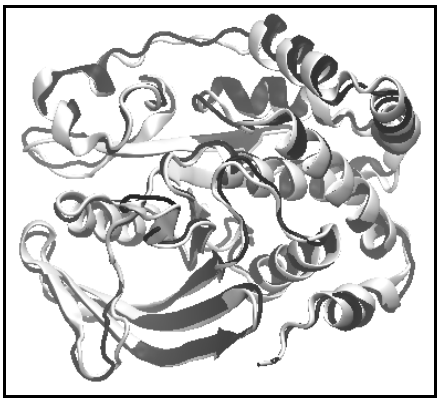

Figure 4.5. The cartoon representation of the average conformation of the MD_{closed} simulation shown in black and the crystal structure (PDB IDs: 1SUG) shown in white.

4.1.4. Residue Mobility in MD_{closed} simulation

Figure 4.6 shows the residue based RMSF of the MD_{closed} simulation structures. The average RMSF of MD_{closed} simulation is 0.72 Å. The regions that shows highest mobility in MD_{closed} simulation is α2', α1, L2, L8, L9, L12, L14 and α7. Correlation between MD_{closed} simulation and scaled B-factors of WPD_{closed} crystal structure (PDB IDs: 1SUG) is calculated as 0.50, slightly lower than the correlation of MD_{open} simulation with scaled B-factors of WPD_{open} crystal structure.

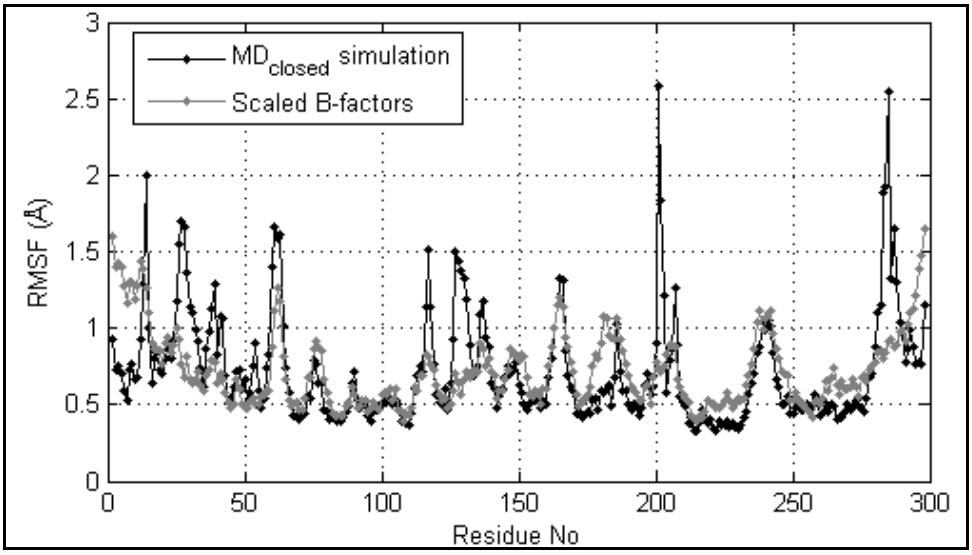

Figure 4.6. RMSF of the residues during the 40 ns MD_{closed} simulation.

4.1.5. Comparison of Residue Mobility in MD_{open} and MD_{closed} simulations

Correlation of the residue based RMSF of the MD_{open} and MD_{closed} simulations is 0.61. While mobility of most of the regions in both simulations are nearly identical, some regions show clear differences (Figure 4.7). L11, WPD loop and α7 have higher mobility in MD_{open} simulation, and S-loop has a higher mobility in MD_{closed} simulation, as highlighted in Figure 4.7.

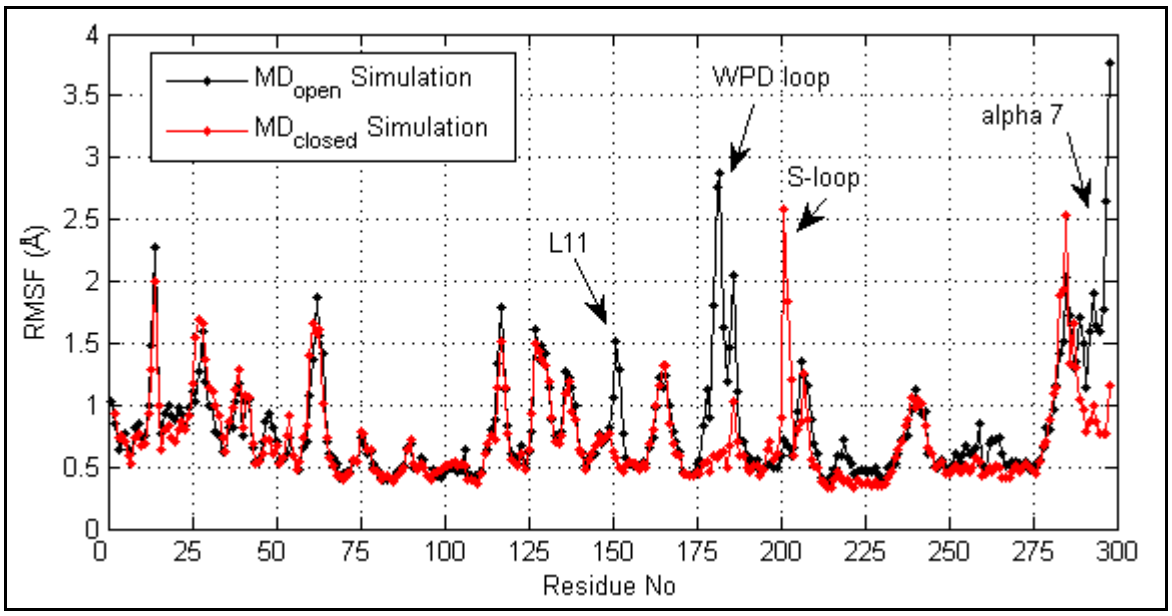

Figure 4.7. Comparison of RMSFs of MD_{open} (black) and MD_{closed} (red) simulations.

4.2. Structural Changes in TMD1 Simulation

Multiple TMD simulations were performed between the WPD_{open} and WPD_{closed} conformations of PTP1B. Initial and target structures for TMD simulations were obtained from MD_{open}, and MD_{closed} (equilibrium) simulations, respectively. Targeting force constant in all TMD simulations is taken to be equal to 3000 kcal/mol/$Å^2$, which gradually forces the starting conformations towards the target structure. TMD force is applied on all atoms of both R-loop (Gly110 to Gly120) and the WPD loop (Thr176 to Ser187) between the initial and target structures of PTP1B. Force constant and the region of application of TMD force on the protein were determined by trial and error, and using previous studies

[52]. Preparation of TMD simulation is explained in Appendix A. TMD simulation protocol is identical to those of the MD simulations (Section 4.1).

Once the first conformational transition from WPD_{open} to WPD_{closed} structures was achieved, then the initial structure (in the open conformation of WPD loop) was taken as the target structure, and the final structure of the previous simulation was taken as the initial structure in the following TMD simulation. This procedure was repeated multiple times, resulting in a series of TMD simulations, in which the target structure was alternated between WPD_{open} and WPD_{closed} conformations. Application of the TMD force between the two different conformations of PTP1B (open and closed) periodically results in a periodic signal. Periodic shifting of the target structure in TMD simulations is demonstrated in Figure 4.8; the length of a single TMD simulation is 2.5 ns, causing the loop to transit between one conformation to another, and thus the period of a whole cycle is 5 ns (=0.2 GHz).

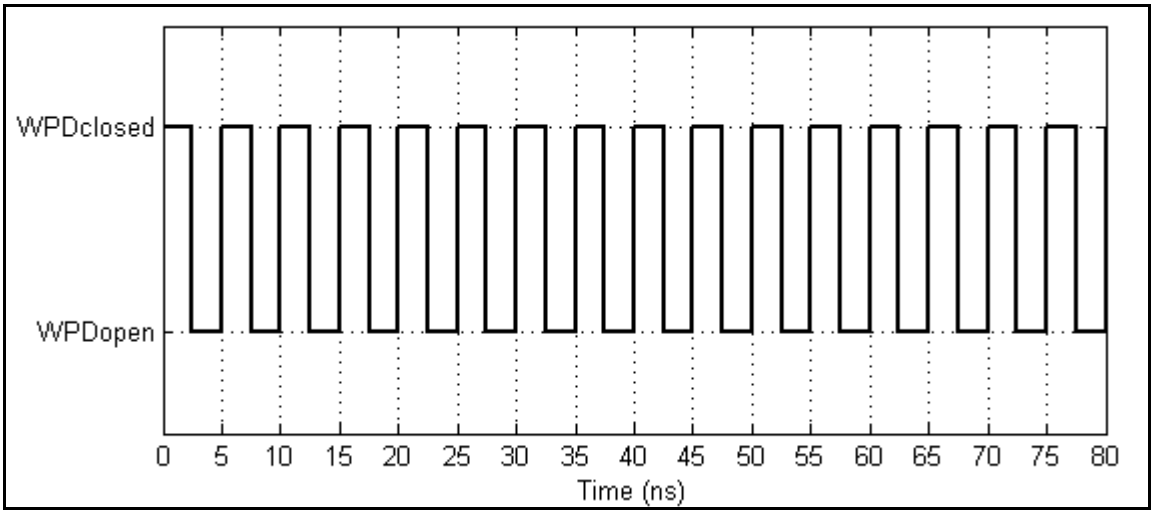

Figure 4.8. Periodic input (target) function at a frequency of 0.2 GHz.

TMD simulations with five different input frequencies (one closing-opening motion of WPD loop) are applied between the WPD_{open} and WPD_{closed} conformations of PTP1B (Table 4.2). For instance, while the WPD loop closure takes place in 2.5 ns with the constraining force 3000 kcal/mol/$Å^2$ for one set of TMD1 simulation, the closure of the WPD loop occurs in 5 ns with the same applied force in TMD2 simulation.

In order to verify the stability of the simulation and to examine the overall structural changes in the protein, RMSD of TMD1 simulation was calculated. To determine these properties, C_α atoms of TMD1 simulation structures were aligned, excluding α7 to eliminate the effects of conformational changes of this highly flexible region, with respect to the WPD_{open} crystal structure and RMSD of C_α atoms between these two structures was calculated (Figure 4.9). It is seen that the RMSD stays between 1.0 Å and 1.8 Å during 80 ns in TMD1 simulation, showing the stability of the overall protein structure.

Table 4.2. List of the TMD simulations.

Name of TMD simulation	Length of each simulation (ns)	Total simulation time (ns)
TMD1	2.5	80
TMD2	5	70
TMD3	10	60
TMD4	0.25	4.5
TMD5	0.05	1.6

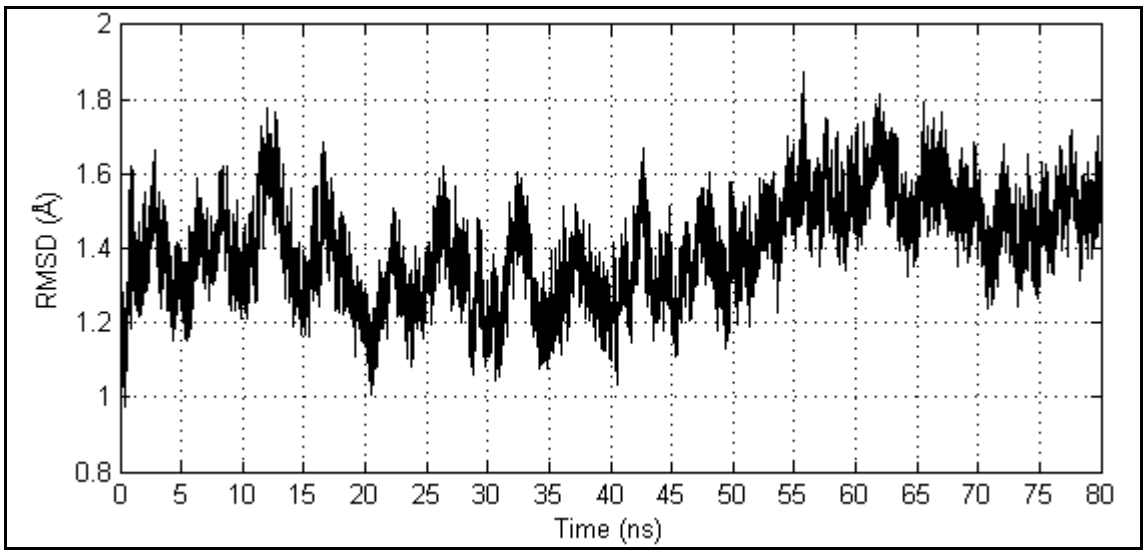

Figure 4.9. RMSD of C_α atoms of TMD1 simulation with respect to the WPD_{open} crystal structure, aligned by excluding α7 helix.

To see the effects of the applied periodic force on WPD loop during TMD1 in detail, time evolution of the RMSDs of the WPD loop trace relative to the WPD_{closed} crystal structure of the TMD1 simulation is examined in Figure 4.10. 16 cycles of the WPD loop in the forward (open to closed) and reverse (closed to open) directions can be clearly observed in Figure 4.10. As shown, WPD loop closes or opens in every 2.5 ns of TMD1 successfully, and has a full cycle in every 5 ns.

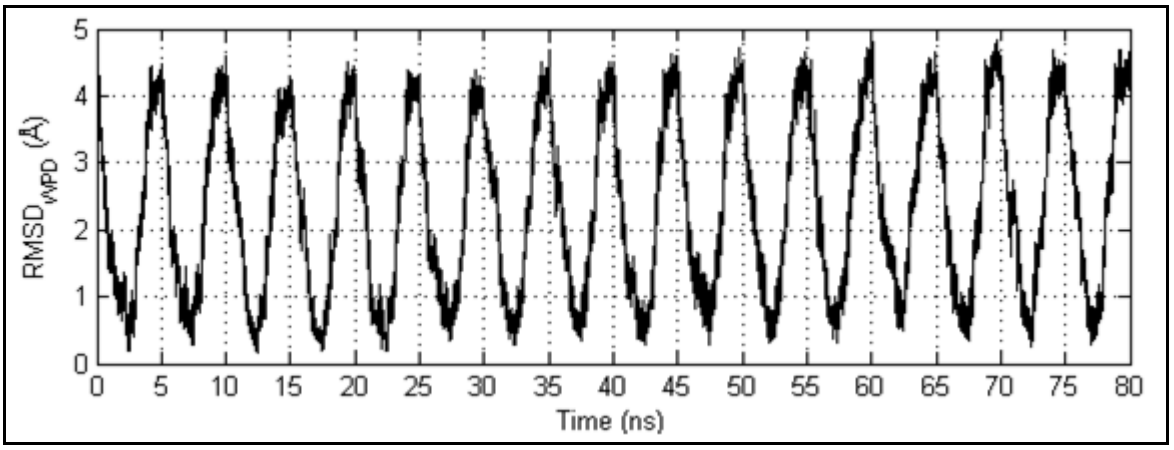

Figure 4.10. The RMSD trajectory of WPD loop w.r.t the WPD_{closed} crystal structure.

Distance of the WPD loop from the active site is a significant criterion for the closure of the WPD loop. Distance between carbonyl oxygen of Trp179 and sidechain nitrogen of the Arg221 was monitored for the analysis of WPD loop motion in TMD1 simulation (Figure 4.11). Distance between the Trp179:O and Arg221:NH1 is decreased from ~7.0 Å to ~2.5 Å, and a hydrogen bond is formed during the closing period of TMD1 run (in the first 2.5 ns) and then this bond is disrupted with WPD loop opening. Only in the 5th run (between 10-12.5 ns) in TMD1, hydrogen bond between Trp179:O and Arg221:NH1 is not formed, while moving from WPD_{open} to WPD_{closed} conformation.

Backbone ψ and sidechain $\chi 1$ angles of Trp179 take drastically different values in the open and closed conformation of the WPD loop, therefore these angles monitored to track the WPD loop conformational transition (Figure 4.12). It should be noted that Trp179 is known to mediate the movement of the WPD loop and mutation of this residue was found to disable catalysis [53]. While backbone ψ dihedral angle of Trp179 is ~150° in

WPD_{open} conformation, the same dihedral angle is ~80° in the WPD_{closed} conformation (Figure 4.12a). Sidechain of Trp179 is to adopt a new value for the closure of the WPD loop: this angle changes ~100° while WPD moves closer to the active site pocket (Figure 4.12b). With Trp179 side chain plunged into the active site pocket, an H-bond is formed between Trp179 and Arg221. While the side-chain of Trp179 approaches the active site pocket, Arg221 sidechain has to turn and move up at the same time.

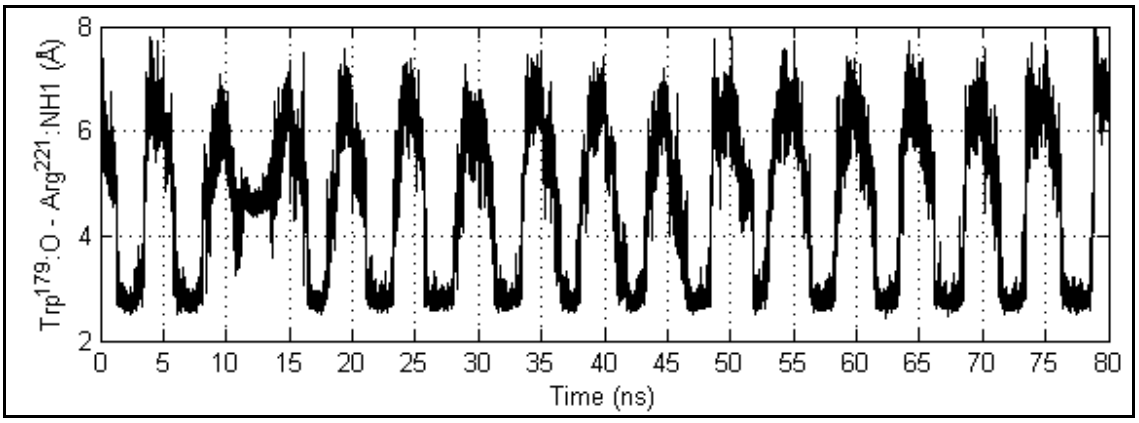

Figure 4.11. Time evolution of Trp179:O and Arg221:NH2 distance during the TMD1 simulation.

TMD force is also applied to the R-loop (Gly110 to Gly120). In order to observe the dynamics of the R-loop and its interactions with the WPD loop, interactions between residues Lys116 and Asp181 are monitored (Figure 4.13a). It is seen that the hydrogen bond between Lys116:NZ and Asp181:O is more stable compared to the Lys116:N and Asp181:side chain O. Asp181:OD1 forms H-bond with Lys116 backbone nitrogen, and Asp181 backbone oxygen atom forms a H-bond with sidechain nitrogen of Lys116 in the open conformation of the WPD loop, but these interactions are disrupted in the closed conformation of the WPD loop. Furthermore, H-bond between Trp179 backbone oxygen and Arg112 side-chain nitrogen should be disrupted (Figure 4.13b) in order to form H-bond between Trp179:O and Arg221:NH1.

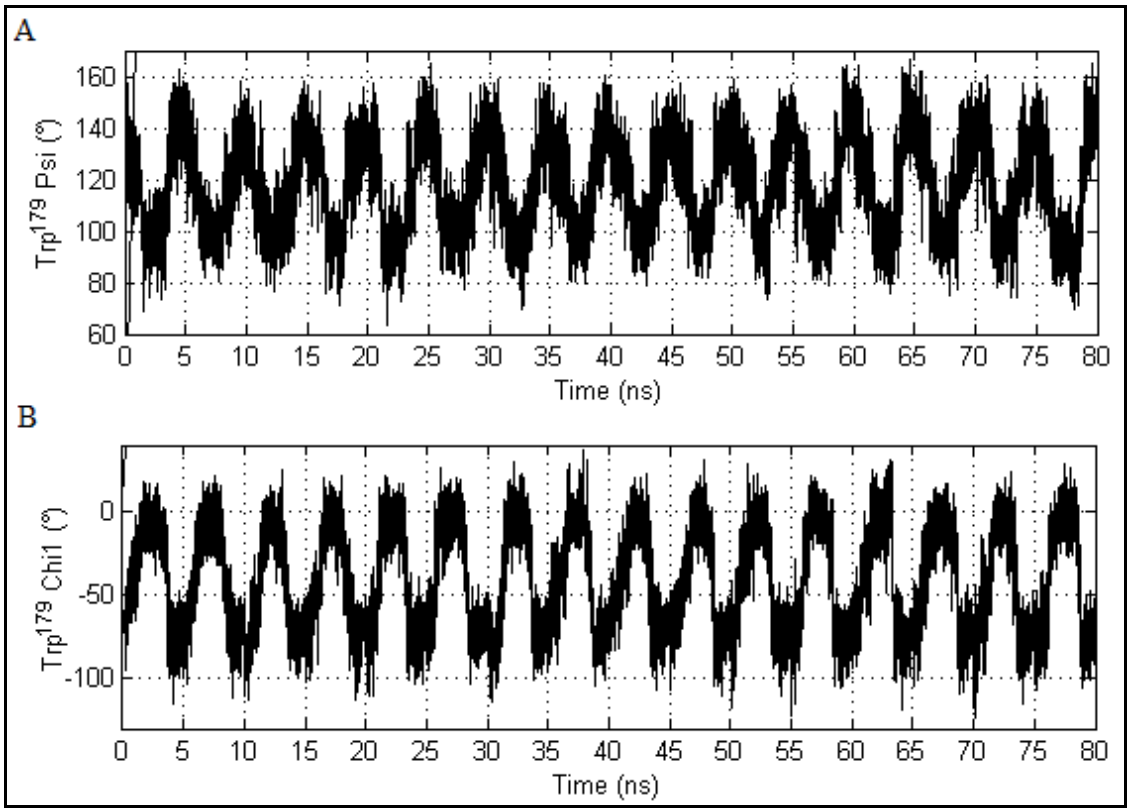

Figure 4.12. (a) Backbone ψ , and (b) sidechain χ1 angles of Trp179 in TMD1 simulation.

Asp181 backbone oxygen has to rotate in order to complete the opening/closing motion of the WPD loop. However, Asp181 backbone oxygen did not rotate to its required conformation in all TMD simulations, as seen from Asp181 backbone ψ angle, showing that there may be a high-energy barrier associated with this conformational transition. (Figure 4.14)

As a summary, one can say that WPD loop opening and closing was successfully achieved in TMD1 simulation, in which most of the interactions stabilizing the WPD loop in the open and closed conformations were formed and disrupted in correct order during the simulations.

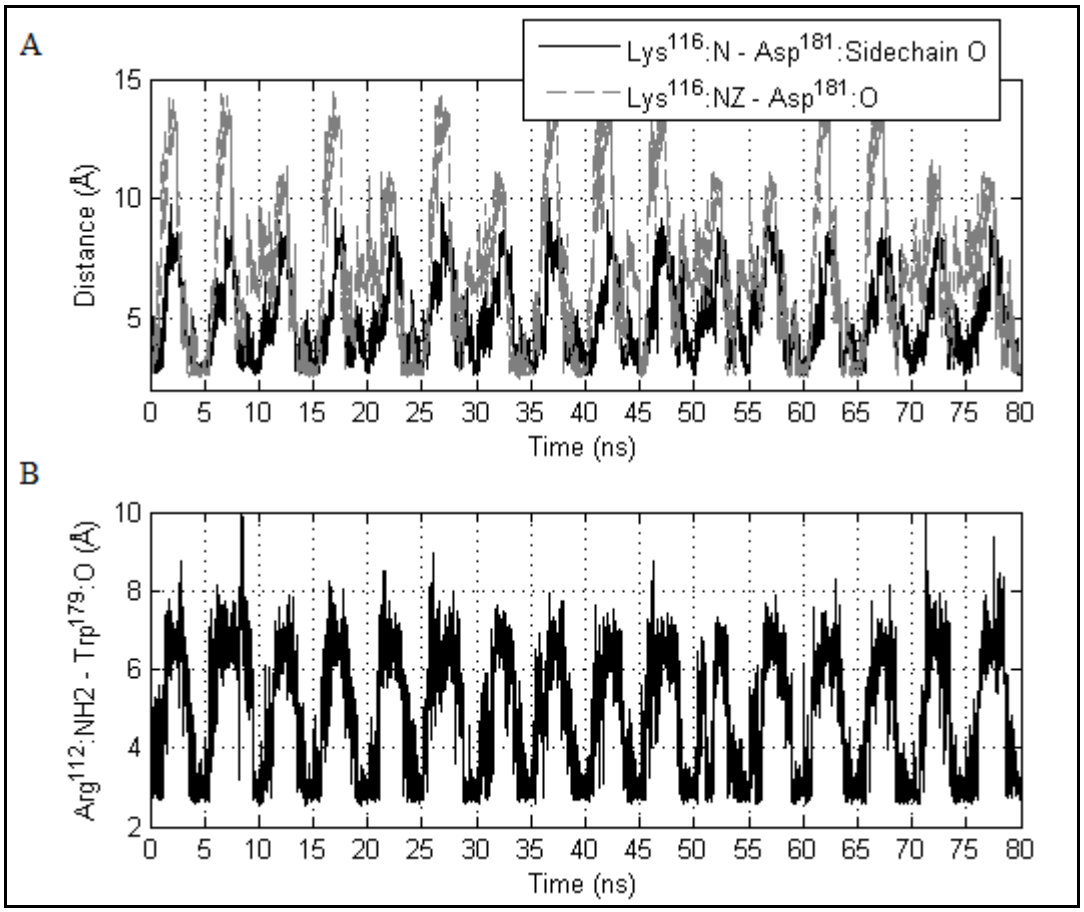

Figure 4.13. The RMSD trajectory of (a) Lys116:N and Asp181: OD1, Lys116:NZ and Asp181:O, and (b) Arg112:NZ and Trp179:O distances during the course of the TMD1 simulation.

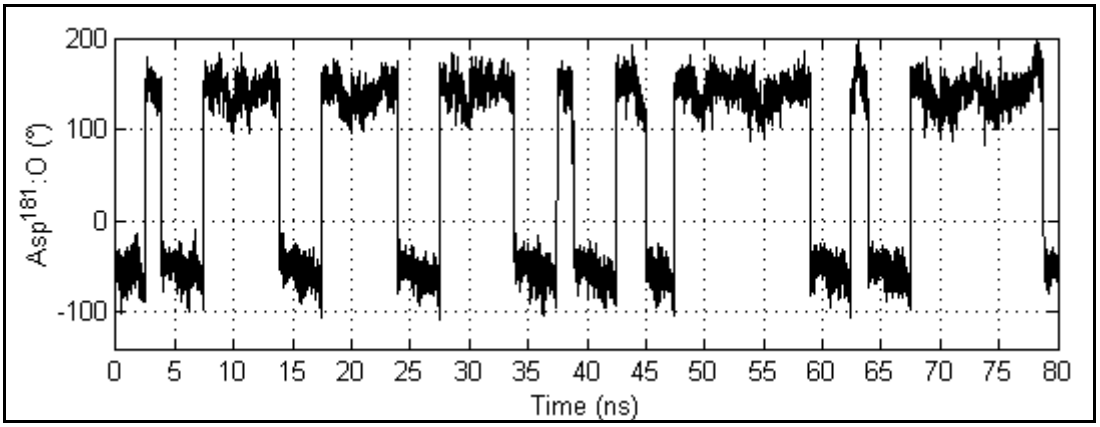

Figure 4.14. Dihedral angle ψ of Asp181 during the course of the TMD1 simulation.

5. FREQUENCY ANALYSIS OF MD AND TMD SIMULATIONS

In the previous section, standard analyses of equilibrium MD and TMD simulations were performed. In the current section of the study, frequency response analysis was performed in TMD1 simulation: WPD loop of the protein is perturbed and the response of the rest of PTP1B is analyzed. The aim is to investigate the communication pathways between the binding region, particularly the WPD loop, and the functionally important distant portions, such as $\alpha 7$ and S-loop, in PTP1B. In order to separate the random and periodic fluctuations during the analysis, TMD simulations are compared with WPD_{open} and WPD_{closed} equilibrium MD simulations.

5.1. Frequency Analysis of WPD Loop Fluctuations

Frequencies responsible for the highest variance of the WPD loop are determined by Discrete Fourier Transform (DFT) on MATLAB. It is seen that the contribution of the fundamental frequency (0.2 GHz) is considerably higher compared to upper harmonics, since the force applied on WPD loop (from Thr176 to Ser187) is at this frequency, while contribution of the first few harmonics, especially 0.4 and 0.6 GHz, to the motion of the WPD loop is not negligible (Figure 5.1). In general, while fundamental frequency (0.2 GHz) has dominance on the N-terminus, significance of the harmonics of the base frequency seems increases on the C-terminus.

When the average Fourier coefficients of the WPD loop in TMD1 simulation were found and compared with those obtained from equilibrium simulations, vibrational motions at the base frequency and its upper harmonics are seen more clearly (Figure 5.2).

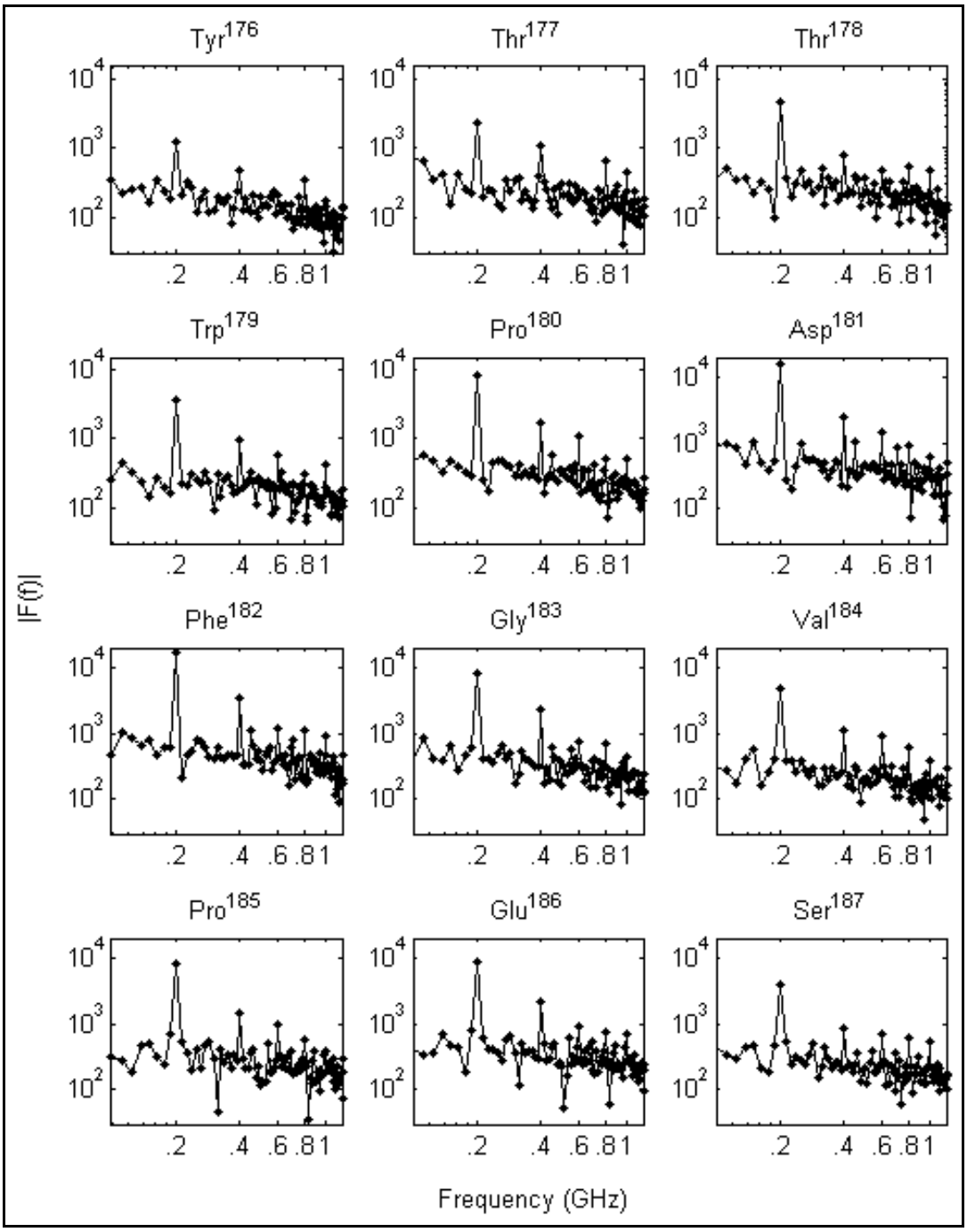

Figure 5.1. Frequency content of the WPD loop residues, as determined by DFT.

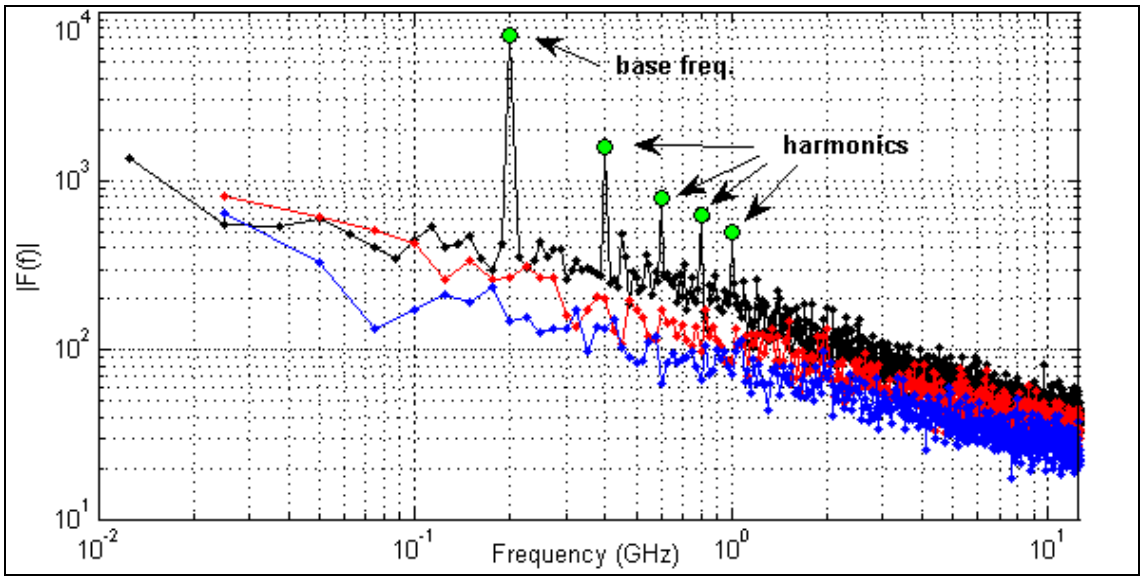

Figure 5.2. Frequency content of the WPD loop in TMD1 (black, green dots show the fundamental frequency and its harmonics), MD_{open} (red) and MD_{closed} (blue) simulations.

5.2. Comparison of Raw and Filtered Coordinates of TMD Simulations

In this section, some examples are given to make a visual comparison of raw and filtered C_α displacement coordinates of TMD simulation. Filtering is performed in the frequency domain to keep some desired frequencies and to suppress the rest, by making the amplitude of the rest of the frequencies zero.

Frequency content of WPD loop in TMD1 trajectory has shown the presence of the fundamental frequency and its upper harmonics. Therefore, this analysis is performed at the five characteristic frequencies (base frequency 0.2 GHz and higher harmonics, 0.4, 0.6, 0.8, 1.0 GHz), detected from the Fourier spectrum of the TMD1. In order to compare the raw and filtered coordinates of WPD loop and some other regions, components of the fundamental frequency and higher harmonics are plotted, and the resulting filtered trajectories are more representative of the protein response, due to the periodic TMD force applied.

Figure 5.3a-c shows C_α atomic displacements of Asp181 on the WPD loop in x, y and z directions in TMD1 (grey line). It is easy to see a periodic behavior consisting of

sixteen cycles (32 TMD runs), corresponding to the opening and closing of the WPD loop in each TMD run. To focus on the WPD loop opening/closing frequency (0.2 GHz), DFT of each atomic displacement is computed as shown on the same Figure 5.3 (black dashed line). The filtered coordinates at 0.2 GHz show the amplitude and phase angle of the oscillations, dictated by the TMD force, more clearly. It is seen that Asp181 fluctuations are not exactly in phase with the beginning of TMD simulations, i.e. each TMD cycle starts every 5 ns, but the minimum of the Asp181 coordinates is reached about 4.6 ps before this time. The percent variance explained by the single base frequency values are also given on each plot. For instance, the fluctuations of Asp181 in x-direction at 0.2 GHz alone is sufficient to explain 90% of the total variance as opposed to the same frequency component explaining only 0.6% in WPD_{open} and 1.3% in WPD_{closed} equilibrium simulations, in which there is not targeted force acting on this frequency.

The same procedure is applied on the C_α atomic displacements of Gln262 (Figure 5.3d-f). Raw coordinates do not clearly show the periodic fluctuations at the fundamental frequency, which can be clearly be seen in the filtered coordinates. Though the percent variance explained by the single base frequency values are smaller than those of Asp181, percent explanation of ~9% is significant when compared to 3% in WPD_{open} and 0.2% WPD_{closed} form simulations at 0.2 GHz.

Contributions of higher harmonics of the base frequency on the C_α atomic displacements are shown in Figure 5.4. In Figure 5.4a, fluctuations at 0.4 GHz, which is the first harmonic of the base frequency, may not seem to represent the movement of the C_α atomic displacement of Asp181 on x direction well (~6.5% explained). However, superposition of 0.2 GHz (base frequency) and 0.4 GHz fluctuations (Figure 5.4b) improves the conformity of the raw coordinates with the filtered data (85% explained), as can be seen with comparing with Figure 5.3a. Figure 5.4c shows the C_α atomic displacements of Ser151 only at base frequency in z-direction. Though the base frequency is sufficient to explain the fluctuations of the raw trajectory data to some extent (17%), it is possible to better capture the characteristics of the motion (21%) with the inclusion of the higher harmonics (0.4, 0.6 and 0.8 GHz). Therefore, a single frequency may not be sufficient to explain the fluctuations of some residues. It is important to emphasize that the conformational transitions seen at ~2 ns and ~26 ns (Figure 5.4c, d), responsible for most

of the observed MSF, can be explained by lower frequencies, which are, quite possibly, not related with the opening/closing motions of the WPD loop, but a consequence of stochastic dynamics of the protein. The advantage of the method suggested in this study is that analysis of the fluctuations on the base frequencies, derived from a sufficient number of WPD loop opening/closing motions, would separate these random fluctuations from the deliberately perturbed frequencies.

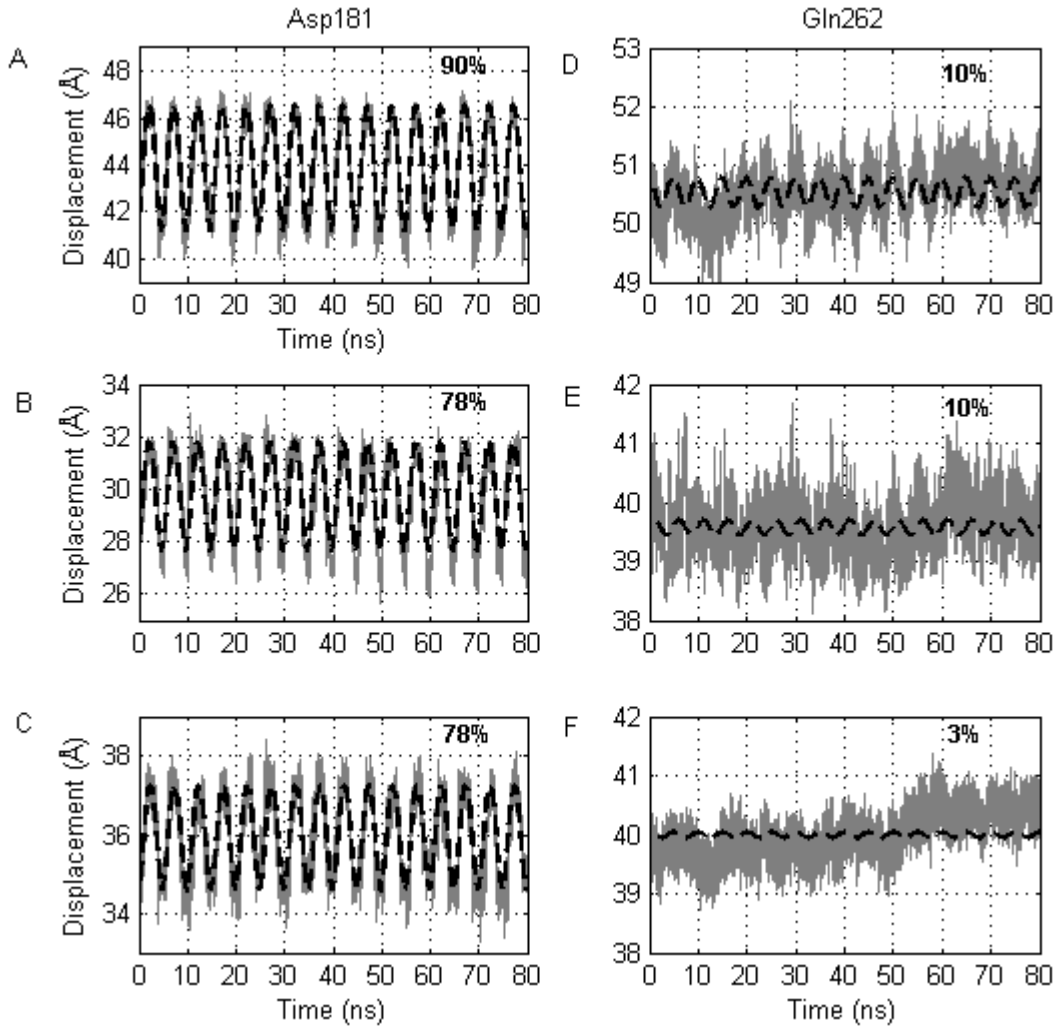

Figure 5.3. C_α atomic displacements of Asp181 on the WPD loop in (a) x, (b) y and (c) z directions, respectively (grey line) at 0.2 GHz. Cα atomic displacements of Gln262 in (d) x, (e) y and (f) z directions, respectively (grey line) at 0.2 GHz. Component of atomic displacement at 0.2 GHz is shown with black dashed lines in all figures.

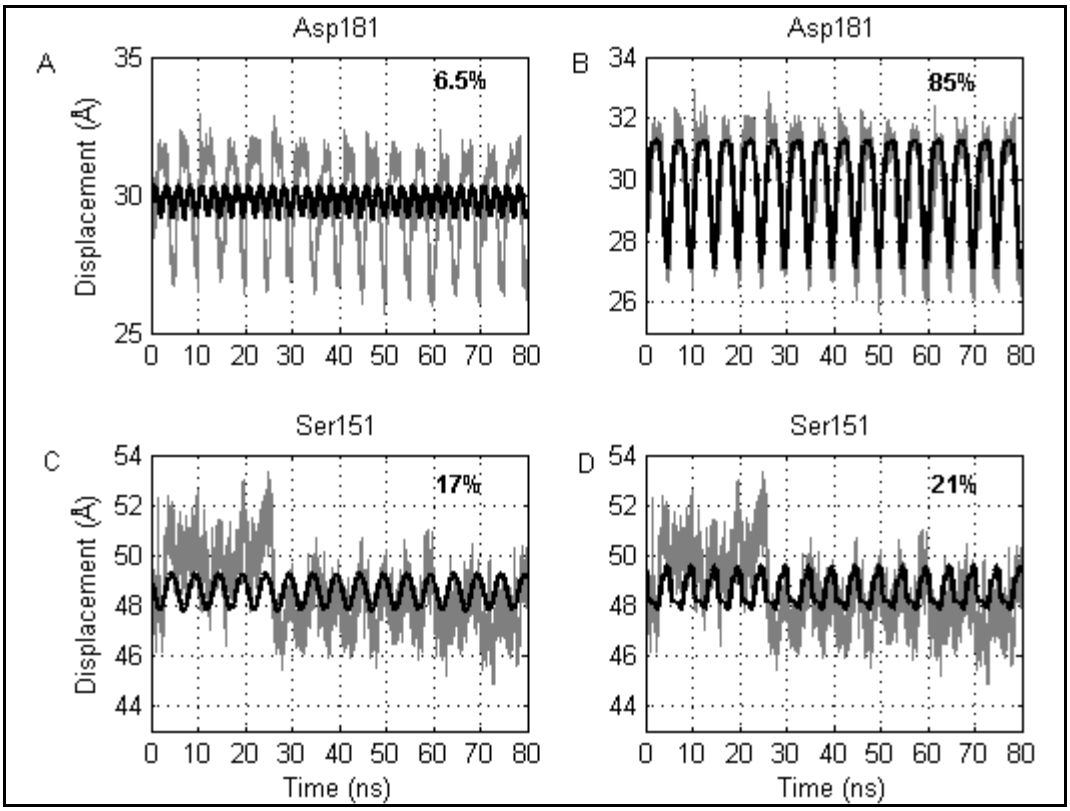

Figure 5.4. Contributions of higher harmonics of the base frequency on the C_α atomic displacements. (a) C_α atomic displacement of Asp181 in x-direction (grey line), and component at 0.4 GHz (black line). (b) Superposition of 0.2 GHz (base frequency) and 0.4 GHz fluctuations on the x direction of Asp 181. (c) Cα atomic displacements of Ser151 at 0.2 GHz in z-direction (d) Superposition of 0.2 GHz, and higher harmonics 0.4, 0.6, and 1.0 GHz fluctuations in z-direction of Ser151.

5.3. Convergence of Residue Fluctuations at Base and Upper Harmonic Frequencies

In order to investigate whether the atomic fluctuations on the perturbed frequencies in TMD simulations converge or not, 80 ns TMD1 simulation is divided into 10 ns segments. In the analysis, only C_α atomic coordinates of the protein are considered. The base frequency of 10 ns segments of TMD1 is 0.2 GHz (=1 cycle/5 ns). Initially, raw coordinates of C_α atoms of each segment were filtered at 0.2 GHz and the mean and range of residue based RMSF were plotted (Figure 5.5). RMSF of different segments show similarity for most of the residues, showing the homogeneity of segments of TMD1. Helix α1'-α2' and α7 are the only regions showing significant difference between different

segments. Standard deviation of the RMSF in different simulation segments of α7 is ~0.4-1.0 Å, while RMSF of the rest of the protein reaches a maximum value of 0.2 Å. This indicates that the dynamics of α7 is not homogeneous in time, confirming the previous finding that dynamics of α7 do not converge in ns scale simulations [12]. The α7 helix is found to be disordered also in crystallographic studies [5].

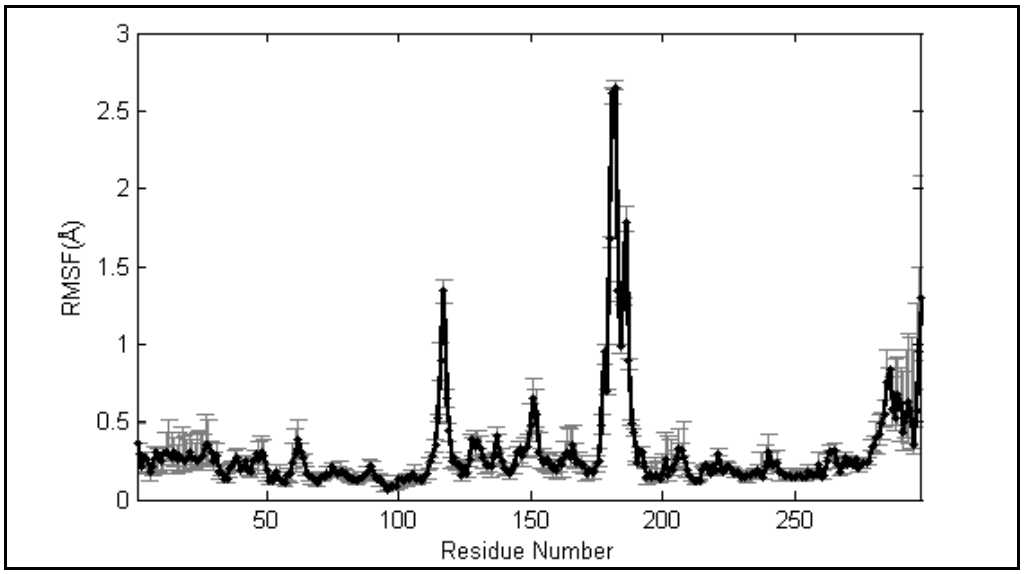

Figure 5.5. Mean (solid black lines) and range (thin grey lines) of residue based RMSF of 10 ns segments of TMD1 simulation at 0.2 GHz.

RMSF on the second and third harmonics of the fundamental frequency were also examined. Standard deviations of residues for upper harmonics showed (figure not shown) a similar result to what has been found for base frequency: RMSF of α7 differ for different time segments on the second and third harmonics. Therefore, residues after Phe280 were not taken into consideration in the following correlation analysis.

To examine the convergence of residue fluctuations at different frequencies, correlations of RMSF in different time segments are computed and shown as correlation maps (Figure 5.6). Figure 5.6a, and Figure 5.6d represent the analysis at 0.2 GHz; Figure 5.6b, and Figure 5.6e show the analysis at 0.4 GHz; Figure 5.6c, and Figure 5.6f indicate the analysis at 0.6 GHz. Segment number 1 corresponds to TMD1 simulation between 0-10 ns, segment 2 corresponds to TMD1 between 10-20 ns, and so on. A correlation value

close to unity shows that the mobility of residues at that given frequency is similar for two different time segments, and is a measure of convergence. Correlation analysis was performed for the whole protein (excluding α7, as previously stated) and was repeated excluding the R-loop and WPD loop. These two regions were directly steered in TMD simulations, thus they are likely to inflate the correlation values, irrespective of the motions of the rest of the protein.

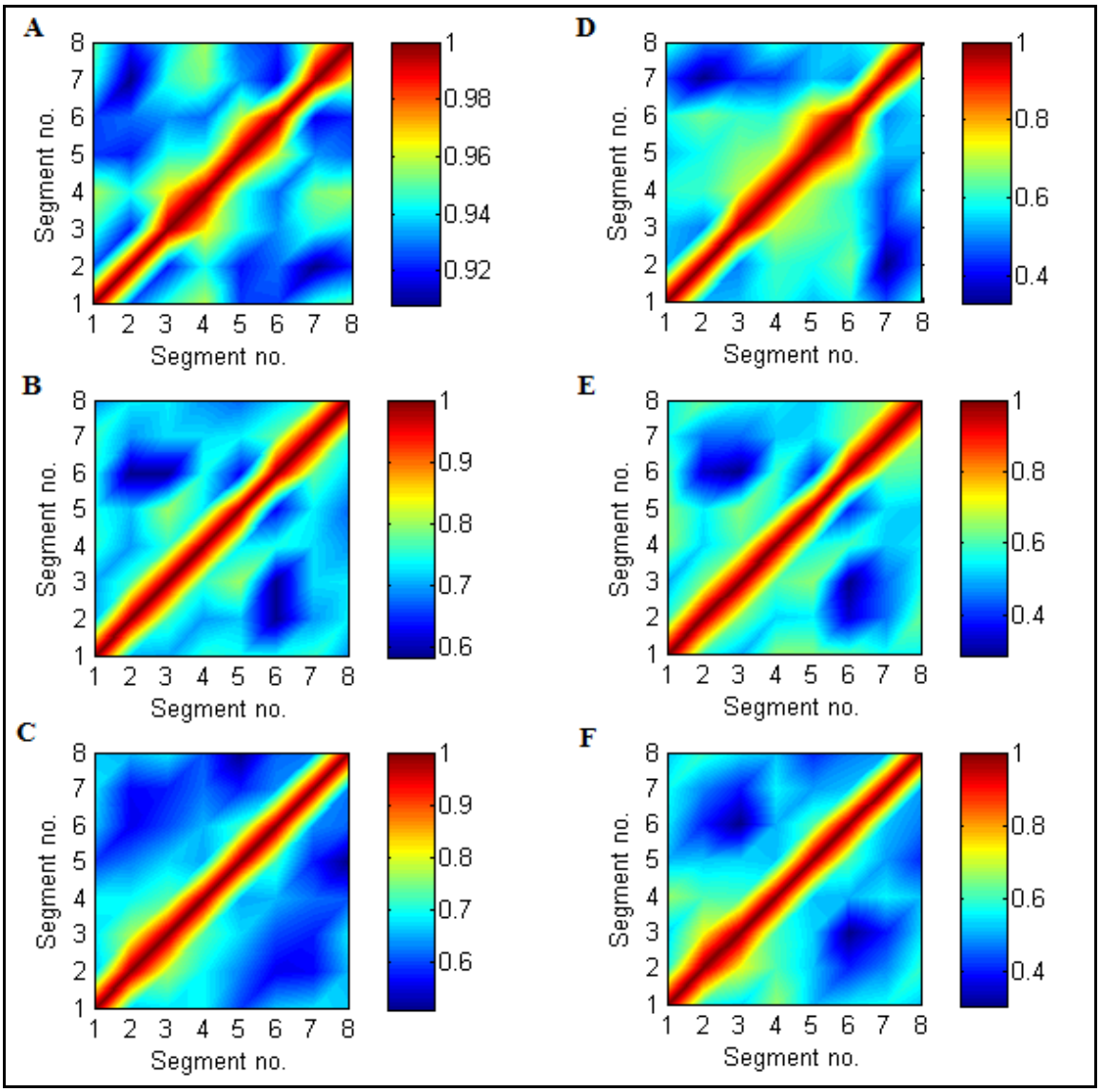

Figure 5.6. Correlation maps of residues from Met1 to Phe280 at (a) 0.2 GHz, (b) 0.4 GHz, (c) 0.6 GHz, correlation maps of the same residues except R-loop, WPD loop and α7 at (d) 0.2 GHz, (e) 0.4 GHz, and (f) 0.6 GHz.

The first column in Figure 5.6 (Figure 5.6a-c) shows the correlation analysis including WPD and R-loops, while the second column in Figure 5.6 (Figure 5.6d-f) represents the correlation analysis excluding R-loop and WPD loop. Comparisons of the two columns indicate that correlations of time segments show a higher degree of similarity when WPD and R-loops are included. Unfortunately, one cannot see a good convergence in correlation values (~0.5) as higher indexed time segments are considered, showing that 10 ns (2 cycles) is too short to obtain reliable information about the residue dynamics. Moreover, correlation values between segments indexed 6, 7 and 8 (within 60-80 ns) are lower compared to other parts of TMD1.

Visual examination of the simulation shows that α7 makes a large conformational jump away from its conformations in WPD_{open} and WPD_{closed} crystal structures after 60 ns (Figure 5.7), and it may be suggested that this conformational transition may have affected the dynamics of the rest of the protein. RMSD of α7 with respect to WPD_{open} and WPD_{closed} crystal structures for TMD2 and TMD3 simulations can also be found in Appendix C, Figure C.1 and Figure C.2).

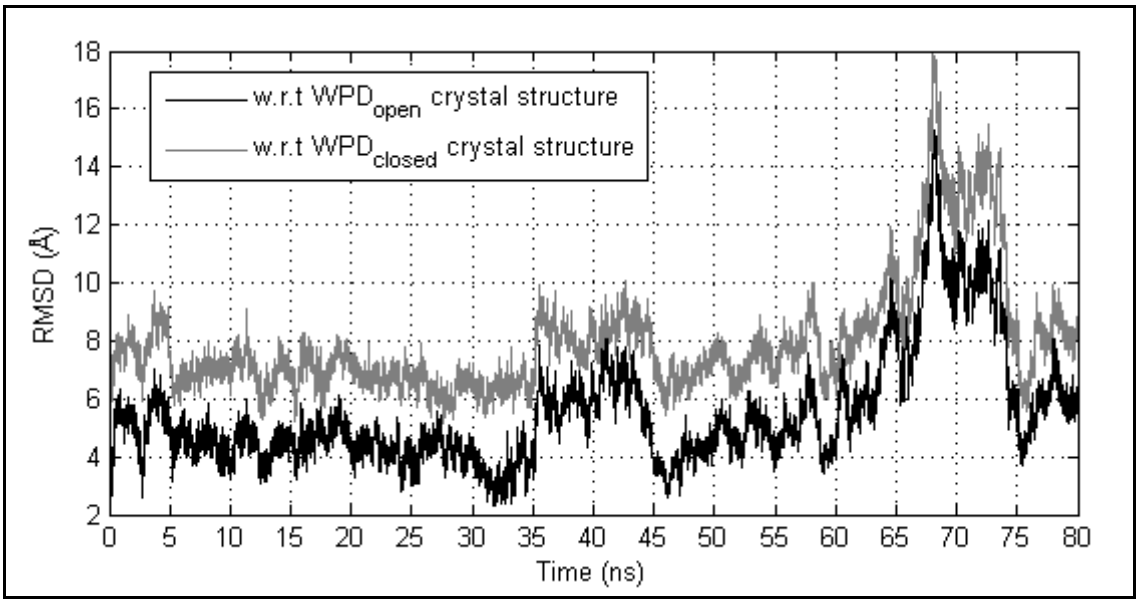

Figure 5.7. RMSD of α7 with respect to WPD_{open} and WPD_{closed} crystal structures.

The average of the correlation values between different time segments of residues is shown in Table 5.1, summarizing the results from Figure 5.6. Repeating the same analysis

for 20 ns segments shows that correlation values of residue RMSF on the first three frequency harmonics increase compared to those for 10 ns segments (Table 5.2), showing that convergence of the dynamics is improved during 20 ns periods.

Table 5.1. Average of the correlation values for 10 ns segments of TMD1 at the fundamental frequency and its second and third harmonics.

Frequency	Mean Correlation Values for 10 ns segments	
	Met1 to Phe280	Met1 to Phe280 (excluding R-loop and WPD loop)
0.2 GHz	0.93	0.56
0.4 GHz	0.72	0.54
0.6 GHz	0.64	0.53

Table 5.2. Average of the correlation values for 20 ns segments of TMD1 at the fundamental frequency and its second and third harmonics.

Frequency	Mean Correlation Values for 20 ns segments	
	Met1 to Phe280	Met1 to Phe280 (excluding R-loop and WPD loop)
0.2 GHz	0.96	0.66
0.4 GHz	0.84	0.64
0.6 GHz	0.69	0.53

If the whole TMD simulation is divided into two (each segment is 40 ns), residue RMSF on the first two harmonics increase substantially, showing a convergence of these motions in ns time scale (Figure 5.8a, b). It is seen that RMSF of α1', α2' region is also quite similar, showing that this region may show different responses in each WPD loop

opening/closing motions, but 40 ns is sufficient to span the conformational space of this region. In order to verify the convergence of the TMD1 simulation in 40 ns, the RMSF of the whole TMD1 (red line) was also plotted on the two segments of TMD1 simulation (Figure 5.8). It is seen that RMSF values are highly similar. Therefore, it can be said that protein dynamics converge during the course of the TMD1 simulation. On the other hand, it is seen that mobility of α7 is different on the base frequency for the two halves of the simulation (as shown in Figure 5.7). Residue fluctuations on the second harmonic are also quite similar for the two 40 ns halves. The third harmonic shows a lower convergence (Figure 5.8c). Correlation of the residue fluctuations between the two simulation halves are shown in Table 5.3 for the first five harmonics. It is important to note that correlation values generally decrease as higher harmonics are considered, showing that noise corrupts higher harmonics more easily.

Table 5.3. Average correlation values for 40 ns segments of TMD1 at the fundamental frequency and its higher harmonics.

Frequency	Mean Correlation Values for 40 ns segments	
	Met1 to Phe280	Met1 to Phe280 (excluding R-loop and WPD loop)
0.2 GHz	0.98	0.81
0.4 GHz	0.91	0.71
0.6 GHz	0.74	0.49
0.8 GHz	0.73	0.54
1.0 GHz	0.77	0.65

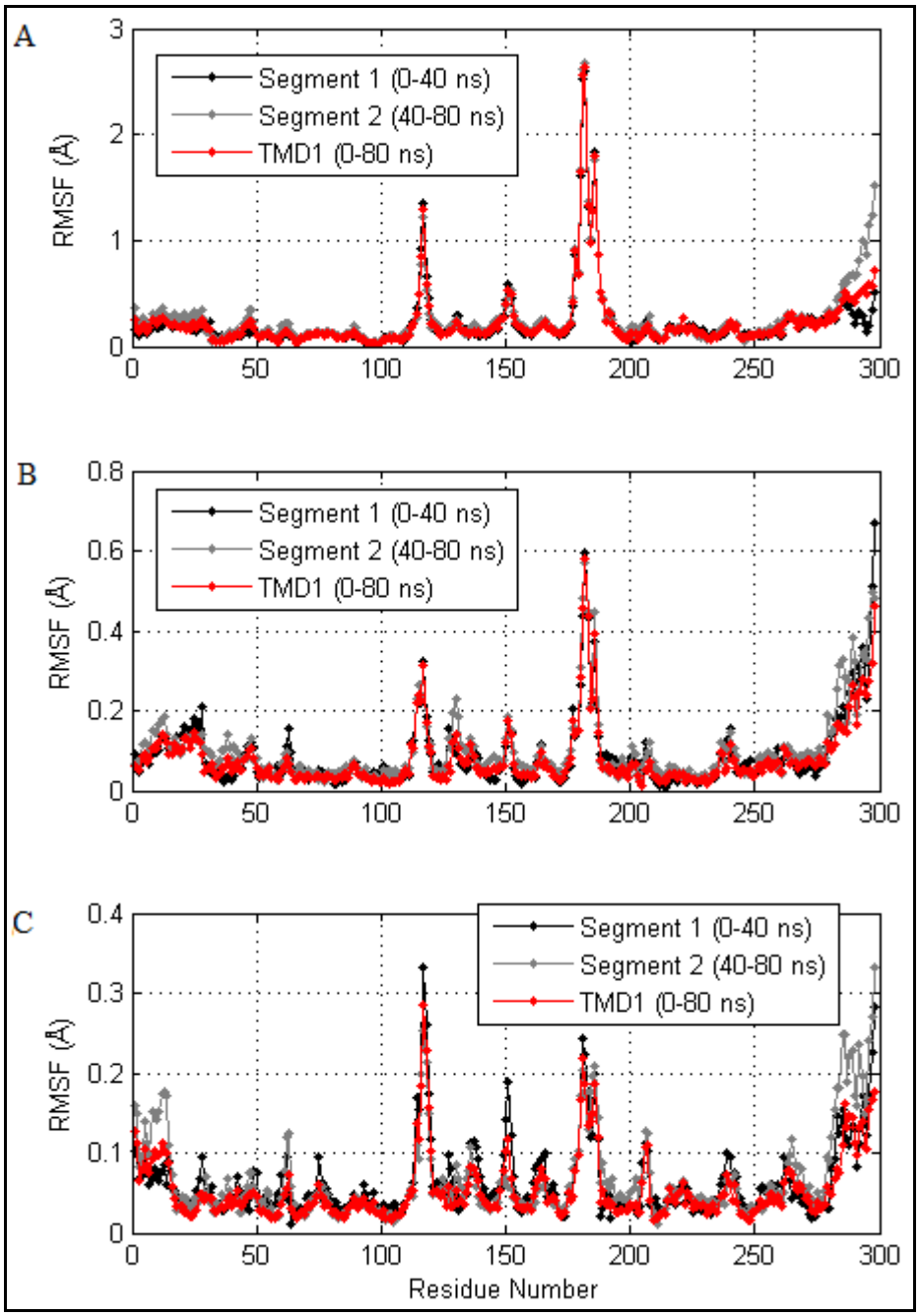

Figure 5.8. Residue RMSF of 40 ns segments of TMD1 simulation on the fundamental frequency (a) 0.2 GHz, (b) its second 0.4 GHz, and (c) its third 0.6 GHz harmonics.

These results show that long TMD simulations are required to obtain reliable estimates of the residue dynamics on the frequency harmonics, and it is not required to truncate an initial portion of the TMD simulation, similar to the truncation of a transient period until steady state is reached for conventional industrial processes.

Correlation analysis was also applied on the WPD loop to examine the fluctuations on the perturbed and harmonic frequencies. It is seen that the highest correlation is obtained at the fundamental frequency 0.2 GHz for each time segments, since the force is applied on WPD loop at this frequency. Correlation of WPD loop decreases with higher harmonics. On the other hand, there is still a high convergence at the higher harmonics showing that magnitudes of residue displacements on 0.4 and 0.6 GHz are similar, particularly in the two halves of the 40 ns length TMD simulations (Table 5.4). This suggests the possibility that the current TMD simulation (80 ns) is able to span the functionally important conformational space of the WPD loop residues.

Table 5.4. Average correlation values of WPD loop for different time segments of TMD1 at the main frequency and its second and third harmonics.

Frequency	Mean Correlation Values for WPD Loop		
	10 ns segments	20 ns segments	40 ns segments
0.2 GHz	0.99	1.00	1.00
0.4 GHz	0.88	0.96	0.98
0.6 GHz	0.61	0.63	0.83

5.4. Analysis of residue fluctuations in PTP1B on the harmonic frequencies

While perturbing the protein by forcing the WPD loop and R-loop in PTP1B, other regions of the protein are expected to be perturbed by this periodic input. One of the consequences of the periodic input on PTP1B may be to increase of residue displacements on the harmonic frequencies. In order to analyze the residue fluctuations in regions other than WPD loop and R-loop, RMSF of all C_α atoms in TMD1 were plotted on fundamental frequency and upper harmonics, and compared with those obtained from equilibrium MD_{open} and MD_{closed} simulations. Increase in the mobility of residues in TMD1 compared to equilibrium MD_{open} and MD_{closed} simulations may be attributed to the perturbation of the dynamics of these residues by the opening/closing motions of the WPD loop.

Figure 5.9 shows the RMSF of all C_α atoms at the fundamental frequency (0.2 GHz) for TMD1 (black), MD_{open} (red) and MD_{closed} (blue) simulations. Neighboring frequencies (0.19 GHz and 0.21 GHz) are taken into consideration to obtain more reliable estimates of RMSF of residues on these frequencies. RMSF graphs are plotted in logarithmic scale to make the visual examination in finer detail. It is seen that highest RMSF values are obtained for the R-loop and WPD loop, as expected. It is interesting that there are numerous regions showing higher mobility in TMD1, such as α1', α2', pTyr recognition loop, R-loop, L11, WPD loop, α3, P-loop, Q-loop and α6. The increase of fluctuations in these regions in TMD1 may be due to the information flow from the R-loop and WPD loop to the rest of the protein during WPD opening/closing. In other words, the effects of WPD loop opening/closing motion is not confined to a local active site region; the force applied on local regions due to ligand binding (mimicked by loop opening/closing) spreads through the whole protein and has global impacts on the protein.

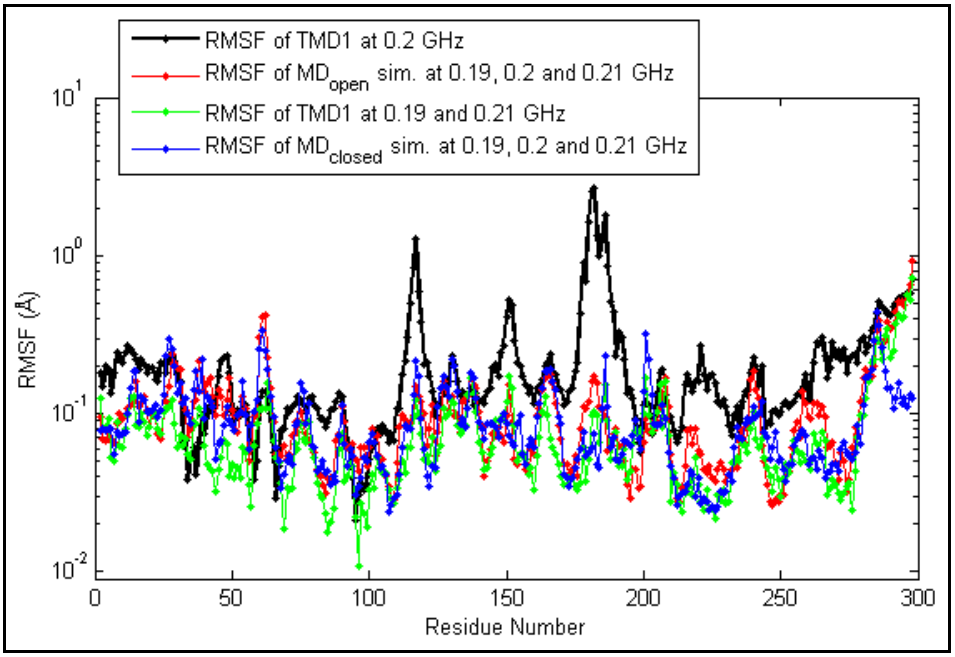

Figure 5.9. RMSF of C_α atoms at the base frequency 0.2 GHz in TMD1 (black), in MD_{open} (red) and MD_{closed} (blue) simulations at 0.2 GHz, and neighboring frequencies in TMD1 (green).

To identify the highly fluctuating regions of PTP1B due to WPD loop opening/closing, RMSF of residues in TMD1 were subtracted from the maximum RMSF

obtained in equilibrium MD simulations (Figure B.1). To see the distribution of residues with increased mobility on the three-dimensional structure of the protein, and to compare the magnitudes of these increases, structure of PTP1B was colored with respect to the increase of RMSF (Figure 5.10). Color transition is from blue (high increase in fluctuations) to red (negligible increase in fluctuations), and white represents moderate increase in fluctuations. The highest increase in fluctuations is seen in WPD loop and R-loop (totally colored with blue), since TMD force is directly applied on these regions. It is seen that regions in the vicinity of WPD loop usually have higher mobility in the presence of TMD force (WPD opening/closing). Of these regions, L11, N-terminus of α3 have the highest mobility increase, followed by α6 and N-terminus of α7. A moderate increase (colored with white) in mobility is seen in pTyr recognition loop, P-loop, and Q-loop, which surround the active site. This shows that a force applied only on WDP loop and R-loop increases the fluctuations of all residues in the active site.

In general, as one moves away from the active site, the effect of periodic force diminishes. The β strands on the N-terminus of the WPD loop, namely β9, β10, β11, have increasing mobility, as one approach to the WPD loop (colored from red to blue). On the other hand, regions on the outer part of PTP1B, (distant from WPD loop, and particularly behind the P-loop) namely α1, β1, β2, α2, regions of β7 in the vicinity of loop regions, and loops connecting these elements, have negligible or no increase in their fluctuations (red colored). It should be noted that there is no significant increase in the mobility of S-loop, either. One may, at first sight, tend to think that regions with negligible or no mobility increase are not affected by WPD loop opening/closing motions. However, as will be shown later, mobility increase may not be the only indicator of information transfer through the protein.

Another important issue that should be noted is the increase in residue fluctuations in distant region of PTP1B, namely α1', α2', β3, α4 and α5 regions. These regions are not directly in contact with the active site, i.e. the nearest distance between the C_α atoms of WPD loop and Lys12 (on α1'), which has a significant increase in mobility in TMD1 simulation, is ~17 Å. On the other hand, these regions reside on the right side of the WPD loop (from the given perspective in Figure 4.10), and all of these regions have nonbonded interactions with the P-loop, α6 and each other (Figure 5.11). Thus, it may be argued that

nonbonded interactions between these regions and those at the active site, directly in contact with the WPD loop and R-loop, are responsible from information transfer between these regions.

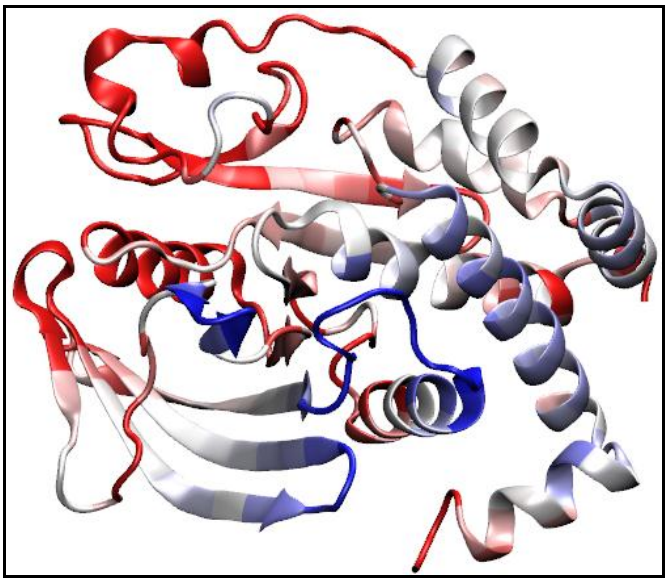

Figure 5.10. Distribution of residues with increased mobility on the three-dimensional structure of PTP1B at 0.2 GHz. Residues, which show an increase greater than 0.4 Å in RMSF is shown in blue, and those smaller than 0.02 Å in shown red, and the mapping of the increase in RMSF to blue-red coloring is logarithmic.

The same analysis was repeated for the second and upper harmonics. Figure 5.12 shows RMSF of the C_α atoms at 0.4 GHz, the second harmonic of the fundamental frequency. RMSF of C_α atoms vibrating at 0.4 GHz are generally smaller than those at 0.2 GHz. The highest RMSF are observed at WPD loop and R-loop. Other regions showing an increase in RMSF are α1', α2', L11, Q-loop and α6. Similar to Figure 5.10, residues with increased mobility at 0.4 GHz are shown on the three-dimensional structure of the protein (Figure 5.13). It is seen that residues with increased mobility at 0.4 GHz are, as in Figure 5.10, centered around the WPD loop, but the range of the TMD force effective at 0.4 GHz spans a smaller region compared to those in Figure 5.10 (at 0.2 GHz). This shows that low frequency motions may spread more efficiently in proteins. Regions α1' and α2' require further attention, for these two alpha helices are the only regions, which have higher mobility at both 0.2 GHz and 0.4 GHz, and do not reside in the vicinity of the WPD loop

and R-loop. This may indicate a correlation between the mobility of α1' and α2' and WPD loop motions.

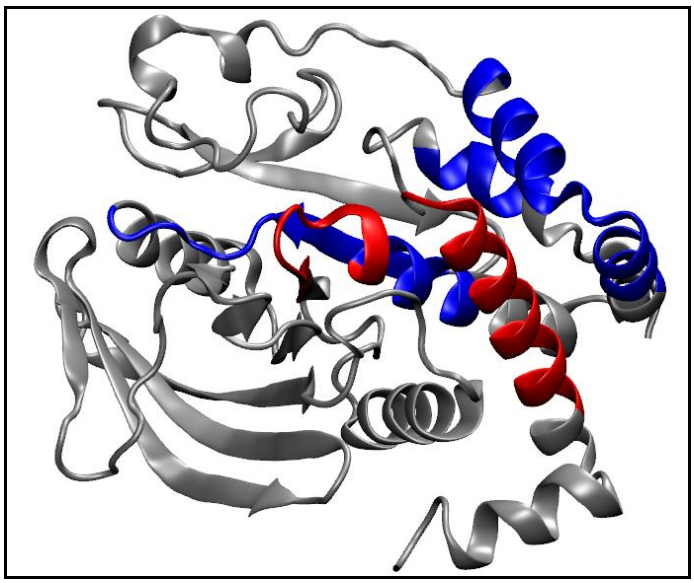

Figure 5.11. Representation of regions on the three-dimensional structure of PTP1B that have nonbonded interactions with the P-loop, α6 and each other. Red denotes P-loop, α6, while blue shows the regions making nonbonded interactions with these two regions.

Figure 5.14 shows the sum of the RMSF of TMD1, MD_{open} and MD_{closed} simulations at upper harmonics (0.6, 0.8 and 1.0 GHz), and Figure 5.15 shows the representation of these regions on the structure of PTP1B. Similar to what is seen at 0.2 GHz, increase in residue fluctuations are seen at α1', α2', R-loop, L11, WPD loop, P-loop and α6. Similar to the character of fluctuations seen at 0.4 GHz, the range of the TMD force effective at 0.6-1 GHz spans a smaller region compared to that at 0.2 GHz. Different from 0.4 GHz, active site of PTP1B has higher increase in its mobility.

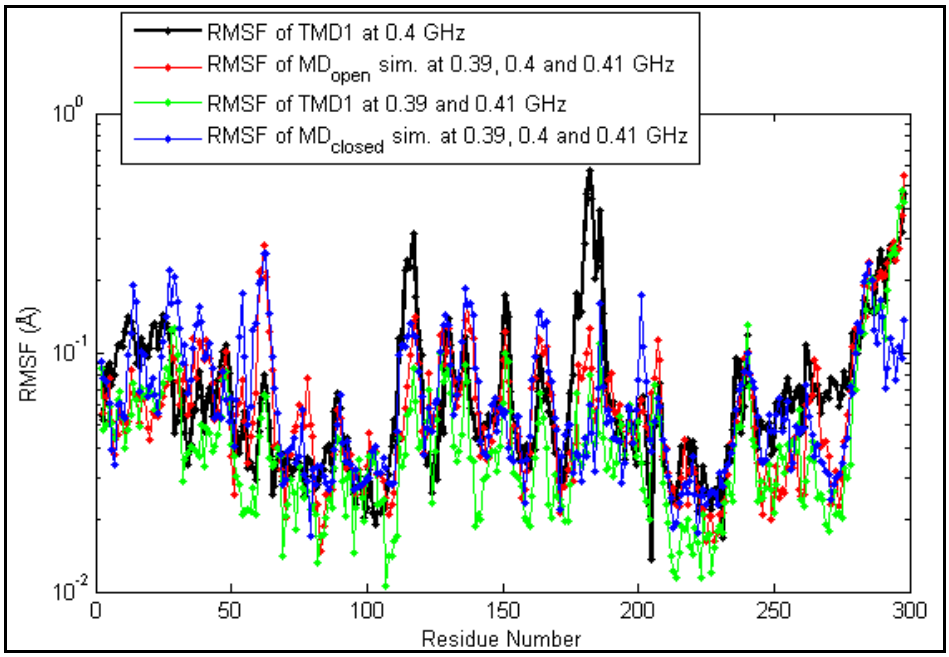

Figure 5.12. RMSF of C_α atoms at the second harmonic 0.4 GHz in TMD1 (black), in MD_{open} (red) and MD_{closed} (blue) simulations at 0.4 GHz and neighboring frequencies in TMD1 (green).

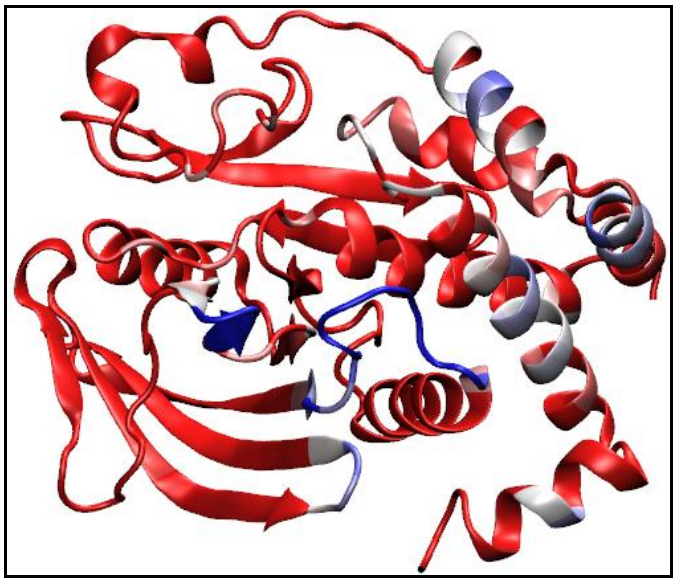

Figure 5.13. Distribution of residues with increased mobility on the three-dimensional structure of PTP1B at 0.4 GHz. Residues, which show an increase greater than 0.1 Å in RMSF is shown in blue, and those smaller than 0.01 Å in shown red, and the mapping of the increase in RMSF to blue-red coloring is logarithmic.

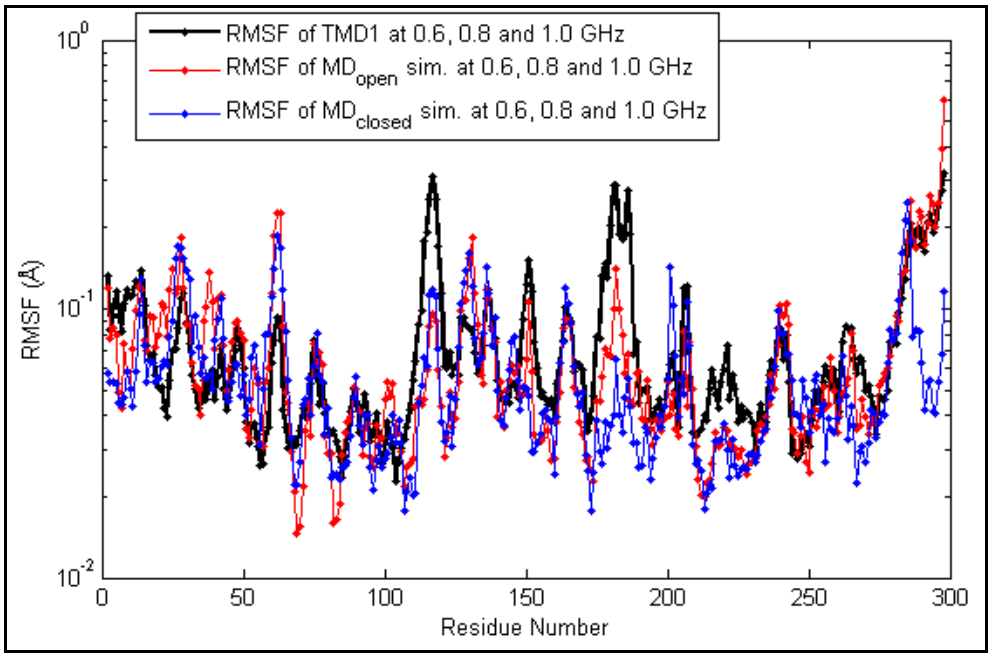

Figure 5.14. RMSF of C_α atoms for TMD1 (black), MD_{open} (red) and MD_{closed} (blue) simulations at 0.6, 0.8 and 1.0 GHz.

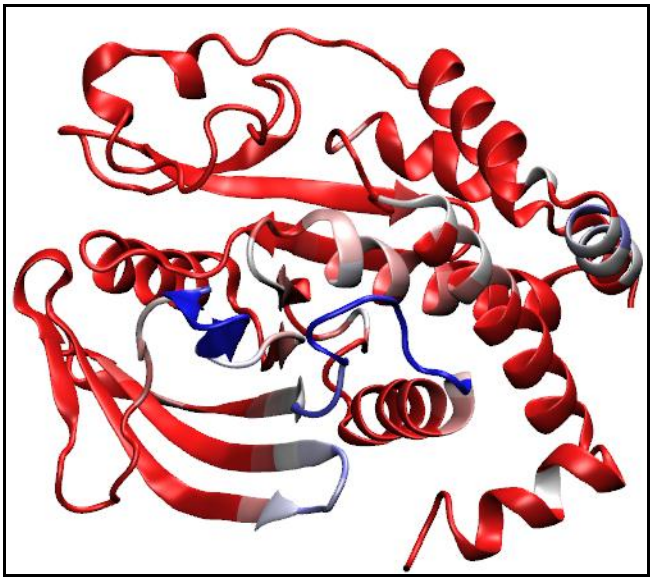

Figure 5.15. Distribution of residues with increased mobility on the three-dimensional structure of PTP1B at 0.6, 0.8, and 1.0 GHz. Residues, which show an increase greater than 0.1 Å in RMSF is shown in blue, and those smaller than 0.01 Å in shown red, and the mapping of the increase in RMSF to blue-red coloring is logarithmic.

In summary, α1', α2', R-loop, L11, WPD loop and α6 increase their fluctuations on the harmonic frequencies examined as a consequence of the periodic force applied on local regions of PTP1B. It is important to note that higher fluctuations are seen not only in the vicinity of WPD loop and R- loop, but also at some distant parts of PTP1B, not directly in contact with the active site or WPD loop, i.e. α1', α2'. RMSF of C_α atoms for TMD2, MD_{open} and MD_{closed} simulations at the first five harmonics can also be found in Appendix C, Figure C.3, Figure C.4, and Figure C.5.

5.5. Comparison of Residue Displacements in Experimental and TMD Structures

In the previous section, increased residue fluctuations due to TMD force at 0.2 GHz and at upper harmonics were determined. However, these fluctuation predictions were not tested with the available experimental data. In this section, computational results obtained by the current frequency analysis are to be compared with experimental findings. Residue displacements between WPD_{open} and WPD_{closed} crystal structures were compared with those obtained from TMD1 simulation only at 0.2 GHz, which represents a full cycle of WPD loop opening/closing motion (Figure 5.16 in logarithmic scale, Figure B.2 in decimal scale). It is seen that most of the residue displacements (between open and closed conformations) at 0.2 GHz in TMD1 are similar to those between crystal structures. Mean residue displacement of TMD1 simulation, and the correlation between experimental and simulation structures of TMD1 is 0.58 and 0.54 at 0.2 GHz, respectively. Excluding the R-loop, and WPD loop, not to increase correlation artificially, α7, due to the nonconvergent dynamics, and Arg238 to Pro241, in which a large difference is seen between the displacements of crystal structures, average residue displacements in TMD1 and X-ray structures are 0.43 Å and 0.38 Å, respectively, and correlation between experimental and computational displacements is 0.56. These results show that the magnitudes of the residue displacements during WPD loop opening (or closing) predicted by the frequency analysis of TMD1 were moderately in agreement with the crystal structures displacements. On the other hand, it should be mentioned that estimated displacements for α2' and P-loop are found to be slightly higher (~0.2 Å) compared to crystal structures.

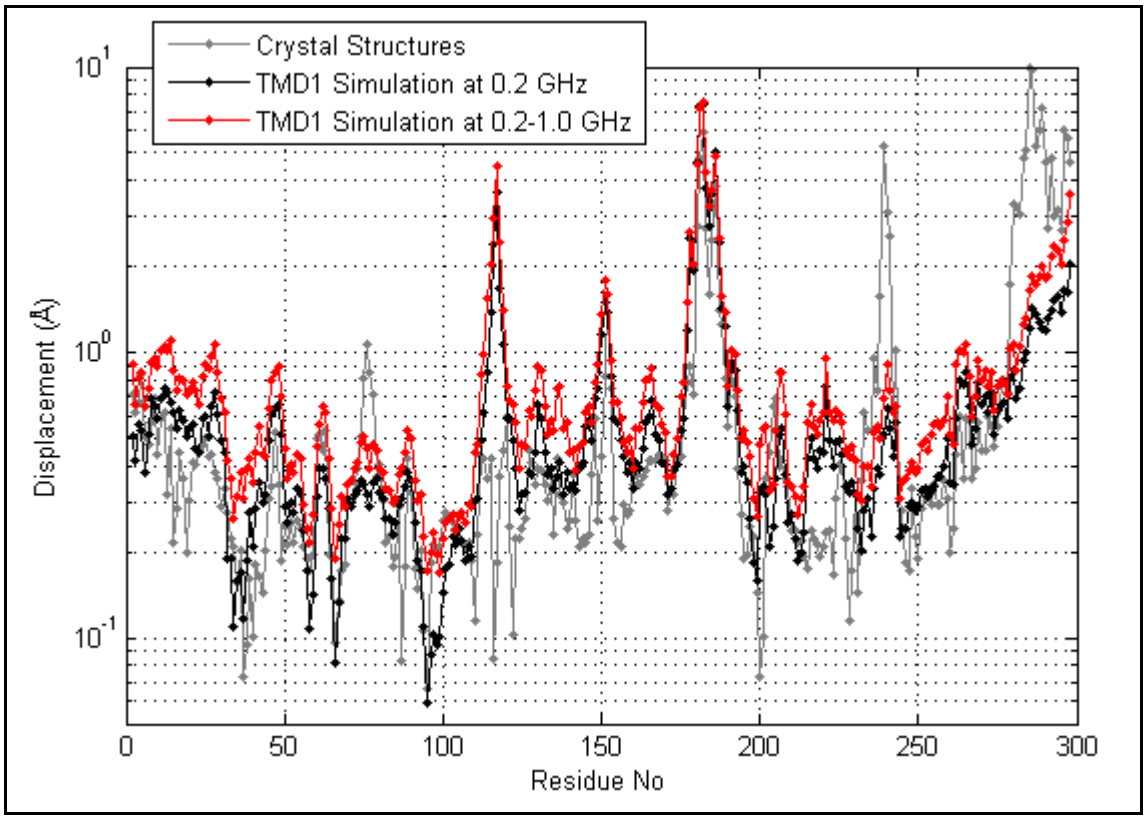

Figure 5.16. Comparison of residue displacements in experimental crystal structures (grey) at 0.2 GHz and TMD1 (black) simulation structures at 0.2-1.0 GHz.

It is also important to check whether the residue displacements at 0.2 GHz are in the same direction with those in the crystal structures. To obtain a single dominating global motion of the protein at 0.2 GHz, PCA was applied on the filtered (at 0.2 GHz) trajectories. A single principal component which explains 86% of the direction of residue displacements is obtained. Then, dot products of the displacement vectors obtained from the first principal component and experimental structures were calculated and divided by the corresponding magnitudes for each residue. Figure 5.17 shows the cosine of angle between the displacement vector of crystal structures and first eigenvector for each residue. Figure B.4 shows the angle cosines of residues between crystal structure displacements and those obtained from the eigenvectors at 0.2, 0.4, 0.6, 0.8 and 1.0 GHz in TMD1. While cosine values close to unity shows that direction of the two vectors are similar, cosine values close to -1 show that the eigenvector and crystal structures are in the opposite direction. Residues showing higher mobility at 0.2 GHz (see Figure 5.9) are marked with different symbols in Figure 5.17. It is seen that the residue displacement vectors of α2',

pTyr recognition loop, L11, P-loop, Q-loop and α6 predicted by frequency analysis are practically in the same direction with those of the crystal structures. Only α1' moves in a relatively different direction to that in seen the crystal structures.

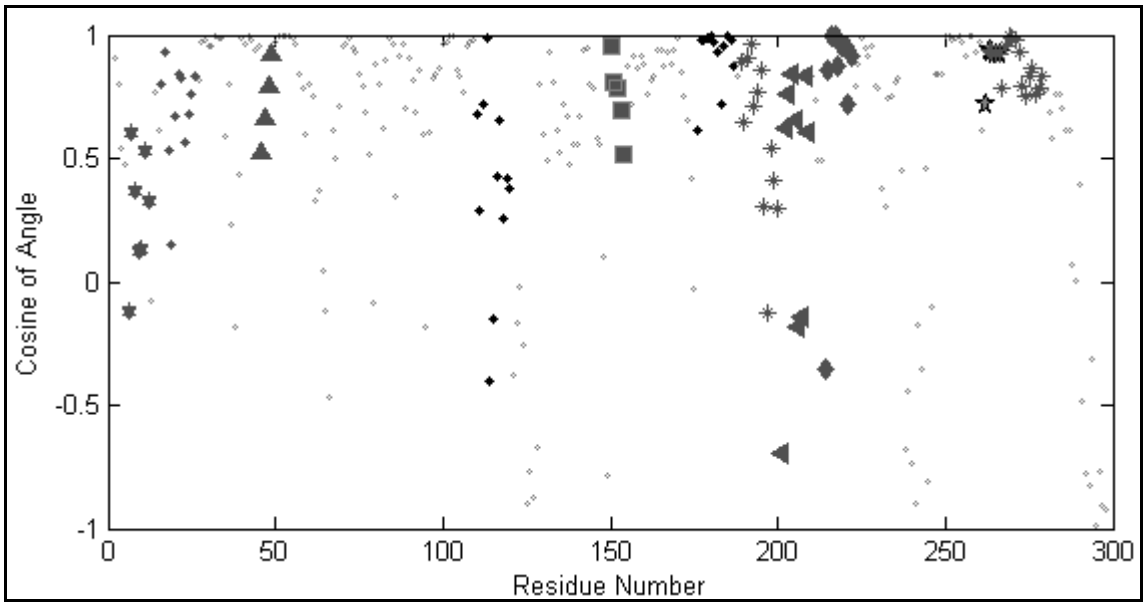

Figure 5.17. Angle cosines of residues between crystal structure displacements and those obtained from the first eigenvector at 0.2 GHz in TMD1.

Probability distribution of the angle cosines of residues (Figure 5.17) between crystal structure displacements and those obtained from the first eigenvector at 0.2 GHz in TMD1, is shown in Figure 5.18a. To see if the local TMD force "tunes" the global protein fluctuations to crystal structure displacements, probability distributions of the angle cosines between the crystal structure displacements and the first eigenvectors obtained from the MD_{open} and MD_{closed} simulations at 0.2 GHz are also shown in Figure 5.18b-c, respectively. It should be noted that R-loop, WPD loop and α7 are excluded from computation of histograms. Probability distribution of the angle cosines of all residues of PTP1B for the first eigenvector of TMD1, MD_{open} and MD_{closed} at 0.2 GHz can be found in Appendix B, Figure B.3. While the angle cosine between crystal structure displacements and TMD1 is more than 0.5 in 78% of the residues (Figure 5.18a), this percentage drops down to 56% and 30% between crystal structure displacements and MD_{open} (Figure 5.18b), and MD_{closed} (Figure 5.18c) simulations, respectively. To check whether an improvement in the direction of the collective motions in the equilibrium MD simulations is obtained in

TMD simulation, PCA is applied to the whole equilibrium MD trajectories. Figure 5.19 shows the probability distribution of the angle cosines of residues between crystal structure displacements and those obtained from the first three eigenvectors in MD_{open} (first column, Figure 5.19a-c) and MD_{closed} (second column, Figure 5.19d-f) simulations. It is seen that similarity of collective motions in MD_{open} simulation to the direction of crystal structure displacements is higher than those in MD_{closed} simulation. On the other hand, even the collective motion on the first eigenvector in MD_{open} simulation is much less similar to the crystal structure displacements, compared to those obtained in TMD simulations (Figure 5.18a). One may say that the local force applied on WPD loop helps "steering" the rest of the protein to its conformations seen in crystal structures.

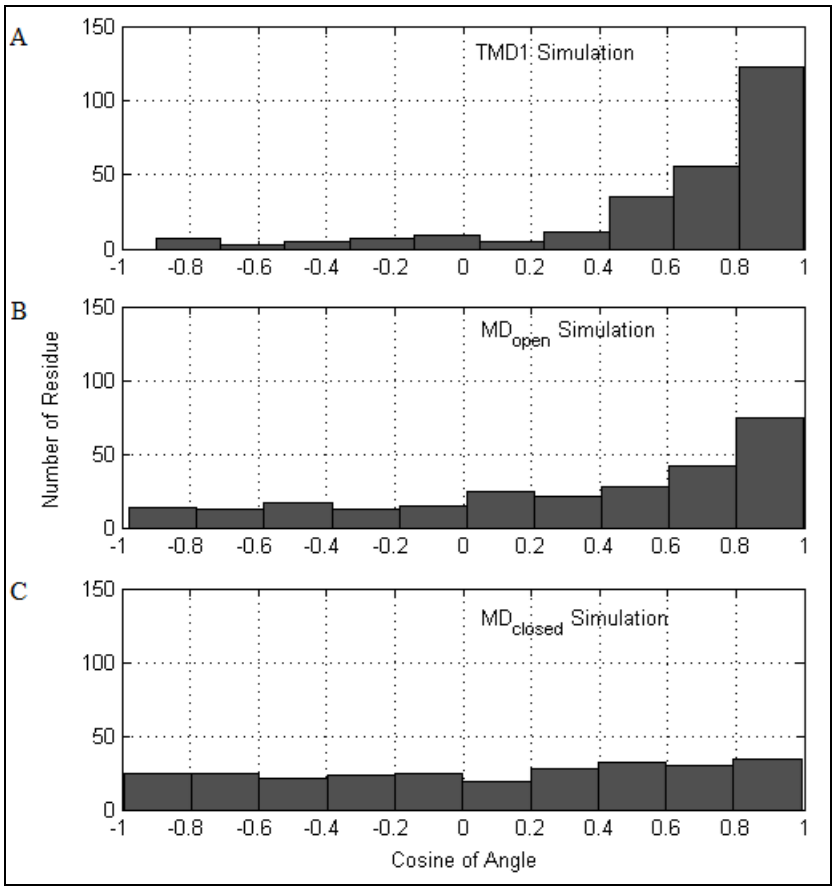

Figure 5.18. Probability distribution of the angle cosines (R-loop, WPD loop and α7 are excluded) between (a) first eigenvector of TMD1 (b) MD_{open} (c) MD_{closed} at 0.2 GHz and crystal structure displacements.

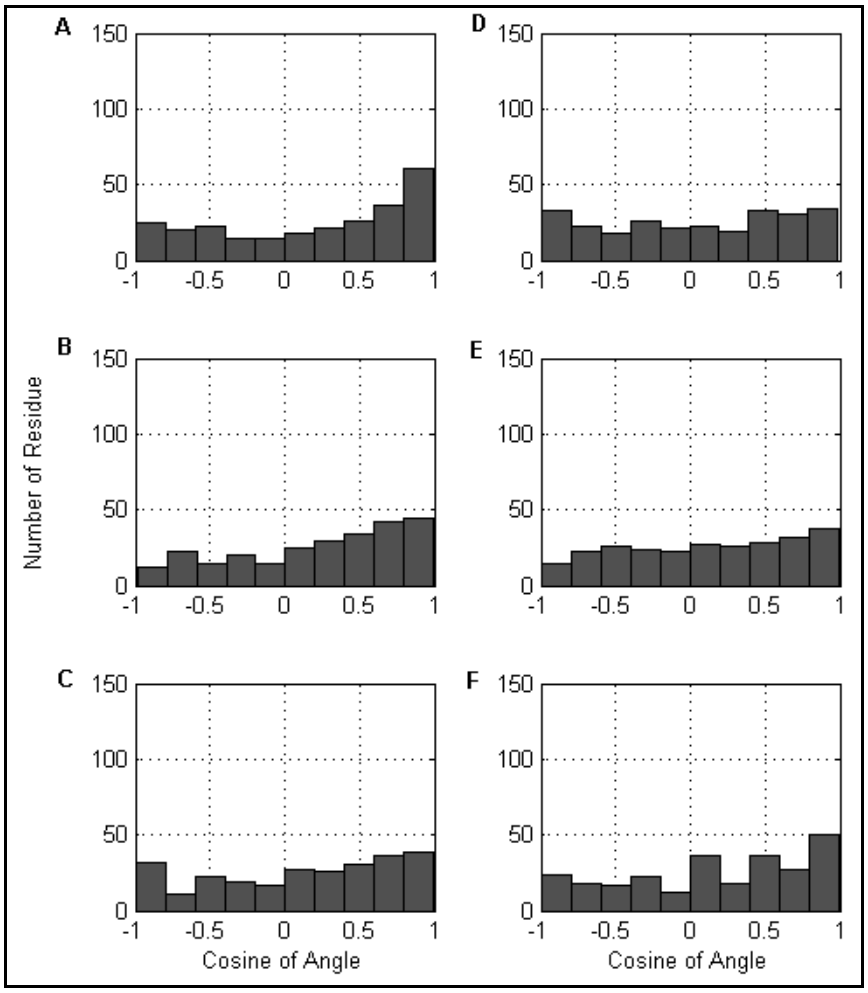

Figure 5.19. Probability distribution of the angle cosines (R-loop, WPD loop and α7 are excluded) between (a) the first, (b) second, and (c) third eigenvectors (explaining 24%, 11% and 7% of the whole fluctuations, respectively) in MD_{open}, and (d) the first, (e) second, and (f) third eigenvectors (explaining 20%, 11% and 5% of the whole fluctuations, respectively) in MD_{closed} simulations, and crystal structure displacements.

The significance of this result can be seen especially in the S-loop, a functionally important region. It should be noted that although there is no mobility increase in S-loop at 0.2 GHz with respect to equilibrium MD simulations (Figure 5.9), opening/closing of WPD loop makes S-loop displacement on the first eigenvector at 0.2 GHz more aligned with that between the crystal structures (Figure 5.20). In this figure, angle cosines between crystal structure displacements and those obtained from the first eigenvector at 0.2 GHz of TMD1, MD_{open} and MD_{closed} simulations of C-terminus of α3 and S-loop are shown. While the mean of the cosine of angles is 0.44 for TMD1, this value is -0.04 and -0.62 for MD_{open}

and MD_{closed} simulations at 0.2 GHz, respectively. This value is 0.16 for the first eigenvector, which represents the highest collective behavior in the crystal structures displacement direction (see Figure 5.19a), of the whole coordinates of MD_{open} simulation. This result shows that the applied periodic force on WPD loop does not increase the fluctuations of S-loop, but it can steer S-loop in the direction of the displacements seen crystal structures.

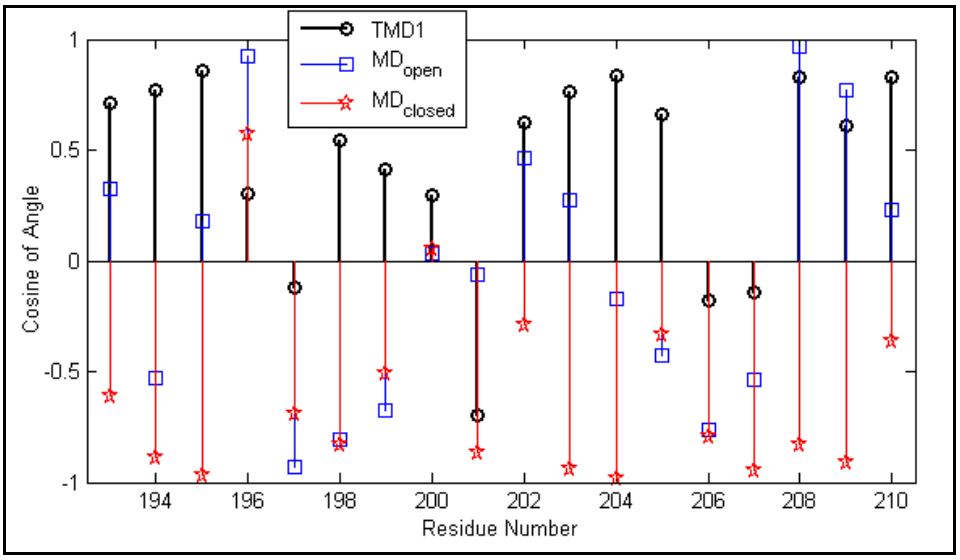

Figure 5.20. Angle cosines of S-loop between crystal structure displacements and those obtained from the first eigenvector in TMD1 (black), MD_{open} (blue) and MD_{closed} (red) simulations at 0.2 GHz.

To see the impacts of the WPD loop opening/closing on the behavior of the whole protein in terms of their direction, angles cosines of the first eigenvector is subtracted from those found in the MD_{open} simulation (Figure 5.21). By this way, it is possible to have an insight about the residues, which are steered in the direction of the crystal structure displacement, or already moving in the opposite or the same direction with the crystal structures. Regions that have angle cosines differences close to zero means they already move in the direction of crystal structures during equilibrium fluctuations. For the residues that are on the positive side of the Figure 5.21, it may be argued that these regions are steered in the direction of the crystal structures more than that in equilibrium MD simulations. It should be noted that, most of the regions of PTP1B, except α1' (from Glu6

to Lys12), moves in the same direction with the crystal structures at 0.2 GHz in TMD1 simulations. Figure 5.22 shows the distribution of residues with respect to the increased/decreased similarity of their displacement directions to those between crystal structures on the three-dimensional structure of PTP1B. In this figure, blue regions represent the residues, whose displacements have a higher dot product with those in crystal structures in TMD1, and displacements of white regions are not different in TMD1 and equilibrium simulations. Red colored residues (rarely seen), on the other hand, correspond to regions which moves more aligned with the crystal structure displacements in equilibrium MD simulations. It should be emphasized that most of the regions with no increased mobility on PTP1B during WPD opening/closing (red colored regions in Figure 5.10) are colored in blue in Figure 5.22. This means, although the applied TMD force does not increase the residue mobility, it tunes the residue motions to the displacements seen between crystal structures. For instance, α1, L1, B1, L2, α1, β7, L9, β8, L10, β9, β10, and S-loop are regions steered in the direction of crystal structures displacements during WPD loop movement. WPD loop, α6, Q-loop, P-loop and some parts of the L11 move in the same direction with crystal structures in both TMD1 and equilibrium MD_{open} simulation, so that is the reason why they are colored white.

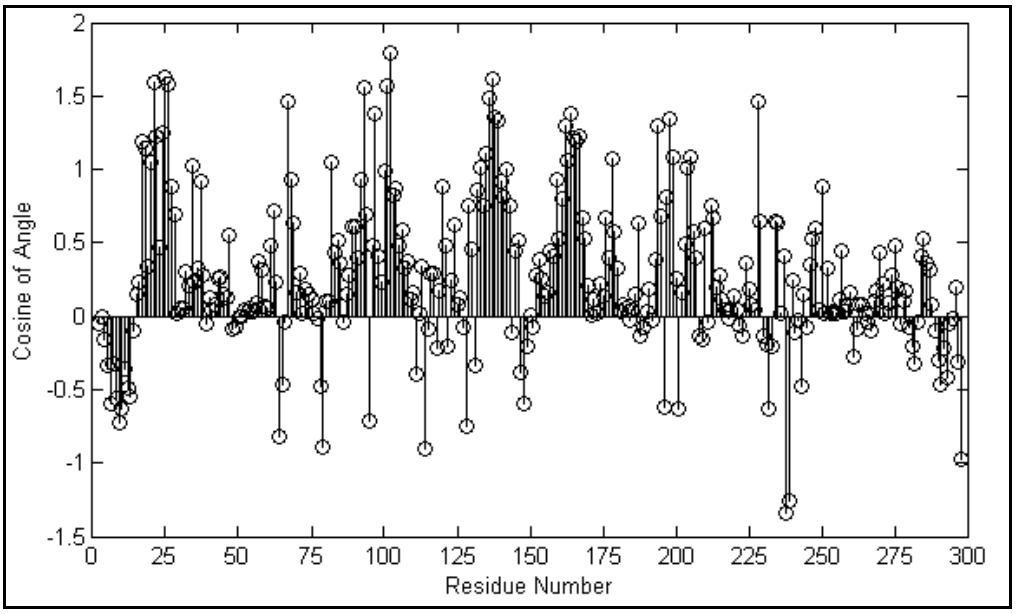

Figure 5.21. Difference of cosine of angles between TMD1 and MD_{open} simulations with displacements in X-ray structures.

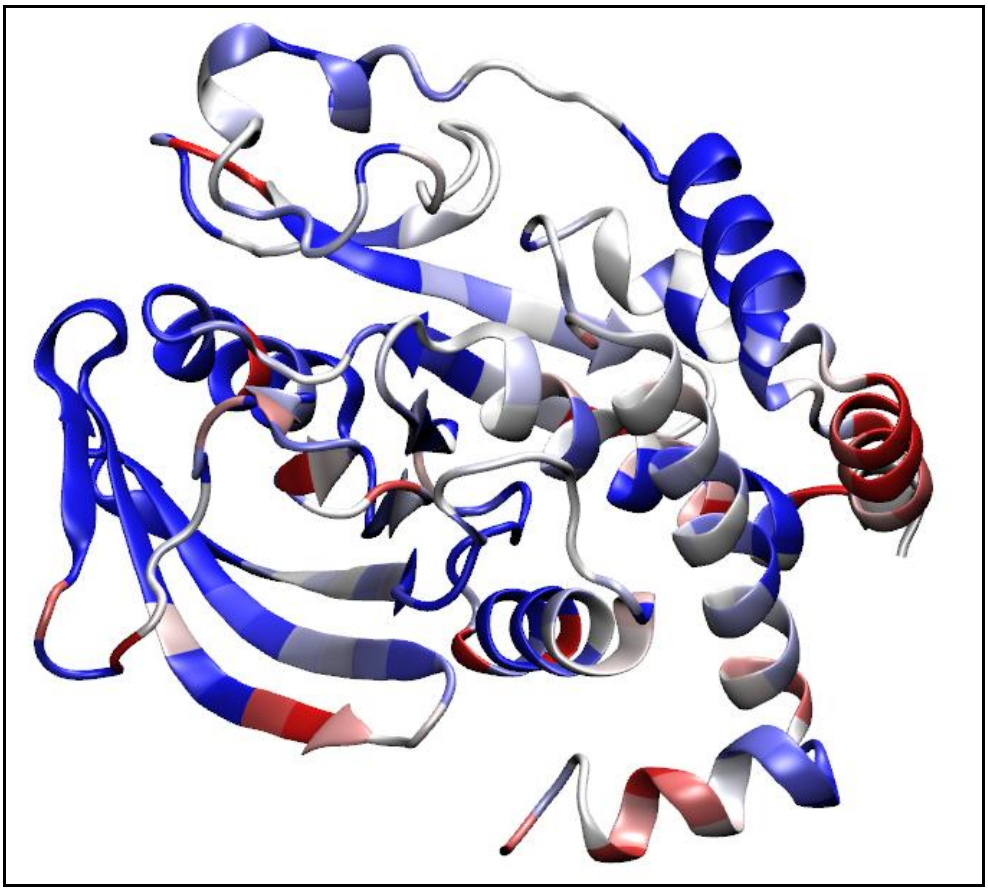

Figure 5.22. Distribution of residues with respect to the increased/decreased similarity of their displacement directions to those between crystal structures on the three-dimensional structure of PTP1B at 0.2 GHz. Blue, white, and red colors represent regions, whose displacements between crystal structure conformations, are explained in TMD1 in a higher, identical, and lower amounts, respectively, compared to MD_{open}.

5.6. Analysis of residue fluctuations in TMD2 and TMD3 on the fundamental and upper harmonic frequencies

In the previous section, TMD1 has been analyzed by various methods, and it is seen that while a force was applied on the WPD loop and R-loop, other regions of PTP1B were also affected by this periodic input. Residue mobility was increased, and residue displacement directions on the fundamental frequency were changed in accordance with the displacements seen in the crystal structures. The aim of this part is to see whether results obtained from the analysis of TMD1 will change, when the frequency of periodic

function changes. Therefore, TMD simulations were repeated by applying targeting forces on R-loop and WPD loop with the same force constant, but at different frequencies.

Fundamental frequencies of TMD2 and TMD3 are 0.1 and 0.05 GHz, respectively. To see the dominant frequencies on the WPD loop motion (fundamental and the upper harmonics) for TMD2 (Figure 5.23) and TMD3 (Figure 5.24), DFT was applied on the C_α coordinates of the WPD loop, and average Fourier coefficients were found. It is seen that, like that seen for TMD1, the first few harmonics affect the WPD loop motion in TMD2 and TMD3 simulations.

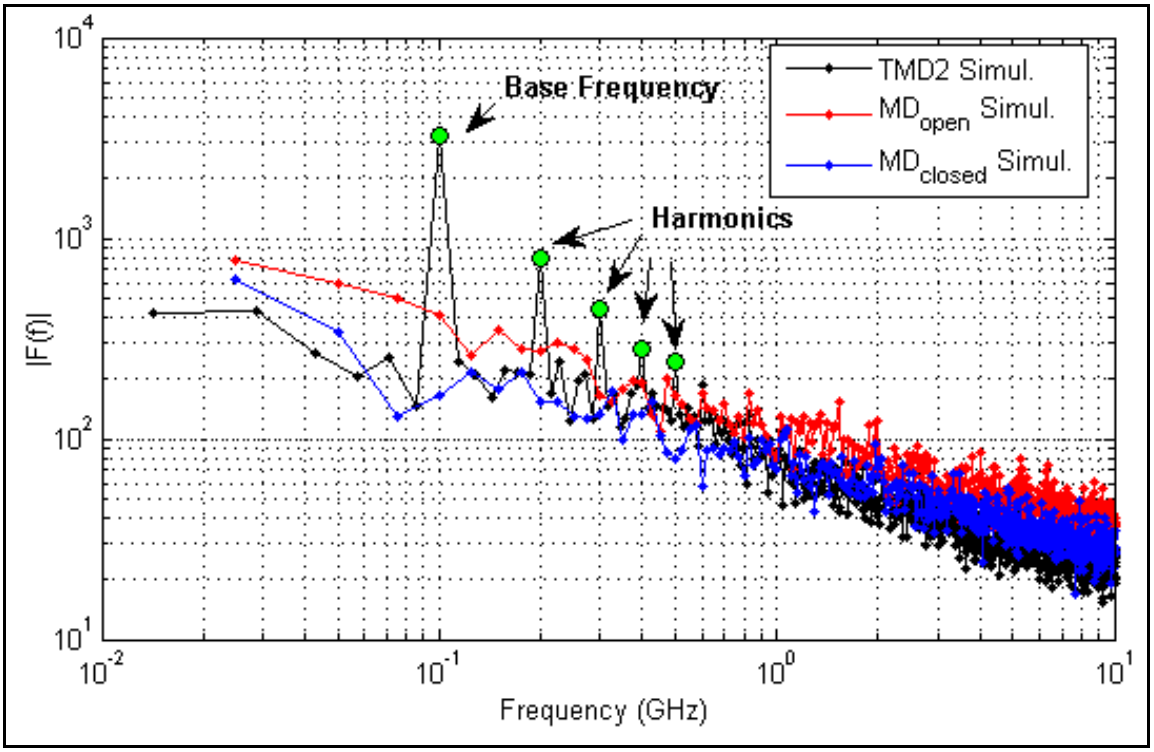

Figure 5.23. Frequency content of the WPD loop in TMD2 simulation (green dots show the significant frequencies).

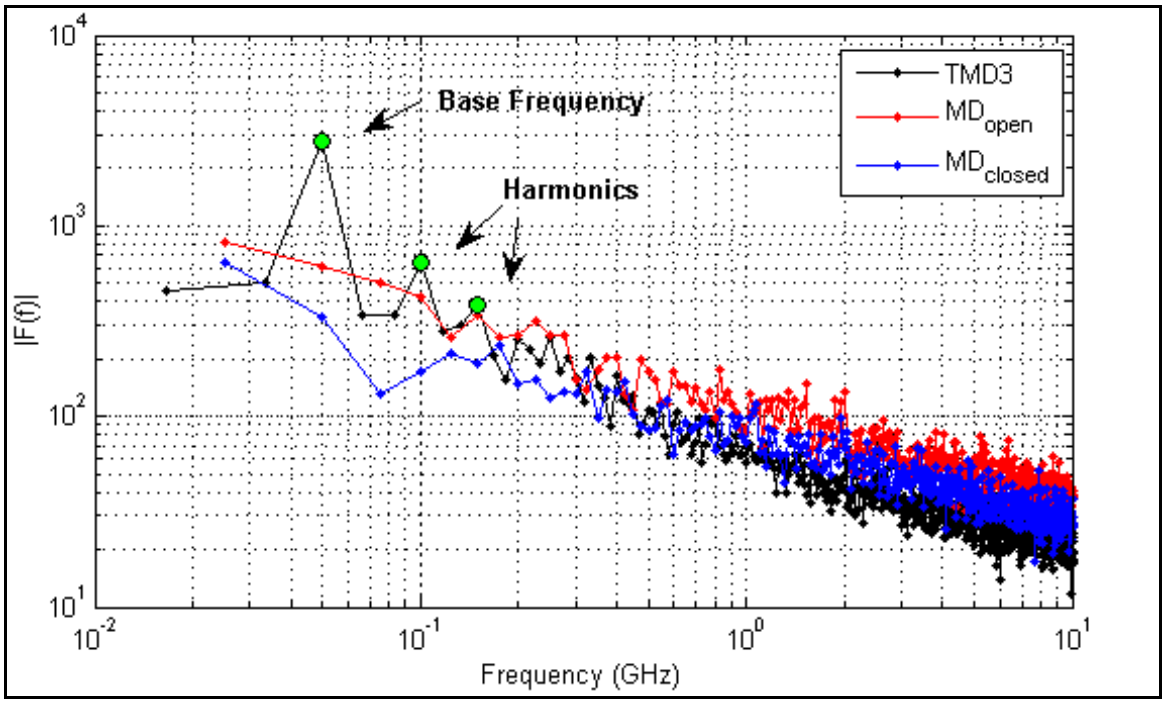

Figure 5.24. Frequency content of the WPD loop in TMD3 simulation (green dots show the significant frequencies).

Residue fluctuations in PTP1B on the harmonic frequencies of TMD2 and TMD3 were analyzed. Figure 5.25 shows the RMSF of C_α atoms in TMD1, TMD2 and TMD3 at their fundamental frequencies of 0,2, 0,1 and 0.05 GHz, respectively. It is seen that regions showing mobility increase in TMD2 and TMD3 are similar with those in TMD1. Correlation between the RMSF of residues in TMD1 and TMD2; TMD1 and TMD3; TMD2 and TMD3 are all very close to unity. There are quite high correlations between RMSF values of TMD simulations. Another important thing is that highest residue fluctuations are seen in TMD3 among all TMD simulation at their base frequencies. This is due to the fact that the fundamental frequency of TMD3 is the lowest, so random fluctuations are more likely to be reflected at lower frequencies.

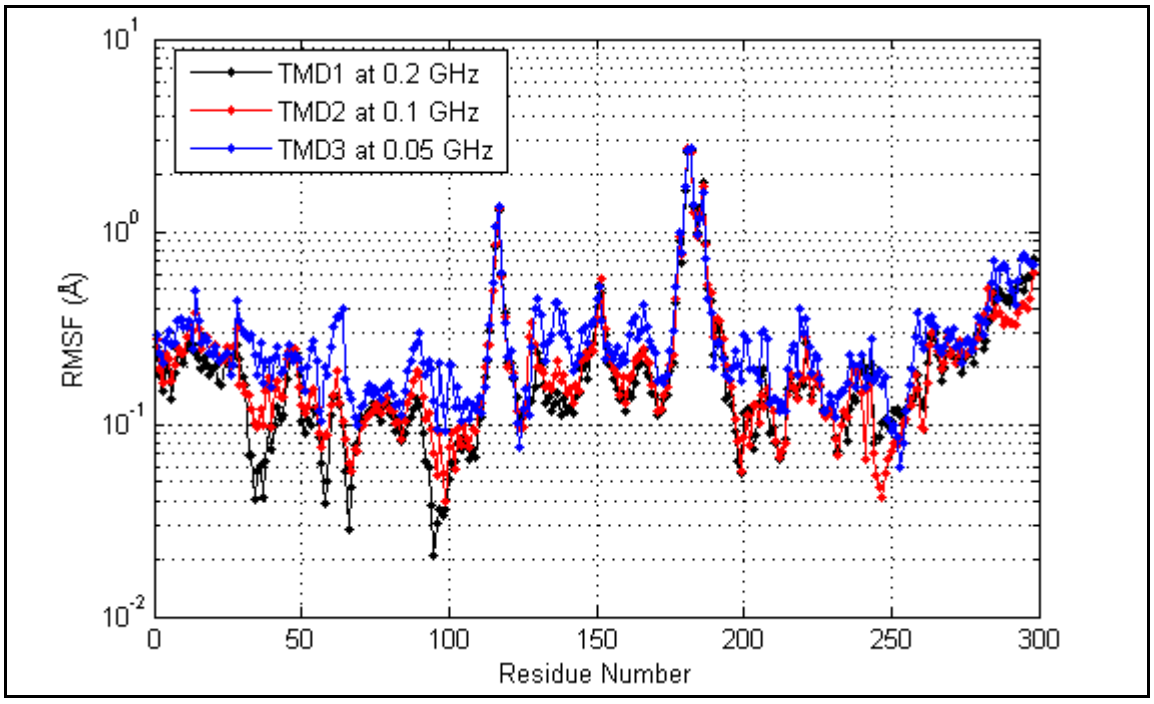

Figure 5.25. RMSF of C_α atoms for TMD1 (black) at 0.2 GHz, TMD2 (red) at 0.1 GHz and TMD3 (blue) at 0.05 GHz.

The same analysis was repeated for the second and upper harmonics in TMD2 and TMD3. Figure 5.26 shows RMSF of the C_α atoms of TMD1, TMD2 and TMD3 at 0.4, 0.2 and 0.1 GHz, respectively (second harmonic of their fundamental frequencies). It is seen that RMSF of C_α atoms vibrating at low frequencies (0.1 GHz) are generally higher than those at 0.4 and 0.2 GHz. In addition, correlation values between TMD simulations also decrease at the second harmonics. Correlation between the RMSF of TMD1 and TMD2; TMD1 and TMD3; TMD2 and TMD3 are 0.83, 0.79 and 0.75, respectively. Figure 5.27 shows the sum of the RMSF of TMD1, TMD2 and TMD3 at upper harmonics (third, fourth and fifth harmonics). Similar to what is seen at the previous frequencies; highest residue fluctuations are seen in TMD3. Correlation between the RMSF of TMD1 and TMD2; TMD1 and TMD3; TMD2 and TMD3 at upper harmonics are 0.90, 0.83 and 0.82, respectively.

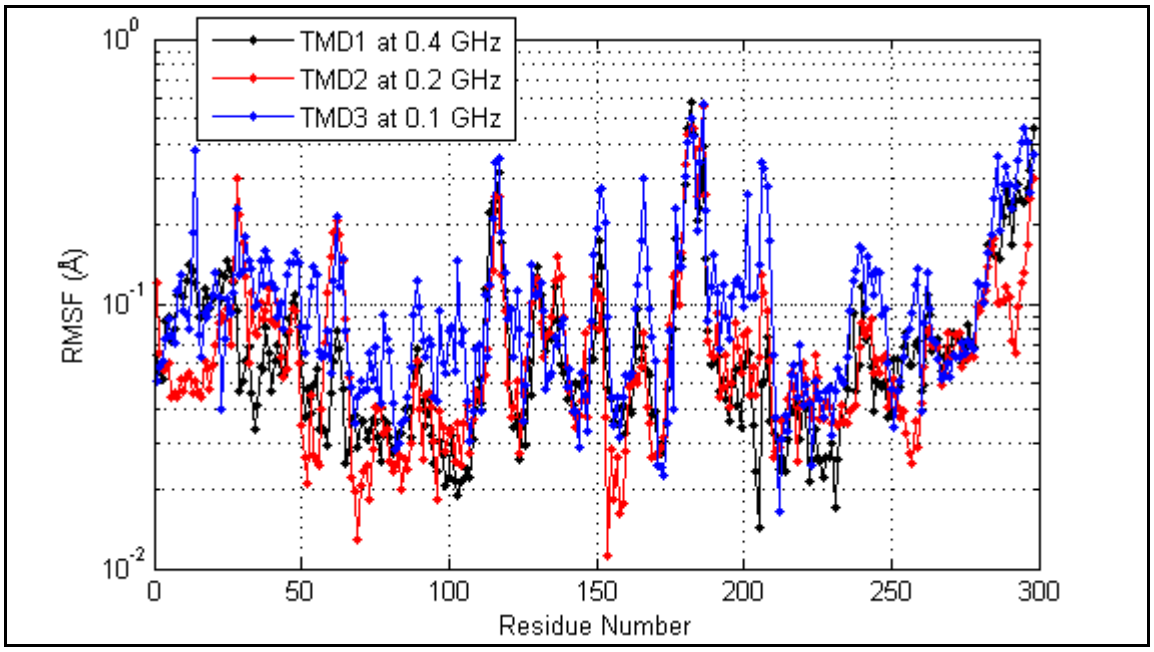

Figure 5.26. RMSF of C_α atoms for TMD1 (black) at 0.4 GHz, TMD2 (red) at 0.2 GHz and TMD3 (blue) at 0.1 GHz.

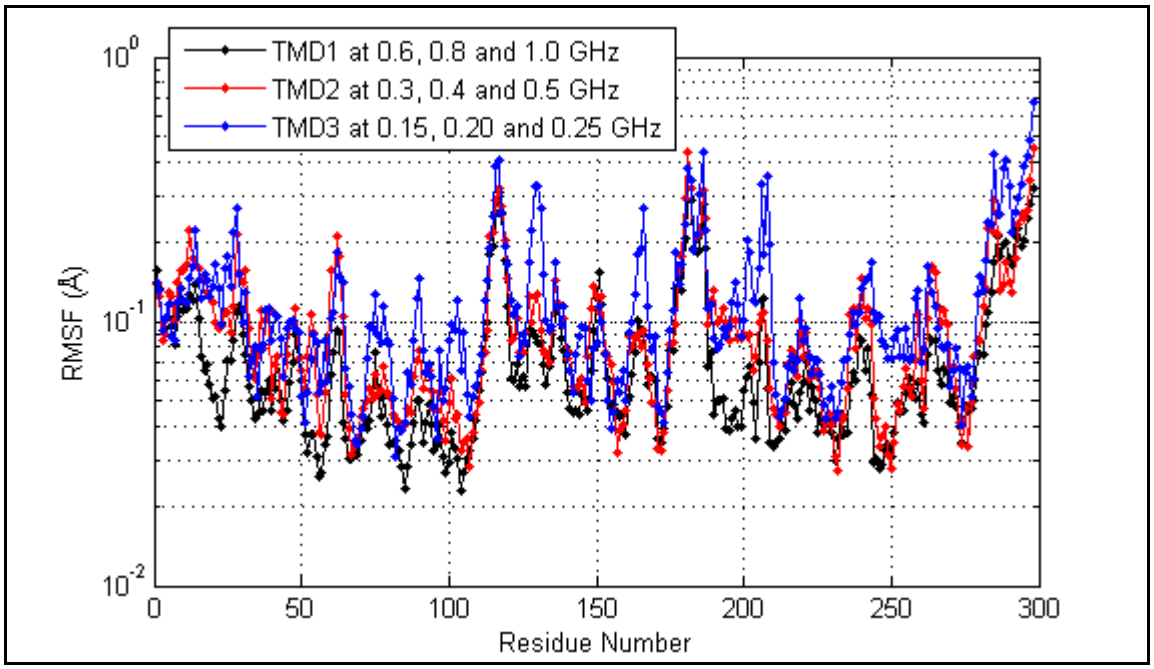

Figure 5.27. RMSF of C_α atoms for TMD1 (black), TMD2 (red) and TMD3 (blue) at third, fourth and fifth harmonics of the fundamental frequency.

In summary, it can be said that overall residue mobility is quite similar at different frequencies of TMD simulations, however as the frequency of the applied force decreases, residue mobility of the PTP1B increases. Therefore, TMD simulations at higher frequencies seem to be more reliable in analyzing perturbation of distant regions via WPD loop motions in PTPB. By this way, effects of the WPD loop opening/closing on the rest of the protein can be investigated more clearly, since random conformational changes of the protein will be manifested less at higher frequencies. At lower frequencies, RMSF of the TMD simulations are inflated due to the random conformational jumps in different regions of the protein, quite possibly, not related with the WPD loop conformational transition.

Direction of the residue displacements were examined in TMD2 and TMD3 simulation, to see whether they are in the same direction with the TMD1 simulation at their own fundamental frequencies. PCA was applied on the trajectory component of TMD2 at 0.1 GHz, and TMD3 at 0.05 GHz, to extract the dominating global motions. While first eigenvector of TMD2 explains 89% of the direction of residue displacements, this value decreases to 81% for TMD3 simulation. To examine the direction of residue displacements, angle cosines between the displacement vector of TMD1 and TMD2; TMD1 and TMD3; TMD2 and TMD3 were calculated (Figure 5.28, Figure 5.29, and Figure 5.30). It is seen that regions showing higher residue mobility have similar angle cosines with the TMD1 simulation at the fundamental frequencies (α2', pTyr recognition loop, L11, P-loop, Q-loop and α6). It should be noted that α1' and addition to this one, S-loop moves relatively less in the direction of crystal structures just as for the first eigenvector of TMD1. While 78% of the residues have more than 0.5 angle cosine between crystal structure displacements and TMD1, these results are 68% and 66% for TMD2 and TMD3, respectively. The assertion that TMD force “tunes” the global protein fluctuations to alternative crystal structures can be also confirmed with TMD2 and TMD3, however, this tuning is decreased at lower frequency TMD simulations.

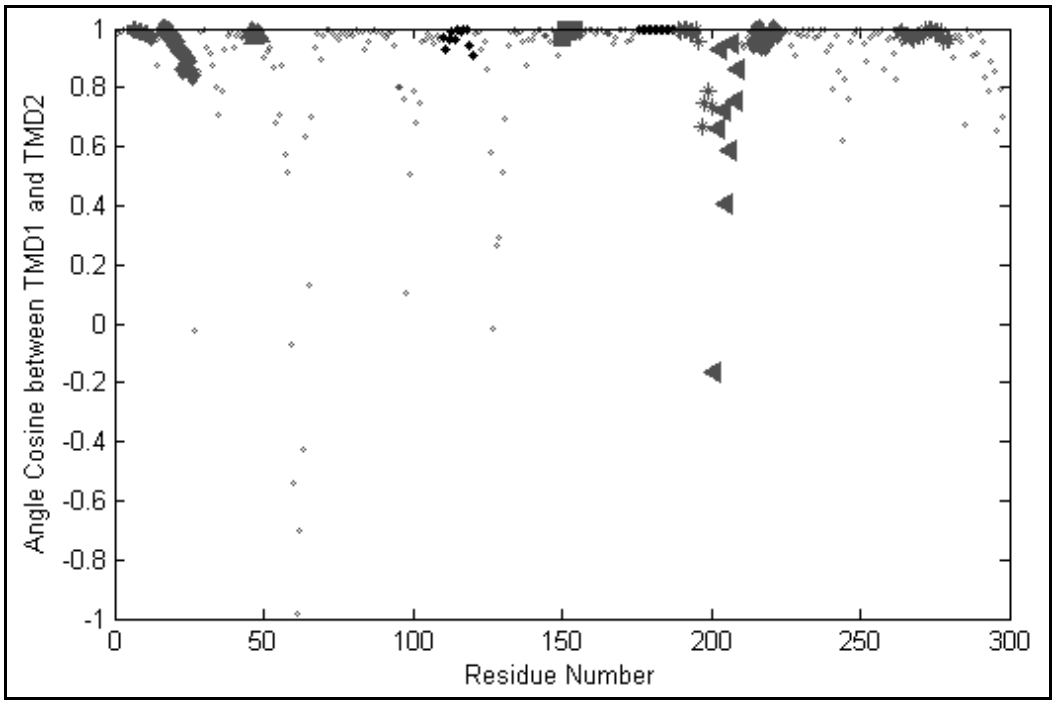

Figure 5.28. Angle cosines of residues between TMD1 at 0.2 GHz and TMD2 at 0.1 GHz simulation structure displacements.

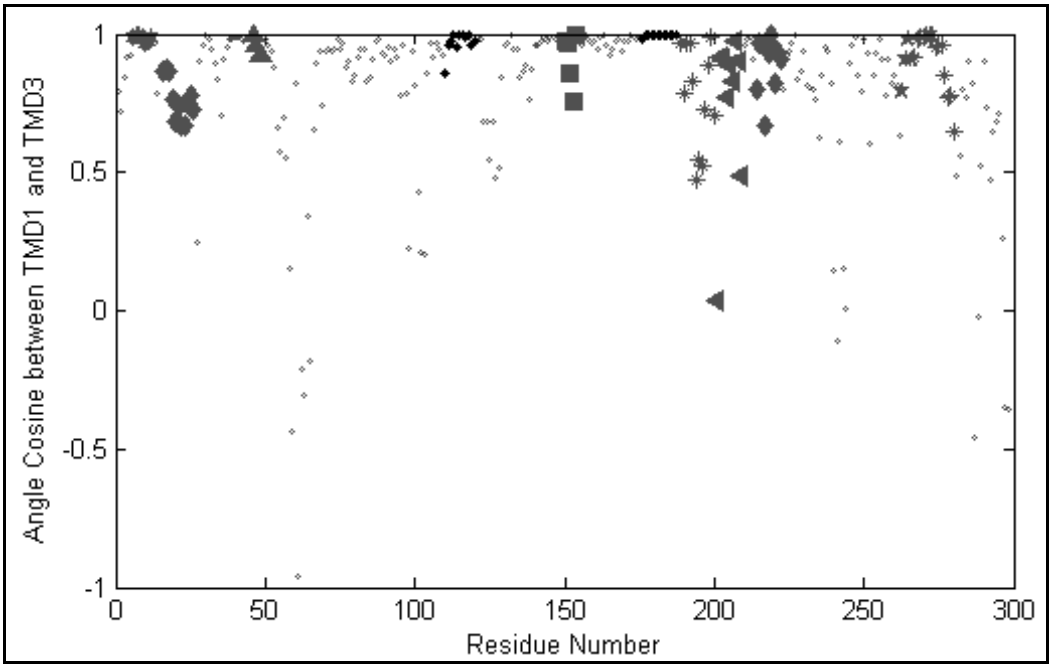

Figure 5.29. Angle cosines of residues between TMD1 at 0.2 GHz and TMD3 at 0.05 GHz simulation structure displacements.

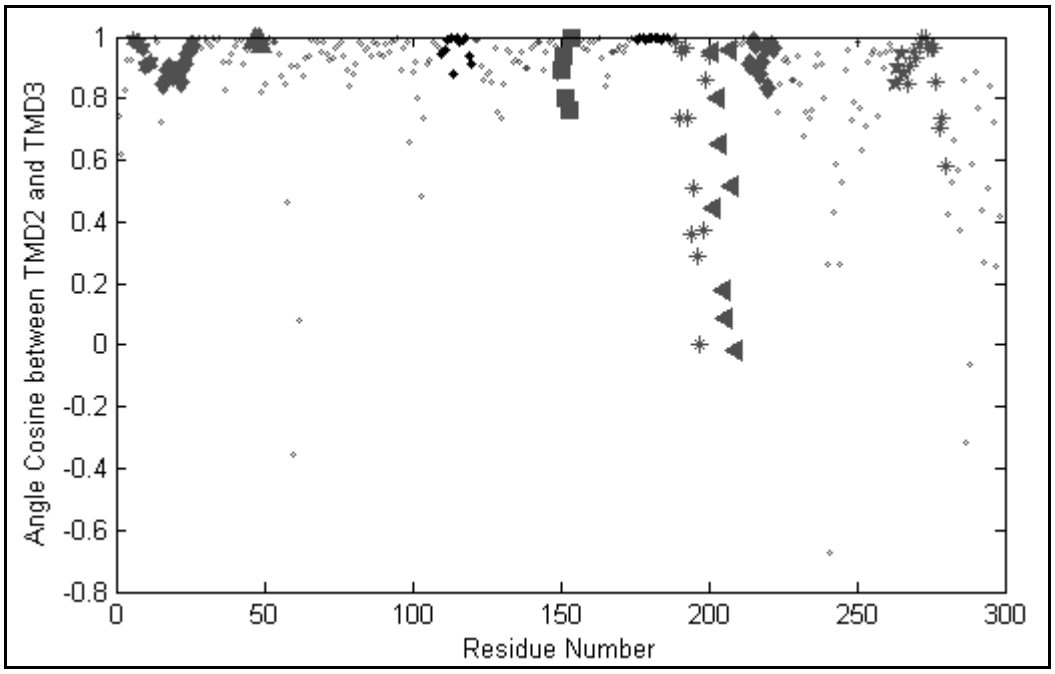

Figure 5.30. Angle cosines of residues between TMD2 at 0.1 GHz and TMD3 at 0.05 GHz simulation structure displacements.

Mobility and the direction of S-loop was also analyzed and it is seen that although there is no significant mobility increase in S-loop in TMD2 at 0.1 GHz and TMD3 at 0.05 GHz during WPD loop opening/closing, it becomes more aligned with that between the crystal structures as for TMD1. It should be emphasized that the mean of the angle cosines of the S-loop is 0.53 for TMD2 and 0.37 for TMD3. There is still considerable improvement in the direction of the S-loop compared to MD_{open} and MD_{closed} simulations. This results also supports the idea of periodic force applied on WPD loop can steer residues in the direction of crystal structures regardless of their mobility increase. Mobility increase is not the only indicator for determining the catalytically active regions while WPD loop movement.

5.7. Analysis of TMD Simulations at the same Fourier Coefficient

Up to this section, the effects of the fundamental frequencies of TMD simulations and their harmonics were investigated. In this part, the effects of the forcing function frequencies on all TMD simulations were analyzed at the same Fourier coefficient. Forcing

function frequencies of all TMD simulation were taken as if they can perform seven cycles of WPD loop opening/closing motion. It is expected that RMSFs of TMD1 and TMD2 at the same Fourier coefficient would be very similar. The reason of this similarity is that in TMD1 and in TMD2, WPD loop movements are so slow that it allows the transmittal of the WPD loop closure information to the other significant regions that take part in the communication pathways of PTP1B. Even TMD1 which has faster WPD loop movement (0.2 GHz) can transfer information to all regions of PTP1B within this time period. Since the frequency introduced at WPD loop of TMD1 is slower than the intramolecular communication speed required for the transmittal of the WPD loop closing information.

In order to analyze the impacts of the same Fourier coefficient on the RMSFs of the residues, two TMD simulations with a WPD loop movement, possibly faster than the breakeven frequency, were performed. Forcing function frequencies of TMD4 and TMD5 are 2 GHz and 10 GHz, respectively. Such fast WPD loop movement is expected not to provide the transmission of information through the PTP1B in this small time interval. This also means RMSF of residues at some regions are not expected to increase during time interval.

The amplitude ratios were used in order to compare the TMD simulations. As shown in Figure 5.31a, the amplitude ratio oscillates around one which means RMSFs of TMD1 and TMD2 are almost identical at the same Fourier coefficient. This finding allows us to use the mean values of RMSFs of TMD1 and TMD2 simulations to compare the TMD4 and TMD5. In Figure 5.31b, it is seen that amplitude ratio decreases at higher frequencies. The lowest amplitude ratio is obtained for TMD5. This may imply that the information of the WPD loop movement will not be transmitted to some regions of PTP1B. This also gives us information about the time necessary for affecting the regions. The amplitude ratios of WPD loop and R-loop do not change for TMD4 and TMD5 with the force applied on these regions.

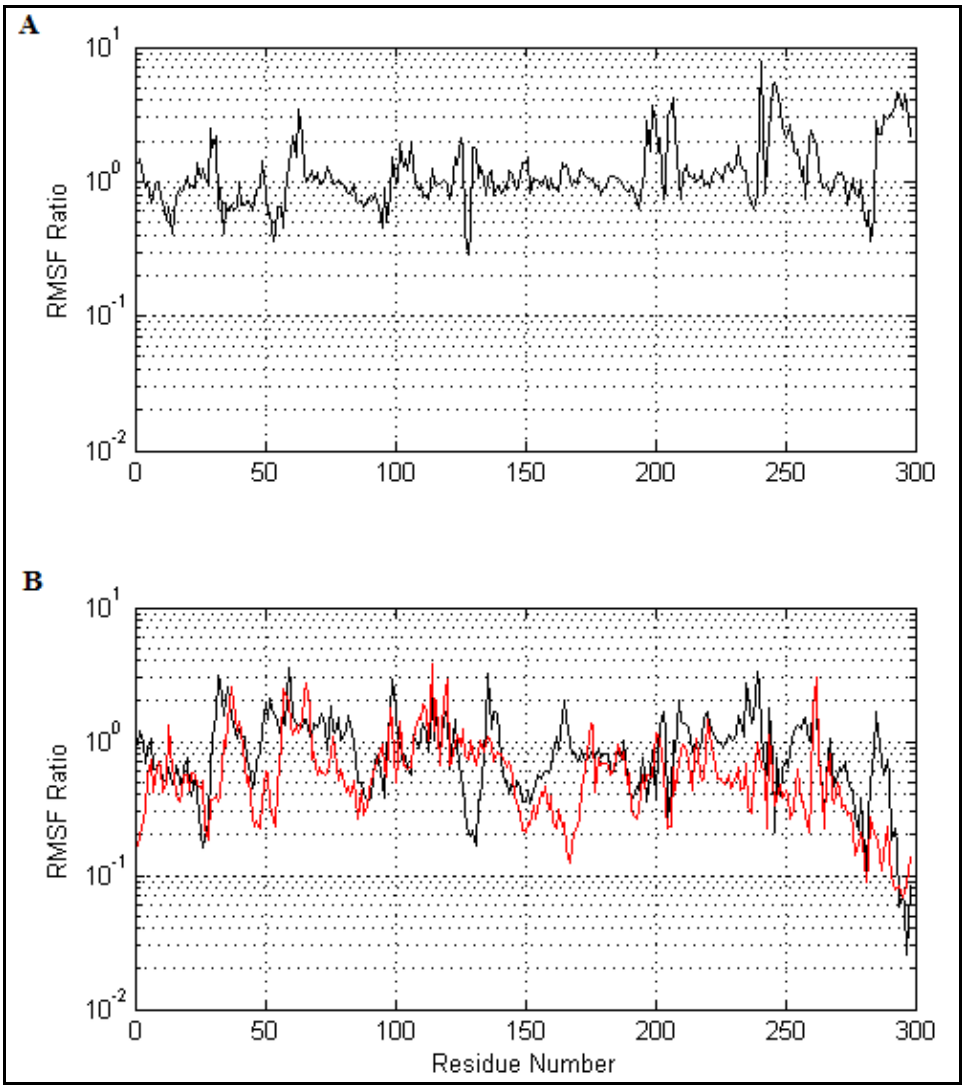

Figure 5.31. (a) Amplitude ratio of TMD1 and TMD2 (b) Amplitude ratio of TMD4 (black) and TMD5 (red) with respect to mean RMSFs of TMD1 and TMD2.

Figure 5.32 shows the distribution of residues according to their amplitude ratios. It is seen that, most of the amplitude ratios of the residues are around one for the mean RMSF values of TMD1 and TMD2 simulations (Figure 5.32a). Figure 5.32b represents the distribution of the amplitude ratio of residues in TMD4, in which it is seen that distribution of amplitude ratio is shifted to the left of unity, compared to TMD1. In Figure 5.32c, distribution is further shifted to the left, showing that fluctuations of the residues decrease at higher frequency of WPD loop motions.

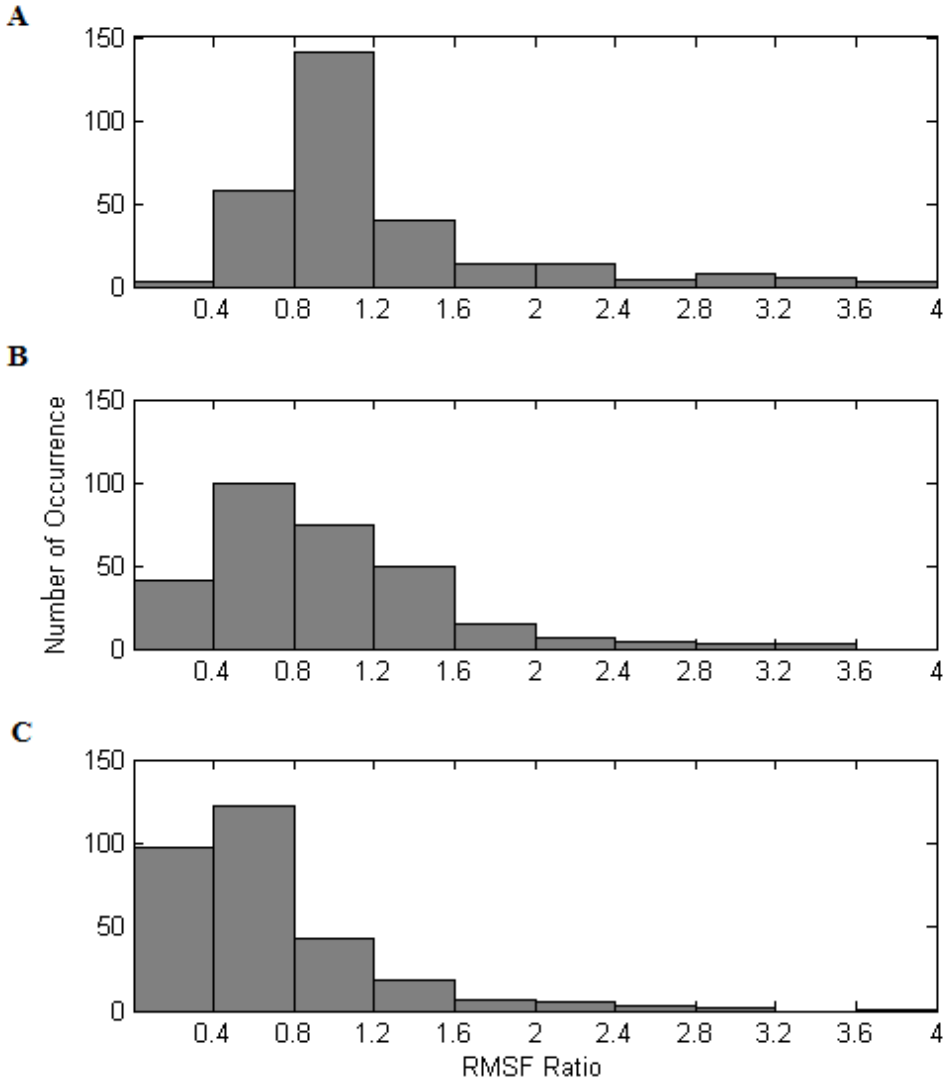

Figure 5.32. Probability distribution of the amplitude ratios of residues in (a) TMD1 over TMD2, (b) TMD4 over mean of TMD1 and TMD2, and (c) TMD5 over mean of TMD1 and TMD2 at their fundamental frequencies.

In order to compare the TMD4 and TMD5 according to their amplitude ratios, and also to have an insight about the transmission time between regions of PTP1B, residues that have amplitude ratio smaller than 0.6 are determined for each simulation and they are represented on three dimensional structure of PTP1B (Figure 5.33). The blue and red colors represent the residues of TMD4 and TMD5, respectively. Most of the regions that have lower amplitude ratio in TMD4 lacked information transmission in TMD5 also. As is seen in Figure 5.33, α1', α2', α6, α7, L11, α3, p-Tyr recognition loop have lower amplitudes at higher input frequencies, showing that their time constants may be higher

than 0.5 ns (= 2 GHz), i.e. it takes more than 0.5 ns for a change in WPD loop to be transmitted to these regions to a significant degree. . On the other hand, mobility of P-loop and residues flanking the active site is similar in TMD1, TMD2, and TMD4 simulations, while it is lower in TMD5. This shows that breakeven frequency of these regions is between 2 and 10 GHz, corresponding to a time constant between 0.1 ns and 0.5 ns. It is interesting that a change in WPD loop is faster transmitted to pTyr recognition loop, β2, β3, P-loop, α4, α5, and Q-loop while it takes longer for this information to be transmitted to L11, α3, and α6.

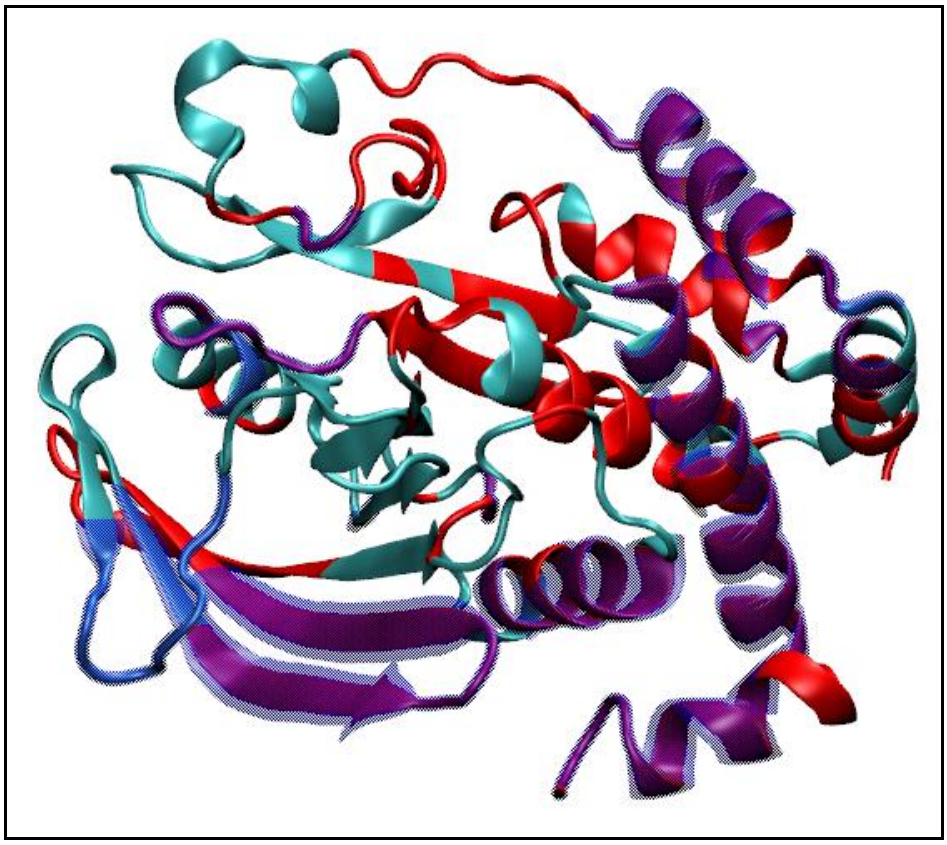

Figure 5.33. Three-dimensional structure of PTP1B colored by blue and red are residues with the amplitude ratio smaller than 0.6 in TMD4 and TMD5, respectively.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

In this study, targeted molecular dynamics simulations (TMD) between WPD_{open} and WPD_{closed} conformations of PTP1B and equilibrium molecular dynamics (MD) simulations were used to perform a detailed examination of WPD loop opening/closing mechanism. Initial and final target structures for the TMD simulations were obtained from equilibrium MD simulations of PTP1B in open and closed conformations. TMD force was applied on all atoms of two important regions: the R-loop (Gly110 to Gly120), which helps substrate bind to the active site, and the WPD loop (Thr176 to Ser187), which is highly essential for the catalytic activity of this enzyme. Target structure was alternated between WPD_{open} and WPD_{closed} conformations, resulting in a series of TMD simulations. Application of this periodic TMD force between the two different conformations of PTP1B (open and closed) results in a periodic signal and changes in the residue coordinates were recorded. Five different periods of TMD simulation cycles (one closing-opening motion of WPD loop) were applied between the WPD_{open} and WPD_{closed} conformations of PTP1B (TMD1: 0.2 GHz; TMD2: 0.1 GHz; TMD3: 0.05 GHz; TMD4: 2 GHz; TMD5: 10 GHz).

For each simulation system, the stability of the simulation systems and the overall structural changes in the protein were examined by calculating RMSD of the C_α atoms of PTP1B with respect to their initial simulation structure. In addition, interactions playing important roles on WPD loop conformational transition were monitored. It was seen that overall protein structure is stable during WPD loop opening/closing motions and WPD loop movement was successfully achieved in TMD1 simulation, in which most of the interactions stabilizing the WPD loop in the open and closed conformations were formed and disrupted in correct order throughout the simulations.

By long time-scale Targeted Molecular Dynamics (TMD) simulations and frequency response analysis, the significant structural motions in the vicinity of the WPD loop were analyzed with the purpose of investigating how functionally important distant portions are perturbed via conformational changes in the binding region, particularly the WPD loop in

PTP1B. Frequency response analysis relies on applying direct forces on some functionally important regions of PTP1B and recording the resulting relative changes in the residue coordinates. Therefore, WPD loop and R-loop of PTP1B were perturbed periodically, and the response of the rest of the protein at the fundamental frequency and the upper harmonics was analyzed using DFT on MATLAB.

During the whole analyses of TMD simulations, DFT was used to filter the raw C_α atomic coordinates, i.e. the fundamental frequency (frequency of WPD loop conformational transitions) and its first four detectable upper harmonics were kept, while the rest of the frequency components was suppressed, assuming that fluctuations at these frequencies are random conformational changes, which are not directly related with the input frequencies. This approach makes it easier to concentrate on residue fluctuations in PTP1B stemming from the active site fluctuations. Although the base frequency is sufficient to explain the fluctuations of the raw trajectory data to some extent, the inclusion of the higher harmonics improves the capture of the characteristics of the motion.

Before a detailed analysis of the residue fluctuations in PTP1B on the perturbed frequencies, convergence of residue fluctuations was checked at different frequencies for 80 ns TMD1 simulation by dividing into different time segments like 10 ns, 20 ns and 40 ns. RMSF of different segments of 40 ns show similarity for most of the residues, showing the homogeneity of segments of TMD1, except α7 helix. It should be recalled from the previous studies, α7 does not converge in ns scale simulations [12]. This result points out that long TMD simulation are more reliable to investigate the residue dynamics in frequency spectra. In addition, truncation of the initial portion of the TMD simulations are not required since correlations of time segments show a high degree of similarity.

Applied periodic force on WPD loop and R-loop perturbs other regions of the protein, where the force is not directly applied. In order to see the additional effect of the applied periodic force on WPD loop and R-loop, RMSF of all C_α atoms in TMD1 on the fundamental frequency, which is equal to the frequency of WPD loop conformational transition, and upper harmonics, which are realized due to the nonlinearity of the "protein machine", were compared with those obtained from equilibrium MD_{open} and MD_{closed} simulations. It is expected that regions affected by the periodic force should show a higher

amplitude of fluctuations on the fundamental frequency and its upper harmonics, compared to those in equilibrium simulations. As expected, highest RMSF values are on the R-loop and WPD loop, since TMD force is directly applied on these regions. Surprisingly, many other regions, such as α1', α2', pTyr recognition loop, R-loop, L11, WPD loop, α3, P-loop, Q-loop and α6, showed higher residue mobility in TMD1. The increased fluctuations in these regions point out the possibility that information flows from the R-loop and WPD loop to the rest of the protein during WPD opening/closing, and WPD loop motions has global effects on the protein.

When the three-dimensional structure of PTP1B is colored with respect to its mobility increase due to periodic force, it is seen that the vicinity of WPD loop, especially L11 and N-terminus of α3, usually have the highest mobility increase, and followed by α6 and N-terminus of α7. In addition, a moderate increase in mobility is seen in pTyr recognition loop, P-loop, and Q-loop, which surround the active site. This result shows that a force applied only on WDP loop and R-loop increases the fluctuations of all the residues in the vicinity of the active site. Effect of periodic force diminishes for regions distant from WPD loop: particularly regions behind the P-loop, and surprisingly S-loop have negligible or no increase in their fluctuations. On the other hand, regions not directly in contact with the active site, reaching a distance up to ~17 Å, namely α1', α2', β3, α4 and α5 regions have a significant increase in mobility in TMD1 simulation. Similar analysis was also repeated for the upper harmonic frequencies of TMD1 simulation and α1', α2', R-loop, L11, WPD loop and α6 increase their fluctuations on the harmonic frequencies examined as a consequence of the periodic force applied on local regions of PTP1B. It is important to note that higher fluctuations are seen not only in the vicinity of WPD loop and R- loop, but also at some distant parts of PTP1B, not directly in contact with the active site or WPD loop, i.e. α1', α2'.

Magnitudes of the residue displacements during WPD loop opening (or closing) predicted by the frequency analysis of TMD1 were in reasonable agreement with the crystal structures displacements. It should be noted that estimated displacements for α2' and P-loop are found to be slightly higher (~0.2 Å) compared to crystal structures. Repeating TMD simulations by applying same force constant on R-loop and WPD loop but at different frequencies at 0.1 and 0.05 GHz showed that overall residue displacements

profile is quite similar at different frequencies of TMD simulations. However as the frequency of the applied force decreases, residue displacements increase, since random fluctuations are more likely to be highly reflected at the lowest frequencies. Therefore, it can be said that TMD simulations at higher frequencies are more reliable in analyzing perturbation of distant regions by WPD loop motions in PTPB.

Response of the protein due to the applied periodic force on WPD loop can also be characterized in terms of the directionality of the residue displacements. Direction of residue displacements at 0.2 GHz was checked whether they are in the same direction with those in the crystal structures. It is seen that the residue displacement vectors of α2', pTyr recognition loop, L11, P-loop, Q-loop and α6 predicted by frequency analysis are practically in the same direction with those of the crystal structures. It is found that 78% of the residues in TMD1 simulation move approximately in the same direction with the displacements in the crystal structures. Collective motions extracted by PCA in MD_{open} and MD_{closed} simulations are much less aligned with the displacements between the crystal structures, thus one may argue that ligand binding and the following local conformational change in PTP1B tunes the direction of global protein fluctuations to crystal structure displacements. One striking example is obtained in the functionally important S-loop region: though there is no mobility increase in S-loop at 0.2 GHz with respect to equilibrium MD simulations, periodic opening/closing motion of WPD loop adjusts the S-loop fluctuations at 0.2 GHz to be aligned with the displacement seen in crystal structures. This result shows that the allostery may be exhibited not only by an increase (or decrease) in the amplitude of residue fluctuations, but also by forcing the direction of these fluctuations. For instance, α1, L1, B1, L2, α1, β7, L9, β8, L10, β9, β10 and S-loop are regions steered in the direction of crystal structures displacements during WPD loop motions, with negligible differences in their MSF. Direction of the residue displacements was similar in TMD2 and TMD3 simulations, but slightly decreased at lower frequency TMD simulations.

In order to see the effects of the forcing function frequency, all TMD simulations were analyzed at the same Fourier coefficient. In addition to TMD1, TMD2, and TMD3, two additional TMD simulations with a WPD loop movements possibly faster than the breakeven frequency were performed. Forcing function frequencies of TMD4 and TMD5

are 2 GHz and 10 GHz, respectively. The reason to perform such fast WPD loop movement was to check whether information of the WPD loop movement can be transmitted through the rest of the PTP1B in such short time durations. It was observed that the lowest amplitude of residue oscillations is seen for TMD5, hence it can be said that amplitude of residue oscillations decreases at higher frequencies, as that would have been expected from a linear system. In TMD5 (at 10 GHz), WPD loop fluctuations does not seem to affect the P-loop, which have been perturbed in TMD4 (at 2 GHz). One may argue that the lack of the transmission of information in the active site may be the main reason for the other regions not to receive information. Some parts of the α1', α2', α6, α7, L11, α3, p-Tyr recognition loop have lower amplitudes at 2 GHz. Therefore, it may be suggested that, the WPD loop closure information is initially transmitted to P-loop, and later to α3, α6, and other functionally important regions.

6.2. Recommendations for Future Studies

As a future study, additional TMD simulations can be performed by changing the initial and target structures to test whether the results obtained from them are similar to those obtained in this study. By this way, it can be tested whether there exists alternative WPD loop opening/closing paths, and the rest of PTP1B may be perturbed in a different way.

In the current study, displacements in the backbone carbon atoms of PTP1B are investigated only. However, it is known that allostery depends also on sidechain motions. As a future study, the periodic changes in the sidechain dihedral angles should be investigated to suggest possible communication pathways between distant regions.

TMD simulations at 0.2 GHz seems to be most consistent with the experimental results, though the displacements observed even at this frequency seem to overpredict the experimental displacements. A method can be devised to use DFT of equilibrium simulations as a reference for equilibrium fluctuations, and subtract these fluctuations from the perturbed ones.

Preliminary results on fast frequencies (2 GHz and 10 GHz) show that there is a significant reduction in the oscillation amplitudes of many regions. This shows that different regions of PTP1B behave similar to a linear filter, whose amplitude ratio diminishes at high frequencies. Repeating these TMD experiments at a whole range of frequencies would enable one to plot Bode plots of residues, and thus predict time constants of different regions. This may be helpful in measuring and comparing how fast information is transmitted from the WPD loop to distant regions.

APPENDIX A: PREPARATION OF TMD SIMULATIONS

Initial and target structures for TMD simulations were obtained from MD_{open}, and MD_{closed} (equilibrium) simulations, respectively. In order to start the TMD simulation, the initial and target crystal structures should include same number of atoms. Therefore, all water molecules and ions of the initial structure (WPD_{open} structure) were copied to the target structure (WPD_{closed}) to have the same number of atoms. Then, by the help of the psfgen module of VMD, missing atoms in the target structure were built and coordinates were guessed. The occupancy of the target structure atoms which the force should be applied in TMD runs was set to 1 (nonzero), while other atoms occupancy is kept at zero. After the target structure was prepared, TMD simulation was performed using the configuration file given below.

A.1. NAMD Configuration File for TMD Simulation

```
#############################################################
#############################################################
### TMD simulation of 2f6f in a water box ###
#############################################################
#############################################################

set pdb 1sug
set outputfile 1sug.1sugto2f6f
set temperature 300

#############################################################
## INPUT FILES (and TEMPERATURE) ##
#############################################################

structure /home/balakent/Desktop/dilek/thesis/1sug_TMD2kapaliacik/ionized.psf
coordinates /home/balakent/Desktop/dilek/thesis/1sug_TMD2kapaliacik/ionized.pdb
bincoordinates 2f6f.2f6fto1sug.coor
binvelocities 2f6f.2f6fto1sug.vel
extendedSystem 2f6f.2f6fto1sug.xsc

#############################################################
## SIMULATION PARAMETERS ##
#############################################################

paraTypeCharmm       on
```

```
parameters /home/balakent/Desktop/dilek/thesis/TMD_prm_inp/par_all27_prot_lipid.prm
parameters /home/balakent/Desktop/dilek/thesis/TMD_prm_inp/thiolate.prm



#############################################################
## STANDARD OUTPUT ##
#############################################################

outputname $outputfile
binaryoutput yes
outputEnergies 500
outputPressure 500

#############################################################
## FORCEFIELD PARAMETERS ##
#############################################################

# Related with cut-offs
exclude scaled1-4
1-4scaling 1.0
cutoff 12.

switching on
switchdist 10.

pairlistdist 14.
stepspercycle 10
nonbondedFreq 1
fullElectFrequency 2

# No ShakeH and SETTLE for any molecule
rigidBonds none
useSettle       off


#############################################################
## PERIODIC BOUNDARY CONDITIONS ##
#############################################################

# Wrap the water molecules
margin 2.5
wrapWater      on
wrapAll on
wrapNearest    off


# PME (for full-system periodic electrostatics
# Grid sizes: 48 54 60 64 72 75 80 81 90
```

```
PME yes
PMEGridSizeX 90
PMEGridSizeY 80
PMEGridSizeZ 64


#############################################################
## INTEGRATION PARAMETERS - related to MD sim. ##
#############################################################
firsttimestep 0
timestep 1

#############################################################
## OUTPUT FILES - related to MD sim. ##
#############################################################
restartname restart14
restartfreq 1000
DCDfile 1sugto2f6f_k3000.dcd
DCDfreq 1000
XSTfile celldim.md14.xst
XSTfreq 1000


#############################################################
## TEMPERATURE CONTROL (Langevin) - related to MD sim. ##
#############################################################

langevin on
langevinDamping 5.0
langevinTemp $temperature
langevinHydrogen off


#############################################################
## PRESSURE CONTROL (Langevin) - related to MD sim. ##
#############################################################

useGroupPressure yes ;# needed for rigidBonds
useFlexibleCell no
useConstantArea no

langevinPiston on
langevinPistonTarget 1.01325 ;# in bar -> 1 atm
langevinPistonPeriod 100.
langevinPistonDecay 50.
LangevinPistonTemp $temperature


#############################################################
```

```
## EXECUTION SCRIPT ##
#############################################################
# TMD
 TMD on
 TMDk 3000
 TMDOutputFreq 1

TMDFile
/home/balakent/Desktop/dilek/thesis/1sug_TMD2kapaliacik/target_tmdready.pdb

TMDFirstStep 0
TMDLastStep 5000000
TMDInitialRMSD 0
TMDFinalRMSD 0

#############################################################
## Molecular Dynamics ##
#############################################################
run 5000000
```

APPENDIX B: ADDITIONAL ANALYSES OF TMD1 SIMULATION

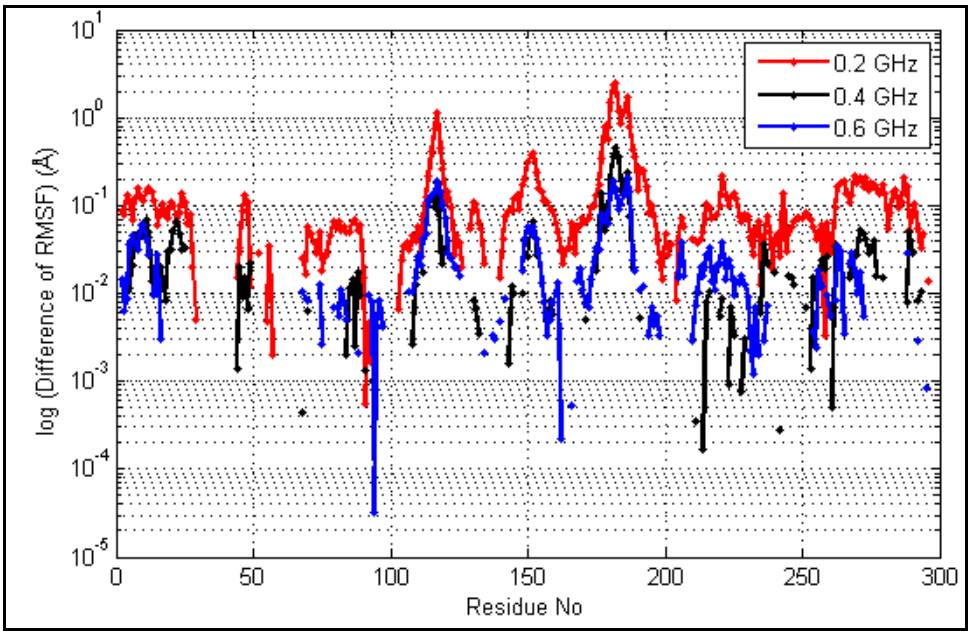

Figure B.1. Difference of RMSF values of TMD1 simulation from equilibrium MD_{open} and MD_{closed} at 0.2 GHz (red), at 0.4 GHz (black), at 0.6 GHz (blue) in logarithmic scale.

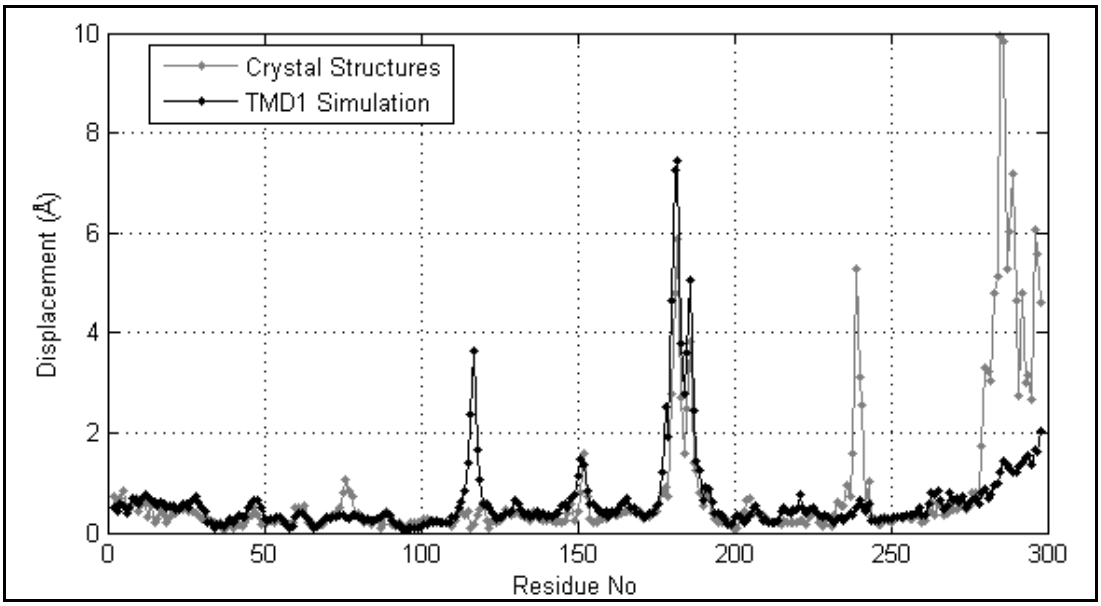

Figure B.2. Comparison of residue displacements in experimental crystal structures (grey) at 0.2 GHz and TMD1 (black) simulation structures at 0.2-1.0 GHz in decimal scale.

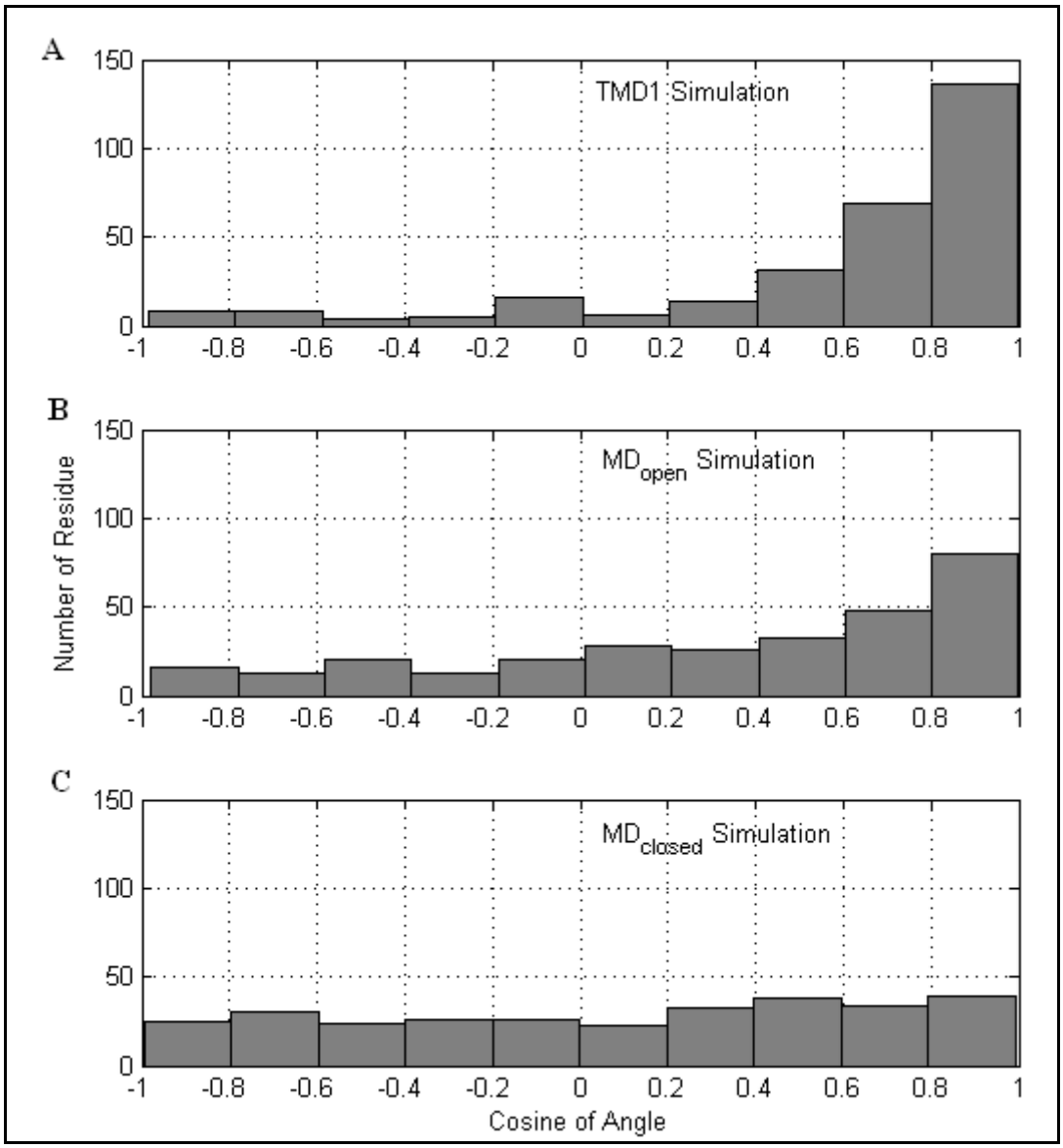

Figure B.3. Probability distribution of the angle cosines of all residues of PTP1B for (a) first eigenvector of TMD1 (b) MD_{open} (c) MD_{closed} at 0.2 GHz.

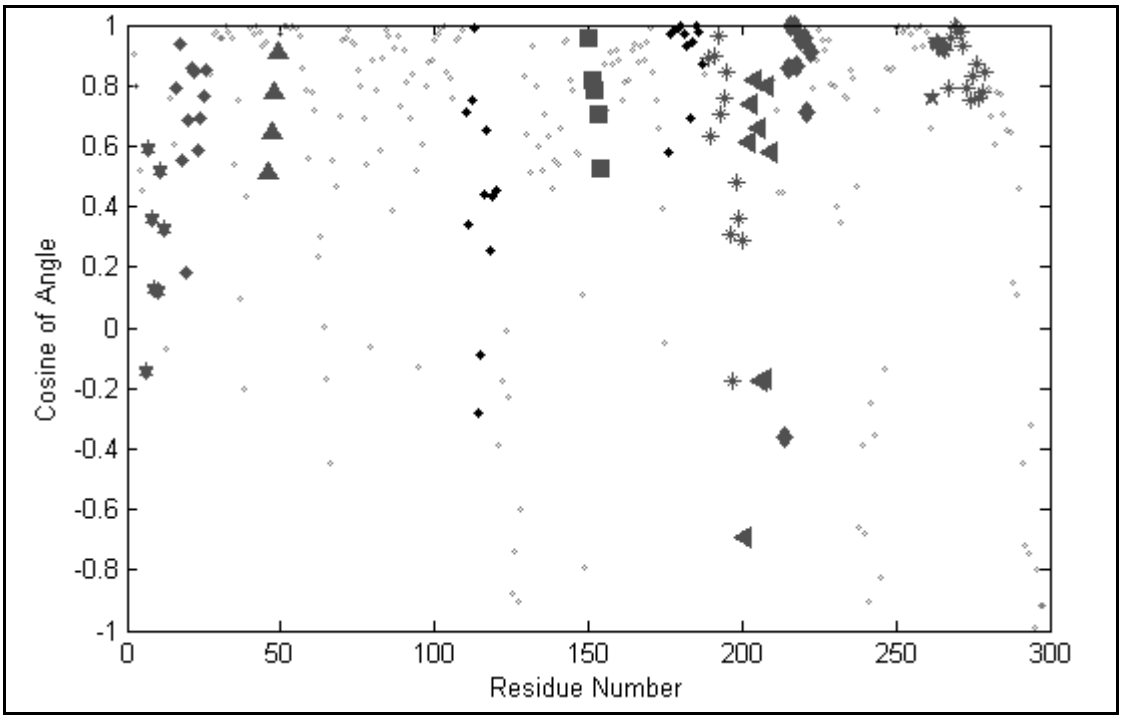

Figure B.4. Angle cosines of residues between crystal structure displacements and those obtained from the eigenvectors at 0.2, 0.4, 0.6, 0.8, and 1.0 GHz in TMD1.

APPENDIX C: ADDITIONAL ANALYSIS OF TMD2 AND TMD3 SIMULATIONS

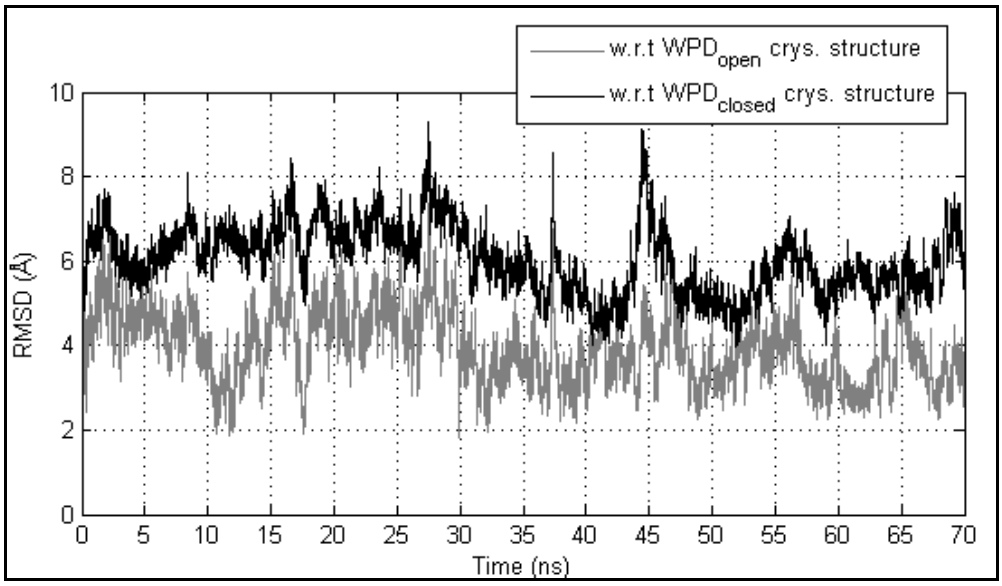

Figure C.1. RMSD of α7 with respect to WPD_{open} and WPD_{closed} crystal structures for TMD2.

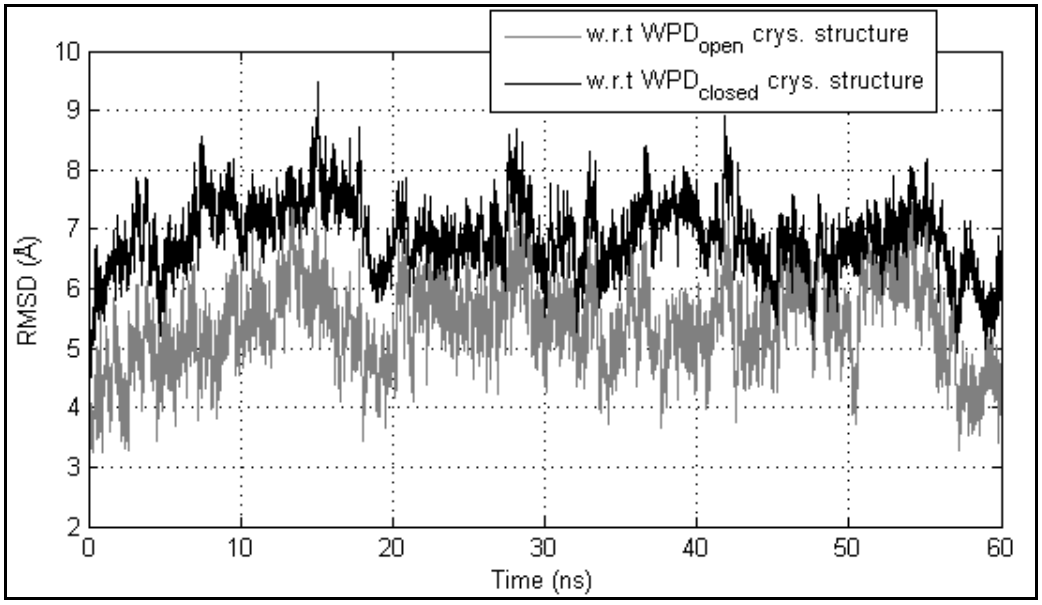

Figure C.2. RMSD of α7 with respect to WPD_{open} and WPD_{closed} crystal structures for TMD3.

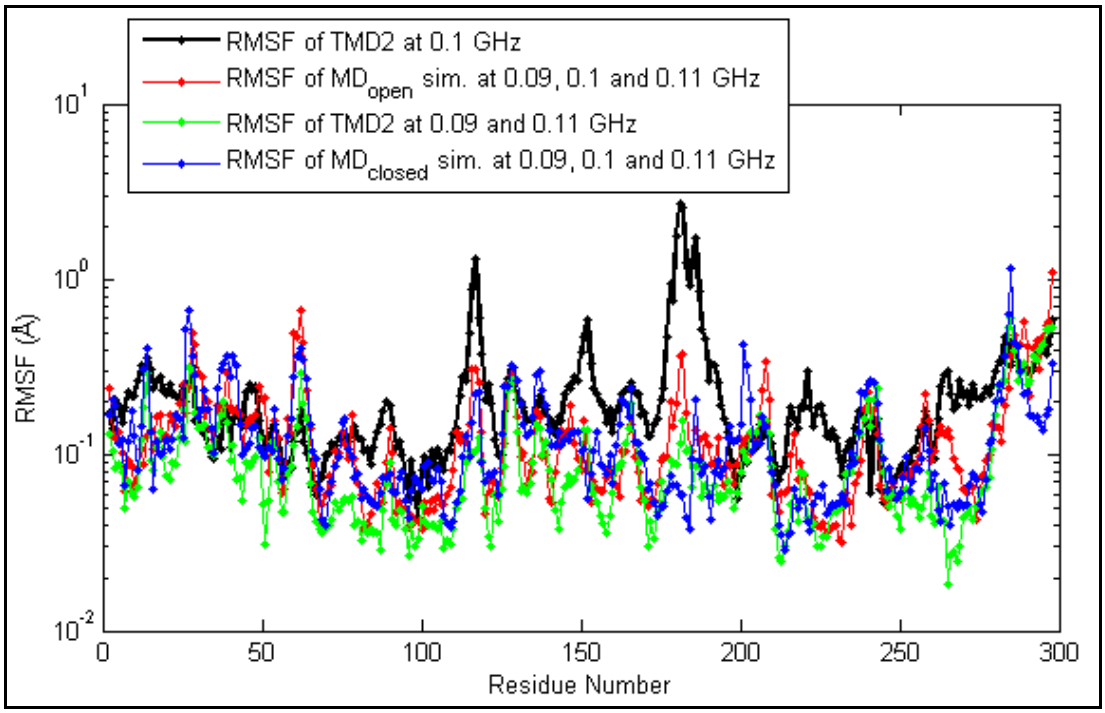

Figure C.3. RMSF of C_α atoms at the base frequency 0.1 GHz for TMD2 (black), for MD_{open} (red) and MD_{closed} (blue) simulations at 0.1 GHz and neighboring frequencies.

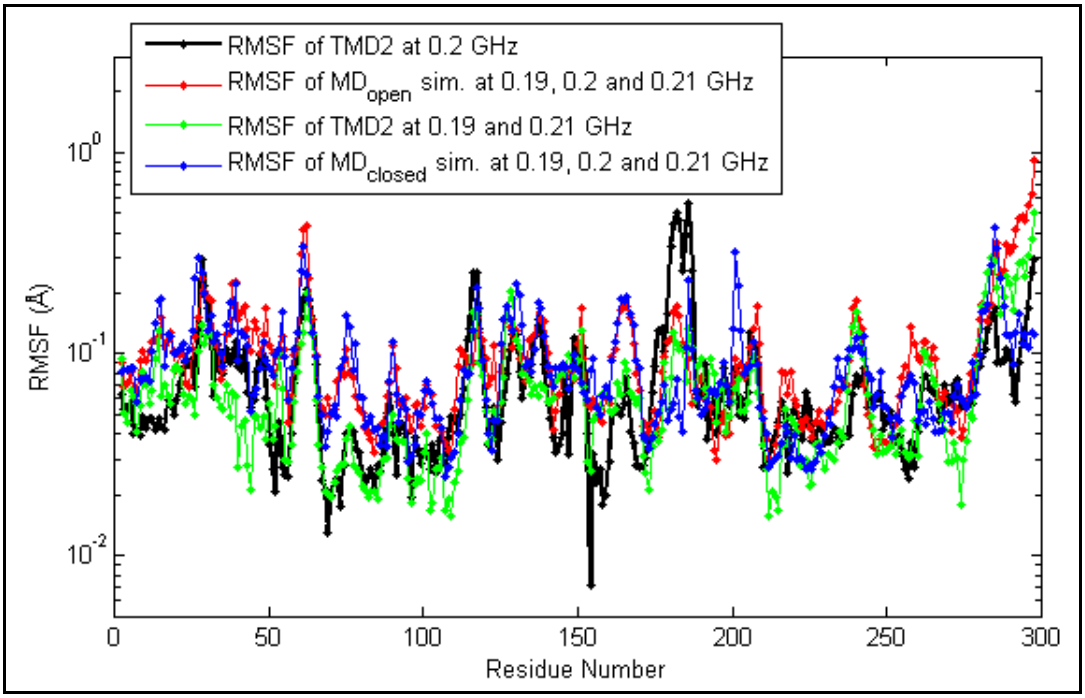

Figure C.4. RMSF of C_α atoms at the second harmonic 0.2 GHz for TMD2 (black), for MD_{open} (red) and MD_{closed} (blue) simulations at 0.2 GHz and neighboring frequencies.

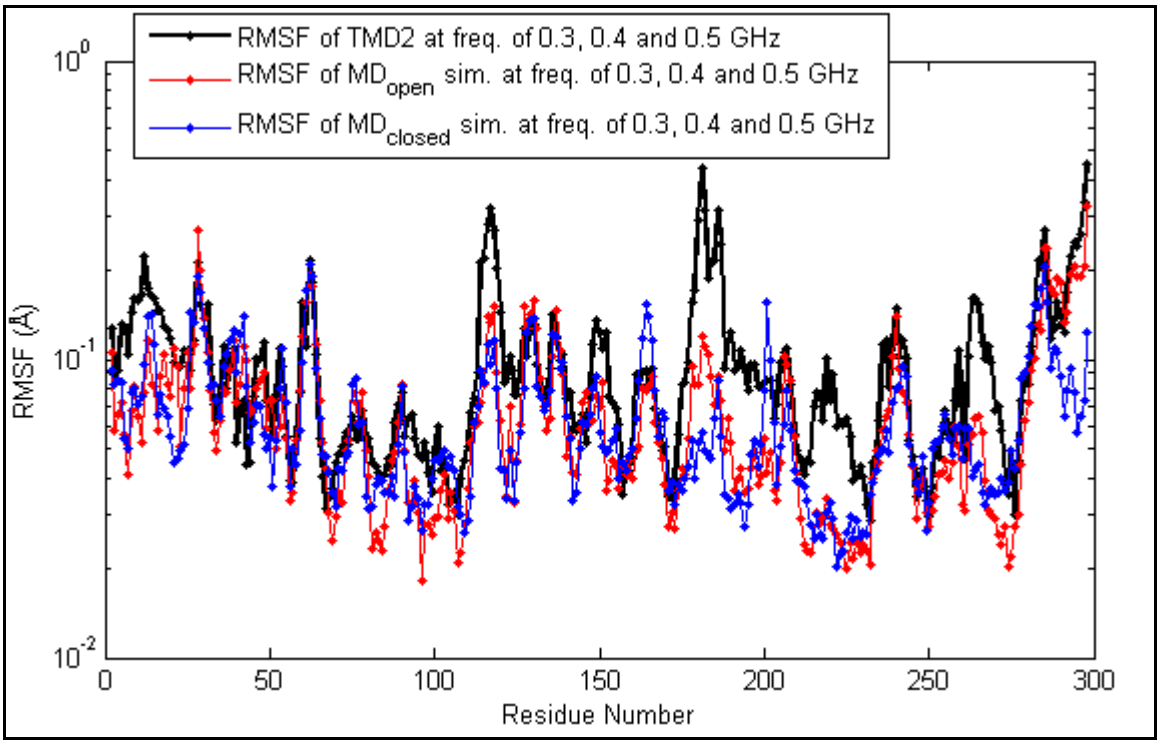

Figure C.5. RMSF of C_α atoms for TMD2 (black), MD_{open} (red) and MD_{closed} (blue) at 0.3, 0.4 and 0.5 GHz.

REFERENCES

1. Secko, D., *Protein Phosphorylation: A Global Regulator of Cellular Activity*, 2003, http://www.scq.ubc.ca/protein-phosphorylation-a-global-regulator-of-cellular-activity, accessed at May 2012.

2. Joachim, L., T. Shen, and P. G. Wolynes, "Conformational Switching upon Phosphorylation: A Predictive Framework Based on Energy Landscape Principles", *Biochemistry*, Vol. 47, pp. 2110-2122, 2008.

3. Soulsby, M. and A. M. Bennett, "Physiological Signaling Specificity by Protein Tyrosine Phosphatases", *Physiology*, Vol. 24, pp. 281-289, 2009.

4. Pannifer, A. D. B., A. J. Flint, N. K. Tonks, and D. Barford, "Visualization of the Cysteinyl-phosphate Intermediate of a Protein-tyrosine Phosphatase by X-ray Crystalography", *The Journal of Biological Chemistry*, Vol. 273, pp. 10454-10462, 1998.

5. Jia, Z., D. Barford, A. J. Flint, and N. K. Tonks, "Structural Basis for Phosphotyrosine Peptide Recognition by Protein Tyrosine Phosphatase 1b", *Science*, Vol. 268, pp. 1754-1758, 1995.

6. Iversen, L. F., K. B. Moller, A. K. Pedersen, G. H. Peters, A. S. Petersen, H. S. Andersen, S. Branner, S. B. Mortensen, and N. P. H. Moller, "Structure Determination of T Cell Protein-tyrosine Phosphatase", *The Journal of Biological Chemistry*, Vol. 277, No. 22, pp. 19982-19990, 2002.

7. Murray, R., D. K. Granner, P. A. Mayes and V. W. Rodwell, *Harper's Biochemistry*, 25th Ed., McGraw-Hill, New York, USA, 2000.

8. Zhang, S., and Z. Y. Zhang, "PTP1B as a Drug Target: Recent Developments in PTP1B Inhibitor Discovery", *Drug Discovery Today*, Vol. 12, pp. 373-381, 2007.

9. Tonks, N. K., "PTP1B: From the Sidelines to the Front Lines!", *FEBS Letters*, Vol. 546, pp. 140-148, 2003.

10. Montalinbet, J., and B. P. Kennedy, "Therapeutic Strategies for Targeting PTP1B in Diabetes", *Elsevier*, Vol. 2, pp. 129-135, 2005.

11. Ozkaral, B., A. Ozcan, B. Alakent, and E. Özkırımlı, "Analysis of Phosphatase 1B WPD Loop Closure", *Health Informatics and Bioformatics (HIBIT)*, Antalya, April 2010, pp. 162-166, 2010.

12. Olmez, E. O., and B. Alakent, "Alpha7 Helix Plays an Important Role in the Conformational Stability of PTP1B", *Journal of Bimolecular Structure & Dynamics*, Vol. 28, pp. 675-693., 2011.

13. Barford, D., A. J. Flint, and N. K. Tonks, "Crystal Structure of Human Protein Tyrosine Phosphatase 1B", *Science*, Vol. 263, pp. 1397-1404, 1994.

14. Ozcan, A., *Effects of Protonation States of Catalytically Important Residues and Active Site Water Molecules on PTP1B Conformation*, M.S. Thesis, Boğaziçi University, 2011.

15. Tabernero, L., A. R. Aricescu, E. Y. Jones, and S. E. Szedlacsek, "Protein Tyrosine Phosphatases: Structure-Function Relationships", *FEBS Journal*, Vol. 275, pp. 867-882, 2008.

16. Wiesmann, C., K. J. Barr, J. Kung, J. Zhu, D. A. Erlanson, W. Shen, B. J. Fahr, M. Zhong, L. Taylor, M. Randal, R. S. McDowell, and S. K. Hansen, "Allosteric Inhibition of Protein Tyrosine Phosphatase 1B", *Nature Structural and Molecular Biology*, Vol. 11, pp. 730-737, 2004.

17. Kamerlin, S. C. L., R. Rucker, and S. Boresch, "A Molecular Dynamics Study of WPD-loop Flexibility in PTP1B", *Biochemical and Biophysical Research Communications*, Vol. 356, pp. 1011-1016, 2007.

18. Kamerlin, S. C. L., R. Rucker, and S. Boresch, "A targeted Molecular Dynamics Study of WPD Loop Movement in PTP1B", *Elsevier*, Vol. 11, pp. 1161-1166, 2006.

19. Ikeguchi, M., J. Ueno, M. Sato, and A. Kidera, "Protein Structural Change upon Ligand Binding: Linear Response Theory", *Physical Review Letters*, Vol. 94, pp. 078102-078104, 2005.

20. Ghosh, A., and S. Vishveshwara, "Variations in Clique and Community Patterns in Protein Structures during Allosteric Communication: Investigation of Dynamically Equilibrated Structures of Methionyl tRNA Synthetase Complexes", *Biochemistry*, Vol. 47, pp. 11398-11407, 2008.

21. Tsai, C., A. D. Sol, and R. Nussinov, "Allostery: Absence Of A Change In Shape Does Not Imply That Allostery is Not At Play", *Journal of Molecular Biology*, Vol. 378, pp. 378-388, 2008.

22. Hilser, V. J., and E. Freire, "Structure-Based Calculation of the Equilibrium Folding Pahway of Proteins. Correlation with Hydrogen Exchange Protection Factors", *Journal of Molecular Biology*, Vol. 262, pp. 756-772, 1996.

23. Baysal, C., and A. R. Atilgan, "Coordination Topology and Stability for the Native and Binding Conformers of Chymotrypsin Inhibitor 2", *Wiley-Liss*, Vol. 45, pp. 62-70, 2001.

24. Latzer, J., T. Shen, and P. G. Wolynes, "Conformational Switching upon Phosphorylation: A Predictive Framework Based on Energy Landscape Principles", *Biochemistry*, Vol. 47, pp. 2110-2122, 2008.

25. Zhuravleva, A., D. M. Korzhnev, S. B. Nolde, L. E. Kay, A. S. Arseniev, M. Billeter, and V. Y. Orekhov, "Propogation of Dynamic Changes in Barnase Upon Binding of Barstar: An NMR and Computational Study", *Journal of Molecular Biology*, Vol. 367, pp. 1079-1092, 2007.

26. Allen, M. P., and D. J Tildesley, *Computer Simulation of Liquid*, Oxford Science Publications, New York, USA, 1991.

27. Allen, M. P., "Introduction to Molecular Dynamics Simulation", *NIC Series*, Vol. 23, pp. 1-28, 2004.

28. Leach, A., *Molecular Modelling: Principles and Applications*, 2nd Edition, Prentice Hall, New Jersey, USA, 2001.

29. Alder, B. J. and T. E. Wainwright, "Studies in Molecular Dynamics 1. General Method", *Journal of Chemical Physics*, Vol. 31, pp. 459-466, 1959.

30. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, "NAMD2: Greater Scalability for Parallel Molecular Dynamics", *Journal of Computational Physics*, Vol. 151, pp. 283-312, 1999.

31. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *Journal of Computational Chemistry*, Vol. 4, No. 2, pp. 187-217, 1983.

32. MacKerell, A.D., Jr. D. Bashford, M. Bellott, R. L. Dunbrack, Jr. J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of proteins", *The Journal of physical Chemistry*, Vol. 102, pp. 3586-3616, 1998.

33. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable Molecular Dynamics with NAMD", *Journal of Computational Chemistry*, Vol. 26, No. 16, pp. 1781-1785, 2005.

34. Maiorov, V. N., and G. M. Crippen, "Significance of Root-Mean-Square Deviation in Comparing Three-Dimensional Structures of Globular Proteins", *Journal of Molecular Biology* Vol. 235, pp. 625-624, 1994.

35. Cohen, F. E., and M. J. E. Sternberg, "On the Prediction of Protein Structure: The Significance of the Root-Mean-Square Deviation", *Journal of Molecular Biology*, Vol. 138, pp. 321-333, 1980.

36. Berendsen, H. J., and S. Hayward, "Collective Protein Dynamics in Relation to Function", *Current Opinion in Structural Biology*, Vol. 10, pp. 165-169, 2000.

37. Kitao, A., and N. Go, "Investigating Protein Dynamics in Collective Coordinate Space", *Current Opinion in Structural Biology*, Vol. 9, pp. 164-169, 1999.

38. Robinson, S., *Production of Membrane Proteins-Strategies for Expression and Isolation*, John Wiley & Sons, USA, 2010.

39. Maisuradze, G. G., A. Liwo and H. A. Scheraga, "Principal Component Analysis for Protein Folding Dynamics", *Journal of Molecular Biology*, Vol. 385, pp. 312-329, 2008.

40. Johnson, R. A., and W. Wichern, *Applied Multivariate Statistical Analysis*, 6th Edition, Prentice Hall, New Jersey, 2008.

41. Kong, Y., and Karpus, M., "The Signaling Pathway of Rhodopsin", *Cell*, Vol. 15, pp. 611-623, 2007.

42. Kong, Y., and Karplus, M., "Signaling Pathways of Pdz2 Domain: A Molecular Dynamics Interactions Correlation Analysis", *Proteins*, Vol. 74, pp. 145-154, 2009.

43. Hyeon, C., Jennings, P. A., Adams, J. A., and Onuchic, J. N., "Ligand-Induced Global Transitions in the Catalytic Domain of Protein Kinase A", *Proceeding of the National Academy of Sciences of the United States of America*, Vol. 106, pp. 3023-3028, 2009.

44. Kamberaj, H., and van der Vaart, A., "Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations", *Biophysical Journal*, Vol. 97, pp. 1747-1755, 2009.

45. Ogunnaike, B. A., and W. H. Ray, *Process Dynamics, Modeling, and Control*, 3rd Edition, Oxford University Press, USA, 1994.

46. Luyben, W. L., *Process Modeling Simulation, and Control for Chemical Engineers*, 2nd Edition, McGraw Hill, New York, USA, 1999.

47. Zhang, Z. Y., and J. E. Dixon, "Active Site Labeling of the Yersinia Protein Tyrosine Phosphatase: The Determination of the Pka of the Active Site Cysteine and the Function of the Conserved Histidine 402", *Biochemistry*, Vol. 32, pp. 9340-9345, 1993.

48. Hoff, R. H., A. C. Hengge, L. Wu, Y. F. Keng, and Z. Y. Zhang, "Effects On General Acid Catalysis from Mutations of the Invariant Tryptophan and Arginine Residues in the Protein Tyrosine Phosphatase from Yersinia", *Biochemistry*, Vol. 39, pp. 46-54, 2000.

49. Lohse, D. L., J. M. Denu, N. Santoro, and J. E. Dixon, "Roles of Aspartic Acid-181 and Serine-222 in Intermediate Formation and Hydrolysis of the Mammalian Protein-Tyrosine-Phosphatase PTP1B", *Biochemistry*, Vol. 36, pp. 4568-4575, 1997.

50. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A Smooth Particle Mesh Ewald Method", *The Journal of Chemical Physics*, Vol. 103, pp. 8577-8593, 1995.

51. Brooks, B. R., C.L. Brooks, III, A.D. MacKerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus,

"CHARMM: The Biomolecular Simulation Program", *Journal of Computational Chemistry*, Vol. 30, pp. 1545-1614, 2009.

52. Ozkaral B., *Investigation of PTP1B Activation Mechanism by Computational Methods*, M.S. Thesis, Boğaziçi University, 2010.

53. Schubert, H. L., E. Fauman, and J. A. Stuckey, and J. E. Dixon, "A Ligand-induced Conformational Change in the Yersinia Protein tyrosine Phosphatase", *Protein Science*, Vol. 4, pp. 1904-1913, 1995.