

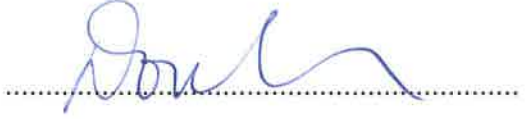
MODELING AND SIMULATION OF PROCESSES REGULATING
PROTEIN BINDING IN VARIOUS ENVIRONMENTS

APPROVED BY:

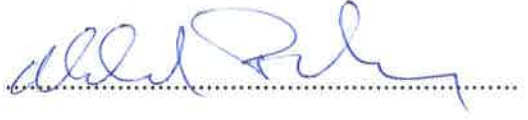
Prof. Dr. Canan Atilgan
(Thesis Supervisor)



Prof. Dr. Pemra Doruker Turgut



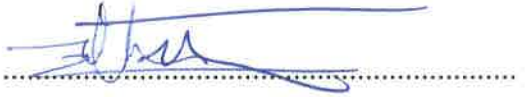
Assoc. Prof. Dr. Melih Papila



Asst. Prof. Dr. Deniz Sezer



Asst. Prof. Dr. Ozge Sensoy



DATE OF APPROVAL: 24/04/2017

**MODELING AND SIMULATION OF PROCESSES REGULATING PROTEIN
BINDING IN VARIOUS ENVIRONMENTS**

by

HALEH ABDIZADEH

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

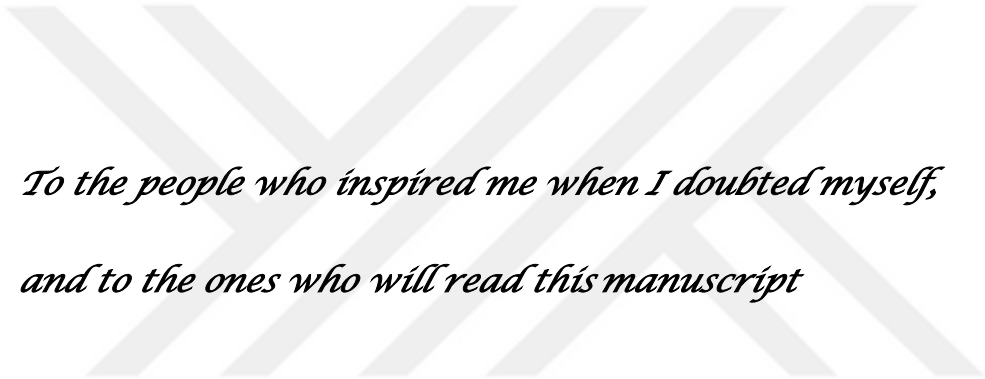
Sabanci University

Spring 2017



© Haleh Abdizadeh 2017

All Rights Reserved



*To the people who inspired me when I doubted myself,
and to the ones who will read this manuscript*

تقدیم به همه کسانی که مشوقم بودند

Modeling and simulation of processes regulating protein binding in various environments

Haleh ABDIZADEH

Materials Science and Engineering, Ph.D. Dissertation, 2017

Thesis Supervisor: Prof. Dr. Canan Atilgan

Co-Advisor: Prof. Dr. Ali Rana Atilgan

Keywords: Molecular dynamics, Anisotropic network model, Perturbation response scanning, Conformational changes, Transferrin, Allostery, Dihydrofolate reductase, Trimethoprim, Antibiotic resistance

ABSTRACT

Conformational changes between structurally different forms of proteins are essential for their equilibrium and kinetic properties. Multiple conformers of the protein scaffolds are triggered by flexibility of the active site, local motions of the loops or allosteric response initiated by distal perturbations. It is our view that the conformational plasticity and allosteric regulations are essential contributors to protein stability, precise specificity and catalytic activity of the enzymes. The large-scale allosteric regulations are believed to involve some

critical sites that mediate the conformational movements between different regions of the free energy landscape. In the first part of the thesis, we have employed Perturbation- Response Scanning (PRS) method unified with elastic network models to predict key residues that induce the dissociated state of 25 non-redundant protein-protein complexes over their bound states. We find that for a group of protein pairs, the functional sites are scattered on the protein surface while for another group the important residues are mainly confined to highly specific regions. We observe a rich mixture of both conservation and variability within the identified residues which we relate to their ability to alter the electrostatic potential distributions and their presence in helix-turn-helix motifs. In the second part of the thesis, we to provide insights into the dynamics of human transferrin (hTf) that transports ferric ions in blood stream with a high affinity and delivers them to cells via receptor mediated endocytosis. We study the effect of pH and salt via molecular dynamics (MD) simulations of total 3 μ s, delineating large conformational changes occurring before and after the iron release process. Via PRS, we show that dynamics within a selected local minimum can lend clues on the global, large scale conformational transition of hTf alone and in its complexes with the human and bacterial receptors. Our studies provide a molecular level understanding of the full cycle of hTf-iron binding-release dynamics while resolving residues directly participating in each event in the cycle. In the third part of the thesis, we discuss the dynamics of dihydrofolate reductase, a ubiquitous enzyme with an essential role in cell growth and an important target by several drugs that act as inhibitors by competing with the ligand, dihydrofolate (DHF). We characterize *Escherichia coli* DHFR in its wild type (WT) and L28R survivor mutant in the presence of the substrate and its inhibitor (trimethoprim). We use extensive MD simulations to determine the conformational space, catalytic relevance of DHFR loop dynamics and hydrogen bond distributions near the active site. The study lays the framework to study the molecular mechanisms that control DHFR mutants conferring resistance to trimethoprim.

ÖZET

Proteinlerin farklı formları arasında üç boyutlu yapı geçişlerinin olması, proteinlerin kinetik ve termodinamik özellikleri açısından elzemdir. Protein üç boyutlu yapıları arası değişim; aktif bölgenin esnekliği, ilmek bölgelerinin yerel hareketleri ve uzak sarsımların başlattığı alosterik cevap gibi etmenlerle tetiklenebilir. Bizim görüşümüze göre, üç boyutlu yapı esnekliği ve alosterik düzenlemeler; protein kararlılığı, hassas özgüllük ve enzimlerin katalitik aktivitesine katkı gösteren önemli elementlerdir. Geniş çaplı alosterik düzenlemelerin, serbest enerji tepelerinin farklı alanlarına tekabül eden üç boyutlu yapı değişimlerini kontrol eden bazı kritik noktaları içerdiği düşünülmektedir. Tezin ilk kısmında, Etki-Tepki Taraması (ETT) ve bu metotla ilintili elastik ağyapı modeli; 25 önemli protein-protein kompleksinin bağlı halinden ayrılmış haline geçişine sebep olan anahtar amino asitleri tahmin etmekte kullanıldı. Protein çiftlerinden oluşan bir grupta, fonksiyonel bölgelerin protein yapı yüzeyine dağıldığını; diğer grupta ise mühim amino asitlerin çok özgül bazı bölgelere gizlendiğini bulduk. Tanımlanmış amino asitlerde zengin bir korunum ve değişkenlik gözlemledik; bu durumu, amino asitlerin elektrostatik potansiyel dağılımları etkileyebilmeleri ve sarmal-ilmek-sarmal motiflerindeki varlıkları ile ilişkilendirdik. Tezin ikinci kısmında; ferrik iyonları yüksek çekim kuvveti ile dolaşıma taşıyan ve reseptörle endositoz vasıtasıyla hücrelere teslim eden insan transferinin (hTf) dinamiğine bir açıklama getirdik. Toplam üç mikrosaniye uzunluğundaki moleküler dinamik (MD) benzetimleriyle, demir ayrılma sürecinden evvel ve müteakip büyük üç boyutlu yapı değişimlerini betimleyerek, pH ve tuz etkisini araştırdık. ETT kullanarak; yalnızca hTf, ve onun insan ve bakteri reseptörleri ile olan komplekslerinde; yerel minimumlarda oluşan dinamiklerin, genel ve büyük ölçekli üç boyutlu yapı değişimleri ile ilgili bilgi verdiğini gösterdik. Bizim çalışmamız, hTf demir bağlanma ve salınım dinamiklerine moleküler seviyede bir anlayış getirmekte; ve bu çevirimdeki her olaya direkt katılan amino asitleri bulmaktadır. Tezin üçüncü kısmında; çok yönlü bir enzim olan, hücre büyümesinde elzem bir role sahip ve dihidrofolat (DHF) ligantı ile rekabetçi ve inhibe edici bir yapı gösteren bazı ilaçların hedefi dihidrofolat redüktazı (DHFR) tartıştık. *Escherichia coli* DHFR'nin yabanıl (WT) ve L28R hayatta kalan mutant yapılarını; substrat ve inhibitör (trimetoprim) varlığında karakterize ettik. Üç boyutlu yapı uzayına, DHFR ilmeğinin katalitik alacasına ve aktif bölge yakınındaki hidrojen bağı dağılımlarına karar vermek için kapsamlı MD benzetimleri kullandık. Bu çalışma DHFR mutantlarının

trimetoprim direnci geliřtirmesinin moleküler mekanizmalarına genel bir çerçeve oluřturmaktadır.

Haleh ABDIZADEH

Malzeme Bilimi ve Mühendislięi, Doktora Tezi, 2017

Tez Danıřmanı: Prof. Dr. Canan Atilgan

İkincil Danıřman: Prof. Dr. Ali Rana Atilgan

Anahtar kelimeler: Moleküler dinamik, Elastik aęyapı model, Etki-Tepki Taraması, Proteinlerin farklı formları, Transferin, Alosteri, Dihidrofolat redüktaz, Trimetoprim, Antibiyotik direnci

ACKNOWLEDGEMENTS

I would like to take the opportunity and express my gratitude to many people who provided me with the possibility to complete this huge, fulfilling and sometimes frustrating task.

First, I would like to acknowledge and thank my supervisors, Dr. Canan Atilgan and Dr. Ali Rana Atilgan for lending me their patient guidance and never frowning on my failures. Canan hocam, you were the person who accepted me to your lab, introduced me to the amazing biophysics research and encouraged me when I doubted my abilities. You were the first one to teach me I could do anything if I had the will. Ali Rana hocam, I appreciated our discussions after every one of which I felt smarter. It was with the inspirations from you that I got to experience science as a passion and not as a frustration. I am sincerely grateful to both of you and very proud to be your student.

In the MIDST lab, Gizem Ozbaykal, Caroline Ross, Omer Acar, Sofia Piepoli, Furkan Guclu, Goksin Liu, Nazli Kocatug and Aygul Zengin were the ones with whom I discussed science and our life events. Upon my arrival in Turkey, I could not speak Turkish but they never made me feel I am a stranger among them. I like to thank them for creating such a friendly lab environment.

I am deeply indebted to my parents and my dear sister for their immeasurable love and care. They have always supported me to explore my potentials and pursue my dreams. Without them, it would have been impossible for me to reach this stage and I salute them for their selfless love and sacrifices that shaped my life.

I owe my loving thanks to a very special person, my long time companion and best friend Jamal Seyyed Monfared Zanjani for his genuine care, constant support and understanding, patient help, keeping me motivated when I wanted to give up and stimulating scientific discussions. Jamal, you always believed in me and I am truly grateful to have you in my life.

Finally, I would like to acknowledge financial support from the Scientific and Technical Research Council of Turkey (TUBITAK) with the project numbers of 110T624, 113Z408 and 114Z527.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZET	vi
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xi
LIST OF TABLES	xxii
Chapter 1. Introduction and overview of the thesis	1
1.1. Overview	1
1.2. Modeling the long-term dynamics of protein complexes by coarse-grained methodologies	2
1.3. Modeling the equilibrium and dynamical properties of iron transport proteins	5
1.4. Modeling the dynamics and stability of dihydrofolate and trimethoprim binding to <i>E.Coli</i> dihydrofolate reductase	10
Chapter 2. Methodology	14
2.1. Molecular Dynamics (MD) Simulations: Theory	14
2.1.1. MD simulation protocol used for hTf systems	15
2.1.2. MD simulation protocol used for DHFR systems	19
2.2. Steered MD (SMD) simulation: Theory	23
2.2.1. SMD simulation protocol	24
2.3. Anisotropic Network Model (ANM): Theory	25
2.4. Perturbation Response Scanning (PRS): Theory and Protocol	27
Chapter 3. Modeling the long-term dynamics of protein complexes by coarse-grained methodologies 30	
3.1. Electrostatics in remote and mechanical control of protein-protein interaction ..	30
3.1.1. Allosteric control of disassociation is coupled to electrostatic effects.....	45
3.1.2. Auto-controlled conformational transitions is correlated to mechanical organization of the protein	50
3.1.3. Combined perturbation of two sides of the protein pair is essential for dissociation in group IV of protein complexes	51
3.1.1. Mechanics of the protein motifs explain the structural preference in PRS residues	52
3.2. An example of PRS at work: PRS specifies key regions in Subtilisin serine protease for both function and stability	53
3.2.1. Conformational differences between the inhibited and free enzyme structures occur along two flexible loops	54
3.2.2. PRS analysis of the enzyme reveals single residues that control the conformational change	55
3.2.3. Biological relevance of the residues specified by PRS method	57
Chapter 4. Modeling the equilibrium and dynamical properties of iron transport proteins: The human serum transferrin	60
4.1 Detailed molecular dynamics simulations of human transferrin provide insights on iron release dynamics at serum and endosomal pH	60
4.1.1. Domain opening is coupled to the protonation states of iron coordinating tyrosines	63
4.1.2. The partially open state is a stable intermediate of hTf	66
4.1.3. The dilysine trigger is not strong enough to destabilize the Fe ⁺³ binding cleft and induce domain opening	71
4.1.4. At serum pH, fluctuations of the N-lobe are communicated to the C-lobe through	

the peptide linker	73
4.1.5. At serum pH, relative fluctuations of the N- and the C-lobes differentiate iron release kinetics	75
4.1.6. Endosomal pH prompts hTf to an open conformation	77
4.2 Predicting long term cooperativity and specific modulators of receptor interactions in human transferrin from dynamics within a Single Microstate	78
4.2.1. D63 mutants affect iron binding propensity	83
4.1.1. The two domains of holo hTf are perfectly correlated at serum pH, resisting iron release	84
4.1.2. Endosomal pH leads to a loss of cooperativity between the two domains of holo hTf and readily promotes iron release	86
4.1.3. hTf – receptor complexes	88
4.2 Conformational dynamics of human serum transferrin is regulated by salt concentration	90
4.2.1. Conformational dynamics of human serum transferrin is regulated by salt concentration	91
4.2.1.1. Conformational preferences of hTf under different conditions.....	91
4.2.1.2. Addition of salts initiates a balanced electrostatic potential distribution on the protein surface	94
4.2.1.3. Upon triggering lobe opening, conformational preferences of the two lobes become IS dependent	97
4.2.1.4. Physiological salt concentration is optimal for iron release from the N-lobe....	101
4.2.1.5. R124 is the major KISAB site in the N-lobe.....	102
4.2.1.6. Exposed iron cannot dissociate without chelators.....	106
Chapter 5. Increased Substrate Affinity in the Escherichia Coli L28R Dihydrofolate Reductase Mutant Causes Trimethoprim Resistance	108
5.1. Protonated state of trimethoprim in WT enzyme and L28R mutant	109
5.2. Conformational preferences of DHF-bound DHFR.....	114
5.3. L28R mutation disrupts water organization at the trimethoprim – enzyme interface.	120
Chapter 6. CONCLUSIONS and future outline	124
6.1. Conclusions	124
6.2. Future outlook	128
APPENDIX A	130
BIBLIOGRAPHY	135

LIST OF FIGURES

Figure 1.1 (a) Crystal structure of holo hTf; N-lobe is displayed in blue surface and the C-lobe is displayed in gray surface. The flexible linker bridging the two lobes is in green. In (b) and (c), the crystal structure of partially open and apo forms of hTf are shown. Color coding is the same as in (a). (d) The octahedral coordination of iron in the N-lobe is shown in vdW representation. The iron coordinating residues from the N ₁ subdomain are in ice blue while the residue from the N ₂ subdomain is in cyan. (e) The iron coordinating residues from the C1 subdomain are in silver and the sole iron coordinating residue from the C ₂ subdomain is in gray. In both lobes, iron is displayed in red CPK and carbonate is in yellow.	7
Figure 1.2 Cartoon representation of E.coli DHFR based on PDB structure 1rx2 is displayed in transparent gray. The characteristic loops of the enzyme are shaded in orange, purple, gray and salmon indicating M20, CD, FG and GH loops, respectively. DHFR is complexed with DHF and NADPH in its native form.	12
Figure 2.1 Binding site residues of TMP/TMP ⁺ bound DHFR. Atom names involved in hydrogen bond formation are in red and atom names involved in torsional angles are in green.	21
Figure 2.2 Binding site residues of DHF bound DHFR. Atom names involved in hydrogen bond formation in the original structure are in red and atom names involved in torsional angles are in green.	21
Figure 2.3 Rotational movements of DHF p-aminobenzoyl glutamate tail in (a) WT ^{DHF} and (b) L28R ^{DHF} systems. The trajectory of the unit vector connecting N10 to CD atom position is shifted to (0,0,0) for better visualization of the tail movements which are projected on the x-y plane on the upper panels; color variations denote z-coordinate changes. The volume swept by the CD atom positions are displayed as a gray surface in the lower panels, where one of the poses of DHF is shown in stick representation. Principal axis contribution is 0.45, 0.33 and 0.22 for WT ^{DHF} , which demonstrates a nearly free rotation of the tail in the binding site. Principal axis contribution is 0.69, 0.23 and 0.08 for L28R ^{DHF} revealing a stiffened tail with the motion predominantly confined to the major axis.	22
Figure 2.4 Constant velocity SMD simulation	23

Figure 3.1 Electrostatics isocontours of α -amylase/ α -amylase inhibitor drawn at ± 0.5 kBT/e. Blue is positive and red is negative. (a) α -amylase in the unbound form, with overall negative electrostatic potential distribution along the surface. (b) α -amylase inhibitor in its unbound form with a mixed pattern of negative and positive negative electrostatic potential distribution along the surface. (c) α -amylase/ α -amylase inhibitor complex with overall negative electrostatic potential distribution. The spatial orientation of the proteins in the complex is kept the same as respective unbound forms. Dashed circle displays positioning of the inhibitor in the complex. Rest of the surface belongs to α -amylase protein. The signaling protein, α -amylase, maintains its electrostatic potential distribution while α -amylase inhibitor displays altered electrostatic potential distribution. 46

Figure 3.2 Electrostatics isocontours of Efb-C and its complement target C3d.drawns at ± 0.5 kBT/e. Blue is positive and red is negative. (a) C3d form, with overall negative/positive electrostatic potential distribution along the surface. (b) Efb-C in its unbound form with negative/positive, mostly positive, electrostatic potential distribution along the surface. (c) Efb-C/ C3d complex. Dashed circle displays Efb-C protein in the complex and the rest of the surface belongs to C3d protein. Both proteins display a mixture of negative/positive electrostatic potential distribution along the surface. Mostly, positive surface of the Efb-C displays grown negative areas upon complex formation. The spatial orientation of the proteins in the complex is kept the same as respective unbound forms..... 47

Figure 3.3 (a) Ribbon representation of C3bot-RalA complex as an example of type I proteins, C3bot is in purple and RalA is in silver. PRS rated residues that control the disassociation of the complex are displayed in VDW and all are confined to RalA compartment. Residues controlling the disassociation of RalA are in yellow while the residues controlling disassociation of C3bot are in green. (b) Ribbon representation of chemotaxis CheA:CheY complex as an example of type II proteins; CheY is in tan and CheA is in blue. Residues controlling the disassociation of CheY are located on CheA and are displayed in green while residues controlling the disassociation of CheA are located on CheY and are displayed in cyan. (c) Ribbon representation of alkaline protease and its cognate inhibitor as an example of type III proteins. Protease is represented as white cartoon and the inhibitor is in ice blue. Residues controlling the disassociation of protease are located on it and displayed in blue. (d) Ribbon representation of dI2dIII1complex of proton-translocating transhydrogenase as an example of type IV proteins. Residues that control the disassociation of the dI subdomain are

located on both subunits and are displayed in red. Residues that control the disassociation of dIII subunit are displayed in green. 49

Figure 3.4 (a) schematic of the RLR motifs. Length of the rigid structure is L and the angle between two rigid structures connected by a flexible loop is θ . (b) Distribution of PRS residues located on rigid structures in yellow and PRS residues located on flexible loops in purple. Analysis pool is displayed in gray dots. 53

Figure 3.5 Subtilisin alignment in two states, 1OYV chain B representing the inhibited structure in light gray, and 1SBC for the apo form in dark gray are shown in cartoon representation. The overall structures are well-aligned while the two loops spanning Phe50-Gly61 and Ser159-Ile165 which display higher RMSD are shown in black. 55

Figure 3.6 PRS map of the interconversion of TI-II/subtilisin complex to apo subtilisin. The displacements of the affected residues are displayed for the case which gives the best correlation for a given perturbed residue. Note that the displacements are on a relative scale, since PRS relies on linear response theory; i.e. if forces in a selected direction are modified by a factor r , then all displacements are also modified by the same factor r . Lower panel displays the relevant maximum Pearson correlation value for all residues. We select the residues that give the largest correlations as candidates controlling the conformational change. 56

Figure 3.7 Residues specified by PRS as being effective in the transition of inhibitor bound subtilisin to apo subtilisin are displayed on the TI-II and subtilisin complex. The pronounced regions are displayed as black cartoons, while the residues with the largest correlations in a given region are further emphasized in ball-and-stick representation and labelled. TI-II is represented as gray colored transparent ribbon and subtilisin as silver colored cartoon. 58

Figure 4.1 RMSD curves for H^{**} , H^* and H systems, labeled as defined in Table 2.3. The overall RMSD (black) as well as that of the N- and the C-lobes are reported (blue and gray, respectively). When both of the iron coordinating Tyr residues in the lobes (a) or only from each lobe (b) are negatively charged, the lobes are sealed tight. The N-lobe fluctuations are slightly more pronounced in (b) than in (a). The overall RMSD of the protein, as well as that of either lobe is below 3 Å throughout the time window of observation. (c) When all Tyr are in their standard protonation states, the N-lobe begins to open up within 10 ns while the C-lobe remains closed. 64

Figure 4.2 RMSD curves for H, PO and A systems (respectively, holo, partially-open and apo hTf at physiological pH) and H[†] system (holo hTf at endosomal pH). The overall RMSD (black) as well as that of the N- and the C-lobes (blue and gray, respectively) are reported. Insets display surface representation of crystal structure for each system. Horizontal dotted lines represent the overall RMSD averaged over last 50 ns of each simulation. (a) In H, N-lobe displays large rearrangements from the beginning of the MD simulation. After 100 ns, these are communicated to the C-lobe. (b) In PO, both the N- and C-lobe stays near the initial conformer. The overall structure displays larger RMSD values due to relative placement of the lobes, communicated by the linker. (c) In A, the C-lobe is more mobile than the N-lobe throughout. It oscillates around the equilibrium state in a hinge like motion (see text for details). (d) In H[†] where the lower pH of the endosome is mimicked, large rearrangements occur simultaneously in both lobes. 65

Figure 4.3 Time evolution of the RMSD of H system and its lobes from different crystal structures. (a) The original overall RMSD curve is shown in black (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (brown and green, respectively). Once the initial conformational change is completed within 10 ns, the newly attained conformation displays lower RMSD from the PO x-ray structure. (b) The original N-lobe RMSD curve is shown in dark blue (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (shown as lighter shades of blue). N-lobe of H system displays excellent resemblance to the N-lobe of partially open structure. (c) The original C-lobe RMSD curve is shown in dark gray (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (shown as lighter shades of gray). Since the RMSD between the x-ray structures for the C-lobe of H and PO structures is very low (0.6 Å, Table 1), the curve for H→PO comparison is nearly identical to the original curve. The difference from the x-ray structure of A diminishes throughout the trajectory. 67

Figure 4.4 Eigenvectors displaying the dominant mode of motion in the C-lobe of apo the structure, obtained from four 50 ns–long portions of the A simulation, superposed on the average structure of each portion. While the RMSD from the crystal structure gets as large as 8 Å in these simulations, the average structures are within 2.8 Å of each other. Simulations displays the same large hinge bending motions around the average structure as shown, and does not exhibit a conformational shift. 69

Figure 4.5 Schematics of hTf displaying the color code used throughout the text (blue for N-lobe, green for linker, gray for C-lobe; pink circles represent Fe^{+3} ions), the points used to define the reduced degrees of freedom (shown by $\otimes$) are the centers of mass of the N- and C-lobes, the beginning and end points of the linker, and the hinge residues between the subdomains of each domain, Y95 and Y426. We define the lobe bending angles θ_N and θ_C , measured by the angle between the two centers of mass and the hinge residue between each subdomain of the respective N- and C-lobes. The lobe torsional angle φ , is defined by the four points shown. These are the major degrees of freedom on which we project all trajectories to monitor the sampled conformations. 70

Figure 4.6 Conformations sampled by H, PO, A, $\text{H}^\dagger$ displayed by projecting on the reduced degrees of freedom, defined in Figure 4.5. The first column shows the joint probability distribution of the θ_N, θ_C intra-lobe bending angles, therefore the differences in their sampled regions. The initial position is shown by the red dot: While the inequality is expected for the PO structure in the x-ray structure, even the holo and the apo forms are positioned at slightly different values; the N-lobe is slightly more open in both cases. The second column shows the distribution of the torsional angle (φ) quantifying the relative positioning of the two lobes. Initially they are at $-15^\circ, -30^\circ, -40^\circ$, and -5° , respectively for H, PO, A, $\text{H}^\dagger$. The relative position of the iron loaded forms at serum pH is ca. 50° different than the iron free form. The endosomal pH system is more labile and samples both positions frequented by the apo and holo systems at higher pH. Shown on the right are the superposed structures of the corresponding systems, 50 snapshots spaced at 4 ns intervals. 71

Figure 4.7 Relative positioning of the subdomains and the schematics of the exit pathways of holo hTf. The two lobes are positioned such that the exit paths whose approximate directions are shown by the arrows are nearly orthogonal. The linker is shown in green. The loops that display large differences in the RMSF profiles are shown in blue on the N-lobe (residues 159-189 and 274-295) and dark gray in the C-lobe (residues 488-519); see Figure 4.9 and the main text for details. 73

Figure 4.8 RMSD profiles of the (a) iron loaded isolated lobe H_N and H_C runs compared to those from (b) fully loaded H. The N-lobe is colored in blue, and the C-lobe in gray. The N-lobe has propensity of open in both the full structure and the single lobe case. The C-lobe, on

the other hand, has an intrinsic stability that prevents it from opening unless triggered by the conformational changes in the N-lobe. 74

Figure 4.9 C_{α} root mean square fluctuations averaged over equally spaced trajectory pieces of 50 ns each. a) RMSF of H system in black, b) RMSF of HN system is on the left (residues 1-331) and HC on the right (residues 339-677). Regions with more pronounced fluctuations in the N-lobe, linker and the C-lobe are highlighted in blue, green and gray, respectively. In the case of H system, in the first 50 ns, more pronounced fluctuations belong to the N-lobe. After 50 ns, the linker initiates fluctuations and the delayed communication between the lobes prompts the C-lobe to go through larger fluctuations after 100 ns. For holo hTf at pH=5.6, large fluctuations begin from the starting point of simulation. 75

Figure 4.10 RMSD profiles of the (a,b) monoferric full length hTf, H^{Fe+N} and H^{Fe+C} , runs compared to those from (c) fully loaded H. The N-lobe is colored in blue, and the C-lobe in gray. Originally, the N-lobe is more mobile than the C-lobe in all cases. In monoferric form with iron in the C-lobe, after 80 ns, the mobility of the C-lobe gets similar to that of the N-lobe despite the associating role of iron present in the former..... 77

Figure 4.11 (a) RMSD of the A^C system from its initial structure. RMSD of the protein, the N-lobe and the C-lobe is displayed in black, blue and gray, respectively. N-lobe is equilibrated in the first 5 ns and does not display large fluctuations. However, RMSD of the C-lobe increases up to ca. 5 Å and affects the overall structure. In the inset, RMSD of the apo hTf in the absence of the carbonate anion (A system) is shown. Equilibrated fluctuations of the apo hTf in the presence of carbonate anion in comparison to the large fluctuation in its absence show how carbonate stabilizes the apo form of hTf. (b) RMSD of AC system from the holo hTf structure. Color coding is the same as in (a). The initial large RMSD decreases with time, demonstrating the closing effect of carbonate binding to the apo conformation. (c) Distance profile between residues F186 and K291, for the AC (black) and A (red) systems, respectively. Upon entry of the synergistic anion, the distance between the tip residues residing on the loops enclosing the cavity decreases from 18.2 ± 2.1 Å to 10.6 ± 1.6 Å (averaged after the first 20 ns of each trajectory). (d) Superposition of the initial (translucent) and final (transparent) structure of the AC system at 100 ns for the N-domain (blue) and C-domain (gray); best fitting of the structures is to the N2 and C2 subdomains, respectively, so that the closing of the other subdomain over the cavity (shown by the arrows) is better visualized. C_{α}

atoms of F186 and K291 whose separation is monitored in part (c) are displayed as red spheres..... 80

Figure 4.12 Electrostatics isocontours (top row) and cross-correlation maps (eqn (S2), bottom row) for holo hTf, at (a) serum pH, Hs and (b) endosomal pH, He. Maps are normalized so that the largest correlation is 1 (black). The regions corresponding to N-lobe – N-lobe interactions are boxed in purple and that for the intra-C-lobe are boxed in green. The N-lobe – C-lobe cross correlations appear in the remaining regions and display nearly complete lack of cooperativity between the two subdomains at endosomal pH, while pointing out residues organizing cooperativity at serum pH. Residues acting cooperatively through all regions in the latter are highlighted by red strips, also indicated by the arrows on the symmetric map. Above each system, Hs and He, snapshots of electrostatics isocontours of hTf drawn at $\pm 0.5kBT/e$ are displayed. Blue is positive and red is negative. For each case, hTf is shown in two orientations: On the left, front views of the protein are displayed where the linker is placed at the back; on the right, a 180° rotation along the vertical axis is displayed with the linker in view. At serum pH, protein surface is predominantly negatively charged. At endosomal pH the two lobes reorient and the subdomains of either lobe are opened to expose the iron binding cleft. The charge distribution on the surface is now predominantly positive, especially on the C-lobe..... 85

Figure 4.13 (a) Slowest three modes of the C matrices obtained for holo hTf at serum pH (Hs). Protein structure is reoriented to display the mode shapes indicated by the arrows. The motions within each lobe are unified for the whole domain while the N- and the C-lobes have opposite sense of direction in all three modes. There are no sub-domain separation motions, consistent with the slow release kinetics at serum pH. (b) Slowest three modes of the C matrices obtained for holo hTf at endosomal pH (He). Protein structure is displayed in the same direction for all three cases. Each lobe moves independent of the other, In the slowest mode of motion, both lobes are mobile and correlated within the domain, but there is lack of correlated motions between the two domains. In the next two modes, N-lobe motions are dominant and there is little to no motion in the C-lobe. Mode II displays hinge-like opening around the iron binding cleft of the N-lobe, consistent with iron removal at this pH..... 86

Figure 4.14 Complexes of hTf with receptor from human and bacteria. (a) hTf-TfR complex. Slowest mode of motion of the complex (contributing 49% of the motion) is displayed by the

arrows. TfR (yellow) interacts with both lobes, although the C-lobe interface has more cohesive interactions. (b-c) hTf-TbpA complex. The two slowest modes of motion contributing ~13 % each have mainly N-lobe motions; these lack domain separation consistent with iron removal by TbpA only from the C-lobe. 89

Figure 4.15 Residues emerging as conformational switch from TbpA bound to free conformation of hTf appear as two patches that display large C_i values (orange). The direct switch contacts the plug loop in the pathogenic protein responsible for sequestering iron from the C-lobe of hTf is (red). The proposed allosteric switch is 30 Å away from the plug loop. Both switches are complemented by residues harboring rapidly evolving residues (blue) (170). 89

Figure 4.16 (a) Three-dimensional structure of holo hTf based on PDB code 3v83. The N-lobe (residues 1-331), linker (residues 332-338) and the C-lobe (residue 339-677) are displayed in blue, red and gray, respectively. This color coding of the regions is used throughout the text. The irons are displayed as pink spheres. (b) The identical ferric ion binding site consist of an Asp, two Tyr, and a His residues along with the synergistic carbonate anion. (c) Schematic of hTf displaying the color code used in (a), the points used to define the reduced degrees of freedom are $\otimes$ to show the centers of mass of the N- and C-lobes and $\blacklozenge$ display the beginning and end points of the linker, and the hinge residues between the subdomains of each domain, Y95 and Y426. We define the lobe bending angles θ_N and θ_C , measured by the angle between the two centers of mass and the hinge residue between each subdomain of the respective N- and C-lobes. The lobe torsional angle ϕ , is defined by the four points shown. These are the major degrees of freedom on which we project all trajectories to monitor the sampled conformations. (d) Heavy atom RMSD curves for the full structure of hTf in H group at various IS values. Systems at zero, 50, 130 and 210 mM are displayed in green, blue, black and red, respectively. The same color coding is used throughout the text for the various IS.. 92

Figure 4.17 RMSD curves for H systems are displayed. The overall RMSD (a) as well as that of the N-lobe (b) and the C-lobe (c) are reported (green, blue, black and red, respectively for H_0 , H_L , H_P and H_H). 93

Figure 4.18 RMSD curves for H^{**} systems are displayed. The overall RMSD (a) as well as that of the N-lobe (b) and the C-lobe (c) are reported (green, blue, black and red, respectively for H_0^{**} , H_L^{**} , H_P^{**} and H_H^{**}). 93

Figure 4.19 Electrostatic isocontours of H^{**} group (at 0 mM top row and 130 mM bottom row). 50 ns time point snapshots of electrostatic isocontours of hTf are drawn at $0.5 k_B T/e$ are displayed. Blue is positive and red is negative. For each case, hTf is shown in two orientations. 96

Figure 4.20 Radial distribution functions (RDF, $g(r)$) between side-chain atoms and nonspecific ions in the solvent (a), (b) Comparing distribution of Na^+ around side chain atoms of Asp and Glu residues of the protein in H_L^{**} (blue), H_P^{**} (black), H_H^{**} (red) systems. (c), (d) Comparing distribution of Cl^- ions around side chain atoms of Arg and Lys residues. Na^+ ions are accumulated around protein surface while Cl^- are more prone to be scattered in the solvent. 97

Figure 4.21 C_α root mean square fluctuations averaged over 30 ns long trajectories for **H** group systems. Error bars reflect the uncertainty obtained from averaging six sets of 30 ns length after removing the first 20 ns of the simulations for equilibration. Residues that display more than 50% change in standard deviation at different IS are indicated with arrows. The linker region (residues 332-338, shaded in transparent red) is less mobile in zero IS as compared to systems with larger IS. 98

Figure 4.22 Conformations sampled by (a) H_0 , (b) H_L , (c) H_P and (d) H_H systems calculated by projecting on the reduced degrees of freedom, defined in 99

Figure 4.23 Conformations sampled by (a) H_0^{**} , (b) H_L^{**} , (c) H_P^{**} and (d) H_H^{**} systems is calculated by projecting on the reduced degrees of freedom, defined in 100

Figure 4.24 Radial distribution functions (RDF, $g(r)$) between side-chain atoms and nonspecific ions in the solvent (a), (b) Comparing distribution of Na^+ around side chain atoms of Asp and Glu residues of the protein in H_L (blue), H_P (black), H_H (red) systems. (c), (d) Comparing distribution of Cl^- ions around side chain atoms of Arg and Lys residues. Na^+ ions are accumulated around protein surface while Cl^- are more prone to be scattered in the solvent. 103

Figure 4.25 A Cl^- ion (cyan sphere) enters into the N-lobe binding cleft and establishes a persistent electrostatic interaction with residues R124 (tan stick), K206 and K296 (purple

sticks). The distances between amino group of each residue and the Cl^- ion are labeled on the figure. Only N-lobe (transparent cartoon) is shown for clarity. 105

Figure 5.1. RMSD profiles for the 210 ns long MD simulations. On the left column, DHF, TMP^+ and TMP bound systems in WT are reported in black and gray, respectively for the first and second runs. On the right column, DHF, TMP^+ and TMP bound systems in the L28R mutant are displayed. 108

Figure 5.2 (a) Cartoon representation of *E.coli* DHFR for the PDB structure 1rx2 is displayed in transparent gray. The characteristic loops of the enzyme are shaded in orange, purple, dark gray and red for the M20, CD, FG and GH loops, respectively. DHFR is complexed with DHF (black) and NADPH (lavender) in its native form. (b) RMSF profiles of WT^{DHF} , WT^{TMP^+} , WT^{TMP} , L28R^{DHF} , $\text{L28R}^{\text{TMP}^+}$ and L28R^{TMP} averaged over all equally spaced chunks of 40 ns are displayed; error bars for each residue are shown in gray. The average fluctuations obtained from crystallographic B-factors are displayed in bottom panel. The regions of the characteristic loops of the enzyme are shaded in orange, purple, gray and red indicating M20, CD, FG and GH loops, respectively. 110

Figure 5.3 (a), (b) Schematic representation of hydrogen bonds in TMP^+ and TMP binding site in *E.coli* DHFR. Hydrogen bonds to the indicated amino acid on DHFR (gray circles) are shown by the colored arrows. (c-f) Average hydrogen bond distances formed between trimethoprim and binding site for WT^{TMP^+} , WT^{TMP} , $\text{L28R}^{\text{TMP}^+}$, L28R^{TMP} , respectively. Standard deviations are shown by vertical lines. Color coding follows the colors used in the arrows in (a) and (b). 112

Figure 5.4 Average distances of characteristic binding pocket hydrogen bonds recorded in 50 ns trajectories of the D27N and D27S mutants for TMP^+ and TMP bound cases. Standard deviations are shown by vertical lines; color coding is the same as in figure 2. The more stable form of the drug is the one which maintains the characteristic hydrogen bonds. For D27N, this is the TMP bound form in contrast to the WT which accommodates TMP^+ (Figure 5.3); the D27S mutant binds both forms. 114

Figure 5.5 (a) Average structure of WT^{DHF} (gray) is superposed on that of L28R^{DHF} (red). The main difference is observed in M20, FG and GH loops. (b) The residue-residue cross correlation maps between WT^{DHF} (lower diagonal) and L28R^{DHF} (upper diagonal), averaged

over chunks of 40 ns. Correlations are much stronger, extending over the whole protein in the WT compared to the L28R mutant. (c,d) Superposed structures of WT^{DHF} and L28R^{DHF}, clearly display the mobility of the loops. 20 equally spaced snapshots are shown..... 115

Figure 5.6 (a) Schematic representation of hydrogen bonds in DHF binding site in *E.coli* DHFR. Hydrogen bonds to the indicated amino acid on DHFR (gray circles) are shown by the colored arrows. (b), (c) Average hydrogen bond distances formed between DHF and binding site for WT^{DHF} and L28R^{DHF}, respectively. Standard deviations are shown by vertical lines. Color coding follows those used in the arrows in (a). (d) Sample donor-acceptor distance between the methyl group of residue 28 and DHF. For the WT, L28 makes no hydrogen bonds with the p-aminobenzoyl glutamate tail of the substrate. Sample donor-acceptor distances between methyl group of R28 and DHF indicates there are sustained hydrogen bonds between the α -amino group of R28 and p-aminobenzoyl glutamate tail of DHF in the L28R mutant. (e), (f) sample pose of DHF relative to residue 28 in WT^{DHF} and L28R^{DHF}, respectively. R28 interacts with DHF p-aminobenzoyl glutamate tail via its side chain and further stabilizes the ligand. 119

Figure 5.7 (a) Average structure of WT^{TMP⁺} (gray) is superposed on that of L28R^{TMP⁺} (red). The main difference is observed in the CD loop. (b) The residue-residue cross correlation maps between WT^{TMP⁺} (lower diagonal) and L28R^{TMP⁺} (upper diagonal), averaged over chunks of 40 ns. Correlations are subsided in TMP⁺ bound forms, slightly more so in the L28R mutant than the WT. (c,d) Superposed structures of WT^{TMP⁺} and L28R^{TMP⁺} clearly display loop motions. 20 equally spaced snapshots are shown. 121

Figure 5.8 Isothermal Titration Calorimetry results for TMP binding are shown for WT and L28R DHFR. Each experiment is repeated in triplicate. All K_D , ΔH , ΔS and ΔG values are significantly different between the WT and the L28R mutant. 122

LIST OF TABLES

Table 2.1 RMSD between x-ray structure pairs*	17
Table 2.2 pK_a values ¹ for holo hTf residues calculated by H++ and PROPKA servers	18
Table 2.3 Summary of system conditions for different simulations	19
Table 3.1 Summary of PRS results	31
Table 3.2 Amino acid type analysis of PRS identified residues	43
Table 3.3 Structural analysis of PRS identified residues	44
Table 3.4 Structural locations of PRS identified residues based on RLR motif analysis	53
Table 3.5 Summary of PRS results for $r_{opt}=8$ Å. Hot residues are the ten residues with highest Pearson correlation values in comparison to whole sequence	57
Table 4.1 Distances from coordinating residues ¹ to iron in the beginning ² /end ³ of simulations.	66
Table 4.2 Summary of PRS results	82
Table 4.3 Average ferric ion distances from interacting side chain of coordinating residues (all values in Å). Distances are measured between OD1, OH and NE2 atoms of ASP, TYR and HIS residues, respectively and ferric ion. Errors are ~0.01 Å. SASA has been calculated for residues surrounding ferric ion (D63,Y95,Y188 and H249 for the N-lobe; D392,Y426,Y517 and H585 for the C-lobe). We did not include water molecules in the calculations.....	94
Table 4.4 Estimated free energy differences upon ferric ion binding in H and H ^{**} systems in the N- and the C-lobes (all values in kcal/mol.).....	107
Table 5.1 Loop and overall RMSD (Å) in the systems with respect to initial structure, averaged over the 210 ns trajectories.	116
Table 5.2 Experimental results for the competitive inhibition of DHFR.....	119

CHAPTER 1. INTRODUCTION AND OVERVIEW OF THE THESIS

1.1. Overview

Early view of proteins and nucleic acids as being relatively static and rigid structures has now been revamped by highly dynamic entities whose internal and collective movements are essential to their function (1). The regulation of biological molecules' function is now treated as being modulated by molecular motions occurring on different time and length scales (2). Likewise, chemical and physical processes within assemblies of biomolecules in the cellular milieu are accomplished based on the transitions between stable conformations where mutations, protein and ligand binding scenarios and interaction with solvent molecules shift their equilibrium distribution of conformational substates. NMR structure determination and relaxation rate measurements (3-5), tryptophan fluorescence depolarization (6), inelastic neutron scattering (7) and time resolved X-ray structure refinements (8, 9) are commonly used experimental methods to analyze protein motions. However, these methods have low time and spatial resolution and report biological properties averaged over ensembles. Recent developments in three-dimensional structure determination techniques and computer simulations operative on different time and size scales supplement experimentally obtained biochemical knowledge that fall short of reporting precise nature of individual molecules movements. From modeling the active site mechanisms using quantum mechanics only (10) and quantum mechanical molecular mechanics (QM/MM) calculations (11), to the study of protein motions on the nano-microsecond time scales through molecular dynamics (MD) simulations (2), a plethora of information that helps explain many experimental observations is now routinely obtained. Coarse-grained methodologies have also become useful tools in detecting protein motions on the global scale as well as for discovering sites of allosteric communication.(12, 13).

In this thesis, we combine different computational approaches to interpret equilibrium, thermodynamic and kinetic properties of proteins in interaction with small ions, inhibitors and proteins with reasonable accuracy. In the first part, we outline coarse-grained models to analyze long-term modulators of conformational change propensity between pre-defined states of protein complexes. Specific residues that affect global dissociation of protein

complexes are introduced and categorized structurally. The focus of the second part is to describe the state of the art classical and biased MD simulations along with coarse-grained analytical techniques in determining environmental and allosteric effects in iron binding, transport and release by transferrins. In the third part, we discuss a classical drug target, *E. Coli* dihydrofolate reductase enzyme and rationalize the dynamic features of antibiotic resistance developed in the mutant strains via MD simulations and non-equilibrium work approaches.

1.2. Modeling the long-term dynamics of protein complexes by coarse-grained methodologies

Coarse-grained (CG) models with several atoms of the protein grouped into one bead and in the absence of atomic details of the solvent molecules have been developed to supplement extremely expensive modeling of large scale motions of bimolecular aggregates (14). CG models have proved to be able to gain sampling efficiency, predict allosteric regulations (15) and describe the conformational transition pathways (16, 17). One useful measure of large-scale protein mechanics in the context of CG models is the elastic network model (ENM) (18, 19). ENM is based on the assumption that potential energy of the system is approximated harmonically about a single minimum energy conformation. Methodologically, in ENM, the atomic details of the biomolecule structure is reduced to a network of nodes (typical one site per residue) and harmonic springs. Since, all springs are at their equilibrium positions in the network, no energy minimization is required in comparison to normal mode analysis in which an expensive initial energy minimization is required prior to calculating the hessian matrix. Alternative approaches predict an ENM model with a resolution lower than standard form for large biomolecules and multi-protein complexes (20-24). The gross representation of large assemblies has proven to predict dynamics of the rigid and flexible parts of the proteins properly. Anisotropic network mode (ANM) and Gaussian network model (GNM) are most powerful and prevalent examples of ENM-based methods (25, 26). While GNM is applied to produce the amplitudes of isotropic thermal motions, ANM describes both amplitudes and directionality of anisotropic motions. Increased amount of data for proteins of different forms (free, liganded, or complexed), elucidates the correlation between protein function observed in experiments and the global motions predicted by ANM/GNM analyses. In recent years, numerous studies have used ENM-based models to explore structural dynamics; namely, identifying and visualizing collective motions (27),

predict modes of motion underlying function (28-31), explain details of conformational changes of various types and amplitudes (32-34) and as a means to improve steered molecular dynamics (MD) simulations of larger time and length scales and explore non-equilibrium dynamics. ENM may be applied to structural refinement of medium to low-resolution structures of electron microscopy density maps of large macromolecular complexes or molecular envelopes derived from small-angle x-ray scattering data (35, 36). Within the concept of ENM, many groups have also developed different approaches to generate feasible pathways for conformational transitions between two end conformers (16, 37, 38). Therefore, expensive MD simulations and the need to all-atom empirical force fields to set up intermediate conformations are removed. ENM is also applied to determine the main evolutionary transformations of structural changes among the homologous proteins (39, 40). In such an approach, for a given set of proteins, evolutionary direction is compared to be a combination of small subspace projected by a few low frequency modes imposed by inter-residue contact topology.

While ENM is easy to implement and provides insights into wide variety of biomolecular functions, it has been subject of various refinements and extensions. For example, the ENM potential has been extended to include three-and four-atom interactions in addition to the conventional two-atom stretching, thus providing a reliable ENM modeling of any three-dimensional graph irrespective of the atom connectivity and eliminating the need to consider large cut-off distances (41). ENM accuracy may decline when modeling large conformational changes (42). Hence, these methods could be found useful for low-resolution structures of proteins where ENM fails to predict the large-scale motions. Application of ENM has further broadened by introducing residue-based force constants as an alternative to single force interactions between neighboring and non-bonded residues (43). Application of heterogeneous spring constants eliminates the need to use any cutoff distance constraint on CG interactions, especially when coarse-graining resolution is lower than one C_α representation per location (44). Therefore, parameterization of the ENM model in such an approach accounts for the ruggedness of the potential surface and severe solvent damping present in biomolecular conformational transitions (44, 45)

Our group has extended ENM to analyze allosterically significant residues and function-related motions of proteins via a technique termed Perturbation-Response Scanning (PRS) (46, 47). The methodology sets the system slightly out of equilibrium and monitors the

evolution of the response using linear response theory (48). In vivo, the perturbation may arrive in the form of changing environmental scalars such as pH, or it may act directly on the chain as in pulling (49, 50) or other single-molecule experiments, as well as through mutations or ligand binding. PRS serves as a tool to gain insight into molecular origins of mechanical feedbacks of bimolecular structures through recording response to each inserted force on each residue of a protein (46, 47, 51, 52). It is further capable of recognizing how directionality of the inserted force may coordinate the response of the protein in a functional motion. It requires two distinct conformations of a protein, determined, e.g. by x-ray crystallography, as input; and relies on linear response theory to relate virtually exerted external forces acting on a protein to the perturbed positions of the residues (48). In PRS, one scans a protein structure residue by residue through applying forces in a diverse number of directions and records the subset of responses that most resemble the conformational change towards another known conformation of the protein. Thus, one can map the regions on the protein surface whose perturbation leads to the expected conformational change. Besides mapping active site residues that are prime regions for invoking conformational transitions, this approach also has the potential of pointing out allosteric locations and drug target regions. For example, previously, our group has studied the proteins calmodulin (52) and ferric binding protein (46) via PRS. By mutating those residues that were implicated in allosteric communication it was later verified through classical molecular dynamics (MD) simulations that they affect the conformation distributions (53, 54).

In the first part of the thesis, we study 26 set of protein-protein structures using PRS in conjunction with the ENM harmonic potential. It is of utmost interest to relate the physical effects of the internal dynamics of protein complexes, rather than knowledge of the interacting subunits, on the motions involved in their dissociation. We show that PRS maps residues that may alone dictate the structural change between the bound and unbound forms during association/dissociation processes of the protein complexes under deformations in numerous directions. The variability of the number and location of the residues that reproduce the displacement profile yields clues to how that particular protein functions, as shown by case studies.

We next move to apply the PRS method to an enzyme in complex with its inhibitor for the first time and show how it can be utilized to map the sites leading to the inhibited conformations. For this purpose, we study serine protease family of enzymes where mutants

of distinct sites might help overcome energy barriers in the catalytic mechanism. Understanding the catalytic power at the molecular level requires knowledge on the dynamic nature and internal motions of these enzymes. Ligand binding to proteins causes significant conformational changes and enzymes are believed to regulate their catalytic functions through such motions (55). We choose the complex of the two-headed tomato inhibitor-II (TI-II) and subtilisin as our model system. TI-II is a 123 residue potent inhibitor of subtilisin ($K_i \sim 9$ nM) that possesses a remarkable dual specificity toward chymotrypsin ($K_i \sim 30$ nM) and trypsin ($K_i \sim 80$ nM) (56). Subtilisin, on the other hand, is a 274 residue monomeric protein with two α/β domains which are composed of seven strands of parallel β -sheets surrounded by five α -helices (57). Apparently, single residue mutations may induce large effects on the functioning of this enzyme. The question we address is how we may manipulate domain motions to follow a conformational transition between such different functional states of subtilisin. To effectively represent the mechanistic coupling between different distributions of protein structure at the atomic level, we apply PRS (46). We make a systematic study on the shift of subtilisin between two of its native states. We analyze the transition of inhibitor bound subtilisin to unbound form of subtilisin, which we call apo subtilisin hereafter, by applying PRS to the available crystal structures of the initial and target conformers. We then map those residues that are capable of inducing the conformational change and discuss their biological significance in relation to experimental observations. We also demonstrate the efficiency of the PRS method by utilizing MD simulations to obtain the kernel between the perturbation and the response (the Hessian matrix, $\mathbf{H}$) and then showing that a simple network-based approach to constructing $\mathbf{H}$ leads to essentially the same results (12, 13).

1.3. Modeling the equilibrium and dynamical properties of iron transport proteins

After analyzing the general properties of protein pairs sharing large interfaces using coarse-grained methods, we move on to explore the detailed dynamics of a protein that binds the smallest possible ligand, a multivalent cation. We utilize a spectrum of methods to study various aspects of this problem, including PRS and ENMs described in the previous section, as well as all atom molecular dynamics simulations.

All mammals require iron in fundamental metabolic processes to fulfill the key roles of enzymatic electron transfer, proliferation of iron-sulfur clusters and oxygen delivery (58). Cellular iron deficiency prevents cell growth and leads to cell death. However, excess iron

engages in Fenton reaction and facilitates formation of destructive hydroxyl radicals that attack lipids, proteins and nucleic acids. Hence, iron is a biologically vital, but potentially dangerous presence in the human body. To control iron levels, sophisticated mechanisms have evolved to acquire, store and deliver iron (59-62). Specific proteins, transferrin and ferritin, have evolved to transport and store iron, respectively, while others purpose iron for additional functions. Transferrin (Tf) family of proteins found in multicellular organisms from insects to humans carry out iron circulation in the body. The class includes serum transferrin, ovotransferrin and lactoferrin. This is a group of nonheme iron-binding glycoproteins with a single peptide chain of nearly 700 amino acids (~80 kDa). Serum Tf plays the role of Fe^{+3} transport from the sites of intake to the circulatory system, cells and tissue (63).

X-ray crystal structures of transferrin family members reveal that human serum transferrin (hTf) exhibits three different conformations (Figure 1.1 a, b, c), the iron-free fully open apo conformation (64), the fully closed iron loaded holo conformation (65) and iron loaded partially open conformation exhibiting a structure intermediate to that of the apo and holo hTf (66). The conformational difference between apo and holo forms involves the opening at the two hinges between the N_1/N_2 - and C_1/C_2 -subdomains, and iron release making use of their twist and bending (64, 67). The C-lobe in partially open form has the same fold as holo. However, the N-lobe adopts a unique structure with a partially open cleft (66). The overall structure of transferrins is divided into two lobes designated as the N-lobe and the C-lobe. In hTf, the N-lobe consists of residues 1-331 and the C-lobe contains residues 339-679. Intra/inter-lobe interactions and an unstructured peptide linker (332-338) stabilize the tertiary structure of the protein where the linker provides hTf an overall flexibility (66). The overall topology in the lobes of transferrins consists of two subdomains that close in on iron with a hinge bending motion. The subdomain are termed N_1 -(1-95 and 247-331), N_2 -(96-246), C_1 -(339-425 and 573-679) and C_2 (426-572). The central hinge region of each lobe is a mixed β -sheet, surrounded by α -helices. Iron coordination in transferrins have similar octahedral arrangement (Figure 1.1c,d). These consist of two tyrosines, a histidine and an aspartic acid residue residing on the protein. The coordination is completed by a synergistic anion, a bidentate carbonate. The iron binding moiety is well-preserved in the known structures of transferrins, implying that the octahedral geometry provides the most advantageous iron binding mode, leading to very tight binding to Fe^{+3} (binding constants on the order of 10^{20} M^{-1} for transferrin (68), but simultaneously allowing for dissociation under suitable conditions).

The conservation of this local structure is quantified by the CONSURF database that provides evolutionary conservation profiles of proteins with known PDB structures on a scale of 1 to 9, residues with a score of 9 having the highest conservation across species.(69) In CONSURF, unique homologs collected from the SWISS-PROT database (70) are subjected to multiple sequence alignment with 219 organisms for transferrin. The algorithm assigning the final scores takes into account the phylogenetic relations between the aligned proteins and the stochastic nature of the evolutionary process. The iron coordinating residues of hTf are highly conserved, more so on the N-lobe than on the C-lobe, as follows (residue index followed by CONSURF score in parentheses): D63 (9), Y95 (9), Y188 (8), H249 (8), D392 (7), Y426 (8), Y517 (6), and H585 (8).

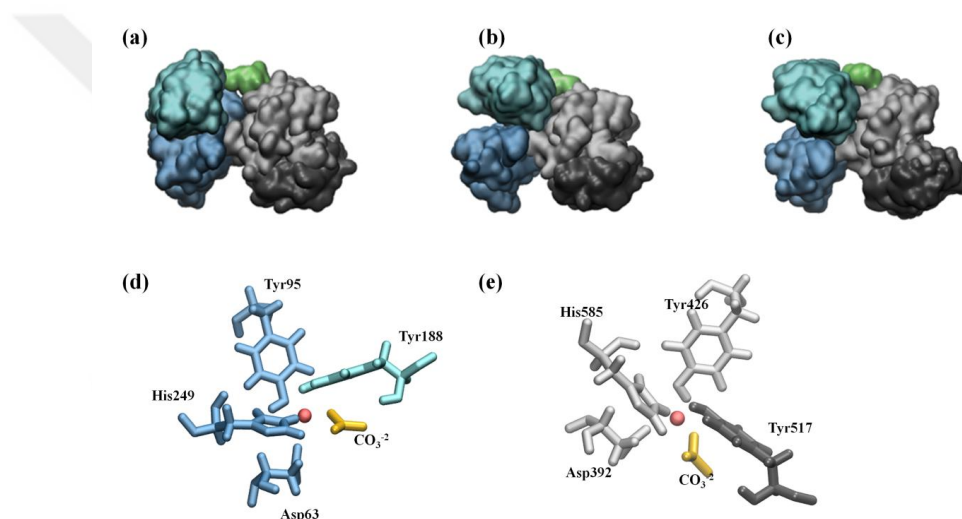


Figure 1.1 (a) Crystal structure of holo hTf; N-lobe is displayed in blue surface and the C-lobe is displayed in gray surface. The flexible linker bridging the two lobes is in green. In (b) and (c), the crystal structure of partially open and apo forms of hTf are shown. Color coding is the same as in (a). (d) The octahedral coordination of iron in the N-lobe is shown in vdW representation. The iron coordinating residues from the N₁ subdomain are in ice blue while the residue from the N₂ subdomain is in cyan. (e) The iron coordinating residues from the C1 subdomain are in silver and the sole iron coordinating residue from the C₂ subdomain is in gray. In both lobes, iron is displayed in red CPK and carbonate is in yellow.

In most cell types, iron intake is initialized through a process known as the transferrin cycle. The ligand-receptor interaction between hTf/transferrin receptor (TfR) complex induces clathrin-mediated endocytosis (71). The lower pH of the endosome leads to a conformational shift in both hTf and the TfR and promotes iron release. Subsequently, released iron is reduced to Fe⁺² by members of the STEAP family of metalloreductases.

Meanwhile, apo-hTf and TfR are recycled to the cell surface for reuse (59, 60). The tertiary structure of hTf is crucial for TfR recognition. The higher affinity of TfR towards holo-hTf in the extracellular milieu (pH $\sim$ 7.4) leads to localization of hTf into the endosome. In the endosome, TfR remains bound to apo-hTf, resulting in the trafficking of apo-hTf from the endosome to extracellular membrane where hTf is finally released (see figure 1 in ref. (71)).

An important question concerning hTf is the mechanism of iron binding with nanomolar affinity ($K_d \sim 10^{-22}$ M) and yet its prompt release at the reduced pH of endosome. Specifically, the conformational changes prior to and during the iron release process are not fully understood despite considerable efforts devoted to study the molecular structure and kinetics of hTf. Furthermore, understanding the molecular basis of the N- and C-lobe characteristics and major differences between their dynamics may help develop a model for the iron release mechanism. In the part section of the thesis, we opt to make use of a range of computational methods to gain molecular level insights into the various parameters affecting the distribution of conformational substates of hTf in its different conformations.

First, we analyze conformational dynamics of the full hTf structure to understand how hTf functions in atomistic detail. We communicate results of a series of extensive MD simulations (of minimum duration of 60 ns) performed on the three resolved conformations (holo, apo and partially open) of the protein; we also manipulate the pH sensitive residues. To examine the routes that may be accounted for the control of coordination between the lobes, we have studied monoferric full length hTf, as well as the separate N- and C-domains. Analysis of the resulting trajectories offers details on the conformational shift and iron binding mechanism. We analyze the structural dynamics through identifiers based on reduced degrees of freedom defined specifically for hTf, defining both the opening modes of motion within the separate lobes and the relative positioning of the lobes. Key events leading to, or preventing, conformational change are discussed. We monitor the path sampled between different conformational states identified by MD simulations and we evaluate the effect of changing the charge states of key residues. The molecular mechanisms that lead to the observed effects, their relationship to the experimental data, and the consequences of the observations that enhance our understanding of the dynamics and function of hTf are presented.

Next, we aim to effectively represent the atomistic details of coupling between the two lobes of holo hTf at the physiological pH and various salt concentrations in the absence of TfR. Within the scheme of molecular dynamics simulations, we study the factors contributing to the ionic strength (IS) sensitivity of the protein. The effect of non-synergistic anions on the properties of hTf by EPR experiments has been pointed out before (72-76). These non-synergistic anions facilitate iron removal from hTf by the aid of various chelators and differentially affect the thermodynamic and kinetic behavior of the metal-binding sites in the absence of TfR (76, 77). While, synergistic anion binding is an absolute requirement of ferric ion binding in hTf, non-synergistic anions interact with specific sites on the protein surface and support iron release. The simulations reported here provide the distributions of individual ionic species on the protein surface and give information about specific ion adsorption to different amino acid residues of hTf. We report the allosterically induced conformational changes upon non-synergistic anion binding by introducing KInetically Significant Anion Binding or KISAB sites in hTf protein. Furthermore, we make use of Steered MD (SMD) simulations to explore the energetics of ferric ion binding to hTf in various environmental conditions. For this purposes, we have analyzed the potential of mean force necessary to dissociate the ligands from the protein and report their relevance to molecular motions of hTf.

We then apply computationally simple and efficient methods sampling local states of hTf to extract residues that might exhibit unique contributions to iron metabolism. We combine perturbation response scanning (PRS) technique with MD and ANM to provide a set of powerful tools that unravel novel local and remote modulators, which are effective on the ion binding site conformations. The bilobal structure of hTf features an excellent model to study how the two lobes accomplish orchestrated motions necessary for different ligand binding/release dynamics observed in various environments. In the PRS approach, relatively short lengths of equilibrated substates sampled by MD simulations are employed to decipher the local dynamical contributions leading to global rearrangements of the protein structure towards pre-selected conformations (12). PRS method is influential in that, it selects residues whose perturbation is expected to shift the conformational surface towards the alternate state while eliminating heuristic biases for focusing on regions that are directly in the active site of a protein. Via PRS, we have been able to determine sites on the apo form whose perturbation is expected to shift the conformation distribution towards holo-like structures. We have then been able to test this idea with conventional MD simulations where these locations are

perturbed via synergistic anion docking, which is the natural ligand expected to enter the iron binding cavity and prepare it for iron uptake. The simulations show that such a local perturbation leads to global motions on both lobes. PRS is also able to explain the lack of iron release propensity at serum pH and its ease of release at endosomal pH.

1.4. Modeling the dynamics and stability of dihydrofolate and trimethoprim binding to *E.Coli* dihydrofolate reductase

After the study of remote and local conformational manipulations in proteins bearing large contact interfaces in part I, and the effect of the environment on a protein with a very small contact surface with the ligand in part II of the thesis, in part III we focus on how the dynamics of an enzyme contacting moderate sized ligands may be manipulated via mutations that preserve function while conferring resistance towards a potent drug.

Multiple conformers of the enzyme scaffolds detected upon complex formation in different stages of a kinetic reaction are triggered by flexibility of the active site, local motions of the loops or allosteric response initiated by a distal perturbation (78, 79). Correlated fluctuations over a wide range of time/displacement scales that are involved in the dynamic shifts between feasible states link the collective motions to the effective chemical reactions of the enzymes in the wild type form and mutated strains (80, 81). In our view, conformational plasticity and allosteric regulations are essential contributors to enzyme stability, precise specificity and catalytic activity (82-84). A comparative study of enzyme dynamics in the presence of its substrate and inhibitor should reveal the changes in the dynamics associated with inhibition (3).

In the third part of this thesis, we study *E. Coli* dihydrofolate reductase (DHFR) in the presence of its functional ligand, and a competitor inhibitor for both the wild type (WT) and L28R mutant that emerges in forms resistant to the drug trimethoprim, and are by far the most beneficial mutations in the WT background (85). DHFR catalyzes the stereospecific reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) exploiting nicotinamide adenine dinucleotide phosphate (NADPH) molecules as the cofactor. The reduction involves a proton transfer from water, followed by hydride transfer where NADPH protonates DHF. The reaction is essential for maintaining cellular reserves of THF and its derivatives, vital precursors for the biosynthesis of DNA bases and amino acids (86). Steady-state turnover of the enzyme under the cellular concentrations of substrate and the cofactor is catalyzed through

five kinetic intermediates, two of which are the Michaelis complex (DHFR/DHF/NADPH) and the complex with the product (DHFR/THF/NADP). Following the hydride transfer, the release of the oxidized cofactor (NADP) and rebinding of NADPH assists the release of the product (THF), which is the rate-limiting step (87). In this manner, the enzyme in the apo state is not available under physiological conditions, remaining prepared for the next cycle of catalysis.

Numerous structures have been reported for complexes of the enzyme from various organisms, in the presence of a wide range of ligands, including the substrate, product, and cofactor, as well as inhibitors (88-90). *E. coli* DHFR is a small (~18 kD) protein formed from a mixture of eight stranded β -sheets and four contiguous α -helices (see Figure 1.2) (88). The protein is divided by the active site cleft into two structural subdomains: the adenosine binding subdomain and the major subdomain. The adenosine binding subdomain (residues 38–88) provides the binding site for the adenosine moiety of the cofactor (NADPH). The major subdomain consists of ~100 residues from the N- and C-termini and is dominated by a set of three loops on the ligand binding face that surround the active site. These loops construct approximately half of the major subdomain, hence they are named as the “loop” subdomain. These loops are termed as M20 (residues 9–24), FG (residues 116–132), and GH (residues 142–150) loops. In addition, the CD loop (residues 64–71) on the adenosine binding subdomain is also deemed important for the DHFR function. The M20 loop is located directly over the active site, protecting it from the solvent, and is primarily responsible for regulating the active site (88). Three well-defined conformations of the M20 loop in the crystal are named as the open, occluded, and closed states (88, 91). In our work, we have used the structure with the protein data bank (PDB) code 1rx2 that has the closed M20 loop conformation (88).

Because of its central role, DHFR is used as the target for several important anticancer (e.g. methotrexate), antimalarial (e.g. Pyrimethamine), and antimicrobial (e.g. trimethoprim) drugs. Trimethoprim is a well-studied competitive inhibitor of bacterial DHFRs and is extensively used across the world because of its relatively low cost. Therefore, resistance to trimethoprim in pathogenic bacteria is very common and renders this important drug ineffective against several infections. Trimethoprim resistance generally emerges as a result of amplification or overexpression of the gene encoding DHFR and amino acid replacements in the DHFR enzyme (92-94).

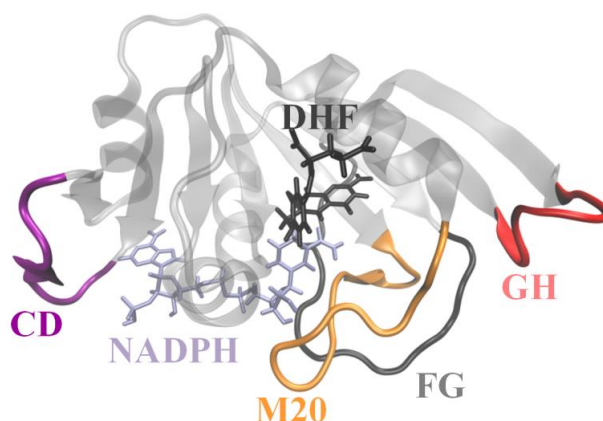


Figure 1.2 Cartoon representation of *E. coli* DHFR based on PDB structure 1rx2 is displayed in transparent gray. The characteristic loops of the enzyme are shaded in orange, purple, gray and salmon indicating M20, CD, FG and GH loops, respectively. DHFR is complexed with DHF and NADPH in its native form.

Much has been discussed on the working principles of *E. Coli* DHFR enzyme in terms of kinetics (95), dynamics (81, 96, 97) and simulations (98, 99). Analysis of several complexes of DHFR with various ligands provides insight to the coupling between the short-lived dynamics and functional motions of the enzyme complexes in various states of the catalytic cycle. For example, the time scale of loop motions in complexes of *E. coli* DHFR at different stages of the kinetic cycle have been measured by nuclear magnetic resonance (NMR) experiments (97, 100, 101). Further NMR studies elucidate how local loop fluctuations are coupled to kinetics and thermodynamics of DHFR (97, 102). Particularly, the amide backbone dynamics alterations represent different physiologically relevant time scales for the enzyme kinetics (100, 103). Meanwhile, molecular dynamics (MD) simulations explain the molecular level effects of different ligands on the network of coupled motions between spatially separated regions of DHFR (98, 104). 10 ns long MD simulations suggest that the secondary structure arrangements of the enzyme in complex with different ligands are intact, while flexible loops move in large conformational changes (99, 105). Other experiments involving kinetic measurements show that multiple mutations in and distal to the active site significantly affect the conformational space sampled by the enzyme (106). Long-range perturbations of the enzyme's structure alter the kinetic parameters followed by reduced

or enhanced activities. Non-additive effects of two different mutations are an additional example of long range coupling between distant regions of DHFR (99).

While there is comprehensive experimental and theoretical data available for DHFR structure and function (98, 104, 107, 108), we do not yet have a dynamical/mechanical picture of how the enzyme may alter its conformation as it binds to an inhibitor such as trimethoprim despite its clinical importance. Previous MD simulations of *E.coli* DHFR report unbiased sampling of the unfolding pathway at temperatures up to 600 K (109). Therein, DHFR in complex with NADP and folate is studied via MD simulations of one nanosecond length at increasing temperatures. MD simulation on the order of 10 ns of the wild type (WT) *E. coli* DHFR and different mutants also investigate the sampled conformational space and find the effect of mutations on the correlated motions, hydrogen bond distributions and hydride transfer reaction under the non-additive effects of mutations (99, 105). Combined with empirical valance bond calculations, MD simulations on the mutated DHFR structures show that the overall conformation of the enzyme is maintained while altering the hydride transfer reaction free energy barriers (108).

Recognizing the significance of dynamical properties and flexibility in enzymatic reactions, here, we study WT and L28R mutant DHFR bound to DHF or trimethoprim through extensive MD simulations. L28R survives the natural selection in the systematic experimental approach carried out in the morbidostat under the selection pressure induced by trimethoprim (93). We explain the coupling between enzyme fluctuation and catalytic function by reporting the details of the molecular motions of DHFR binding to its substrate (DHF) and its inhibitor (trimethoprim) in two protonation states with sampling of equilibrium conformations on the time scales of 210 ns. Examination of the six structures together provides a dynamical picture for conformational changes in the presence of various ligands, and the catalytic relevance of loop motions in WT and mutated forms. To explore the energetics of DHF and trimethoprim binding to DHFR, we analyze the free energy changes in response to ligand dissociation from the enzyme. The results are supported by trimethoprim binding constant measurements for WT and the mutant. We present a structural view of how trimethoprim binding inhibits DHFR function, which may provide guidelines for future drug design efforts. We also report CharmM22 parameters for neutral and protonated trimethoprim for the first time in the literature.

CHAPTER 2. METHODOLOGY

2.1. Molecular Dynamics (MD) Simulations: Theory

In molecular dynamics, the integration of classical Newton's equation is solved to generate successive trajectories of the system in time (110). Positions and velocities of particles in time, can be obtained from the Newton's second law:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{f}_i}{m_i} \quad (2.1)$$

where $\mathbf{f}_i, m_i$, and $\mathbf{r}_i$ are force exerted on particle i , mass and position of particle i , respectively. The force is the gradient of potential on particle i , which is defined as the sum over all the effective interactions of all other particles on i :

$$\mathbf{f}_i = -\frac{\partial U(r_1 \dots r_N)}{\partial r_i} \quad (2.2)$$

The solution of equation 2.1 reproduces a trajectory of atomic coordinates and velocities. Since the molecular potentials have complex forms, there is no analytical solution of the equations of motions. Thus, a need to numerical methods and algorithms arises to obtain the trajectories of particles. The potential energy, U , can be separated into non-bonded and bonded interactions:

$$U(r_1 \dots r_N) = U_{\text{bonded}}(r_1 \dots r_N) + U_{\text{non-bonded}}(r_1 \dots r_N) \quad (2.3)$$

Non-bonded potential is composed of 1-body, 2-body and higher body terms,

$$U_{\text{non-bonded}}(r_1 \dots r_N) = \sum_i u(r_i) + \sum_i \sum_{j>i} v(r_i, r_j) + .. \quad (2.4)$$

For simplicity, higher order terms are neglected and the pair potential is used. The $U(r)$ term represents the applied external potential, and the two body interaction potential, $v(r_i, r_j)$ equals to $v(r_{ij})$. There are numerous experimental and theoretical models on how to define these potentials, Lennard-Jones potential being the most commonly used:

$$v_{LJ}(r) = \frac{a}{r^6} - \frac{b}{r^{12}} \quad (2.5)$$

In the presence of charges, a Coulombic interaction is also added.

For simplest intramolecular interaction potential, vibrational harmonic form can be used by including summation over all bonds and bond bending angles in addition to a periodic function of torsional angles:

$$u_{\text{intra}} = \frac{1}{2} \sum_{\text{bonds}} k_{ij}^r (r_{ij} - r_{eq})^2 + \frac{1}{2} \sum_{\substack{\text{bond} \\ \text{angles}}} k_{ijk}^\theta (\theta_{ijk} - \theta_{eq})^2 + \frac{1}{2} \sum_{\substack{\text{torsional} \\ \text{angles}}} \sum_m k_{ijkl}^{\phi, m} (1 + \cos(m\phi_{ijkl} - \gamma)) \quad (2.6)$$

where k_r , k_θ , and k_ϕ are constants that depend on the identity of the atoms participating in the interaction, r_{eq} and θ_{eq} are the average bond length and angle, m and γ depend on the rotameric states of the torsional angle.

A reliable simulation force-field package has the specification of the strength parameters and constants or other additional terms that is obtained by matching experimental and/or quantum mechanical data, and is tested for a wide variety of systems.

2.1.1. MD simulation protocol used for hTf systems

We use NAMD package to model the dynamics of the protein-water systems (111). Solvation is achieved via the VMD 1.9.1 program with solvate plug-in version 1.2 (112). The protein is soaked in a cubic solvent box such that there is at least a 10 Å layer of solvent in each direction from any atom of the protein to the edge of the box. Charge neutrality is obtained by adding an adequate number of Na⁺ and Cl⁻ ions. The CharmM22 all-atom force field is used to parameterize the topology of the protein atoms (113). The force field parameters for synergistic carbonate ion, DHF and Fe⁺³ reported in the literature are adopted (46, 114, 115). We note that the DHF parameters can be used for the molecular state whereby the proton transfer from water has taken place, but hydride transfer has not occurred. TIP3P model is used to describe water molecules (116). During the simulations, periodic boundary conditions were used. Long range electrostatic interactions are calculated by the particle mesh Ewald method (117), with a cutoff distance of 12 Å and a switching function at 10 Å. RATTLE algorithm (118) is applied and a time step size of 2 fs in the Verlet algorithm is used. Temperature control is carried out by Langevin dynamics with a dampening coefficient

of 5/ps. Pressure control is attained by a Langevin piston. All systems were first subjected to energy minimization with the conjugate gradients algorithm. The resulting structures are then run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations stabilize and the desired average pressure is maintained.

For MD simulations of Fe^{+3} -loaded hTf, we select closed or a partially open full structure forms that exist in the protein data bank (PDB). For the fully loaded holo form, we select the relatively recent structure with PDB code 3V83 (65), because it has minimal number of missing atoms in chain B. We shall call this structure the holo form (H) in the rest of the manuscript. The open form has the 2HAV code (64) and the first three residues are not present in the crystal structure; we shall refer to this structure as the apo form (A) throughout. The partially open (PO) form is represented by the 3QYT coded (66) structure, and has 679 reported residues; while its integrity might be questionable due to the presence of sulphate in the structure, we nevertheless consider this as representing a possible (meta)stable structure of fully loaded hTf. These structures have been determined at 2.1, 2.8 and 2.7 Å resolution, and were crystallized at pH of 7.5, 7.4 and 7.0, respectively. We report the RMSD comparison for the overall structure as well as the N- and C-lobes of the three x-ray structures used in this study in Table 2.1. In particular, we find that the C-lobe of the PO state is similar to that of H (0.5 Å RMSD), while its N-lobe is similar to that of A (1.7 Å RMSD).

We mimic the different pH environments by assigning the most probable states of ionizable groups according to their pK_a values which were calculated by PROPKA and H++ servers (119, 120). PROPKA relies on the accurate calculation of the shifts in free energies and it is based on an improved description of the desolvation and the dielectric response of the protein. The H++ server uses a different approach in which the titration curves are directly represented as a weighted sum of Henderson-Hasselbalch curves of decoupled quasi-sites. We report values for the server settings of 150 mM salinity, external dielectric of 80 and internal dielectric of 10. We list the values in Table 2.2 only for those residues where the calculated pK_a is different from the standard value by more than 2 units. We also report calculated and measured pK_a values from the literature, computational for fully loaded holo hTf (67), and experimental for apo hTf (121-123).

These corroborate PROPKA and H++ server results for all cases except for two Tyr residues, whose role we will further discuss in this manuscript.

Table 2.1 RMSD between x-ray structure pairs*

	holo (3v83)	apo (2hav)	PO (3qyt)
holo (3V83)	<u>3.2</u>	6.4 , 6.6	5.2 , <i>0.5</i>
apo (2HAV)	8.2	<u>3.4</u>	1.7 , 6.6
PO (3QYT)	4.9	6.3	<u>6.1</u>

* Upper diagonal, RMSD between N lobes (**Bold**) and C lobes (*Italic*); Diagonal (underlined), RMSD between N lobe and C lobe within each structure; lower diagonal, overall RMSD between every pair.

Of the acidic Asp and Glu residues, those with upshifted pK_a values into the physiological range occur only in at most one of the two servers. Since the probability of an acidic residue being charged at 7.4 pH is greater than 0.9 for $pK_a < 6.6$, we assume that their most probable state carries negative charge at this pH as well as at endosomal pH. A Similar argument applies to the basic Arg residues and all but one of the Lys; probability of being protonated at this pH is greater than 0.9 for $pK_a > 8.2$, so these are kept positively charged in all simulations. Finally, we observe downshifted values for K206 residues on both servers. We keep it positively charged since the calculations in ref. (67) also indicates it would be charged at the pH values we are interested in here. Note that, since the two Lys residues involved in the proposed dilysine trigger (124) (K206/296) are protonated in all our MD simulations, we will directly test if protonating them leads to the direct conformational switching of hTf. All His residues have either the standard values, or are downshifted, having neutral side chains at pH 7.4. Thus, in the MD simulations of the systems at the higher pH, the standard protonation state is adopted for all His. To mimic the most probable states at endosomal pH, His residues with downshifted pK_a s (those listed in Table 2.2) are kept neutral while the rest of them are protonated. Finally, Tyr residues directly coordinating the iron display downshifted values (Table 2.2). In particular, Y188 and Y517 exhibit a downshift in their values in all pK_a predictions. We will change the protonation states of these residues in various combinations to test if a conformational switch may be triggered in the holo structure by the uptake/release of protons at these sites.

16 sets of simulations were performed with various initial starting conditions and perturbation scenarios (Table 2.3). The different simulation conditions are labeled with the initial structure, H, A, or PO, for the full length holo, apo and PO structures, respectively; PDB codes used for each is listed in Table 2.1.

Table 2.2 pK_a values¹ for holo hTf residues calculated by H++ and PROPKA servers

Residue	PROPKA ^{2,3}	H++	Literature ⁴	Residue	PROPKA ^{2,3}	H++	Literature ⁴	Residue	PROPKA ^{2,3}	H++	Literature ⁴
D58	1.1	<0	0.0 (67)	K18	8.1	>12	>12 (67)	H119	3.9	1.5	5.0 (67)
D63	<0	<0	0.0 (67)	K206	7.9*	4.2	7.4 (67)	H207	2.4	2.5	0.0 (67)
D297	6.4*	2.1	3.8±2(67)	K296	<0*	>12	>12 (67)	H249	<0	<0	0.0 (67)
D392	<0	<0	-	K354	7.8*	>12	-	H300	4.8	5.6	2.6 (67), <6 (123)
D633	6.8*	<0	-	R124	8.9	>12	>12 (67)	H350	4.9	2.3	<6 (123)
D634	<0	<0	-	Y95	>12*	>12	1.2 (67) 0.0 (67), 7.47 (121), 6.91(122)	H451	3.7	1.2	-
E15	6.3	<0	0.0 (67)	Y188	6.0*	0.6	-	H473	5.6	3.4	<6 (123)
E212	7.6*	1.0	1.4 (67)	Y426	10.6*	>12	-	H535	3.7	2.6	-
E351	6.5	<0	-	Y517	4.1*	<0	-	H585	<0	<0	-
E375	5.5	0.2	-					H598	5.2	3.7	<6 (123)
E512	4.4	<0	-								
E573	4.8	2.1	-								

¹ Of the total number of 214 residues with ionizable groups, those with $\Delta pK_a > 2$ for D, E, K, R, Y and those with $pK_a \leq 5.6$ for H are reported. Values in the pH window of 0 – 12 are shown.

² Standard pK_a values of residues in PROPKA server are D: 3.8, E: 4.5, H: 6.5, Y: 10.0, K: 10.5, R: 12.5.

³ Residues marked by * are coupled to others, and hence may display alternate pK_a values.

⁴ Other results from the literature for hTf. Reference (67) reports residues from the N-lobe; we select those for iron loaded, closed form. Others are experimental results for apo hTf, since the presence of iron precludes use of many conventional methods used for specific pK_a measurements in the holo form (121-123).

The number of atoms in each system and equilibrated water box sizes are reported in Table 2.3. We have focused on six major issues: (i) The role of the protonation states of the pair of Tyr residues in contact with the Fe^{+3} ions in holo hTf (H, H* and H** systems. In H, all tyrosines in contact with Fe^{+3} are selected to have their standard protonation states; *i.e.* they are protonated on both lobes. In H*, Y188 and Y517 are deprotonated, and in H** we deprotonate all four tyrosine residues); (ii) the similarity of the conformation triggered by neutralizing the aforementioned Tyr residues in the H state with the A and PO states (H, A and PO systems); (iii) the effect of having a single bound Fe^{+3} ion on the H state (H, H^{Fe+N} and H^{Fe+C} systems); (iv) the role of the communication between the N- and C-lobes on the overall dynamics (H, H_N and H_C systems); and (v) the effect of lowering the pH (H and H[†] systems). (vi) the effect of decreasing and increasing ionic strength on protein motions and iron binding affinities.

Each condition has at least 60 ns of sampling time, and at least 200 ns for those displaying non-equilibrated behavior by the end of 60 ns. For the H system, we have performed an additional, independent run of 150 ns to check the reproducibility of the results.

We have also prolonged the original H simulation to 500 ns in case significant changes are observed in the relative local and global positioning of the two lobes, quantified by the RMSD of different regions (for details, see Results and Discussion).

Table 2.3 Summary of system conditions for different simulations

System label	Perturbation scenario	Number of atoms (Na ⁺ , Cl ⁻)	Equilibrated box size (Å ³)	Simulation length (ns)
H**	All tyrosines in contact with Fe ⁺³			
	deprotonated on both lobes (locked irons), at 150 mM salt concentration	78214 (70,63)	88×107×90	60
H*	One tyrosine contacting Fe ⁺³ from each lobe deprotonated (Y188 and Y517)	77947 (68,63)	88×107×90	60
H	All residues in standard protonation states, at 150 mM salt concentration	78214 (66,63)	88×107×90	500,150
A	All residues in standard protonation states	86426 (73,71)	97×99×96	200
PO	All residues in standard protonation states	79270 (66,65)	88×105×91	200
H ^{Fe+N}	Fe ⁺³ only in N-lobe	77940 (68,63)	88×107×90	200
H ^{Fe+C}	Fe ⁺³ only in C-lobe	77968 (68,63)	88×107×90	200
H _N	N-lobe only; Fe ⁺³ included	34781 (28,28)	85×60×73	200
H _C	C-lobe only; Fe ⁺³ included	36957 (30,30)	78×75×68	200
H [†]	Histidines protonated as described under Models and Methods to mimic endosomal pH.	77947 (63,69)	88×107×90	200
H ₀ **	locked irons at 0 mM of salt concentration	78187 (7,0)	88×107×90	50
H _L **	locked irons at 50 mM of salt concentration	78103 (28,21)	88×107×90	50
H _H **	locked irons at 250 mM of salt concentration	77763 (113,106)	88×107×90	50
H ₀	All residues in standard protonation states, at 0 mM salt concentration	78199 (3,0)	88×107×90	200
H _L	All residues in standard protonation states, at 50 mM salt concentration	78115 (24,21)	88×107×90	200
H _H	All residues in standard protonation states, at 250 mM salt concentration	77803 (109,106)	88×107×90	200

2.1.2. MD simulation protocol used for DHFR systems

MD simulation of the ternary complex of the DHF bound WT system (WT^{DHF}) is constructed based on the crystallographic structure with protein data bank (PDB) code 1rx2 (88). DHFR is complexed with DHF and NADPH in this native form (Figure 1.2). In a

separate set of MD simulations, we study the effect of trimethoprim binding in its unprotonated (TMP) and ground state (TMP⁺) on the WT DHFR conformation (WT^{TMP} and WT^{TMP+}). Since there are no crystal structures of *E. coli* DHFR with trimethoprim, we have docked trimethoprim based on the coordination of equivalent residues of the trimethoprim binding region of *Staphylococcus Aureus* DHFR (PDB code: 2w9g) that has 36.6% sequence identity (89). We have used AutoDock 4.2 for this purpose (73, 125). The center of mass of the docked molecule was selected from the coordinates of the C_α atoms of the trimethoprim contacting residues in *Staphylococcus Aureus* DHFR and in consistency with DHF binding site of *E. Coli* dihydrofolate binding complex. Accordingly, we choose residues 5, 6, 7, 27, 28, 30, 31, 50, 54, 94, 100, 111, and 113. trimethoprim was enclosed in a box with number of grid points 40×40×40 with a grid spacing of 0.375 Å. Lamarckian genetic algorithms, as implemented in AutoDock, were employed to perform docking calculations. All other parameters were set to their default values. For each of the docking cases, the lowest energy docked conformation, according to the Autodock scoring function, was selected as the binding mode.

Available force field parameters for trimethoprim (126) molecule found in the literature are not compatible with CharmM22 force fields due their use of united atom model for CH₂ and CH₃ groups and thus are not useful for the purpose of this study. We therefore parameterized TMP and TMP⁺ molecules using Force Field ToolKit (fftk) plugin implemented in VMD (127). For necessary quantum mechanical calculations, Gaussian package was utilized (128). We followed the protocol developed by Mayne *et al.* (127) to calculate the force field parameters. As the initial three-dimensional structure of the trimethoprim molecule, we adopted the coordinates for trimethoprim bound to *Staphylococcus aureus* DHFR structure. Missing parameter determination, geometrical optimization and bond, angle and dihedral potential optimizations were done by quantum mechanical calculations. The resulting parameters are listed in Appendix A.

For MD simulations of the L28R mutant in their DHF, TMP⁺ and TMP bound forms of DHFR, we mutated the WT structures in silico via BIOVIA Discovery Studio 4.0 package (129). The minimization and equilibration was performed as described for the WT systems. Each condition has at least 210 ns of sampling time. For the WT systems in the presence of DHF and trimethoprim, we have performed an additional, independent run of 210 ns to check the reproducibility of the results.

The mobility of the equilibrium structures of trimethoprim and DHF is quantified by calculating the region spanned by the principal axes of pyrimidine tail of trimethoprim and p-aminobenzoyl glutamate tail of DHF. For TMP, a vector is defined between the NA2 and C7 atoms (NA2-C7 vector) (see Figure 2.1 for atom labeling scheme), while for DHF, a vector is defined between N10 and CD atoms (N10-CD vector) (see Figure 2.2 for atom labeling scheme).

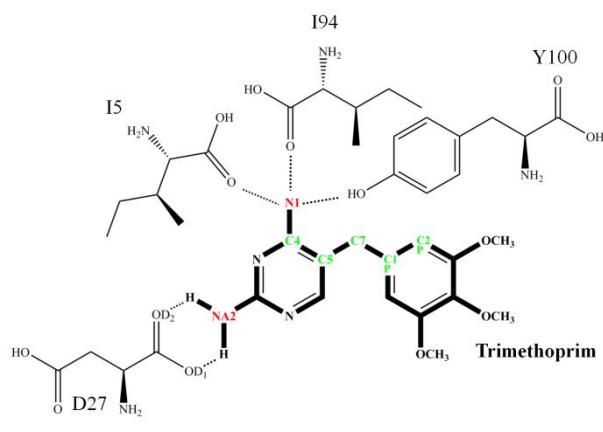


Figure 2.1 Binding site residues of TMP/TMP⁺ bound DHFR. Atom names involved in hydrogen bond formation are in red and atom names involved in torsional angles are in green.

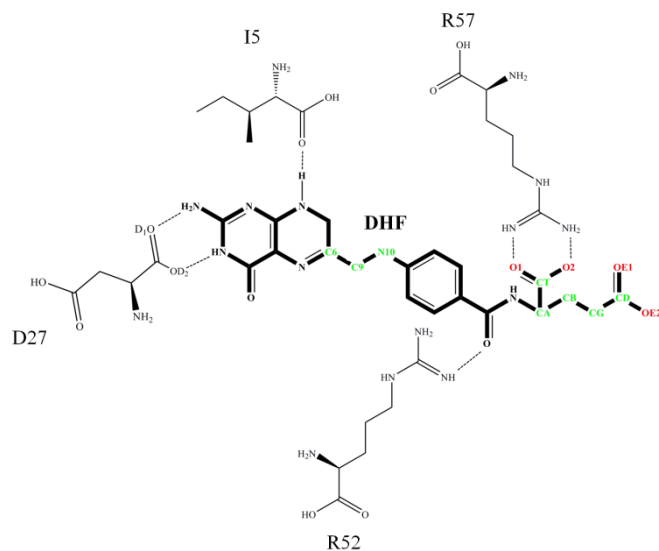


Figure 2.2 Binding site residues of DHF bound DHFR. Atom names involved in hydrogen bond formation in the original structure are in red and atom names involved in torsional angles are in green.

The $A_{3 \times N}$ matrix of the NA2-C7 and N10-CD vectors are formed where N denotes number of snapshots collected during the 210 ns of MD simulation. The three principal axes of the molecules are calculated and the projections of the molecule along these axes are considered as follows:

$$\mathbf{A}_{3 \times N} = \mathbf{U}_{3 \times 3} \mathbf{W}_{3 \times 3} \mathbf{V}_{3 \times N}^T \quad (2.7)$$

Contribution fraction of the shape along the three principal axes is evaluated from the size of the eigenvalues belonging to each principal axis, per equation 2.7. In this context, equal contributions along the three axes represent spanning a spherical region, a large contribution to one axis and insignificant contributions to the remaining two represent a rod like shape, and significant contributions to two axes, and a small contribution to the third represent a planar shape. Examples are provided in Figure S3, where the regions spanned by the NA2-C7 unit vector are plotted for the WT and the L28R mutants with the contribution fractions of the major axis being 0.42 and 0.69, respectively.

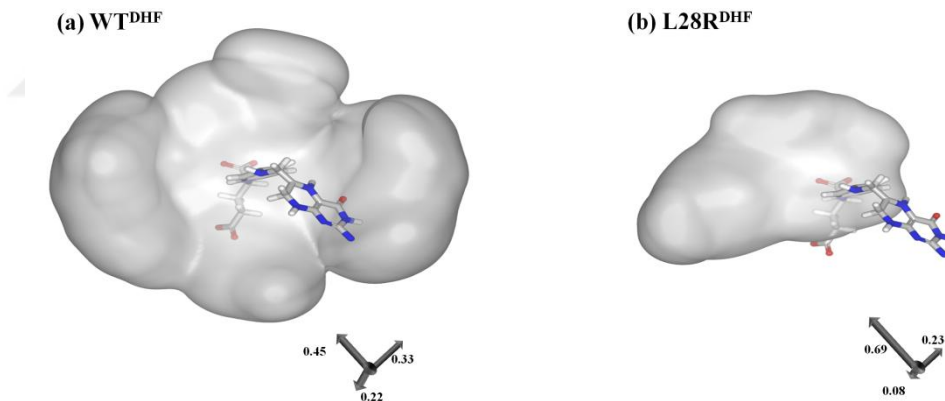


Figure 2.3 Rotational movements of DHF p-aminobenzoyl glutamate tail in (a) WT^{DHF} and (b) L28R^{DHF} systems. The trajectory of the unit vector connecting N10 to CD atom position is shifted to (0,0,0) for better visualization of the tail movements which are projected on the x-y plane on the upper panels; color variations denote z-coordinate changes. The volume swept by the CD atom positions are displayed as a gray surface in the lower panels, where one of the poses of DHF is shown in stick representation. Principal axis contribution is 0.45, 0.33 and 0.22 for WT^{DHF}, which demonstrates a nearly free rotation of the tail in the binding site. Principal axis contribution is 0.69, 0.23 and 0.08 for L28R^{DHF} revealing a stiffened tail with the motion predominantly confined to the major axis.

2.2. Steered MD (SMD) simulation: Theory

The technique can be applied in two ways, either pulling with a constant velocity or constant force. In constant velocity pulling, the atom to be pulled is virtually attached to a dummy atom. Then the dummy atom is moved with a constant velocity and the force between the atom of interest and the dummy atom is measured. Various types of pulling speeds may be induced and consequently, higher speeds lead to stronger forces. The following equation is valid for this scenario.

$$\vec{F}(t) = -\nabla \left(k \left[vt - (\mathbf{r} - \mathbf{r}_0) \cdot \vec{n} \right] \right) \quad (2.8)$$

where k is the spring constant, v is the pulling velocity, t is the time, $\mathbf{r}$ and $\mathbf{r}_0$ are the actual and initial position of the SMD atom, respectively.

As the dummy atom moves with a constant velocity, the SMD atom experiences a force that depends linearly on the distance between the dummy and the SMD atom (Figure 2.4). The key parameters for this technique are the pulling velocity and the spring constant. These values may be determined directly from the experiments (AFM experiments for the system of interest) or by running several trials.

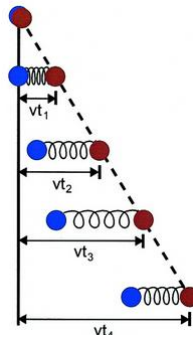


Figure 2.4 Constant velocity SMD simulation

The second way to carry out a SMD simulation is by application of constant force. In this case, the SMD atom experiences a constant force in the direction by the vector that links the fixed and the pulled atoms. In this case there is no dummy atom or virtual spring.

2.2.1. SMD simulation protocol

The constant velocity SMD simulation implemented in NAMD package (111) was carried out by fixing the C_α atoms of ferric binding region (D63, Y95, Y188, H249 and carbonate anion for the N-lobe, D392, Y426, Y517, H585 and carbonate anion for the C-lobe) in hTf systems and by applying external force to ferric ions in the N-lobe and C-lobe. hTf were fixed in ligand binding sites to prevent it from translation and rotation during the pulling experiment. Iron was pulled with a harmonic constraint of $7 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ and constant velocity of $0.25 \text{ \AA}\cdot\text{ps}^{-1}$. For DHFR systems, we fix the C_α atoms of DHF/TMP/TMP⁺ binding region in WT and L28R mutant forms and external force was applied to DHF/TMP/TMP⁺ molecules. For SMD simulations, several snapshots from the equilibrated enzyme described in the previous section is soaked in a cubic solvent box such that there is at least a 40 Å layer of solvent in each direction from any atom of the protein towards the edge of the box. The larger water box is used to dissociate ligands from the protein completely and reach a plateau in the cumulative work profile. DHF/TMP/TMP⁺ was pulled with a harmonic constraint of $20 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ and constant velocity of $0.05 \text{ \AA}\cdot\text{ps}^{-1}$. In both cases, The large value of spring constant satisfies the stiff spring approximation (130) and the pulling speed is slow enough to reduce the non-equilibrium effects. The pulling direction was defined based on a vector from the center of mass of the fixed atoms toward the center of mass of the ligand.

The pulling process was applied to the equilibrated hTf positions in H group systems at 50, 100 and 200 ns frames as the starting structure. For H^{**} group, we take the snapshot at 50 ns time point as the initial structure for the pulling experiment. In DHFR systems, the pulling experiment was applied to the equilibrated DHFR positions at 10, 50, 90, 130, 170 and 210 ns frames as the starting structures. Each SMD simulation lasted for 0.1 and 0.8 ns in the NPT ensemble with the Langevin thermostat and isotropic position scaling, at 310 K and 1 bar for hTf and DHFR systems, respectively. For each initial structure, ten independent SMD simulations were performed to a measure the work done by the external force to expel the ligands from their respective binding site. Therefore, the total number of trajectories to define the work done by the external force to dissociate the ferric ion in is $3 \times 10=30$, $1 \times 10=10$ and $6 \times 10=60$ for H, H^{**} and DHFR systems. The force exerted on the ligand is defined as equation 2.8.

Free energy difference estimator in terms of the potential of mean force (PMF), defined as the free energy profile along a reaction coordinate was initially calculated by Jarzynski's equality to relate the probability distribution of non-equilibrium work and equilibrium free energy differences obtained during the constant velocity SMD processes (131):

$$e^{-\frac{\Delta G}{k_B T}} = \left\langle e^{-\frac{W}{k_B T}} \right\rangle \quad (2.9)$$

where W is the work performed, k_B is Boltzmann's constant, T is the bulk temperature, and ΔG is the free energy difference. Angle brackets denote the ensemble average over ten independent constant velocity SMD trajectories for each starting position. The bias corrected Jarzynski free energy difference estimator (the final ΔG) was then obtained considering cumulative work of 30, 10 and 60 near-equilibrium trajectories of H, H^{**} and DHFR systems (ten independent SMD simulations from three, one and six initial structures) whose probability distribution is Gaussian. Using the dissipated work, the difference between the average work and the free energy difference estimator, large statistical errors and biases evolving due to the sensitivity of the Jarzynski's approach to rare events was calculated as described by Gore et al. (132).

The standard free energy changes were then computed via

$$\Delta G^0 = \Delta G_{bind} + k_B T \ln\left(\frac{C^{comp}}{C^0}\right) \quad (2.10)$$

where C^{comp} refers to the concentration at which the calculations are carried out (~2 mM), ΔG_{bind} refers to the negative of the energy estimate calculated via SMD, and C^0 is the constant standard concentration (1 M) (133, 134). The dissociation constants are finally obtained via

$$K_d = \exp\left(\frac{\Delta G^0}{k_B T}\right) \quad (2.11)$$

2.3. Anisotropic Network Model (ANM): Theory

In ANM, protein is represented as an assembly of beads and springs. Each residue is perceived as a bead, which interacts via a harmonic potential with another bead residing in a

distance less than a defined cut-off value. Thus, a molecule having N residues, each with 3 degrees of freedom, will possess $3N$ degrees of freedom. Six purely rotational and translational degrees of freedom are removed since they do not assign any shape changes to the protein. Thus, $3N-6$ degrees of freedom will account for the modes of motion in any given protein. In ANM that is originally developed for proteins (25), each node is represented by the C_α coordinates of the residues in a folded protein, and the interactions between them are considered to be due to uniform harmonic potentials with a force constant γ . The overall potential of the molecule is given by the sum of all harmonic potentials among interacting nodes:

$$V = \frac{\gamma}{2} \sum_i \sum_{j>i} (\Gamma_{ij})(\mathbf{R}_{ij} - \mathbf{R}_{ij}^0)^2 \quad (2.12)$$

Here Γ_{ij} is the ij^{th} element of the Kirchhoff matrix $\mathbf{\Gamma}$ of inter-residue contacts. This term is equal to 1 if the distance between nodes i and j , $\mathbf{R}_{ij}$, is smaller than the cutoff distance r_c , and zero otherwise. $\mathbf{R}_{ij}^0$ is the equilibrium distance between corresponding residues. For a network of N nodes, the Hessian matrix is a $3N \times 3N$ matrix formed by a number of N^2 super elements, $\mathbf{H}_{ij}$. The off-diagonal super elements of $\mathbf{H}_{ij}$ ($i \neq j$), obtained from the second derivative of the total potential with respect to node positions, are given by:

$$\mathbf{H}_{ij} = \frac{\gamma A_{ij}}{(R_{ij}^0)^2} \begin{bmatrix} X_{ij} X_{ij} & X_{ij} X_{ij} & X_{ij} Z_{ij} \\ Y_{ij} X_{ij} & Y_{ij} Y_{ij} & Y_{ij} Z_{ij} \\ Z_{ij} X_{ij} & Z_{ij} Y_{ij} & Z_{ij} Z_{ij} \end{bmatrix} \quad (2.13)$$

where X_{ij} , Y_{ij} , and Z_{ij} are the Cartesian components of the distance vector $\mathbf{R}_{ij}^0$. The diagonal terms of $\mathbf{H}_{ij}$ ($i = j$) are the sum of the off diagonal super elements of $\mathbf{H}_{ij}$ ($i \neq j$). The pseudo-inverse of $\mathbf{H}$ is the $3N \times 3N$ covariance matrix, $\mathbf{C}^{\text{ANM}} \mathbf{C}^{\text{ANM}}$, that can be expressed in terms of the $3N-6$ non-zero eigenvalues λ_k and corresponding eigenvectors $\mathbf{u}_k$ of $\mathbf{H}$ as:

$$\mathbf{C} = \sum_{k=1}^{3N-6} \frac{1}{\lambda_k} \mathbf{u}_k \mathbf{u}_k^T \quad (2.14)$$

The frequency and the spatial dependence (direction) of each mode is represented by the eigenvalues λ_k and eigenvectors $\mathbf{u}_k$, respectively. Smallest non-zero eigenvalue λ_1

corresponding to lowest frequency carries the information on the most collective internal motion.

To benchmark ANM predicted equilibrated fluctuations of a folded protein, mean square fluctuations (msf) of C_α calculated through ANM, might be compared with crystallographic temperature factors (B-factors). B-factors are predicted by the ANM for residue i is calculated from the trace of $\mathbf{C}_{ii}; \mathbf{C}_{ii}^{ANM}$

$$B_i^{ANM} = (8\pi^2 k_B T / 3\gamma) \text{tr}(\mathbf{C}_{ii}) \quad (2.15)$$

2.4. Perturbation Response Scanning (PRS): Theory and Protocol

The PRS method is based on the assumption that the ligand bound state of the protein is described by a perturbation of the Hamiltonian of the unbound state. Under the linear response assumption, the shift in the coordinates is approximated by:

$$\Delta \mathbf{R}_1 = \langle \mathbf{R} \rangle_1 - \langle \mathbf{R} \rangle_0 \cong \frac{1}{3k_B T} \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle_0 \Delta \mathbf{F} = \frac{1}{3k_B T} \mathbf{C} \Delta \mathbf{F} \quad (2.16)$$

where the subscripts 1 and 0 denote perturbed and unperturbed configurations of the protein, the angle brackets denote the ensemble average and superscript T denotes the transpose. k_B is the Boltzmann constant and T is temperature $\Delta \mathbf{F}$ vector contains the components of the externally inserted force vectors on the selected residues. $\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle_0$ is the variance-covariance of the atomic fluctuations in the unperturbed state of the protein. In this study, we will utilize both MD simulations and ANM analysis to represent the variance-covariance matrix, which acts as the kernel between the inserted perturbations and the recorded displacements. Different derivations leading to the above equation may be found in references (48, 135, 136).

Our detailed PRS analysis is based on a systematic application of equation 2.16. We apply a random force to the C_α atom of the initial structure. The protein is scanned consecutively perturbing each residue i by applying the force $\Delta \mathbf{F}$ on C_α atom. Thus, $\Delta \mathbf{F}$ vector is formed in such a way that all the entries, except those corresponding to the residue being perturbed, are equal to zero. For a selected residue i , the random force q_i is $(q_x \ q_y \ q_z)^i$ so that the external force vector is constructed as:

$$\Delta \mathbf{F}^T = \left\{ 0 \ 0 \ 0 \dots (q_x \ q_y \ q_z)^i \dots 0 \ 0 \ 0 \right\}_{1 \times 3N} \quad (2.17)$$

We then compute the resulting changes $(\Delta \mathbf{R})^i$, as a result of the linear response of the protein, through equation 2.16. The predictions of the average displacement of each residue, j , as a response of the system to inserted forces on residue i , $\Delta \mathbf{R}^i$ are compared with the experimental conformational changes $\Delta \mathbf{S}$ between the initial and the target PDB structure, e.g. the apo and the holo forms of a protein. For the latter vector, each pair of experimental structures, the holo is superimposed on the apo form, followed by the computation of the residue displacement vectors, $\Delta \mathbf{S}$. The goodness of the prediction is quantified as overlap coefficient, O^i , or Pearson correlation coefficient, C_i , for each perturbed residue by comparing the predicted and experimental displacements:

$$O^i = \frac{\Delta \mathbf{R}^i \cdot \Delta \mathbf{S}}{\left| (\Delta \mathbf{R} \cdot \Delta \mathbf{R})^i (\Delta \mathbf{S} \cdot \Delta \mathbf{S}) \right|^{1/2}} \quad (2.18)$$

$$C_i = \frac{\sum_{k=1}^N [(\Delta \mathbf{R}_k)^i - (\overline{\Delta \mathbf{R}})^i] (\Delta \mathbf{S}_k - \overline{\Delta \mathbf{S}})}{(N-1)\sigma_R\sigma_S} \quad (2.19)$$

O^i is the dot product of the two vectors, as a measure of the similarity of the direction in the predicted conformational change. In equation 2.19, the overbar indicates the average and σ_S and σ_R are the respective standard deviations for experimental and calculated displacements.

As selection criteria of residues that are effective in directing the protein towards alternative conformations we take several approaches depending on the distribution of the maximum of the C_i values, C_i^{max} , obtained from the 500 perturbations and calculated through eq. 2.18. We first list C_i^{max} in ascending order: (i) If there is a sharp decrease in the Pearson correlation values, we list the top residues until that gap. (ii) If there is a smooth decrease in the Pearson correlation values, we list the top 10 residues. We also check the location of the residues that do not survive this selection cutoff. In all cases, we find that the next 10 residues are spatial neighbors of the listed ones. Therefore, the interpretation of the results does not depend on the cutoff point. (iii) If there are too many residues with $C_i^{max} > 0.8$, we do not list any residues as “identified.” Conversely, we interpret this behavior indicative of a

natural propensity of the protein to interconvert to the target conformation. (iv) If all residues have $C_i^{max} < 0.6$, we interpret this behavior as indicative of resistance to interconvert to the target conformation. However, we still list the top residues according to criterion (i) or (ii).



CHAPTER 3. MODELING THE LONG-TERM DYNAMICS OF PROTEIN COMPLEXES BY COARSE-GRAINED METHODOLOGIES

3.1. Electrostatics in remote and mechanical control of protein-protein interaction

We analyze a set of 25 protein complexes in their bound and unbound forms (Table 3.1). We collected protein pair structures from proteins reported in Benchmark 0.0 of ZDOCK benchmarks (137). The complexes are non-redundant and they have X-ray structure solved at better than 2.90 Å resolution. They include wide variety of function and affinities and belong to different biological families. The constituent unbound forms of all the 25 complexes are available in protein data bank and they have solution NMR or X-ray structure solved at better than 3.50 Å resolution. For the proteins resolved by solution NMR, we always use the first model for all the calculations. More specifically, we have chosen the protein complexes with no less than two missing residues along the protein chain, either in the bound or unbound form. Furthermore, if the number of residues in the bound and unbound components differ, we only analyze the common parts of the bound and unbound proteins. Therefore, in some cases, due to lack of structural information the complete sequence of a protein complex has not been studied.

For each pair of experimental structures, the unbound form is superimposed on the bound form, followed by the residue displacement vectors computation, $\Delta\mathbf{S}$. In this part, we perturb the bound form of each protein by applying a random force to the C_α atom of each residue to induce the dissociation of the proteins. For a given protein, we select the cut-off distance, r_{opt} that yields the closest agreement with the displacement vectors from experiments for at least one residue.

Table 3.1 Summary of PRS results

Protein complex	PDB codes (Bound: chain/Unbound: chain) (X-ray resolution, Å)	Amino acid sequence	N	RMSD (Å)	Regions of motion (Å)	Cut-off (Å)	Correlation	C_i^{max}	$> C_i$	Binding site residues	First shell residues	Remote site residues	
Type I	Alpha amylase:Alpha amylase inhibitor	1CLV:A/1JAE:A (2.00/1.65)	2-471	470	0.5	81-90: 1.8 290-300: 2.5 440-450 :1.7	12	0.92	0.87	0.87	G292A	N/A	N/A
		1CLV:I/1HTX:A (2.00/NMR)	501-532	32	0.9	511-520:1.0 520-525:1.3		0.75	0.76	0.66	W57A, N138A, V151A, G152A	N/A	N/A
	MAP kinase-activated protein kinase 2:Mitogen-activated protein kinase 14	2OZA:A/1KWP:A (2.70/2.80)	51-215	165	2.0	65-70:6.2 70-75:8.0 75-80:3.0	14	0.97	0.68	0.65	L70A	G71A	N/A
		2OZA:B/ 1P38:A (2.70/2.10)	16-169	154	1.7	16-21:1.8 25-30:2.3 30-35:5.3 35-40:2.6 55-58:2.4 97-100:2.2 115-120:2.9 120-125:2.9		0.96	0.80	0.80	N/A	N/A	G171A, Q175A, Y176A

	Trypsin:Trypsin inhibitor	1AVX:A/1QQU:A (1.90/1.63)	16-245	230	0.5	95-100:0.85 165-170:0.85 215-220:0.90 240-245:0.75	13	0.9	0.62	0.59	N/A	N/A	L520B, K552B
		1AVX:B/1BA7:A (1.90/2.50)	501-623	123	0.5	545-550:0.6 595-600:0.8		0.87	0.42	0.40	N/A	N/A	S579B, F580B, A581B, D598B , K611B
	Ras-related protein Ral-A:Mono-ADP-ribosyltransferase C3	2A9K:A/1U90:A (1.73/2.00)	13-178	166	0.6	47-51:0.8 56-62:0.8 70-75:2.1	11	0.78	0.8	0.79	N/A	N/A	Y66B, G67B, L68B, S69B, D112B
		2A9K:B/2C8B:X (1.73/1.70)	45-245	201	1.2	70-82:1.2 140-150:1.5 180-185:2.9 209-216:3.3		0.73	0.68	0.62	N/A	N/A	S181B , F209B , A210B , G211B
	Ribonuclease SA:Barstar	1AY7:A/1RGH:A (1.70/1.20)	1-96	96	0.5	-30:0.7 39-41:0.7 61-65:0.9	13	0.93	0.72	0.69	N/A	N/A	L41B, T42B, G43B, W44B
		1AY7:B/1A19:A (1.70/2.76)	1-89	89	0.6	5-15:0.95 55-66:0.95		0.75	0.52	0.50	N/A	N/A	E8B , E57B , Q58B

	Carboxypeptidase A: Metallo carboxypeptidase inhibitor	4CPA:A/1YME:A (2.50/1.53)	1-307	307	0.5	132- 136:1.8 245- 249:1.2	13	0.97	0.70	0.70	N/A	N/A	S134A
		4CPA:I/1H20:A (2.50/NMR)	4-37	34	1.1	10-16:1.6 17-21:2.1		0.50	0.70	0.66	N/A	N/A	K177A, S199A, I274A
	Ribonuclease A: Ribonuclease inhibitor	1DFJ:E/9RSA:A (2.50/1.80)	1-124	124	0.7	15-17:0.8 90-95:1.2 110- 114:1.0	12	0.76	0.59	0.58	N/A	N/A	G186I, D213I, P450I, G451I
		1DFJ:I/2BNH:A (2.50/2.30)	1-456	456	1.5	1-30:2.9 40-55:2.6 414- 423:2.7 443- 451:3.1		0.80	0.91	0.89	N/A	N/A	L22I, A46I, L47I, R48I, A49I
	superoxide dismutase: copper chaperone for superoxide dismutase	1JK9:A/2JCW:A (2.90/1.70)	1-153	153	0.8	51-62:2.5	13	0.70	0.78	0.76	N/A	I69B	D67B, A68B
		1JK9:B/1QUP:A (2.90/1.80)	3-222	220	4.3	3-10:5.4 30-40:5.3 50-70:6.9 160- 170:5.7		0.70	0.60	0.57	N/A	N/A	C27B, P54B, S55B

	Carboxypeptidase A1: Metallo-carboxypeptidase inhibitor	2ABZ:A/1M4L:A (2.16/1.25)	5-305	301	0.4	245- 250:1.3	13	0.96	0.75	0.73	Q16C,V17C	C18C	N/A
		2ABZ:C/1DTV:A (2.16/NMR)	5-65	61	1.3	12-15:2.1 52-56:1.5		.56	0.75	0.75	N/A	E31C	N/A
	Ferredoxin-NADP reductase: Ferredoxin I	1EWY:A/1GJR:A (2.38/2.10)	9-303	295	1.1	50-54:1.4 67-74:3.3 102- 112:2.0 221- 236:2.3 262- 270:1.4 280- 287:1.2	12	1.00	0.70	0.67	N/A	N/A	V67A, D68A, K69A
		1EWY:C/1CZP:A (2.38/1.17)	1-98	98	0.8	10-15:1.5 50-67:1.5		0.80	0.74	0.72	N/A	I62A	T164A, F183A
	cysteine protease: cysteine protease Inhibitor	1PXV:A/1X9Y:A (1.80/2.50)	223-392	170	2.5	330- 339:9.1 375- 382:3.9	11	0.55	0.9	0.85	N/A	H54C	V10C, Y11C, H44C
		1PXV:C/1NYC:A (1.80/1.40)	0-109	110	0.9	0-5:1.6 91-96:1.3 104- 109:2.7		0.72	0.80	0.80	N/A	N/A	V109C

Type II	Chemotaxis protein CHEY: Chemotaxis protein CHEA	1FFW:A/3CHY:A (2.70/1.66)	2-129	128	0.5	90-92:1.1 110- 112:0.9 124- 130:1.3	12	0.83	0.75	0.	N/A	N/A	K190B, G191B, L195B, A197B
		1FFW:B/1FWP:A (2.70/NMR)	160-226	67	1.8	165- 170:2.6 200- 205:2.6 210- 215:2.3		0.77	0.61	0.54	N/A	N/A	G52A, V54A, D57A, N59A
	Cell division protein FTSZ: Hypothetical protein PA3008	1OFU:A/2VAW:A (2.10/2.90)	11-316	306	0.5	70-72:0.9 122- 125:0.9 202- 208:2.7 231- 235:1.2 268- 273:1.8 288- 293:0.9 299- 306:2.0	12	0.92	0.83	0.82	H89X, R93X	L87X, T88X	N/A
		1OFU:X/1OFT:A (2.10/2.90)	45-160	116	0.8	70-72:1.1 87-90:1.0		0.90	0.58	0.56	I207A	D210A(3.9 5), L271A, S272A	N/A
	Complement C3: Fibrinogen-binding protein	3D5S:A/1C3D:A (2.30/1.80)	8-298	291	0.4	44-51:1.7 165- 171:1.0	12	0.92	0.70	0.69	N/A	N67C, K70C, Q71C	N/A
		3D5S:C/2GOM:A (2.30/1.25)	15-75	61	0.4	15-19:0.7 70-75:0.9		0.78	0.80	0.78	N/A	N/A	R10A, L11A, K12A, H13A, L14A, I15A, V16A, T17A

Type III	Bovine hymotrypsinogen A:human pancreatic secretory trypsin inhibitor (Kazal-type)	1CGI:E/2CGA:A (2.30/1.80)	1-245	245	1.5	31-39:1.2 140- 154:4.5 183- 195:3.5 216- 223:1.6	13	0.60	0.70	0.68	T30I, Y31I, P32I	N/A	N/A
		1CGI:I/1HPT:A (2.30/2.30)	(1-56)	56	1.8	1-5:4.4 10-14:3.2 17-20:1.5		0.48	0.83	0.79	G197E	A179E	N/A
	Elastase:Elafin	1FLE:E/1QNJ:A (1.90/1.10)	16-245	229	0.9	59-62:1.4 96-106:1.6 166- 176:1.7 186- 196:1.3 216- 226:1.7	13 Å	0.96	0.75	0.65	N/A	N/A	T11I, K12I, P13I, L33I, K34I
		1FLE:I/2REL:A (1.90/NMR)	11-57	47	2.8	11-43:3.5		-0.30	0.71	0.66	N/A	N/A	L123E, A208E, V209E
	Interleukin-6 receptor beta chain Leukemia inhibitory factor	1PVH:A/1BQU:B (2.50/2.00)	101-301	201	0.9	130- 135:1.7 210- 215:1.3 240- 245:1.3 295- 301:1.8	12	0.89	0.85	0.83	N/A	N/A	R276A, I277A, E294A, A295A, S296A, G297A
		1PVH:B/1EMR:A (2.50/3.50)	22-180	159	0.9	135- 140:1.3 140- 150:2.0 150- 160:1.6		0.84	0.58	0.54	N/A	N/A	G147B, P148B, D149B, T150B

	Alkaline metalloproteinase:Proteinase inhibitor	1JIW:P/1AKL:A (1.74/2.00)	1-470	470	1.2	20-22:1.5 185- 195:7.3	10	0.98	0.85	0.78	N/A	N191P, A192P, G193P	N/A
		1JIW:I/2RN4:A (1.74/NMR)	8-105	98	1.4	18-25:3.4 50-55:1.2 95-98:2.2		0.80	0.79	0.73	N/A	N/A	E21I, A22I
	Heat shock protein HSP82:HSP90 Co-chaperone CDC37	1US7:A/1AH6:A (2.30/1.80)	2-207	206	0.8	54-57:1.3 91-99:3.5 100- 117:1.7 198- 203:1.3	10	0.80	0.79	0.79	N/A	A97A	N/A
		1US7:B/2W0G:A (2.30/1.88)	148-276	129	1.0	224- 228:1.4 235- 240:1.3 240- 255:2.7 270- 276:1.9		0.92	0.79	0.79	N/A	N/A	A244B
	Trypsinogen:Bowman-Birk proteinase inhibitor precursor	1D6R:A/2TGT:A (2.30/1.70)	19-245	217	0.6	96-99:0.7 144- 146:0.8 188- 194:2.6 217- 220:1.6	11	0.96	0.78	0.78	N/A	N/A	N/A
		1D6R:I/1K9B:A (2.30/2.80)	7-63	57	1.0	30-35:1.3 42-46:1.4 57-59:1.1		0.50	0.70	0.70	N/A	S34I	H33I

Type IV	Human transferrin receptor:apo transferrin	1SUV:CE/1CX8:B (7.50/3.20)	122-760	639	0.3	205- 215:1.2	0.80	0.85	0.75	N/A	N206A	K205A, G207A, R208A, L209A, V210A, Y211A, L212A, V213A
		1SUV:CE/2HAV:B (7.50/2.70)	4-676	674	9.7	N/A	0.44	0.52	0.50	N/A	N/A	S117C, G156C,S157C , C158C(E), A159C, A162C, Q172C, G200C
	Falcipain 2:Chagasin	2OUL:A/2GHU:A (2.20/3.10)	-15-224	240	0.6	1-5:2.1 115- 120:1.0 153- 158:1.2 185- 195:2.6	1.00	0.87	0.83	N/A	D148A, Y37B, G41B	G40B
		2OUL:B/2H7W:A (2.20/1.70)	4-110	107	0.6	60-65:1.7	0.25	0.78	0.76	N/A	I68A	S113A, V114A, D148A, F219A

ATP-dependent Clp protease ATP-binding subunit clpA:ATP-dependent Clp protease adaptor protein clpS	1R6Q:A/1R6B:X (2.35/2.25)	1-141	141	1.0	15-20:1.2 68-76:2.8 91-102:1.6	0.85	0.82	0.77	N/A	E73A, K49C	D45C, L61C
	1R6Q:C/3O1F:A (2.35/1.40)	26-106	81	0.4	37-40:0.6	0.76	0.76	0.74	S118A	N/A	E7A, Y122A
NAD(P) transhydrogenase subunit alpha part 1:NAD(P) transhydrogenase subunit beta	2OOR:A/1L7D:A (2.32/1.81)	1-220	220	0.6	45-50:0.9 80-85:1.3 164- 170:1.6 204- 207:1.2 214- 220:1.1	0.23	0.68	0.68	N/A	N103C, P105C	L214A, T220A
	2OOR:C/1E3T:A (2.32/NMR)	30-201	172	2.2	30-50:2.8 140- 150:2.9 170- 190:2.4	0.52	0.50	0.49	A166A	N/A	M167C

	Alpha-Chymotrypsin:Protease inhibitor LCMI II	1GL1:A/1MTN:F (2.10/2.80)	16-146	131	0.5	71-82:1.3 143-146:1.0	13	1.00	0.76	0.52	N/A	N/A	S76A, S77A
		1GL1:I/1PMC:A (2.10/NMR)	2-33	32	1.6	7-23:2.1		-0.40	0.87	0.76	C58A, C14I	K13I	N/A
	Alpha amylase:Tendamistat	1BVN:P/1PIG:A (2.50/2.20)	2-496	495	0.7	51-55:0.8 108-112:2.3 140-150:1.5 238-243:1.9 303-308:2.7 347-351:1.4	13	0.89	0.64	0.64	A823T	C811T	N/A
		1BVN:T/1HOE:A (2.50/2.00)	804-874	71	0.5	827-830:0.6 838-840:0.7 860-862:0.7		0.68	0.57	0.53	N/A	N/A	Q5P, T6P, Q7P, S8P, R10P, V804T
	Human transferrin receptor:holo transferrin	1SUV:CE/1CX8:B (7.50/3.20)	122-760	639	0.3	205-215:1.2	12	0.80	0.83	0.75	N/A	N/A	K205A, N206A, G207A, R208A,L209A ,V210A,Y211 A, L212A, V213A
		1SUV:CE/3V83:B (7.50/2.10)	3-676	674	4.7	N/A		0.52	0.49	0.47	N/A	N/A	E634A, M635A, G636A, G724A, A725A, F726A, N506E, K531E, S532E, V584E

We verify that the residue indices that provide the best Pearson correlation value are always present within the range perceived as the highest correlation value for all cut-off distances studied. We note that the Pearson correlation values reported in Table 3.1 are not affected for the range of values $r_{opt} \pm 1$ Å for all the 25 protein complexes. Strikingly, although many of these proteins display large Pearson correlation values, the numbers of residues yielding the highest values differ among proteins. For some proteins, there is not much specificity on the residue to be perturbed to reproduce the conformational change. For others, by perturbing a very specific location, the complete conformational change is obtained. The former category is exemplified by fibrinogen-binding protein (pdb code: 3d5s (138)) while serine protease and its proteinaceous inhibitor (pdb code: 1d6r (139)) is an example of a protein complex that we need to perturb in a specific location to maintain dissociation.

The structural difference between the bound and the unbound forms, based on the C_α superposition of the binding partners, show that while for 21 of the cases interface RMSD is less than 1.5 Å; for two cases interface RMSD is 1.5-2.2 Å and two cases have interface RMSD greater than 2.2 Å (140). For the pair of initial and target forms of the proteins present in bound and unbound form, we perform STAMP structural alignment, implemented in VMD 1.9.1 MultiSeq plugin (141). We record the RMSD between the structures of the target forms with the initial structure. The overall RMSD between the initial and target structures vary between 0.4-9.7 Å. Although large motions of side chains and surface loops is always present as a local conformational change, we do not detect any discernible shape change on global scale in proteins in their bound form compared to respective unbound. Thus, for the 25 cases, the main chains essentially have same conformation in bound and unbound forms and the binding partner interact as rigid bodies. However, the superposition of different segments of the proteins in the complex and corresponding unbound forms produces higher RMSDs and can be as large as 9.1 Å (see case of cysteine protease and its inhibitor (142)); the largest deviations being mostly confined to flexible unstructured stretches as in turns and bends. Recording the residues identified by PRS, displays that hot residues encouraging the conformational change from the bound form to unbound form do not always reside on these flexible structures. We have collected 187 hot residues from PRS calculations and 65 of them are located on the flexible loops with large motions and high RMSD values (Table 3.1). The remaining loop residues are located on the loops without any large structural change upon

binding event. We also identify 40 hot residues as part of α -helical structures and 37 residues as β -strands.

Root mean square fluctuations (RMSF) of each protein complex constituents in their bound and unbound form are derived from auto-correlation of the residues in each protein pair. Auto-correlation of residue is calculated by treating the **C** matrix as an $N \times N$ supermatrix, whose ij^{th} element is the 3×3 second moment matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta \mathbf{R}_i$ and $\Delta \mathbf{R}_j$ of residues i and j (82). Residue auto-correlations are calculated by taking $i=j$. The cut-off distance for building the Hessian matrix of each protein complex have the same values as cut-off distances chosen for PRS analysis. In Table 3.1, we report the correlation values between proteins in their bound and unbound form for all the protein pairs. The correlation values vary in 0.23-1.0 range. We observe that there is a significant change in RMSF of binding region residues in at least one of the constituents in each pair. This means that the local fluctuations of interface area varies in at least one protein upon complex formation while local fluctuations of binding protein display the same pattern as the unbound form. However, in an exceptional case of transhydrogenase complex (pdb code :2oor (143)), we observe low correlations, 0.23 and 0.52, between the binding proteins and their unbound form. For this protein complex the RMSF curves of the constituents display a significant transformation of fluctuation pattern upon protein binding.

Among 10126 of total protein residues, 187 of them are selected by PRS. These residues, which are involved in encouraging the unbound over bound form, are distributed among all amino acid types. Total number of each amino acid type in our analysis pool and their contributions in PRS analysis are listed in Table 3.2. PRS does not display any preference over amino acid types and any contribution to PRS selection is corroborated to the population of any given amino acid type in the analysis pool. For example, Glycine with 13% and Alanine with 10% of all PRS top rated residues, are observed more frequently than other types. However, this is attributed to the highest presence of Glycines and Alanines in the analysis pool with 8% distribution each. In the same manner, Methionine and Tryptophan, each with 1% frequency are the least detected residues by PRS. They are also less frequently seen in the analysis pool with 2% population.

Table 3.2 Amino acid type analysis of PRS identified residues

	Non-polar amino acids								Charged amino acids				Polar Amino acids							
	A	V	L	I	M	F	G	W	P	D	E	R	K	N	Q	T	Y	S	H	C
PRS residues	18	14	17	7	2	5	24	2	6	11	7	6	12	7	7	9	7	17	4	5
PRS residues (%)	10	7	9	4	1	3	13	1	3	6	4	3	6	4	4	5	4	9	2	3
All residues	82	70	89	53	17	38	82	15	39	56	59	39	57	53	41	58	36	74	20	27
All residues (%)	2	5	0	3	1	2	1	7	4	8	0	9	0	5	6	7	6	5	3	2

In Table 3.2, we report the local secondary structure attributes of all residues and residues detected by PRS. The secondary structures are assigned by "Timeline" plugin of VMD and are calculated based on STRIDE algorithm (141). Among all the protein complexes, we do not find any π -helix type of structure. Residues in the analysis pool are mostly populated by turns, α -helices and β -strands with 26, 26 and 24% distribution, respectively. However, we find that most preferred regions by PRS defined residues are on coils and turns, each with 29% of all PRS defined residues, although they populate only 44% of the analysis pool. We will discuss the secondary structure properties of the PRS residues later in the text in more detail.

In Table 3.3, we also categorize the hot residues locations and divide protein structure into three zones. The interface of the two proteins present in the complex is defined as C_α all the residues from the two sides of the pair sitting in the smallest possible cut-off distance of each other. The cut-off distance was measured from C_α of each residue. We select the smallest distance as 7 Å if construction of a Hessian matrix based on 7 Å cut-off produces six zero eigen-values corresponding to pure translational and rotational degrees of freedom. Otherwise, we increase the smallest possible distance to 8 Å. First coordination shell residues are defined as residues located in one cut-off distance from interface residues. In definition of first shell residues, same cut off distance has been chosen as in the definition of the binding region. Remote residues are the ones out of these two zones. We observe that except for one case, Alpha amylase and its inhibitor (pdb code: 1clv (144)), PRS selects residues that are not confined only to the interface region (Table 3.1). Accordingly, we conclude that the interface of two proteins is not the only controlling region for dissociation of the two proteins. In fact, in ten cases, we observe that residues with best Pearson correlation values never reside on the

interface. The remaining 16 protein complexes display residues from different parts of the protein in their PRS analysis.

Table 3.3 Structural analysis of PRS identified residues

	α -helix	π -helix	3-10 helix	β -strand	Isolated bridge	turn	coil	Total
PRS residues	36	0.0	4	37	1	54	55	187
PRS residues (%)	19	0.0	2	20	1	29	29	100.0
All residues	2637	0.0	403	2398	171	2683	1834	10126
All residues (%)	26	0.0	4	24	2	26	18	100.0

Remarkably, residues signaling the dissociation of each protein in a given complex are not located on the same protein in all cases. Accordingly, effective sites involved in the dissociation of various protein complexes found in PRS analysis are categorized into four groups based on their responses to the perturbations on the protein. Proteins in which dissociation is signaled through remote compartment of the complex are labeled as type I. In this group, PRS top rated residues are all confined to one of the binding proteins. Thus, residues on this protein also control the binding protein. Type II are the proteins in which each protein controls the interacting bound protein. In this group, residues confined to one of the proteins control the other binding protein and vice versa. Type III are the proteins in which each constituent controls its own dissociation activity, and type IV are the proteins where presence of both proteins are essential for the dissociation reaction and a combination of various characteristics are observed. In this group, residues signaling the dissociation are scattered on both binding partners.

We pay special attention to the interface region of all case studies. The protein-protein interface similarity scores measured by PIFACE server (145) screens a diverse variety of similarities based on structural alignments for all types of protein complexes provided here. We note that there is no interface similarity data available for Carboxypeptidase A/Metalloprotease inhibitor (pdb code : 4cpa) and hTf-TfR (pdb code: 1suv (61)) complexes. We observe that the highest similarity score within one group (0.23 ± 0.18) belongs to type II group. The similarity scores between various groups are 0.17 ± 0.10 the lowest, and 0.21 ± 0.12 the highest. Overall, there is no significant gap among the different types in terms of interface similarity. Evidently, the structural properties of the interfaces may not be the only effective parameter in protein-protein behavior and the collective dynamics of the

complex assigned by local variations, specifically remotely located amino acids may also be in effect in the functional pathway of protein complexes.

3.1.1. Allosteric control of disassociation is coupled to electrostatic effects

In type I group of protein complexes, perturbation of specific sites on only one of the constituents in protein complex modulates a process, namely the dissociation of either itself or the binding partner. The functional amino acids defined by PRS, which are involved in disintegrating the contact network of the interface area, tend to be located on the interface, first shell or distant locations of the protein tertiary structure. Additionally, no specific perturbation location is observed. This means that the local perturbations, which lead to global conformational shifts between bound and unbound states, are not bundled in one region; rather they are scattered around protein structure.

We also model the electrostatic interactions and obtain the electrostatic potential distribution on the biomolecular surface using the APBS package (146). We note that APBS calculations were set to their default values. Biomolecular and solvent dielectric constant is set to 2 and 78.54, respectively. Radius of the solvent molecules and temperature is 1.4 Å and 298.15 K. Cubic B-spline discretization method is used to grid biomolecular point charges. Electrostatic effects play a major role in functionality of this group of protein complexes. Monitoring the electrostatic potential distribution along the protein surfaces in their bound and unbound forms, reveals that any given protein in the complex whose electrostatic potential distribution state is stable in their bound and unbound form will be the signaling protein during the dissociation reaction. The intermolecular forces with electrostatic origins in the binding protein will always display considerable change in our test cases. In Figure 3.1a, we show how the population of charged and polar residues distribution on the surface of the protein changes in the free to bound form.

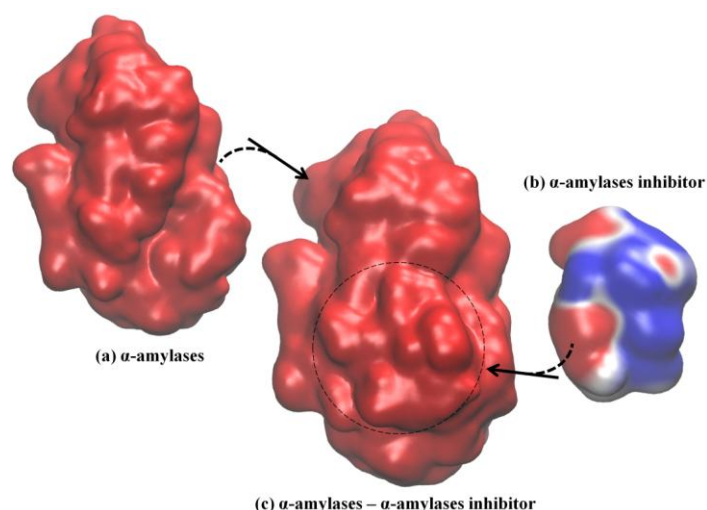


Figure 3.1 Electrostatics isocontours of α -amylase/ α -amylase inhibitor drawn at ± 0.5 kBT/e. Blue is positive and red is negative. (a) α -amylase in the unbound form, with overall negative electrostatic potential distribution along the surface. (b) α -amylase inhibitor in its unbound form with a mixed pattern of negative and positive negative electrostatic potential distribution along the surface. (c) α -amylase/ α -amylase inhibitor complex with overall negative electrostatic potential distribution. The spatial orientation of the proteins in the complex is kept the same as respective unbound forms. Dashed circle displays positioning of the inhibitor in the complex. Rest of the surface belongs to α -amylase protein. The signaling protein, α -amylase, maintains its electrostatic potential distribution while α -amylase inhibitor displays altered electrostatic potential distribution.

Type II group of protein complexes, are another representative of protein complexes with remotely controlled movement from bound to unbound conformation. In this group, dissociation of each element of the complex is controlled by the binding element; that is, the cross-controlled dissociation. We observe that a regional perturbation is required to encourage the interconversion between two conformational endpoints. This means that, as defined by PRS, any point mutation in a stretch of sequential amino acids will promote conformational change. In addition, analysis of the electrostatic potential distribution shows that the proteins interacting with each other possess similar state of charge distribution. Thus, if the unbound forms had a different electrostatic potential distribution, they reorient themselves to the same state upon protein complex formation. In Figure 3.2, we illustrate how electrostatic potential surface is oriented before and after complex formation in Efb-C and its complement target C3d (pdb code: 3d5s (138)). It has been reported that formation and stability of Efb-C binding to C3d is electrostatic in nature (138). Kinetic experiments in salt concentrations of 75-600 mM indicate the sensitivity of association/dissociation phases of wild type and various mutants to ionicity (138). This suggests that overall electrostatic contribution to be of importance in initial complex formation and further stability of complex under environmental

conditions. Our PRS analysis of group I and II, is further improved by these observations such that under various environmental perturbations, disruption of long range and short range electrostatic complementarity seem to affect association/dissociation rates, impair stability and complex formation of binding partners.

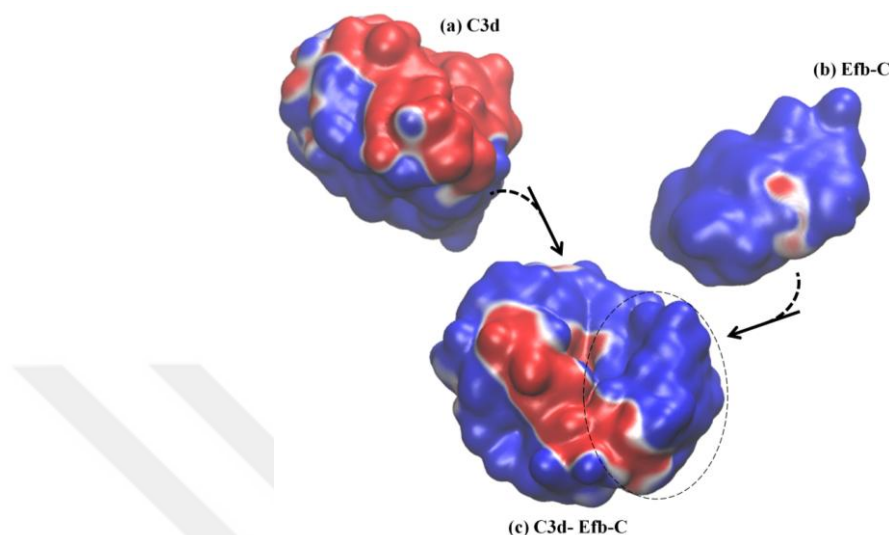


Figure 3.2 Electrostatics isocontours of Efb-C and its complement target C3d. drawn at ± 0.5 kBT/e. Blue is positive and red is negative. (a) C3d form, with overall negative/positive electrostatic potential distribution along the surface. (b) Efb-C in its unbound form with negative/positive, mostly positive, electrostatic potential distribution along the surface. (c) Efb-C/ C3d complex. Dashed circle displays Efb-C protein in the complex and the rest of the surface belongs to C3d protein. Both proteins display a mixture of negative/positive electrostatic potential distribution along the surface. Mostly, positive surface of the Efb-C displays grown negative areas upon complex formation. The spatial orientation of the proteins in the complex is kept the same as respective unbound forms.

Remarkably, electrostatic effects provide an excellent description for the observed pattern in type I and II groups. Electrostatic interactions are the primary factors of pH dependant processes in biochemical reactions. Particularly, we find that among 16 protein complexes included into type I and II groups, 13 of them belong to enzyme family of proteins. Enzymatic activities are known to be pH dependant and protonation state of catalytic and active site residues are effective in potential distribution of binding region. Consequentially, charge distribution of these regions will modulate the interactions between them and the reaction products. It has been reported that, enzymes make use of their preoriented environment to stabilize the transition state and the reduction in catalytic energy is accomplished by electrostatic stabilization of the active site of the enzyme (147, 148). In the same vein, we propose that the electrostatic characteristics of residues top-rated by PRS might

be found essential to structural stability, specificity and ligand binding properties in enzymatic reactions. For example, in C3bot-RalA complex (pdb code:2a9k (149)), displayed in Figure 3.3a, residues 207-214 of C3bot are part of the substrate recognition site and important in catalytic activity. PRS selects residues 209-211 as candidate regions to provoke the dissociation of the complex. Another key region involved in dissociation upon force application is the residue 112 of C3bot, which also contributes to complex formation. In addition, residue 181 is located on a loop responsible in substrate recognition. Accordingly, perturbation of specific sites found by PRS analysis and experimentally verified to be functionally important might reorient the enzyme/substrate dipoles that organize the catalysis and destabilize the charged transition state (150). Thus, preventing the binding event or promoting dissociation after the chemical reaction has terminated.

We also find that effect of the local charge distribution on enzyme functions is not limited to the active site and that the remote locations on the protein are effectively involved in association/dissociation process. Mutations of charged patches remote from either the protein or ligand binding site might alter the binding kinetic rates, shift pK_a s and lead to weak molecular recognition (150). In response to a particular perturbation, more likely exposure to an alternative environment, reorganization of charged atomic groups and dielectric relaxation of the protein affects the electrostatic potential distribution of the interface or active site area considerably, creating the steering torque that guide the dissociation reaction. Thus, a local change of interactions at a remote site leads to a global structural change that modifies the organization of the interface contact network and leads to dissociation of the two proteins. The contribution of this distal mutation on enzyme/inhibitor activity basically may be viewed as a cooperative conformational transition. In both type I and type II groups, the information transmission between remote functional sites on one protein and the entire structure of the interacting partner naturally occurs via interface linkage. We propose that in such proteins, the network of contacts in the interface could form a so-called "conductive" interface. Therefore, the signal from one protein is conducted through the interface to control the functionality of the second binding protein.

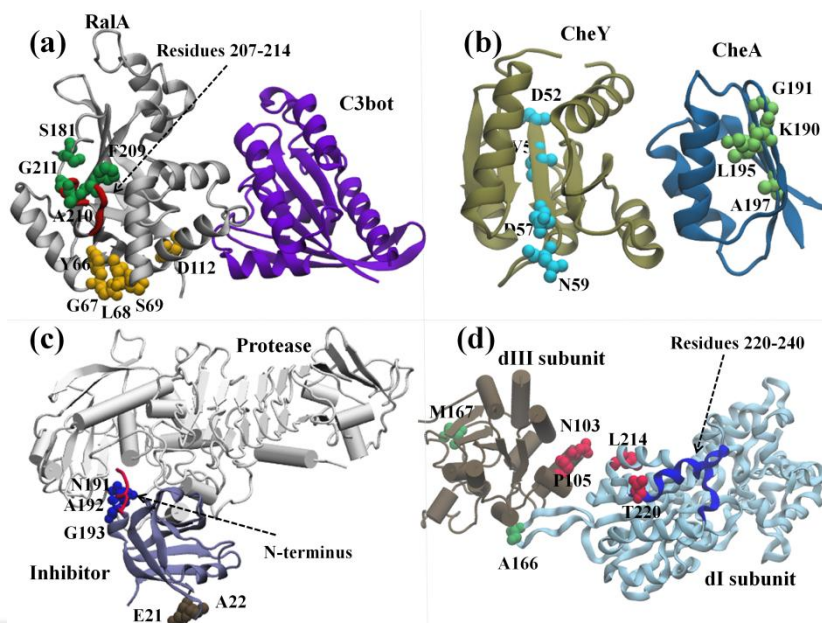


Figure 3.3 (a) Ribbon representation of C3bot-RalA complex as an example of type I proteins, C3bot is in purple and RalA is in silver. PRS rated residues that control the disassociation of the complex are displayed in VDW and all are confined to RalA compartment. Residues controlling the disassociation of RalA are in yellow while the residues controlling disassociation of C3bot are in green. (b) Ribbon representation of chemotaxis CheA:CheY complex as an example of type II proteins; CheY is in tan and CheA is in blue. Residues controlling the disassociation of CheY are located on CheA and are displayed in green while residues controlling the disassociation of CheA are located on CheY and are displayed in cyan. (c) Ribbon representation of alkaline protease and its cognate inhibitor as an example of type III proteins. Protease is represented as white cartoon and the inhibitor is in ice blue. Residues controlling the disassociation of protease are located on it and displayed in blue. (d) Ribbon representation of dII2dIII1 complex of proton-translocating transhydrogenase as an example of type IV proteins. Residues that control the disassociation of the dI subdomain are located on both subunits and are displayed in red. Residues that control the disassociation of dIII subunit are displayed in green.

As one of the test beds, we have studied chemotaxis CheA:CheY complex (pdb code: 1ffw (151)). PRS finds residues 52, 54, 57 and 59 of CheY to play a role in selecting the unbound over bound conformation upon force application (Figure 3.3b). This complex is an example of remote communication between a two-component signal-transducer. pH dependant catalytic activity has also been observed in chemotaxis CheA:CheY, in which ligand binding on CheA is conducted to CheY. Phosphorylation reaction on H48 of CheA subsequently transfers to D57 of CheY and the complex dissociates (152). CheY, itself is incapable of providing an acidic residue during the phosphorylation events and the complex formation with CheA results in a conformational change on CheY so as an acidic residue (D57) donates proton to a phosphodonor in an optimal orientation and the protein-protein phosphotransfer occurs (152).

The pH dependency of the phosphotransfer kinetics in pH range of 7.5-10, studied through two mutants of CheY active site residues simply shows a moderate decrease in rate constants compared to the wild type CheY. This observation suggests that conserved active site residues do not have an essential and direct role in catalysis. Thus, the loss of activity throughout this range for phosphotransfer to wild type CheY is attributed not to the deprotonation events on CheY; rather it is likely due to deprotonations in CheA (152).

Variability of position 59 of CheY as a non-conserved and indirect active site residue in modulating the autophosphorylation of CheY with small molecule phosphodonors shows no detectable binding between the phosphodonor and CheY, validating the significant impact of position 59 on autophosphorylation kinetics(153). Mutation of N59 to R, K, M, L, A, D and E results in both increase and decrease in autophosphorylation rate constants. Substitution with positively charged residues increases the kinetic rates while substitution with negatively charged residues decreases the rate, implying how local electrostatic interactions at position 59 modulate the CheY autophosphorylation (153).

3.1.2. Auto-controlled conformational transitions is correlated to mechanical organization of the protein

In type III group of protein complexes, dissociations of either constituent of the complexes are governed by the local perturbation of respective protein. We label this behavior as auto-controlled dissociation. In this category, all the residues involved in global transformation of bound to unbound forms are located out of interface region, Trypsinogen and its Bowman-Birk proteinase inhibitor precursor (pdb code: 1d6r (139)) being an exception. We observe that the global RMSD between bound and unbound forms of each protein varies between 0.6-1.8 Å. However, regions of high mobility exist, in which the local RMSD increases to as high as 7.3 Å. Such regions belong to unstructured surface exposed loops where they display large deviations upon complex formation. The amino acids identified by PRS in this group belong to these regions and very specific location may be perturbed so as the local fluctuations of the amino acids chooses the conformational switch to the target structure. Thus, mechanical motions of the loops produce essential conformational transitions under such point perturbations. Remarkably, there exists stretches with large RMSDs in type II protein complexes. Nevertheless, those regions cross control the

dissociation of the binding proteins remotely and their own shape changes are not mechanically controlled.

The electrostatic potential distribution along the protein surfaces of type III, reveals that each proteins maintain their electrostatic potential distribution state in their respective bound and unbound forms. This means that for any protein complex in this group, orientation of the negatively and positively charged surfaces of the constituents has the same distribution in their respective free forms. Thus, the analysis identifies an "insulating" interface area that prevents the allosteric communication between interacting proteins and each protein functions independently under various perturbations.

In case of alkaline protease and its cognate inhibitor (pdb code: 1jiw (154)) classified as type III, PRS defines that residues 191-193 of protease are in direct contact with the N-terminus of the inhibitor (Figure 3.3c). The latter has been shown to coordinate the catalytic zinc anion (154). Such an interaction adds to the structural stability and leads to a low dissociation constant. Upon any kind of perturbation applied to residues 191-193 of the protease, disruption of the interactions in this region would change the extended conformation of the N-terminus and modify the zinc coordination and thus might facilitate the dissociation process. Furthermore, pK_a value of the N-terminus has been reported to be an important factor in zinc coordination (154). Any force application such as charge alteration on residues 191-193 might affect the N-terminus pK_a through changing its environment and change the inhibition mode of the inhibitor.

3.1.3. Combined perturbation of two sides of the protein pair is essential for dissociation in group IV of protein complexes

In type IV group of protein complexes, dissociation of the proteins is facilitated by the perturbation of both sides of the protein complex. Therefore, dissociation of the two constituents requires point perturbations on both subunits. In this process, simultaneous perturbation of the two sides of the protein complex is necessary for the dissociation to proceed while in other groups, perturbation of any subunit will mediate the disruption of interactions in the protein pair and dissociation will happen. Different functional subdomains may exist on these proteins, which contribute to binding to diverse set of ligands in their functional pathway and promote association/dissociation reactions. Additionally, protein complexes involved in transmembrane processes also fall into this group of proteins.

Presences of each fragment of the protein pair in different compartments of the cell environment assist their exposure to different perturbation scenarios and support the idea of simultaneous perturbation of both sides of the protein complex in dissociation process. As an example, we focus on dI₂dIII₁ complex of proton-translocating transhydrogenase (pdb code: 2oor (143)). The complex is found in membrane compartment of the bacteria or animal cells and proton transfer across the membranes is facilitated by conformational changes of transhydrogenase. PRS identifies residues A166, L214 and T220 of dI subunit and residues N103, P105 and M167 of dIII subunit to be effective in dissociation of the complex compartments (Figure 3.3d). dI and dIII protrude from the membrane while a third compartment dII spans the membrane. Thus, each part of the protein is involved in different compartments of the cell, making the protein susceptible to different perturbation scenarios. Residue T220 of dI is part of a mobile loop (residues 220-240) that becomes less mobile upon ligand binding due to surface closing of the protein (143). We note that in dI subdomain analysis, we have only considered residues 1-220 due to lack of structural information of the unbound form. Therefore, all the residues of the mobile loop have not been analyzed. Furthermore, M167 of dIII is in the neighboring site of H₂NADH and any perturbation on this site might alter the proton pump reactions due to changes made to interaction network of ligand binding region. Accordingly, we conclude that any perturbation on H₂NADH binding region in dIII subdomain may alter the structural features of dI subdomain through a remote communication.

3.1.1. Mechanics of the protein motifs explain the structural preference in PRS residues

In Table 3.3, we show that PRS top rated residues appear on coil and turn structures more often than other structures. In our analysis pool, coils and turns respectively account for 18 and 26% of all secondary and flexible structures. However, residues defined by PRS, residing on coil and turn structures contribute by 29% each. We attribute the enhanced frequency of coil residues observed among PRS residues to the mechanistic properties of structural motifs. We define molecular motifs that capture the essential components necessary to summarize dissociation of various protein complexes. To identify the molecular motifs, we combine elements of secondary structures in order to partition rigid and flexible parts of the protein fold. We treat various types of helices and strands as rigid bodies and coils and turns as flexible elements (loops). If isolated bridges appear one residue away from a β -strand

connected by a coil/turn residue, we combine them with the β -strand. However, if an isolated bridge is calculated to sit in the middle of coil/turn sequence, we count it as turn/coil residue. Accordingly, structural distribution of all the residues in analysis pool and residues defined by PRS change as reported in Table 3.4. Therefore, we transform the secondary structure fold of any protein to an assembly of rigid-loop-rigid (RLR) elements (Figure 3.4a). In Figure 3.4b, we display the distribution of residues on various RLR elements; we also divide the PRS top rated residues into groups of residues located on loops and residues located on rigid elements. PRS residues found on the loops tend to be located on longer flexible loops with short supporting rigid structures, while PRS residues on rigid elements tend to be connected to shorter stretched loops.

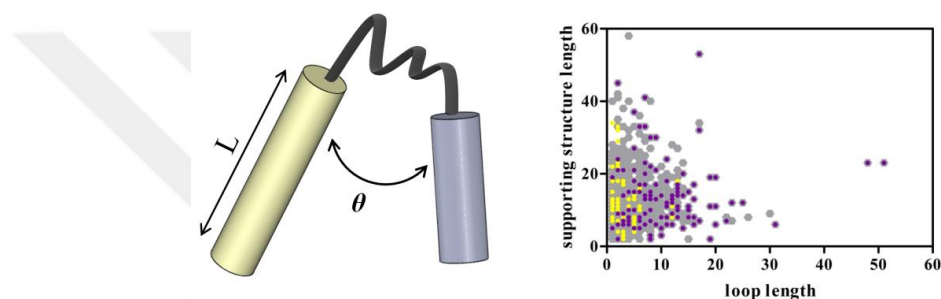


Figure 3.4 (a) schematic of the RLR motifs. Length of the rigid structure is L and the angle between two rigid structures connected by a flexible loop is θ . (b) Distribution of PRS residues located on rigid structures in yellow and PRS residues located on flexible loops in purple. Analysis pool is displayed in gray dots.

Table 3.4 Structural locations of PRS identified residues based on RLR motif analysis

	helix	strand	loop	total
PRS residues	40	37	110	187
PRS residues (%)	21	20	59	100.0
All residues	3040	2443	4643	10126
All residues (%)	30	24	46	100.0

3.2. An example of PRS at work: PRS specifies key regions in Subtilisin serine protease for both function and stability

Subtilisin-like proteases are a class of principle structural enzymes involved in many important metabolic processes (e.g., processing of endocrine (155)) and disease prevention (e.g., diabetes (156)) through their catalytic function. Therefore, a rational approach to stabilize and improve catalytic activity of subtilisin allows exercising biomedical benefits of

this enzyme. Furthermore, subtilisin has been the subject of practical catalysis in industry at low hydration (157). The catalytic activity of the enzyme decreases substantially when exposed to organic solvents compared to that of aqueous solutions. Thus, activation of subtilisin in harsh reaction media is being exploited for enantio-pure drug designs (157). Single residue mutations may induce large effects on the functioning of this enzyme. The question we address is how we may manipulate domain motions to follow a conformational transition between such different functional states of subtilisin. To effectively represent the mechanistic coupling between different distributions of protein structure at the atomic level, we apply PRS (46).

3.2.1. Conformational differences between the inhibited and free enzyme structures occur along two flexible loops

For the pair of initial and target forms of the enzyme present in enzyme/inhibitor complex and apo form, we perform STAMP structural alignment, implemented in VMD 1.9.1 MultiSeq plugin (141). We record the RMSD between the structures of the target forms with the initial structure. We find that the overall RMSD between the initial and target structures is approximately 1.1 Å. The two superimposed structures are shown in Figure 3.5. The active site residues of the catalytic triad are composed of His64 and Ser221 located at the N-terminus of adjacent α -helices and Asp32 situated at the C-terminal end of the central β -sheet (158). The superposition of different segments of the protein pairs yields higher RMSDs and can be as large as 2.3 Å; the largest deviations are in the two loops spanning Phe50 to Gly61 and Ser159 to Ile165, highlighted in black in Figure 3.5. Thus, while the protein secondary structural scaffold remains the same in the inhibited conformation, there are conformational changes involving the local internal motions of the unstructured loops. These regions also display the largest B-factors in the x-ray structure of the unbound form, and include the two cleavage sites susceptible to proteolysis (159).

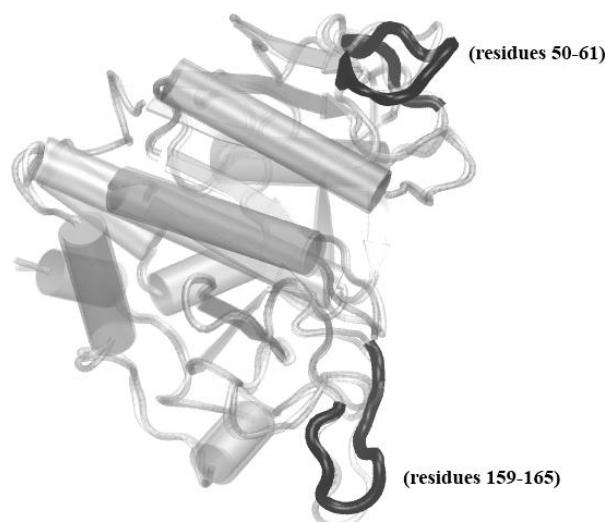


Figure 3.5 Subtilisin alignment in two states, 1OYV chain B representing the inhibited structure in light gray, and 1SBC for the apo form in dark gray are shown in cartoon representation. The overall structures are well-aligned while the two loops spanning Phe50-Gly61 and Ser159-Ile165 which display higher RMSD are shown in black.

3.2.2. PRS analysis of the enzyme reveals single residues that control the conformational change

Protein dynamics may significantly influence enzyme catalysis and specificity. The coupling between reaction kinetics and protein motions may mechanically stress the enzyme/substrate binding or disrupt stabilization (160). Whether there is a remote communication propagating through the protein's amino acid interaction network to the catalytic region is our focal point. We study the conformational change upon inhibitor binding of subtilisin serine protease. For the enzyme/inhibitor complex, we take two approaches: (1) we build the **C** matrix of the enzyme by removing the inhibitor. Therefore, we employ the inhibitor bound conformation of subtilisin, but in the absence of the inhibitor and then with PRS we scan the protein by inserting random forces on all residues sequentially. For each residue, we then record the Pearson correlation between the calculated displacements with those from the experimental conformational change (eq. 2.18). (2) We construct the **C** matrix of the enzyme/inhibitor complex together to capture the conformational change in the enzyme, initiated by the perturbation of the residues in the inhibitor. The PRS map for case (1) is displayed in Figure 3.6. Examination of this figure will clarify how we identify the

residues responsible for controlling the conformational change. By noting that using PRS we apply forces on each residue in 500 random directions, we emphasize the scan collects data from 500 N response vectors for further analysis. In this map, we display for each residue the result that gives the best correspondence out of the 500 perturbation directions. Thus, for each perturbed residue labelled on the x-axis, all other residues are affected, and their displacements are shown on the y-axis. The lower panel records the corresponding Pearson correlation. As a result, those residues that give the largest correlation in any one of the 500 perturbation directions are classified as capable of controlling the conformational change.

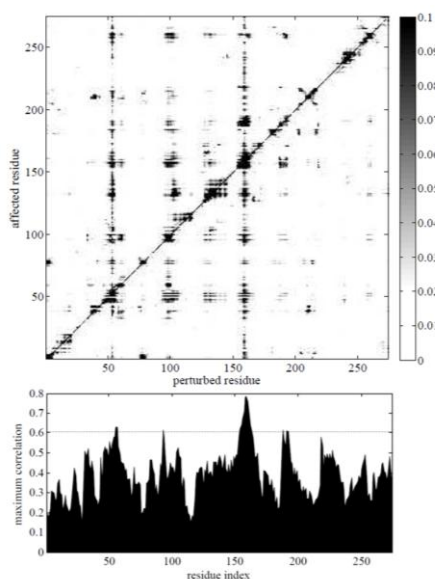


Figure 3.6 PRS map of the interconversion of TI-II/subtilisin complex to apo subtilisin. The displacements of the affected residues are displayed for the case which gives the best correlation for a given perturbed residue. Note that the displacements are on a relative scale, since PRS relies on linear response theory; i.e. if forces in a selected direction are modified by a factor r , then all displacements are also modified by the same factor r . Lower panel displays the relevant maximum Pearson correlation value for all residues. We select the residues that give the largest correlations as candidates controlling the conformational change.

In Table 3.5, we list the residues that are accentuated by the PRS analysis. These are the single residues whose perturbation selects the apo conformation over the inhibited one. We observe that the residues are clustered mainly on three regions on the enzyme: (i) those residing on the flexible loop spanning residues 50-61 in figure 2, (ii) those residing on or near the flexible loop spanning residues 159-165 in figure 2, and (iii) those in region spanning residues 192-194. We note that both the MD derived $\mathbf{C}$ matrix and the ENM derived ones

point to the same regions in both the absence and the presence of the inhibitor. We further note that the cutoff distance noted in Table 3.5 for the ENM analysis is optimized to maximize the Pearson correlations although the regions pointed by different cutoff distances spanning $r_c = 8 - 20 \text{ \AA}$ are the same. A detailed discussion on the invariance of the elastic network model findings on the selection of cutoff distances may be found in ref. (161).

Table 3.5 Summary of PRS results for $r_{\text{opt}} = 8 \text{ \AA}$. Hot residues are the ten residues with highest Pearson correlation values in comparison to whole sequence

Label	Initial→target structure	PDB entries	Maximum correlation	Hot residues
1 (MD)	Inhibited → Apo	1OYV:B→1SBC	0.74	57, 156-162, 164, 193
1 (ENM, $r_c = 9 \text{ \AA}$)	Inhibited → Apo	1OYV:B→1SBC	0.77	54-55, 57-58, 155-164, 193
2 (MD)	Inhibitor bound → Apo	1OYV:BI→1SBC	0.68	158 – 163, 192, 19*
2 (ENM, $r_c = 18 \text{ \AA}$)	Inhibitor bound → Apo	1OYV:BI→1SBC	0.68	53-54, 62, 158–162, 194, 12-13*

3.2.3. Biological relevance of the residues specified by PRS method

The three regions (i) – (iii) accentuated by the PRS analysis of the previous subsection are displayed in Figure 3.7 on the enzyme – inhibitor complex structure. The stretch of residues with high correlations is shown in black, while the hot residues that have the highest correlation are further displayed in ball-and-stick representation. PRS identifies that hot residues do not always reside on Ca^{+2} /substrate binding regions, but rather local motions originating from their perturbations propagate to distal sites and cause stabilization of the uninhibited conformation.

The stretch of residues in (i) residing on the flexible loop (Figure 3.5, top loop) spanning residue indices 50 – 61 are known to significantly affect subtilisin function. Substitution of Ca^{+2} binding residues on the analog loop from the thermophilic thermitase into subtilisin has caused not only additional calcium binding affinity, but also significant pH dependent stabilization of the enzyme (159).

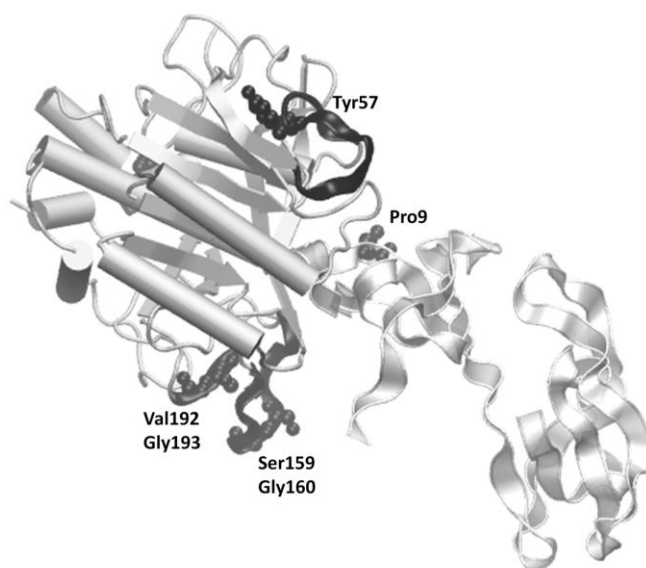


Figure 3.7 Residues specified by PRS as being effective in the transition of inhibitor bound subtilisin to apo subtilisin are displayed on the TI-II and subtilisin complex. The pronounced regions are displayed as black cartoons, while the residues with the largest correlations in a given region are further emphasized in ball-and-stick representation and labelled. TI-II is represented as gray colored transparent ribbon and subtilisin as silver colored cartoon.

Also accentuated by PRS analysis are the group of residues (ii) that reside on or near the second flexible loop (Figure 3.5, bottom loop). Single residue mutations in this region have been shown to significantly affect stability and/or activity. For example, the S161C mutant brings increased stability to subtilisin, while the G166R mutant is both more stable and displays more activity (162). Furthermore, loop mutants 158, 159, 164 and 165 have led to significant loss of stability in subtilisin (159). Amongst the many mutations introduced in (163), the E156S mutation has the largest effect on substrate binding while displaying a minor effect on its stability. Conformational flexibility and protein dynamics play an important role in functional properties; one desired functionality of subtilisin as an industrial enzyme might be the ability to function in non-natural environments such as polar organic solvents. Chen and Arnold have employed random mutagenesis and screening to find amino acid substitutions that recover catalytic activity of subtilisin in polar organic solvents (164). Among 10 mutations of subtilisin described in their work, Glu156 is one of the hot residues identified by PRS, which exhibits enhanced activity in DMF. This improvement includes decrease in Michaelis constant K_m and essentially same catalytic activity in 0% and 20% DMF. Although E156N variant does not display activity enhancement in 20% DMF, it

may be beneficial at higher DMF concentrations. The reason for Glu156 emerging in screening process is related to its slightly stabilizing effect (164).

Also in group (ii) residues, we find that Asn155 is highly conserved and as calculated by Consurf server (69). Critical function of Asn155 is mainly attributed to electrostatic effects. Comparison of the transition state models of native subtilisin and mutated one confirms that Asn155 may selectively stabilize transition state for catalysis by hydrogen bond formation between amide side group of Asn155 and carbonyl oxygen of a peptide substrate. Such a major factor in lowering the free energy of activation in enzyme/substrate complex would be restrained in the alanine mutant. To test this hypothesis, leucine has been introduced to the Asn155 position, which led to a reduction in the catalytic rate constant by a factor of 200-300 (165). Free energy perturbation calculations were performed on the free energy of binding/activation by N155A mutant (166). Differences in free energy of binding is small (0.1 ± 0.8 kcal/mol), but free energy change in catalysis is high enough (3.4 ± 1.1 kcal/mol) for such a site directed variant.

Residues identified by PRS in group (iii) are different from those in (i) and (ii) in that they reside in a location that displays no conformational difference between the apo and inhibited forms of subtilisin. Nevertheless, experimental studies have shown the importance of this region in directed mutagenesis studies. Out of the 10 random mutations occurring in conversion of subtilisin into an enzyme functionally equivalent to its thermophilic homolog thermitase, two are A192T and S194P. The former appears to be neutral, while the latter is very effective in both thermostability and activity of the enzyme.

Finally, we note that several residues on the inhibitor are also capable of inducing the observed conformational change in the enzyme. These include positions 2, 3 that are in direct contact with the enzyme as well as residue 9 that is a second shell residue, at a relatively remote location. The latter is also shown in Figure 3.7.

CHAPTER 4. MODELING THE EQUILIBRIUM AND DYNAMICAL PROPERTIES OF IRON TRANSPORT PROTEINS: THE HUMAN SERUM TRANSFERRIN

4.1 Detailed molecular dynamics simulations of human transferrin provide insights on iron release dynamics at serum and endosomal pH

It is hypothesized that hTf has specific features which differ at physiological and endosomal pH (see e.g. (71)). The structural difference between the two lobes of hTf account for the nuances in their biological functionalities as well as their thermodynamic and kinetic properties in different environments (59). Dissimilarities arise in part because of "second shell" residues. These do not contribute to iron binding, but participate in a complex hydrogen-bonding network with iron coordinating residues. In the N-lobe of hTf, second shell residues include of K206 and K296, forming the so-called "dilysine trigger," positioned in N₂ and N₁ subdomains, respectively (124). In the C-lobe, the dilysine trigger is replaced by a triad of residues consisting of K534 in the C₂-subdomain, R632 and D634 in the C₁-subdomain (167). It has been suggested that the hydrogen bond network created by the second shell residues at physiological pH is disrupted upon protonation processes occurring at the endosomal pH, assisting domain opening (124).

Due to the significance of iron metabolism, biochemistry of hTf has been the subject of many studies. In particular, binding and release of iron and other metals as well as a variety of synergistic anions in the presence or absence of its receptor have been quantified (122, 168-170). Furthermore, many environmental factors affect the distribution of conformational substates of hTf. These factors include pH, presence of chelators, temperature, salt concentration and lobe-lobe interactions, with current active research being focused on the effect of these factors on mutated and wild type hTf (71, 76, 167-169, 171, 172). At pH 7.4, four major populations (namely, diferric, two monoferric and apo hTf) have been suggested, based-on the iron occupancy of the deep cleft formed between the subdomains in each lobe of hTf (173). Lobe-lobe interactions and the ability to capture iron by each lobe have been suggested to play a major role in the dynamics of interconversion between conformational substates (171). Due to the difficulties in the experimental observation and quantification of hTf dynamics, computational tools including all-atom and biased MD simulations along with coarse-grained approaches might help clarify some of these issues. However, the total process

of conformational switching at physiological pH from the holo to the apo form is not expected to be observed on the time scales reachable by conventional MD simulations; the former usually occurs on micro to millisecond time scales while the latter is on time scales of hundreds of nanoseconds for a large protein such as hTf. Molecular level insights into structure and dynamics of non-heme iron binding proteins like transferrin have been gained by making use of a range of computational tools. Three-dimensional structure of transferrin was initially resolved in partial form by X-ray crystallography where only the N-lobe fold was characterized (174, 175). Dynamics of the iron free form of the N-lobe in open and closed conformations was studied by ~2 ns classical molecular dynamics (MD) simulation (67). Therein pKa calculations show that pKa values of iron coordinating residues shifts to that of lower pH upon iron binding. Positive charge of iron will favor negative charge state of D63, Y95 and Y188 while H249 will be in its neutral state. In particular, K206, which is built in to dilysine trigger, displays a very low pKa of less than 7.5 in the closed form of hTf. K206 interacts with K296, which in turn shows a large pKa. Thus, it is predicted that protonation of K206 in lower pH of endosomal region and repulsion between the two lysines is a necessary step to open the subdomains of the N-lobe and release iron. In addition to K206, Y95, Y188, H249 and carbonate anion need to be protonated to assist iron release (67). Nevertheless, series of adequately sampled MD simulations (~ 60 ns) of various forms of holo conformation at physiological and endosomal pH show that protonation of dilysine bridge is not enough to trigger the domain opening and assist iron release (176). In the holo N-lobe conformation with negative charge state of coordinating tyrosines, 60 ns of MD simulation displays a stable RMSD profile and the repulsion between the two lysines on the N-lobe has been shown to be locally accommodated without the need for domain opening (67, 176, 177). It is concluded that while dilysine bridge effects iron release by interacting with carbonate anion, protonation of one the lysines does not destabilize the closed state of hTf to trigger the domain opening. In fact, the protonation of coordinating tyrosine is the essential factor in conformational changes before the release process (176, 177). Rinaldo and Field propose that proton transfer from K296 to Y188 is the likely scenario to initiate iron release (67). Absence of dilysine bridge in the C-lobe and other transferrin family members such as Lactoferrin in addition to various mutants of dilysine bridge which release iron at lower pH than the N-lobe support the trivial effect of dilysine bridge on domain opening (178-183). Further MD simulations on the apo form of hTf and principal component analysis show that the apo hTf goes through hinge bending/twisting motions at both physiological and endosomal pH (67, 176, 177). This

property of hTf enhances its ability to adopt various ligands with different sizes (67). Moreover, Molecular mechanics-Poisson-Boltzmann surface area free energy calculations explain that at the endosomal pH the N-lobe of apo form has lower internal energy due to lack of nonbonding interactions between N1 and N2 subdomains and more favorable entropic contributions due the flexibility of this form. Accordingly, at 5.6 pH apo form of isolated N-lobe is more stable than the respective holo form (67). This accounts for the strong interaction between apo hTf and TfR in intercellular environment while hTf is recycle back to the cell surface. In another study, computational measurements of pKa values of the Lys residues in apo hTf has been used as a method to predict the most susceptible residues to glycation (184). Among all Lys residues, K206 and K534 are the residues with the lowest pKa value and consistent with the glycation order profile determined experimentally. MD simulation comparison of glycated and non-glycated structures shows that glycation brings about large conformational changes to the apo structure. Glycation of K206 and K534 introduces a bulky glucose to the binding cleft and disrupts the interactions within second shell and coordinating residues in the N-lobe and C-lobe, respectively (184). Thus, iron and carbonate access to coordinating residues are blocked sterically and iron binding capacity is lowered by preventing the conformational transition from apo to holo structure. At glucose levels higher than 10 mM, which is observed in diabetic patients, glycation is likely to happen and resultant structural impairments in hTf will increase the risk of iron toxicity.

Allosterically induced conformational transitions upon non-synergistic anion binding have been introduced as the so-called KInetically Significant Anion Binding or KISAB sites in hTf where the location of such a binding site remains debated (76). A molecular modeling strategy was used to identify a possible KISAB site on hTf surface using only the N-lobe structure of hTf (72). Therein, possible anion binding sites were identified based on a heuristic search, calculation of electrostatic potential energies and grid-based strategy. Various anion binding energies and significant residues were identified based on molecular dynamics and molecular mechanics analysis.

hTf is saturated with iron only 30% of its residence time in blood stream (185) and therefore is able to bind many transient and non-transient metals other than iron, such as Bi, V, Ru, Cr, Al, Cd and Ga, which have possible therapeutic benefits (186). Binding of toxic metals facilitate detoxification of various organs while damaging hTf ability to bind iron. Molecular mechanism of Cd binding to hTf was studied through CdCl₂ docking in partially

open full structure of hTf (187). The identified CdCl_2 docking region does not interrupt iron binding sites on either lobes but it is predicted that the residues involved in CdCl_2 binding might contribute to premature iron release from hTf. In high binding affinity sites, no iron release was observed in a stable 1 ns MD simulation as CdCl_2 binding does not interfere with iron binding region. However, upon CdCl_2 binding to low binding affinity sites, interaction with iron binding region concludes in iron release. hTf is also able to pick up 60% of Al in serum (188). 60 ns MD simulation of Al loaded N-lobe of closed hTf at physiological and endosomal pH show that global conformational changes of hTf is similar to iron bound hTf (176). However, local changes especially in the binding site are different for Al binding mode compared to iron binding mode, which results in different release mechanisms of the metals. hTf is also an attractive molecule for many targeted therapies. For example, particular cancerous cells need iron for cell growth and therefore develop a tenfold increase in TfR number on the cell surface (189). Consequentially, hTf might be useful as a vehicle to carry anticancer drugs to such cell. Thr457 was identified as the binding site of a chemotherapy drug, cisplatin, to transferrin (189). A homology model of the C-lobe was generated as full structure of hTf was not available at the time of the experiments and cisplatin was docked to the structure. Model of the cisplatin modified hTf shows that cisplatin binding reduced iron binding affinity of hTf in the C-lobe while being ineffective to iron binding properties of the N-lobe. Thus, a monoferric htf complexed with cisplatin might be able to target cancer cells.

Due to the difficulties in the experimental observation and quantification of hTf dynamics, all-atom MD simulations might help clarify some of these issues. However, the total process of conformational switching at physiological pH from the holo to the apo form is not expected to be observed on the time scales reachable by conventional MD simulations; the former usually occurs on micro to millisecond time scales while the latter is on time scales of hundreds of nanoseconds for a large protein such as hTf. We therefore define perturbation scenarios under which the conformations sampled by hTf may be triggered (Table 2.3).

4.1.1. Domain opening is coupled to the protonation states of iron coordinating tyrosines

In the holo form of hTf, it has been proposed that the iron coordinating Tyr residues have downshifted pK_a values such that they are negatively charged when Fe^{+3} is bound (67, 122, 190), also corroborated by various pK_a calculations, particularly for Y95, Y188, and

Y517 (Table 2.2). Therefore, we first run simulations to test that the systems are indeed stable when the protonation states of Tyr residues are changed (see Figure 4.1 for the RMSD profiles over the course of 60 ns). We find that in H^{**} system, where all the Tyr residues in the two Fe^{+3} binding sites are deprotonated, the initial holo structure is stable with an overall, N-lobe, and C-lobe RMSD of 2.7 ± 0.3 , 2.1 ± 0.1 and 2.4 ± 0.3 Å, respectively, averaged over the last 50 ns. In the H^* system, where Y188/Y517 are deprotonated and Y95/Y426 are protonated, we observe similar behavior to H^{**} in both lobes. The system is stable around the initial conformation with an overall, N-lobe, and C-lobe RMSD of 2.8 ± 0.3 , 2.5 ± 0.2 , and 1.9 ± 0.1 Å, respectively, averaged over the last 50 ns. The iron is still engaged in the binding region. As listed in Table 4.1, the liganding residue – Fe^{+3} distance is within at most 2 Å of the initial, energy minimized structure in both the H^* and H^{**} simulations. This is contrary to the observations in the H run, which we discuss next by assessing how the systems respond to neutralizing iron coordinating Tyr residues in holo hTf.

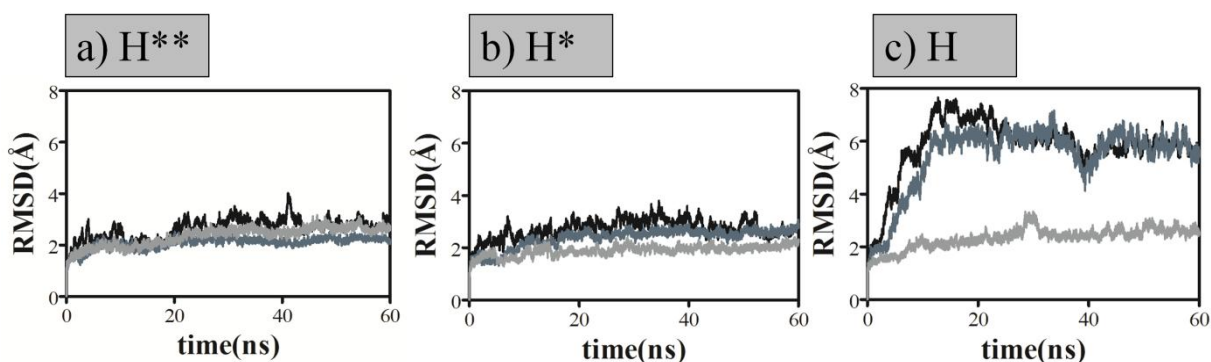


Figure 4.1 RMSD curves for H^{**} , H^* and H systems, labeled as defined in Table 2.3. The overall RMSD (black) as well as that of the N- and the C-lobes are reported (blue and gray, respectively). When both of the iron coordinating Tyr residues in the lobes (a) or only from each lobe (b) are negatively charged, the lobes are sealed tight. The N-lobe fluctuations are slightly more pronounced in (b) than in (a). The overall RMSD of the protein, as well as that of either lobe is below 3 Å throughout the time window of observation. (c) When all Tyr are in their standard protonation states, the N-lobe begins to open up within 10 ns while the C-lobe remains closed.

While it is expected that the Tyr residues in the iron binding site would be neutral in the apo form, their ionization states in the holo form are debated and the pK_a calculations also provide different results for these residues. E.g. PROPKA assigns pK_a values greater than 12.0 to all these Tyr residues in the apo form, compared to 6.0 to Y188 and 4.1 to Y517 in holo form (the holo values are listed in Table 2.2). The RMSD profiles in a 200 ns run of the

H system where all four iron coordinating Tyr residues are neutralized is displayed in Figure 4.2a, and the first 60 ns of this run is also shown in Figure 4.1 for comparison with the H^{**} and H^* runs. Large changes are observed in the whole protein almost instantly, leading to RMSD from the initial structure of up to 6.7 ± 0.5 Å. In the first 60 ns, the curves indicate the N-lobe is more mobile than the C-lobe. In fact, it readily moves towards a more open conformer within the first 10 ns. Thus, the initial conformational change mostly involves motions in the N-lobe rather than global rearrangements, despite the fact that coordinating Tyr residues are neutralized in both lobes. After 100 ns, the C-lobe also displays larger deviations from the initial structure, moving towards a more open state. By the end of 200 ns, all four Tyr are ca. 6 Å away from the Fe^{+3} compared to the ca. 2 Å value in the initial structure (Table 4.1). The Tyr residues in both lobes are the first ones to leave the binding region. The coordinating D63 and D392 in both lobes remain in contact with the iron in the simulations, in conformity with mutagenesis studies (191, 192) which have shown that if Asp residue vacates the coordination shell, the iron will have enough freedom to leave the binding cleft. In addition, the coordinating His on the C-lobe (H585) also remains in contact, while that on the N-lobe (H249) has moved away (Table 4.1). We conclude that the protonation states of the key Tyr residues are responsible for triggering the lobe opening, starting with the N-lobe; this is the first of the multi-step process that eventually leads to iron dissociation.

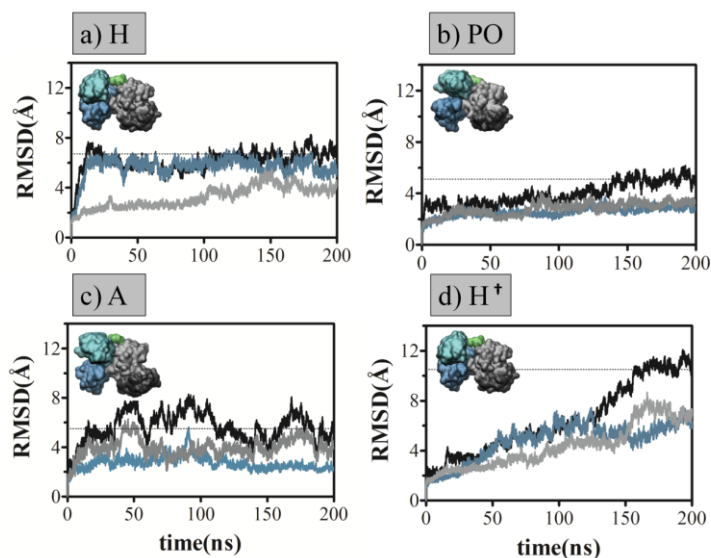


Figure 4.2 RMSD curves for H, PO and A systems (respectively, holo, partially-open and apo hTf at physiological pH) and H^+ system (holo hTf at endosomal pH). The overall RMSD (black) as well as that of the N- and the C-lobes (blue and gray, respectively) are reported. Insets display surface representation of crystal structure for each system. Horizontal dotted

lines represent the overall RMSD averaged over last 50 ns of each simulation. (a) In H, N-lobe displays large rearrangements from the beginning of the MD simulation. After 100 ns, these are communicated to the C-lobe. (b) In PO, both the N- and C-lobe stays near the initial conformer. The overall structure displays larger RMSD values due to relative placement of the lobes, communicated by the linker. (c) In A, the C-lobe is more mobile than the N-lobe throughout. It oscillates around the equilibrium state in a hinge like motion (see text for details). (d) In H[†] where the lower pH of the endosome is mimicked, large rearrangements occur simultaneously in both lobes.

Table 4.1 Distances from coordinating residues¹ to iron in the beginning²/end³ of simulations.

Liganding Residue	H**	H*	H	PO	H ^{Fe+N} /H ^{Fe+C}	H _N /H _C	H [†]
Y95	1.9/1.9	2.2/4.2	2.2/6.0	2.2/6.0	2.2/4.8	2.2/10.7	2.2/9.5
Y188	1.9/2.0	1.9/1.9	2.3/6.4	2.2/8.0	2.8/5.9	2.8/10.6	4.0/10.8
H249	1.9/3.7	3.8/3.7	2.2/4.5	10.1/4.3	4.2/5.2	4.2/5.5	4.3/4.2
D63	3.4/2.1	1.9/2.0	1.9/2.1	6.0/2.1	2.0/2.0	1.9/2.0	2.0/2.1
Y426	1.9/1.9	3.4/3.3	2.2/6.0	2.2/8.7	2.2/8.8	2.2/7.6	2.2/5.1
Y517	1.9/2.0	1.9/2.1	2.2/6.1	2.3/9.0	2.2/6.5	2.2/8.3	2.2/6.9
H585	3.3/4.3	3.8/4.8	1.9/2.3	2.2/5.4	4.9/4.2	4.9/5.9	4.6/4.1
D392	1.9/2.0	1.9/2.0	2.2/2.1	1.9/4.3	1.9/4.2	2.0/2.0	2.0/2.1

¹ measured from carboxylate oxygen of D, imidazole nitrogen of H and phenolate oxygen of Y.

² obtained right after energy minimization of the system.

³ that of the snapshot in the final 50 ns of the trajectory for which average distance between Fe+3 and its four coordinating residues is minimum.

4.1.2. The partially open state is a stable intermediate of hTf

To check if the new conformation attained by H is similar to other experimentally determined N- and/or C-lobe structures, we superimpose the trajectories of the lobes on the PO and apo form x-ray structures (Figure 4.3). We find that the N-lobe is in fact quickly driven towards a conformation similar to that in the apo form (2.9 Å RMSD) or the PO form (2.4 Å RMSD) by the triggering event. The C-lobe, on the other hand, is more similar to that in the holo form structure, despite a larger RMSD of 4.4 Å by the end of 200 ns. Overall, the final structure attained after the triggering event is more similar to the PO structure, with an open N-lobe and a closed C-lobe. We note that no additional events are observed when this trajectory is prolonged to 500 ns (data not shown). In addition, the initial triggering event on the N-lobe is observed within 20 ns in an independent, 150 ns long run carried out on the same system. Thus, our results corroborate experimental studies pointing to the possibility of a stable intermediate between apo and holo conformations of hTf (193, 194).

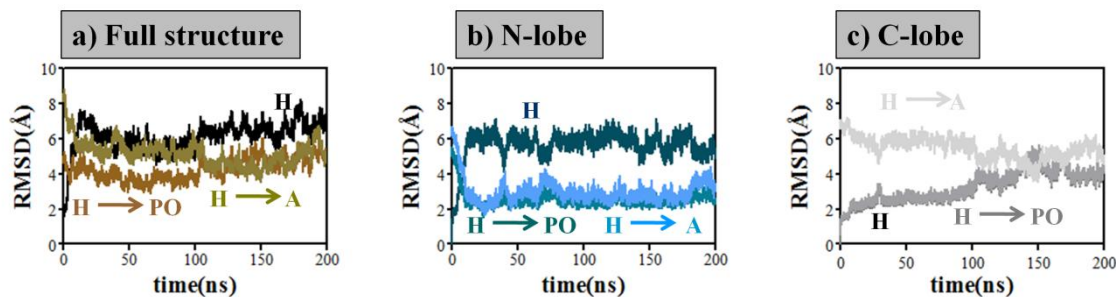


Figure 4.3 Time evolution of the RMSD of H system and its lobes from different crystal structures. (a) The original overall RMSD curve is shown in black (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (brown and green, respectively). Once the initial conformational change is completed within 10 ns, the newly attained conformation displays lower RMSD from the PO x-ray structure. (b) The original N-lobe RMSD curve is shown in dark blue (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (shown as lighter shades of blue). N-lobe of H system displays excellent resemblance to the N-lobe of partially open structure. (c) The original C-lobe RMSD curve is shown in dark gray (a replica of that in Figure 4.2a), while the same trajectory is also superimposed on the PO and A structures (shown as lighter shades of gray). Since the RMSD between the x-ray structures for the C-lobe of H and PO structures is very low (0.6 Å, Table 1), the curve for H→PO comparison is nearly identical to the original curve. The difference from the x-ray structure of A diminishes throughout the trajectory.

Figure 4.2b provides information on the dynamics of the PO system. It fluctuates around equilibrium up to 100-120 ns, and then exhibits a small jump in RMSD. In the new equilibrated state, coordination of iron in the N-lobe binding cleft changes to what we observe by the end of 200 ns simulation of H. H249 and D63, which are 10.1 and 6.0 Å away from the Fe^{+3} in the initial structure, respectively, move to a distance similar to that observed in the end of H simulations (Table 4.1). In accordance with this conformational change, Y95 and Y188 move away to 6.0 and 8.0, respectively. In the case of the C-lobe of PO system, all the iron coordinating residues get more distant from Fe^{+3} due to triggering by tyrosine protonation, resulting in opening of the C-lobe that is originally closed in PO. Most of the change that leads to the 5.1 ± 0.4 Å RMSD in the PO system (Figure 4.2b) is due to the mobility of the linker bridging the two lobes. Thus, within 200 ns, the H structure evolves to a conformation that is also sampled by the equilibrated PO structure and we conclude that the experimentally observed PO state is a stable intermediate of hTf once N-lobe opening is triggered.

Finally, we study the dynamics of apo hTf (Figure 4.2c). We find that, in the absence of iron, the N-lobe is now more restricted in its motions compared to the C-lobe. Moreover,

the overall RMSD of the protein displays large deviations from the initial structure, exceeding 5.5 ± 0.9 Å. However, these fluctuations do not signify sampling of distinct states separated by a well-defined energy barrier. To quantify this observation, we calculate the RMSD between the average structures of the four consecutive 50 ns portions from the A run. The values are as small as 1.4 Å (between the average structures in the 100-150 and 150-200 ns portions) and as large as 2.8 Å (between the 50-100 and 150-200 ns portions). Note that these values are on par with the resolution of the crystal structure of the apo form (2.7 Å). We also monitor the eigenvectors belonging to the slowest mode of motion, obtained from the principle component analysis of the 50 ns subsets of the 200 ns long trajectory. We find that the overlap between the dominant eigenvector of each chunk is larger than 0.5 in all cases, for the overall protein, and for only the C-lobe. As a sample, the dominant eigenvector of the C-lobe is displayed in Figure 4.4 for the four 50 ns trajectory pieces, signifying the same hinge motion around the iron binding region in the absence of the iron. Thus, despite the large RMSD recorded during the MD simulations, the apo form is a loose structure displaying large fluctuations around its experimentally determined state and does not make a conformational jump to another.

According to the RMSD curves, we hypothesize that at serum pH, upon triggering lobe opening, the conformation travels through the partially open form before adopting the open conformation. To further assess the details of this conformational change, we display in Figure 4.6, the locations of conformational space sampled during the MD simulation of different systems projected on the reduced degrees of freedom. We use reduced degrees of freedom to distinguish between the different conformations of hTf by projecting the $3N$ -dimensional conformational space into three visually tractable ones. These are bending angles within the N- and the C-lobes (θ_N and θ_C) and a torsional angle defining the relative positioning of the two lobes (φ). The former is defined as the bending angle between the center of mass of each subdomain in every lobe and a hinge residue; we choose C_α atom of Y95 and Y426 as the hinge residues in the respective lobes, which also bind iron. The torsional angle is defined by four points: the hinge residues in the N-lobe (C_α atom of Y95),

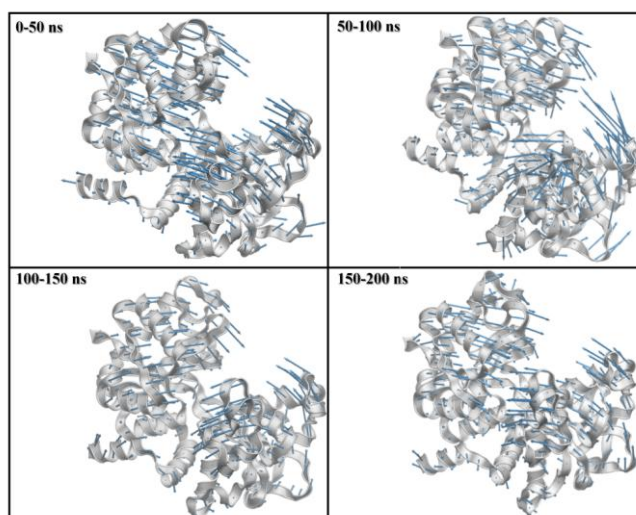


Figure 4.4 Eigenvectors displaying the dominant mode of motion in the C-lobe of apo the structure, obtained from four 50 ns–long portions of the A simulation, superposed on the average structure of each portion. While the RMSD from the crystal structure gets as large as 8 Å in these simulations, the average structures are within 2.8 Å of each other. Simulations displays the same large hinge bending motions around the average structure as shown, and does not exhibit a conformational shift.

two end points of the peptide linker (C_α atoms of residues C331 and C339), and the hinge residue in the C-lobe (C_α atom of Y426). These points and the reduced degrees of freedom are schematically shown in Figure 4.5.

In the Figure 4.6, regions sampled by θ_N - θ_C are displayed on the leftmost column as contour maps to show the more populated areas of these geometric parameters. The position of the initial x-ray structures are marked by the red dots. The center column displays the distribution of the relative positioning of the N- and the C-lobes, projected on the angle φ . Also displayed on the right side of the figure is the superposition of the full structure snapshots of corresponding systems sampled during the 200 ns at 4 ns intervals.

Comparing the distributions of the reduced degrees of freedom for the H and PO systems (Figure 4.6), we find that the N-lobe of both systems samples bending angles in the range $\theta_N = 120 \pm 10^\circ$. Interestingly, the C-lobe of the PO samples a wider range than the N-lobe, although this behavior is not directly evident from the RMSD curves in Figure 4.2. The most populated region by both is nevertheless the same, centered near $\theta_C = 100^\circ$. The relative orientation of the N- and the C-lobes for these systems, quantified by φ are nearly the same by the end of the simulations. The second hump in the φ distribution of H appears later in the simulations, once the conformational change in the C-lobe is also triggered at around 100 ns

time point. The final distribution of φ is centered on $-9\pm 8^\circ$ in both H and PO systems (collected from 180 - 200 ns part of the trajectory). Thus, there is substantial overlap between the regions populating the N-lobe of H and PO systems, corroborating our previous observation that the PO state is a stable intermediate of the H state at serum pH.

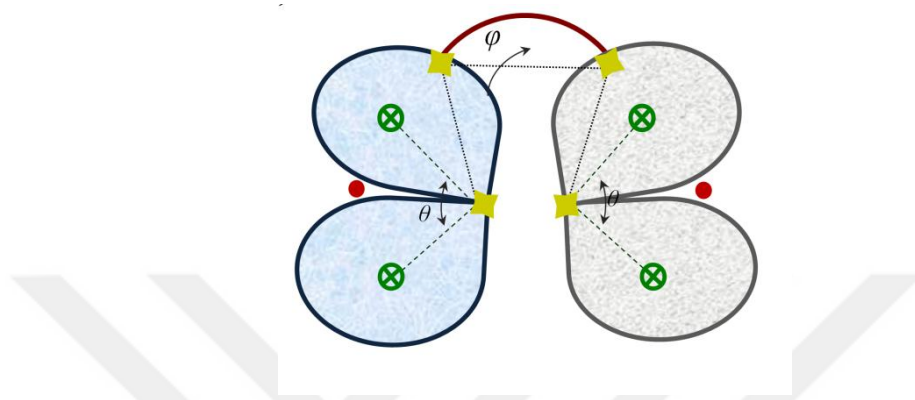


Figure 4.5 Schematics of hTf displaying the color code used throughout the text (blue for N-lobe, green for linker, gray for C-lobe; pink circles represent Fe^{+3} ions), the points used to define the reduced degrees of freedom (shown by $\otimes$) are the centers of mass of the N- and C-lobes, the beginning and end points of the linker, and the hinge residues between the subdomains of each domain, Y95 and Y426. We define the lobe bending angles θ_N and θ_C , measured by the angle between the two centers of mass and the hinge residue between each subdomain of the respective N- and C-lobes. The lobe torsional angle φ , is defined by the four points shown. These are the major degrees of freedom on which we project all trajectories to monitor the sampled conformations.

The above-mentioned results provide a good description of how the H system is driven towards the regions sampled by the PO conformer, once the change is triggered by the initial conditions described earlier. Furthermore, comparison of the θ_N - θ_C maps and φ distributions of the H and A (Figure 4.6) portrays the fact that, in spite of the opening motions of the N-lobe and large mobility of the C-lobe, H has more propensity toward PO than A. The bending angle in the C-lobe of A, θ_C , overlaps significantly with that of either H and PO, while in the N-lobe, θ_N displays a small but discernible shift by 10° centered on the range $\theta_N = 110\pm 10^\circ$. More significantly, the relative positioning of the two lobes, quantified by φ , is displaced by ca. 50° in A compared to either H or PO. Thus, the next step in the conformational switch

requires additional large motion of the two lobes involving large changes in θ and φ .

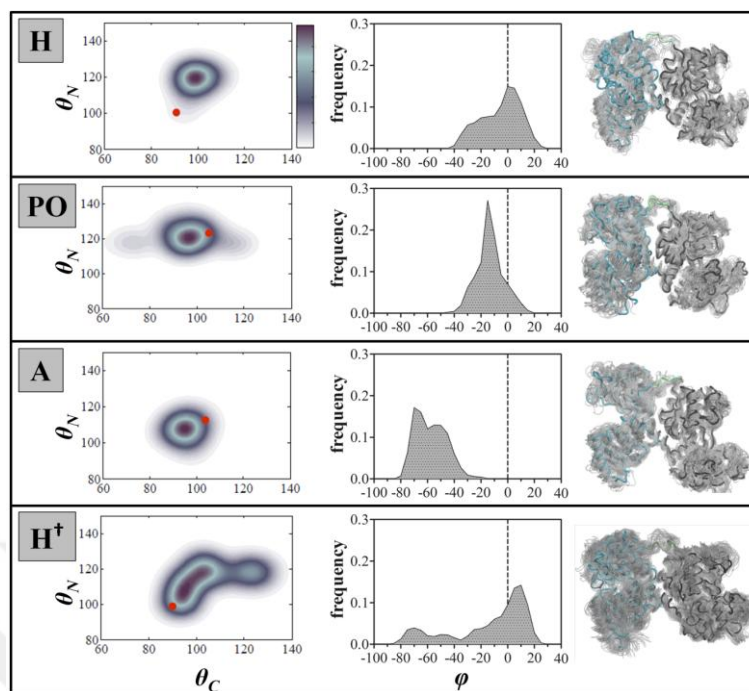


Figure 4.6 Conformations sampled by H, PO, A, H⁺ displayed by projecting on the reduced degrees of freedom, defined in Figure 4.5. The first column shows the joint probability distribution of the θ_N, θ_C intra-lobe bending angles, therefore the differences in their sampled regions. The initial position is shown by the red dot: While the inequality is expected for the PO structure in the x-ray structure, even the holo and the apo forms are positioned at slightly different values; the N-lobe is slightly more open in both cases. The second column shows the distribution of the torsional angle (φ) quantifying the relative positioning of the two lobes. Initially they are at -15, -30, -40, and -5°, respectively for H, PO, A, H⁺. The relative position of the iron loaded forms at serum pH is ca. 50° different than the iron free form. The endosomal pH system is more labile and samples both positions frequented by the apo and holo systems at higher pH. Shown on the right are the superposed structures of the corresponding systems, 50 snapshots spaced at 4 ns intervals.

4.1.3. The dilysine trigger is not strong enough to destabilize the Fe⁺³ binding cleft and induce domain opening

It has been proposed that either K206 or K296 are deprotonated in holo hTf and therefore are able to tightly interact via a hydrogen bond to lock the N₁-N₂ subdomains (124). Therein, it was suggested that the protonation of these lysines would lead to strong enough repulsions in the second shell to force open the iron binding cleft. These residues were therefore called “dilysine trigger.” On the other hand, previous MD simulations of N-lobe hTf suggested that protonation of these lysines was not the sole reason of domain opening (67). However, therein the simulations were carried out on closed form of the apo protein (i.e. iron

was not included in the simulations) and was of length 1.5 ns. Our MD simulations provide a testbed for this discussion since we take the standard protonation state of these two lysines (*i.e.* they are charged and are expected to repel each other; therefore, they might trigger a loosened iron binding cleft).

In H, the two lysines are initially 4.0 Å apart and their separation (measured from the ϵ -amino group) grows up to 19.0 Å, along with N-lobe opening. If a strong repulsion between them were at the route of the trigger, similar observations should be made for the H* and H** systems which have the same Lys protonation states. However, our observation is contrary to this assumption: In H* K206-K296 distance changes from 4.4 Å up to 6.5 Å and in H** from 3.9 Å to 8.4 Å. In neither of these runs does the N-lobe open (see Table 4.1 for the final iron coordination distances in each of these system). Thus, we find that opening of the N-lobe steered by the protonation states of the tyrosine residues is the main reason for the two lysines to grow apart and open the cleft.

The unusually low pK_a of Y188 (see Table 2.2 and ref. (67)) suggests that it might be charged in the holo form at serum pH and therefore might resist losing contact with Fe^{+3} . Kinetic data also provides evidence for contributions from tyrosine residues and mainly report the conformational changes which precede and allow iron release (6). Site-directed mutagenesis of Asp and His residues in the binding region changes the binding properties of the protein, but the protein never loses its ability to bind iron even when the coordinating residue types have been turned off. Conversely, the Y188F mutation completely abolishes the natural ferric complex with carbonate and iron binding ability (190).

These observations emphasize the importance of protonation states of the two liganding tyrosines and the fact that “dilysine trigger,” even when the Fe^{+3} is loosely connected as in the H* run, does not play a key role in opening the iron binding cleft. On the other hand, it has been shown that mutants of K206 and K296 have much lower or even undetectable iron removal rate constants at endosomal pH (181, 195) It is possible that the mutants might affect the pK_a of residues in the region which also include H249 and Y95, contributing to the changes in the iron release rates. The charge states of the residues in the iron binding cleft at serum and endosomal pH should certainly be studied in further detail.

4.1.4. At serum pH, fluctuations of the N-lobe are communicated to the C-lobe through the peptide linker

We next investigate if there is crosstalk between the two lobes. We therefore record the RMSF of the H system in 50 ns time intervals. We display in Figure 4.9a the time evolution of fluctuations resulting in global rearrangements of the two lobes in the full structure. The regions with large RMSF are all confined to unstructured loops. In the first 50 ns MD simulation of the H system, large fluctuations mostly occur in the N-lobe. In the time interval of 50-100 ns, the unstructured linker undergoes larger fluctuations, the effect of which is later communicated to the C-lobe. After 150 ns, we observe propagated dynamics throughout the full structure. In particular, fluctuations in residues 488-519 (shaded in gray) are accentuated over time, becoming the most dominant in the C-lobe as well as the whole structure, while those in the N-lobe residues 159-189 and 274-295 (shaded in blue) are diminished. These residues are colored in Figure 4.7. They control the exit pathways of the Fe^{+3} ion in the lobes (shown by the arrows Figure 4.7). Since previous MD simulations on hTf involved only the holo N-lobe structure in the absence of iron, this is the first work where such cooperativity has been observed.

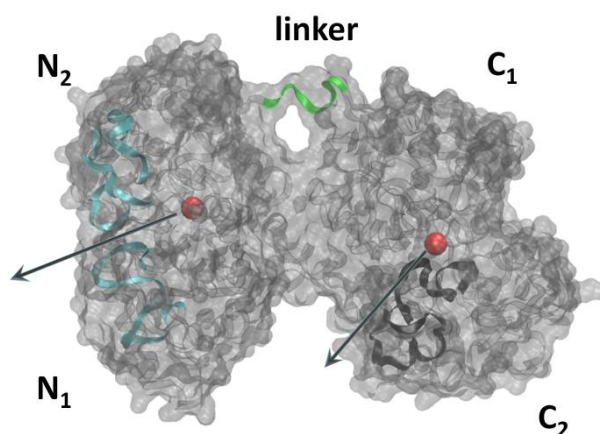


Figure 4.7 Relative positioning of the subdomains and the schematics of the exit pathways of holo hTf. The two lobes are positioned such that the exit paths whose approximate directions are shown by the arrows are nearly orthogonal. The linker is shown in green. The loops that display large differences in the RMSF profiles are shown in blue on the N-lobe (residues 159-189 and 274-295) and dark gray in the C-lobe (residues 488-519); see Figure 4.9 and the main text for details.

Accordingly, we compare the RMSD trajectories of the two lobes in the full structure to those from 200 ns MD simulations of isolated lobes of hTf, labeled as H_N and H_C (Figure 4.8). We note that the lobe opening is “triggered” in these simulations as well by neutralizing the Fe^{+3} coordinating tyrosine residues. The RMSD profile of the H_N is similar to that of the N-lobe in the full structure of H, where both of them advance towards an open conformer starting within the first 20 ns of the simulation. Conversely, based on the RMSD of H_C , we observe a more compact and stable C-lobe in the absence of N-lobe compared to its counterpart in the full structure. The RMSF profiles of these systems are provided in Figure 4.9b. We find that the relative changes in the peak positions of the fluctuations in loops 159-189 and 274-295 also occur in the isolated system H_N . Thus, these conformational dynamics are intrinsic to the N-lobe. On the other hand, the large fluctuations in the 488-519 loop of the C-lobe observed after 100 ns in the full structure (Figure 4.9a gray shaded region) are absent throughout the H_C system simulations (Figure 4.9b gray shaded region), implying that they are triggered by the mobility in the N-lobe in the full structure.

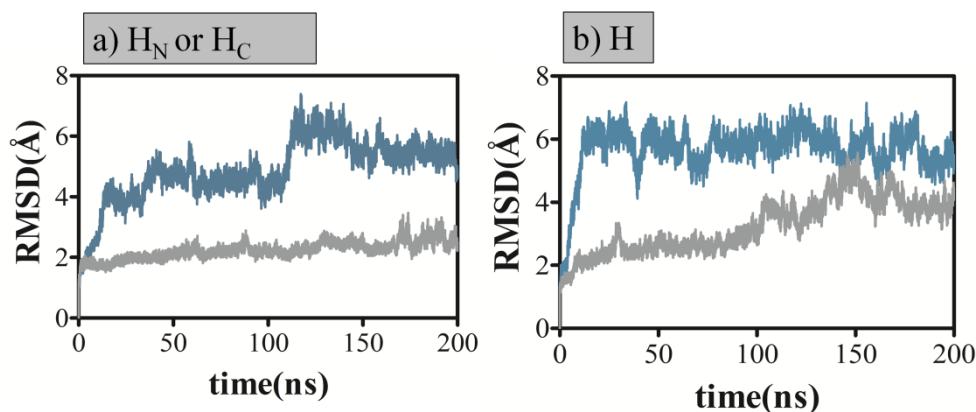


Figure 4.8 RMSD profiles of the (a) iron loaded isolated lobe H_N and H_C runs compared to those from (b) fully loaded H. The N-lobe is colored in blue, and the C-lobe in gray. The N-lobe has propensity of open in both the full structure and the single lobe case. The C-lobe, on the other hand, has an intrinsic stability that prevents it from opening unless triggered by the conformational changes in the N-lobe.

As a final word on the isolated lobe simulations, we note that while the RMSD change is low in the C-lobe, the iron coordination is nevertheless disrupted. In fact, the iron coordination distances observed by 200 ns for the N-lobe only and C-lobe only systems (H_N and H_C) are similar to those from the H system (Table 4.1). That the overall structure remains almost intact while the iron coordination is lost points to the fact that the overall structure of the C-lobe contributes to stabilizing the conformations.

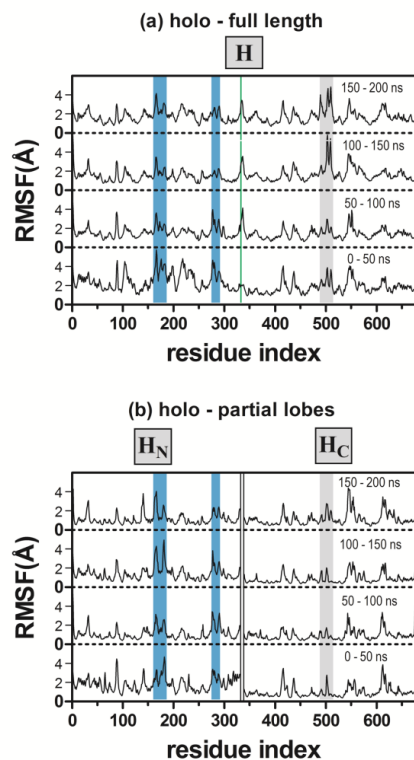


Figure 4.9 C_{α} root mean square fluctuations averaged over equally spaced trajectory pieces of 50 ns each. a) RMSF of H system in black, b) RMSF of HN system is on the left (residues 1-331) and HC on the right (residues 339-677). Regions with more pronounced fluctuations in the N-lobe, linker and the C-lobe are highlighted in blue, green and gray, respectively. In the case of H system, in the first 50 ns, more pronounced fluctuations belong to the N-lobe. After 50 ns, the linker initiates fluctuations and the delayed communication between the lobes prompts the C-lobe to go through larger fluctuations after 100 ns. For holo hTf at pH=5.6, large fluctuations begin from the starting point of simulation.

4.1.5. At serum pH, relative fluctuations of the N- and the C-lobes differentiate iron release kinetics

While it is well-established that the two monoferric forms of hTf have the same affinity for TfR (196), there are differences between their biological behavior (197). Moreover, iron release from the diferric hTf is more rapid than either monoferric form, and the affinity of diferric form to TfR is ten-fold that of the monoferric ones (196, 198). Below we discuss how each lobe might contribute to iron release once cleft opening is triggered by protonation of liganding tyrosines; *i.e.* the conformational change prior to iron release.

Iron release rates depend highly on pH (167), temperature (167, 171), presence of a single vs. two bound irons (171), and the presence/absence of TfR (171). Kinetic experiments show the rate of iron release is generally slower in the C-lobe compared to the N-lobe in full

length hTf in the absence of TfR at endosomal pH, while this order is reversed in its presence (171). Our results in this study are related to the former, and we find that the release rates are significantly affected by the amount of overall lobe opening, measured by the RMSD of a particular lobe following triggering.

Comparison of RMSD of each lobe in H, $H^{\text{Fe+N}}$, $H^{\text{Fe+C}}$, H_N and H_C systems within the first 50 ns specifies that upon triggering opening, the N-lobe always has higher mobility than the C-lobe. In the isolated lobe simulations H_N and H_C (Figure 4.8a), the C-lobe fluctuations are very low (~ 2.5 Å) compared to the N-lobe (~ 6 Å) as discussed in the previous subsection. These observations parallel the 14-fold difference in the measured iron release rates from the single lobes ($22.5 \times 10^{-3} \text{ min}^{-1}$ in the N-lobe (181) and $1.6 \times 10^{-3} \text{ min}^{-1}$ in the C-lobe (199) at serum pH and 25°C.

In H, the C-lobe does not display an identical response to triggering as the N-lobe, despite having the same modification in the charge state of the liganding tyrosines (Figure 4.2a and Figure 4.8b). The N-lobe starts opening from the beginning of the simulations, reaching an RMSD of ~ 7 Å within 10 ns. Conversely, the C-lobe displays RMSD of ~ 4 Å within 150 ns. These simulations mimic the 7.4 pH environment at 37°C, under which conditions iron release rates of $54.7 \times 10^{-3} \text{ min}^{-1}$ and $40.4 \times 10^{-3} \text{ min}^{-1}$, respectively from the N- and the C-lobes have been measured with a 1.4 fold difference in favor of the N-lobe (the difference is 3.5 fold at 25°C) (167).

The stability of the C-lobe over the N-lobe at serum pH is also corroborated by our 200 ns MD simulations on monoferric forms of the full structure ($H^{\text{Fe+N}}$ and $H^{\text{Fe+C}}$ systems in Table 2.3). The RMSD of the iron loaded lobe in these simulations are both on the order of 6 Å, the changes in the $H^{\text{Fe+N}}$ setting in earlier than in $H^{\text{Fe+C}}$ (Figure 4.10). Kinetic data available for iron dissociation in monoferric forms were measured at 25°C, having values of $65 \times 10^{-3} \text{ min}^{-1}$ and $58 \times 10^{-3} \text{ min}^{-1}$, respectively (having only 1.12 fold difference) (197).

Thus, the more similar the amount of fluctuations in the N- and the C-lobes, the closer are the values for iron dissociation rates at serum pH. We note that this observation is not valid at endosomal pH where both lobes readily open (see next subsection) and the measured iron removal rates are 2-3 orders of magnitude larger, signifying that the conformational change associated with lobe opening is not an important contributor to the dynamics at the lower pH.

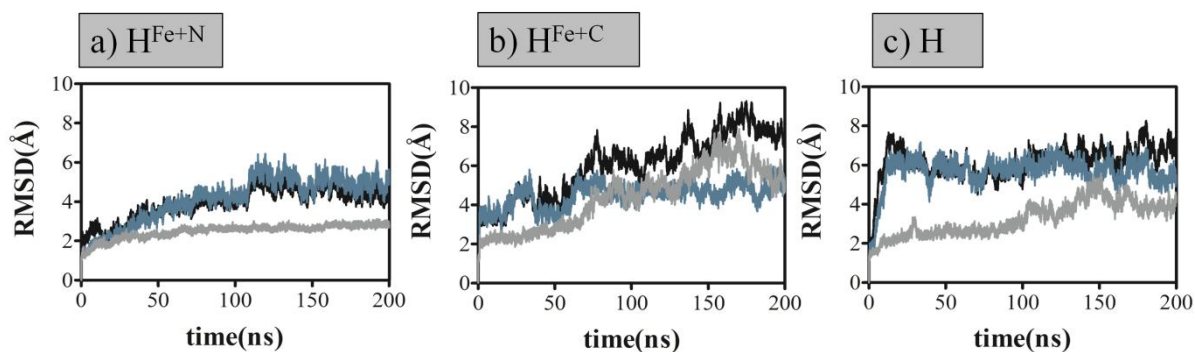


Figure 4.10 RMSD profiles of the (a,b) monoferric full length hTf, H^{Fe+N} and H^{Fe+C} , runs compared to those from (c) fully loaded H. The N-lobe is colored in blue, and the C-lobe in gray. Originally, the N-lobe is more mobile than the C-lobe in all cases. In monoferric form with iron in the C-lobe, after 80 ns, the mobility of the C-lobe gets similar to that of the N-lobe despite the associating role of iron present in the former.

The difference between the behavior of the two lobes is structural in origin: The hinge region in the C-lobe (residue 424-433) is longer than that in the N-lobe (residue 94-100), while the C-lobe as a whole is more rigid than the N-lobe due to the presence of additional disulphide bridges. The control of the binding cleft in the latter is therefore accomplished by the internal linker connecting the rigidified C_1 and C_2 subdomains (6). The C-lobe of hTf is shown to play a major role of hTf interactions with both TfR (61, 168) and bacterial receptors (65). We propose that the low fluctuations in the C-lobe and the lag of its opening behind that of the N-lobe provide the stability necessary for efficient recognition.

4.1.6. Endosomal pH prompts hTf to an open conformation

At the acidic pH of the endosome iron release rates are 2-3 orders of magnitude larger than in the serum, even in the absence of TfR (167, 181). We mimic the endosomal pH of ~ 5.6 in H^+ system by changing the protonation of the pK_a shifted residues mentioned in Table 2.2. We observe some specific structural changes that are different from all other systems studied in this work. The most prominent effect is in the simultaneous opening of the two lobes, as opposed to the delayed opening of the C-lobe observed at serum pH (compare Figure 4.2a and d). Thus, at low pH, hTf proceeds from the closed conformation to the open form directly, without the requirement of passing through the PO structure.

In the H^+ system, Y95 and Y188 in the N-lobe move away from the iron, reaching distances of ca. 10 Å, even larger than that observed in the H system. In the C-lobe, Y517 is

more distant than Y426. Coordinating Asp and His residues never leave the active site, preventing the two subdomains in each lobe to reach the completely open conformation.

The sequence of events in $H^\dagger$ is examined in more detail through the distributions of the reduced degrees of freedom θ_N , θ_C , φ , displayed in Figure 4.6, bottom row. θ_N and θ_C mostly sample the exact same region as A up to 130 ns. Thereafter, the θ_C angle equilibrates at the region around 125° . This extreme opening of the C-lobe, also observed in the PO structure to some extent, exposes iron to chelators for proper dissociation. On the other hand, the φ distribution shifts to the region centered around -70° between 115-125 ns, also sampled by A. It then goes back to the original φ angle region near 0° at 130 ns. Thus, once the pH is lowered, the two lobes are able to make the large motion of moving past each other by ca. 50° multiple times. The lower pH has a lubricating effect on the conformations of hTf. The above-mentioned results provide us with an excellent description of how $H^\dagger$ system is driven directly toward open conformer under endosomal pH within the observation window of 200 ns MD simulations.

4.2 Predicting long term cooperativity and specific modulators of receptor interactions in human transferrin from dynamics within a Single Microstate

Proteins of the transferrin family utilize conformational shifts between different states occurring before or after iron capture/release to function. It has been well established that chelators (200), salt (171), pH (169) and presence of TfR (201) affect the functioning of hTf in different states. The coupling between local and global motions governs the equilibrated sequence of events in iron regulation. For example, iron release from holo hTf, whose iron contacting residues and overall structure is displayed in Figure 1.1, is accompanied by a major conformational change. It involves the two jaws enclosing the binding site moving apart through a hinge motion, accompanied by a $55\text{-}65^\circ$ reorientation of the two subdomains of either lobe (174).

Since it is a computationally prohibitive task to simulate the long-term dynamics of the different conformers of hTf in various environments, we delineate the extent to which computationally inexpensive methods may uncover the long-term response of the protein to environmental changes. Point mutations introduced on either lobe of hTf have been shown to have unique consequences on the function of hTf (these include iron coordinating residues, see e.g. (179, 192), second shell residues, see e.g. (182) and residues at the hTf-TfR interface

(202)). Here we seek regions potentially controlling iron binding/release without a priori assumptions on their location on the protein.

There is a rich literature on the biochemistry of hTf. In particular, binding and release of iron and other metals as well as a variety of synergistic anions in the presence or absence of its receptor have been quantified (see (177) and references cited therein). In this part of the thesis, we study hTf in the absence of TfR. Conveniently, PRS require short enough simulations that may be achieved in moderate computational times. We first build the variance-covariance (**C**) matrices from the 20 to 40 ns piece of the MD trajectories (eq. (2.15)). We note that we have generated longer and multiple trajectories for each of the systems studied here, and we have repeated the analyses described in this work on all equilibrated 20 ns pieces. We have verified that the main results we present are independent of the choice of the piece. The only requirement is that the **C** matrix represents the same microstate within the time window of observations; we have previously shown that 20 ns is adequate for this purpose (12). Too long time windows might allow sampling neighboring microstates in which case the harmonic assumption imposed on the $\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle_0$ matrix in eq. (2.15) fails.

The formation of a stable hTf-iron complex involves a sequence of events (203). Thus, the iron bound state of hTf may not be assumed to be the result of a direct transition between apo and holo forms. Nevertheless, we study the A $\rightarrow$ H transition by applying the PRS analysis to these end points. Note that in this transition we cannot specify the difference between serum versus endosomal pH structures, since their average conformation is the same.

We first note that there is no single dominant mode of motion that describes the majority of the observed conformational change that involves a complex twisting of the two domains relative to each other, accompanied by closing of both lobes. The total correlations between the displacements defined by the eigenvectors for the top five modes and the A $\rightarrow$ H transition is 0.49, 0.56, 0.25, 0.29 and 0.38. We shall see below, however, that select single residue perturbations provide a much better descriptor of the conformational change than any single mode.

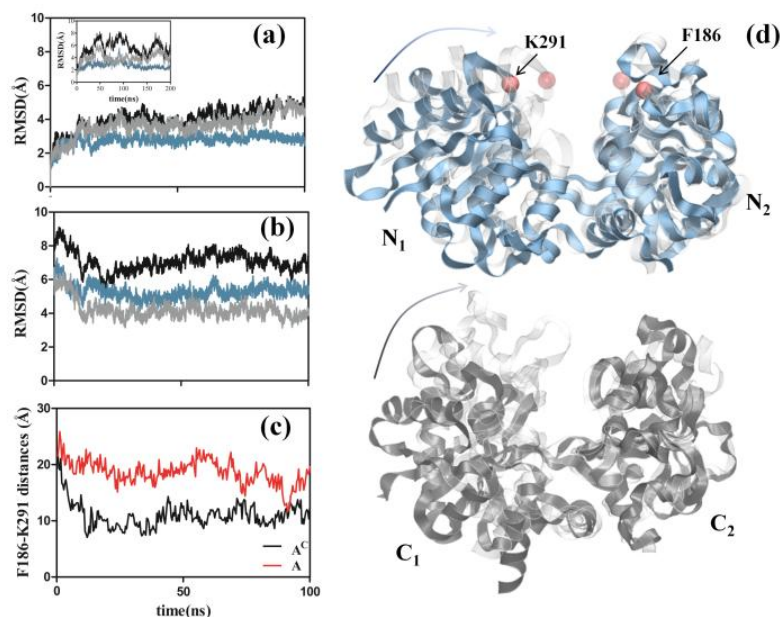


Figure 4.11 (a) RMSD of the A^C system from its initial structure. RMSD of the protein, the N-lobe and the C-lobe is displayed in black, blue and gray, respectively. N-lobe is equilibrated in the first 5 ns and does not display large fluctuations. However, RMSD of the C-lobe increases up to ca. 5 Å and affects the overall structure. In the inset, RMSD of the apo hTf in the absence of the carbonate anion (A system) is shown. Equilibrated fluctuations of the apo hTf in the presence of carbonate anion in comparison to the large fluctuation in its absence show how carbonate stabilizes the apo form of hTf. (b) RMSD of AC system from the holo hTf structure. Color coding is the same as in (a). The initial large RMSD decreases with time, demonstrating the closing effect of carbonate binding to the apo conformation. (c) Distance profile between residues F186 and K291, for the AC (black) and A (red) systems, respectively. Upon entry of the synergistic anion, the distance between the tip residues residing on the loops enclosing the cavity decreases from 18.2 ± 2.1 Å to 10.6 ± 1.6 Å (averaged after the first 20 ns of each trajectory). (d) Superposition of the initial (translucent) and final (transparent) structure of the AC system at 100 ns for the N-domain (blue) and C-domain (gray); best fitting of the structures is to the N2 and C2 subdomains, respectively, so that the closing of the other subdomain over the cavity (shown by the arrows) is better visualized. Cα atoms of F186 and K291 whose separation is monitored in part (c) are displayed as red spheres

In Table 4.2, we list the residues that are accentuated by the PRS analysis. These are the single residues whose perturbation selects the coordinate change towards the target conformation. As selection criteria of residues that are effective in directing the protein towards alternative conformations, we take several approaches depending on the distribution of the maximum of the C_i values, C_i^{max} , obtained from the 200 perturbations and calculated through eq. (2.18). We list PRS identified residues as described in methods section.

Of the 16 residues listed in Table 4.2 and having significant C_i^{max} values, five (D63, S189, H451, T452, G516) are in carbonate binding regions. Thus, perturbing these residues (e.g. by associating with carbonate) might induce displacements that better overlap with the fully iron bound conformation, in conformity with experiments that show the first step towards iron binding is carbonate capture (204). Thus, the PRS analysis from the 50 ns trajectory lets us make the prediction that perturbations due to the entrance of carbonate into the binding cavity are expected to lead to conformational changes encouraging iron uptake from the medium.

We test this idea by showing that docking carbonate anion to the apo protein followed by all atom MD simulations drives the completely open cleft of each lobe in hTf towards an intermediate with a more compact configuration. In the docking procedure residues 95, 120, 124-127 and 188 in the N-lobe and 452, 456-459 and 517 in the C-lobe are used as contacts; among them are the PRS identifies residues T452 (a direct contact) and S189/H451/G516 (residues neighboring the binding site). Following 100 ns of MD simulations of the carbonate bound apo hTf (A^C) system, we observe that a partially closed conformation is promoted. Such an intermediate is expected since until iron enters the binding site, the iron-binding cleft will not close tightly. We note that both carbonates remain in their respective clefts throughout the 100 ns. In Figure 4.11, the RMSD profiles of the A^C system in comparison to its original structure (Figure 4.11a) and to holo hTf (Figure 4.11b) are displayed. In the inset to Figure 4.11a, the RMSD of the A system is also shown from a 200 ns trajectory. We have previously presented results for this simulation and we have verified that it displays large fluctuations, but does not lose its average conformation. The motions of apo hTf upon the addition of carbonate are reduced, accompanied by a closing over the iron binding cavity (Figure 4.11). The presence of carbonate has a stabilizing effect on the system with dampened fluctuations and a net change towards an alternative, intermediate structure, suitable for iron binding. The average RMSD for the last 10 ns of MD simulations are 4.8 ± 0.2 , 2.8 ± 0.1 and 4.7 ± 0.2 Å for the full structure, N-lobe and C-lobe, respectively. Figure 4.11b displays the RMSD of the A^C system with respect to holo hTf where the difference between A^C and H^{**} decreases due to the presence of carbonate. The average RMSD during the last 10 ns of MD simulations decreases from initial values of 8.2, 6.4 and 6.6 Å to 6.8 ± 0.3 , 5.4 ± 0.2 and 3.8 ± 0.2 Å for the full structure, N-lobe and C-lobe, respectively.

We find that the reduction in RMSD is not merely due to local arrangements accompanying carbonate docking, but are rather global changes that require large motions of the loops closing over the iron binding cleft; e.g. as monitored by the distance between F186 and K291 located on the lids of each subdomains of the N-lobe (Figure 4.11c). It reduces from the initial value of 21.5 Å to an average of 18.2 ± 2.1 Å in A while it attains an average value of 10.6 ± 1.6 Å in A^C upon carbonate binding.

Table 4.2 Summary of PRS results

Initial $\rightarrow$ target structure	PDB entries	Maximum correlation (eq. 2.18)	Identified residues	Notes
A $\rightarrow$ H	2HAV:B $\rightarrow$ 3V83:B	0.75	K4 T5 R7 A59 V60 T61 L62 D63 S189 G198 D201 V251 H451 T452 L497 G516	Residues with Cimax = [0.70-0.75]
H** $\rightarrow$ A	3V83:B $\rightarrow$ 2HAV:B	0.52	C498 C523 V525 E526 P549 W550 N553	Residues with Cimax = 0.52
H [†] $\rightarrow$ A	3V83:B $\rightarrow$ 2HAV:B	0.89	All residues	All residues have Cimax > 0.85
hTf-TbpA $\rightarrow$ H	3V8X:B $\rightarrow$ 3V83:B	0.77	S390 V429 A430 R568 K569 V587 R602 Q603 Q605 H606 G609 L621 F631 R632	Residues with Cimax = [0.73-0.77]

The closing motion of carbonate on A is presented in more detail by superposition of the A^C structure at 100 ns on the initial structure (Figure 4.11d). In particular, residues 159-189 and 274-295 in the N-lobe and amino acid sequences 343-369, 488-515 and 610-633 in the C-lobe display large motions resulting in the partially closed structure. These residues also have been shown to be highly fluctuating in the H system and mark the iron exit pathway. Thus, we show that carbonate binding may act as a perturbation to shift the equilibrium towards the holo form. Presence of carbonate restricts the wide microstate sampled by the A system. Iron capture at this step might result in further structural changes to attain full closure of the subdomains.

We have checked the coordination of carbonate at the end of 100 ns. Residues we have originally specified as contact points remain within 6 Å; in addition, iron binding residues 392 and 585 as well as 85 which was identified as an anchor for carbonate in iron bound form (205) is also within 6 Å of the anion following the conformational change, implying the intermediate observed in the simulations is a likely scenario in the iron binding steps of hTf. Thus, carbonate capturing manipulates the overall structure of apo hTf, more so on the C-lobe than the N-lobe since the RMSD profile of the former is more similar to that of the holo form

(Figure 4.11b). This is consistent with the experimental observations that initial iron binding occurs on the C-lobe and further promotes the binding of the second iron on the N-lobe (206).

4.2.1. D63 mutants affect iron binding propensity

While most residues implicated as controlling the apo→holo transition interact with the synergistic anion in the holo form, the top residues L62 and D63 with $C_i^{max} = 0.75$ residing on a β -strand with indices 59–63 on the N_1 subdomain (listed in Table 4.2 along with V251 in direct contact with this strand) are not involved in synergistic ion coordination, but only iron coordination. Direct experimental evidence of their influence on the iron binding propensity of apo hTf has been well provided. This highly conserved sequence (69) is situated in the vicinity of N-lobe binding site and also displays low thermal fluctuations. Of main interest is D63, the only iron-coordinating residue from N_1 subdomain, which coordinates iron through the carboxylate oxygen. D63 provides structure-stabilizing effects in closed form by bridging N_1 and N_2 subdomains (207). Experimentally, this residue has been perturbed through a series of mutations. D63S mutant shows that shorter serine side chain along with the reduced charge of mutant makes iron coordination less favorable (208). D63E mutant has the same overall fold as the native holo hTf despite the weakened iron binding affinity; although it maintains iron coordination, the structure of metal binding site is altered (191). The distorted coordination of iron is followed by the accommodation of the bulk methylene group of the glutamic acid side chain in the binding region and results in the loss of hydrogen bonding with neighboring residues (191).

While iron is initially coordinated to H249 and Y95, domain closure and retaining a closed conformation is maintained by D63. In both D63S and D63E mutants, Fe-H249 and Fe-Y95 distances change and N-lobe gets trapped in a semi-closed form (207). The ability of D63E mutant to retain iron at lower pH also reduces (191). Furthermore, the family of mutants, D63S, D63N, D63E and D63A produce altered interaction with synergistic anion and accelerate iron loss to chelators (192). Thus, for mutants of D63, tight iron binding and apo → holo transition do not take place in a closed, compact form.

As a final note, residues K4, T5 and R7 also emerge as possible remote hot residues affecting apo → holo transition (Table 4.2). Presence of charged residues identified by PRS elucidates the importance of N-terminus in stabilizing hTf dynamics and reorienting it towards another conformation with fully engaged irons. Thus, we propose these N-terminus

residues to be studied via mutagenesis to determine their effect on iron binding propensity of hTf. We next turn our attention to the reverse problem of dissociation of iron and investigate which structural elements facilitate iron release from holo hTf at endosomal pH while preventing it at serum pH.

4.1.1. The two domains of holo hTf are perfectly correlated at serum pH, resisting iron release

The residue-residue correlation map at serum pH, calculated through eq. (2.15) is displayed in Figure 4.12a. Residues K88-P91, S501-L503 and a region centered on N611-T613 are correlated with the whole protein structure (marked by the red strips on the figure and indicated by the arrows). All these residues are located on unstructured loops. The cooperativity between the two lobes is also manifested in the slowest three modes at this pH, displayed in Figure 4.13a. The total contribution of these modes to the overall motion is 67 % of all modes, computed from their eigenvalues, λ_i , as $(\lambda_i/\sum \lambda_i) \times 100$ (47, 14, and 6% for the respective modes). There is no domain opening motion in any of these three modes. They are all circular motions around the N- or the C-domains in counter directions, appearing roughly along the x-z, y-z and x-y, planes, respectively. By inspection of these motions, iron loss at serum pH is not expected to occur. In fact, the partially open state of hTf observed at serum pH (66) is not implicated by these motions, and protonation of liganding tyrosines appear to be a prerequisite for lobe-opening to be observed on time scales accessible by MD simulations as we have shown in earlier in the text via two independent MD simulations of length 500 and 150 ns. While the modes of motion provide information on the general propensity for conformational change in a given structure, understanding molecular details of local and long-range interactions that redistribute the conformations introduce another challenge. Iron release pathway also involves multiple steps consisting of interaction with TfR, protonation steps and interaction with metalloredutases. We explore the first step of dissociation by employing the PRS approach.

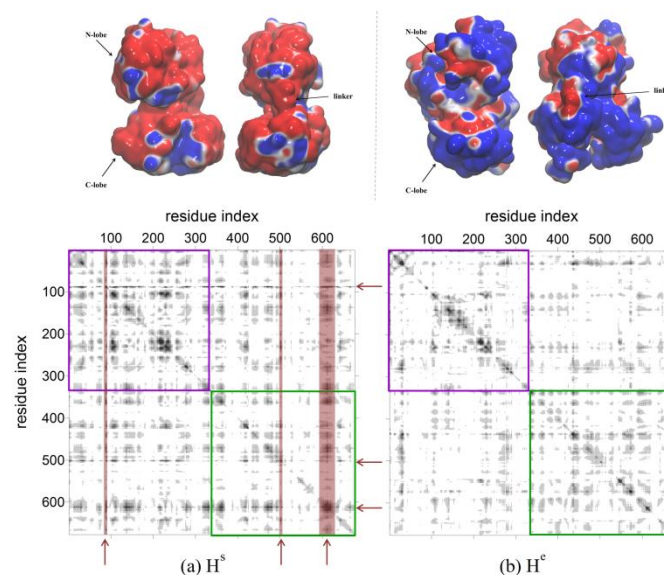


Figure 4.12 Electrostatics isocontours (top row) and cross-correlation maps (eqn (S2), bottom row) for holo hTf, at (a) serum pH, Hs and (b) endosomal pH, He. Maps are normalized so that the largest correlation is 1 (black). The regions corresponding to N-lobe – N-lobe interactions are boxed in purple and that for the intra-C-lobe are boxed in green. The N-lobe – C-lobe cross correlations appear in the remaining regions and display nearly complete lack of cooperativity between the two subdomains at endosomal pH, while pointing out residues organizing cooperativity at serum pH. Residues acting cooperatively through all regions in the latter are highlighted by red strips, also indicated by the arrows on the symmetric map. Above each system, Hs and He, snapshots of electrostatics isocontours of hTf drawn at $\pm 0.5\text{kBT/e}$ are displayed. Blue is positive and red is negative. For each case, hTf is shown in two orientations: On the left, front views of the protein are displayed where the linker is placed at the back; on the right, a 180° rotation along the vertical axis is displayed with the linker in view. At serum pH, protein surface is predominantly negatively charged. At endosomal pH the two lobes reorient and the subdomains of either lobe are opened to expose the iron binding cleft. The charge distribution on the surface is now predominantly positive, especially on the C-lobe.

Residues identified by PRS in $\text{H}^{**} \rightarrow \text{A}$ transition (Table 4.2) display low conformational transition propensity, measured by their correlation coefficient (~ 0.52). Such a low value is attributed to the rare event of releasing iron at serum pH. In serum environment at 37°C , iron release rates of $54.7 \times 10^{-3} \text{ min}^{-1}$ and $40.4 \times 10^{-3} \text{ min}^{-1}$, respectively from the N- and the C-lobes, have been measured (167). Thus, with such slow processes compared to fast endocytosis ($\sim$ minute range (209)), we do not expect the release process to occur. The top residues selected according to PRS are all confined to the C-lobe (those having $C_i^{\text{max}} = 0.52$ are listed in Table 4.2). While these are top residues, even when the hTf-TfR complex is formed as the first step towards iron release process in vivo, the structure will have little

propensity to open at this pH, despite having stronger interactions along the C-lobe interface (202).

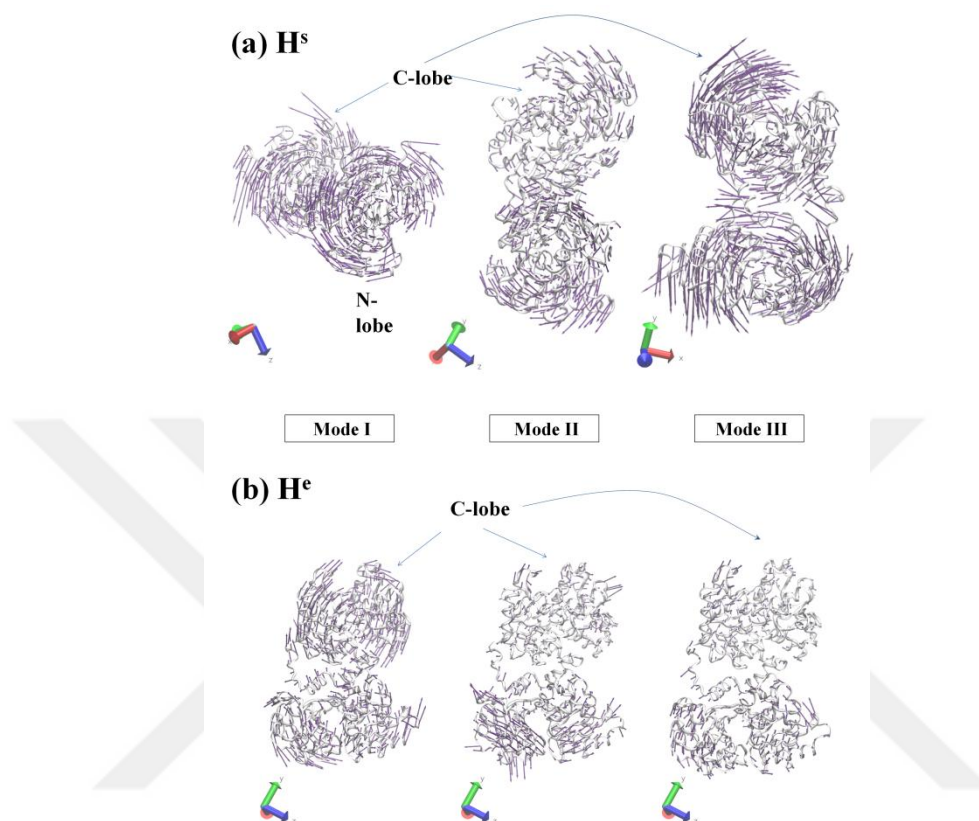


Figure 4.13 (a) Slowest three modes of the C matrices obtained for holo hTf at serum pH (H^s). Protein structure is reoriented to display the mode shapes indicated by the arrows. The motions within each lobe are unified for the whole domain while the N- and the C-lobes have opposite sense of direction in all three modes. There are no sub-domain separation motions, consistent with the slow release kinetics at serum pH. (b) Slowest three modes of the C matrices obtained for holo hTf at endosomal pH (H^e). Protein structure is displayed in the same direction for all three cases. Each lobe moves independent of the other, In the slowest mode of motion, both lobes are mobile and correlated within the domain, but there is lack of correlated motions between the two domains. In the next two modes, N-lobe motions are dominant and there is little to no motion in the C-lobe. Mode II displays hinge-like opening around the iron binding cleft of the N-lobe, consistent with iron removal at this pH.

4.1.2. Endosomal pH leads to a loss of cooperativity between the two domains of holo hTf and readily promotes iron release

The residue-residue correlations obtained from MD simulations at endosomal pH are displayed in Figure 4.12b. In stark contrast to the case at serum pH, here only very weak correlations between the two lobes are observed and the two domains act independent of each

other. Thus, at the lower pH (incorporated in our simulations via protonation of selected histidines and neutralization of iron coordinating tyrosines), the surface loops whose regions are shaded on Figure 4.12a “soften” and lose their ability to orchestrate correlations across domains.

The top three modes, contributing 60 % of the motion, are displayed in Figure 4.13b, with the respective contributions being 39, 13 and 8 %. In the most collective mode (I), each lobe displays correlations within their own domain, but the relative orientation of the motion is orthogonal across the domains (roughly in the y-z plane in the N-lobe and x-y plane in the C-lobe). In modes II and III, dominant motions appear on the N-lobe, which is consistent with the experimentally observed primary iron release from the N-lobe in the absence of TfR. Moreover, mode II is a hinge-like opening motion around the iron binding cleft of the N-lobe, hinting that there is intrinsic propensity for iron dissociation at endosomal pH.

The $H^{\dagger} \rightarrow A$ transition in the absence of TfR is also monitored by PRS. At this pH, iron release from the N-lobe is faster ($17.7 \pm 2.2 \text{ min}^{-1}$) and has precedence relative to the slower release ($0.65 \pm 0.06 \text{ min}^{-1}$) from the C-lobe (71). While there is asymmetry in the release rates from the two lobes, they are both three orders of magnitudes larger than at serum pH. We find that this conformational change may be induced by non-specific perturbations arriving on any of the residues. We make this conclusion because all residues have at least one perturbation for which the correlation is greater than 0.85. Thus, at endosomal pH, the protein has the propensity to open irrespective of the location of perturbation. The large number of residues with high correlations obtained by the PRS approach parallels what we had observed for the PRS of apo $\rightarrow$ holo transition in bacterial FBP (46) whereby the apo form has an intrinsic capability to close, corroborated by the extreme association constants on the order of 10^{17} – 10^{22} M^{-1} depending on the measurement conditions (210, 211).

To establish how point charge alterations affect the overall response of the protein and interpret the mechanistic basis of pH affects iron affinity, we monitor the electrostatic potential distribution along the protein surface using the APBS package set to the default values (146). While lowering the pH significantly alters this distribution, single point mutants or single protonation events might also change it (46). In Figure 4.13, top row, these surfaces are displayed for the H^{**} and $H^{\dagger}$ systems. The H^{**} system is the initial conformation taken from the PDB although we check that repeating these calculations on four other snapshots

collected at equal spacing along the trajectory yields the same picture. For H^+ system, the displayed conformation is that attained at the end of the 200 ns MD simulation; again we verify that conformations selected from 50 ns apart trajectory pieces yield similar distributions even before complete equilibration is attained at 150 ns. The main effect of lowering pH on the electrostatic distribution of the system is the engorged positively charged surface on the protein, particularly on the C-domain. Thus, the uniformly distributed positive charge of the acidic endosome effectively acts as a forcing agent on the protein surface at lower pH. As a result, the repulsion between the subdomains of each lobe drives the iron-binding cleft to open and promotes iron release, more so on the C-lobe than the N-lobe.

4.1.3. hTf – receptor complexes

We next study the effect of the interacting with receptors on the functioning of hTf by analyzing complex structures. To utilize iron in the cells, TfR recruits hTf on the cell surface. In bacteria, a highly sophisticated mechanism for sequestering iron from humans has been developed (212) whereby TbpA is the protein responsible for hijacking iron from hTf in the first step of the process.

We model the hTf – TfR complex with ANM using the complex structure (PDB code: 1SUV) obtained by electron microscopy (61, 213). The slowest mode that contributes 49 % of the motion is displayed in Figure 4.14a. TfR interacts with both lobes, although the C-lobe interface is larger. The complex acts unified, with both domains of hTf closing up on the receptor. This observation is in conformity with the fact that the complex remains bound through endocytosis and the release of iron in the endosome, until they reach cell surface (71). Furthermore, the cooperativity between the two lobes that prevents iron release in isolated hTf at serum pH is maintained in the presence of its receptor (compare Figure 4.13a and Figure 4.14a). While the resolution of this structure (7.5 Å) allows for making general comments on the overall motion, we refrain from using it in the PRS analysis which requires more specific atomic detail. We then model the hTf – TbpA complex by ANM, using the PDB structure 3V8X (65) (2.6 Å resolution). The overall conformation of hTf interacting with TbpA adopts an intermediate between that of holo and apo form and exposes iron to the bacterial receptor. TbpA exclusively interacts with the C-lobe residues. The slowest two modes contribute equally by ~13 % in this case, and they are displayed in Figure 4.14b and c.

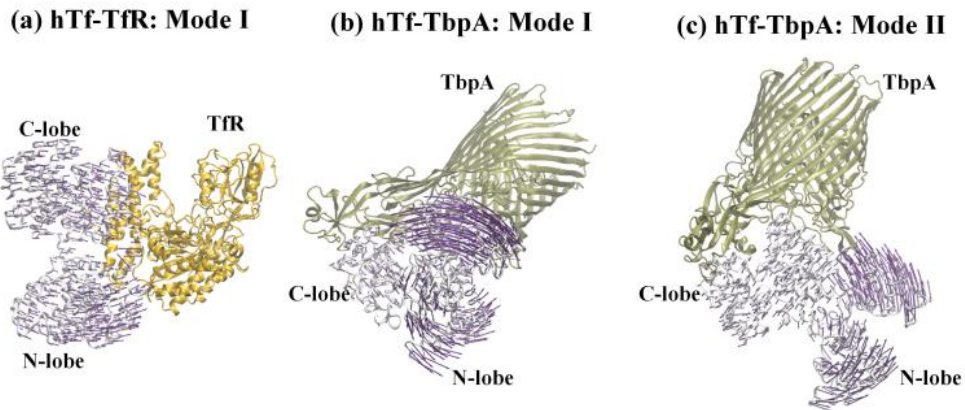


Figure 4.14 Complexes of hTf with receptor from human and bacteria. (a) hTf-TfR complex. Slowest mode of motion of the complex (contributing 49% of the motion) is displayed by the arrows. TfR (yellow) interacts with both lobes, although the C-lobe interface has more cohesive interactions. (b-c) hTf-TbpA complex. The two slowest modes of motion contributing ~13 % each have mainly N-lobe motions; these lack domain separation consistent with iron removal by TbpA only from the C-lobe.

Since the hTf structure is held at the C-lobe, the dominant motions occur on the N-lobe, with shear-like motions in mode I (Figure 4.14b) and rotational motion of N-lobe in mode II (Figure 4.14c). As TbpA sequesters iron from the C-lobe with the aid of a plug loop, hinge-like motions on the N-lobe exposing that iron is not expected to occur.

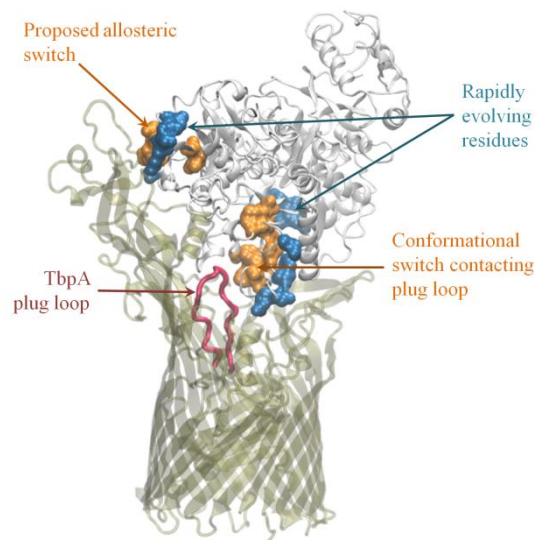


Figure 4.15 Residues emerging as conformational switch from TbpA bound to free conformation of hTf appear as two patches that display large C_i values (orange). The direct switch contacts the plug loop in the pathogenic protein responsible for sequestering iron from the C-lobe of hTf is (red). The proposed allosteric switch is 30 Å away from the plug loop.

Both switches are complemented by residues harboring rapidly evolving residues (blue) (170).

PRS analysis results are mapped onto this structure in Figure 4.15 (also see Table 4.2) where the identified residues are colored in orange. In a recent study, Barber and Elde have determined sites on hTf are subject to pathogen-driven evolution by cloning and sequencing transferrin from 21 different monkey species for phylogenetic analysis (170). Single substitutions in hTf at rapidly evolving sites to counteract bacterial iron piracy determined in that study are highlighted in blue (434-436, 439, 576, 591-593, 614, 617, 619, 625) and the plug loop in TbpA is shown in red. We thus make the observation that a patch of residues directly interacting with the plug loop (S390 V587 R602 Q603 Q605 H606 G609 L621 F631 R632) may be perturbed to access the TbpA-free conformation. Another patch that is ~ 30 Å from the plug loop may also be perturbed to achieve the same conformational change (V429 A430 R568 K569), implicating this as an allosteric regulator of conformational change. During evolution, no direct substitutions in residues of these patches are observed, possibly because the rigidity of the regions must be maintained for mechanical manipulation of the conformations. Based on our observations, we propose that point mutations occurring on residues directly contacting them might offer a perfect solution during the persistent evolutionary conflicts between pathogens and primates.

4.2 Conformational dynamics of human serum transferrin is regulated by salt concentration

We study hTf in the absence of TfR under various salt concentrations using MD simulations of 200 and 50 ns long for H and H^{**} group systems, respectively. We note that the simulations at the physiological IS is prolonged to 500 ns to track if there will be additional changes in the dynamics. In addition, a second run of 150 ns length is conducted to trace reproducibility of the results. Similarly, for the H^{**} system at 150 mM, the run is prolonged to 200 ns, and an additional 50 ns run is conducted. Neither of the additional runs change the properties and statistics reported in this section. Detailed analysis of the dynamical changes and protein motions as contributing factors to iron capture/release mechanisms are made using the descriptors representing the local and global equilibrium fluctuations of each system.

4.2.1. Conformational dynamics of human serum transferrin is regulated by salt concentration

Several factors including pH, chelators, temperature, salt concentration and presence of TfR contribute to the distribution of conformational substates of hTf and the complex iron release process (5, 8, 9, 14, 23, 53, 54). Herein, we study hTf in the absence of TfR under various salt concentrations using MD simulations of 200 (in duplicate) and 50 ns long for H and H^{**} group systems, respectively. We note that one of the simulations at the physiological IS is prolonged to 500 ns to track if there are further changes in the dynamics. For the H^{**} system at 150 mM, the run is prolonged to 200 ns, and an additional 50 ns run is conducted. Neither of the additional runs changes the properties and statistics reported in this paper. Detailed analysis of the dynamical changes and protein motions as contributing factors to iron capture/release mechanisms are made using the descriptors representing the local and global equilibrium fluctuations of each system.

4.2.1.1. Conformational preferences of hTf under different conditions

Eight sets of simulations were performed with various initial starting IS and perturbation scenarios. The different simulation conditions are labeled with the initial structure, H or H^{**}, alongside (assigned as subscripts) the approximate IS as 0 (~0 mM), L (~50 mM), P (~130 mM) and H (~210 mM). The number of ions in each system with their respective labeling is reported in Table 2.3.

In

Figure 4.16d, Figure 4.17, and Figure 4.18, we display the heavy atom root mean square deviation (RMSD) of the whole protein, and separately for the N- and the C-lobes. The average RMSD (from minimized holo crystal structure and after 200 ps) and standard deviations for H₀, H_L, H_P and H_H are 7.4±0.5, 6.4±0.4, 7.3±0.5 and 6.1±0.4 Å for the last 25 ns of MD simulations. In the N-lobe, RMSD values reach 6.1±0.4, 5.0±0.3, 6.0±0.4 and 4.0±0.2 Å after 175 ns, while they are 4.0±0.1, 4.0±0.2, 4.3±0.2, 3.5±0.2 Å in the C-lobe for H₀, H_L, H_P and H_H, respectively, in the same time frame. Thus, neutralization of the tyrosine residues in the binding site brings about large conformational changes compared to the initial conformation irrespective of the IS. These changes are most prominent in the N-lobe of H systems (Figure 4.17b). They are less mobile in the C-lobe compared to the N-lobe based on

their RMSD changes (Figure 4.17b vs. Figure 4.17c). The main difference is the decrease in the N-lobe motions at 250 mM IS that never exceeds 4 Å, compared to all other cases at lower IS where RMSD of 6 Å or more may be attained at some point of the trajectory. In contrast, the RMSDs in the C-lobe are on the order of ~4 Å in all systems studied.

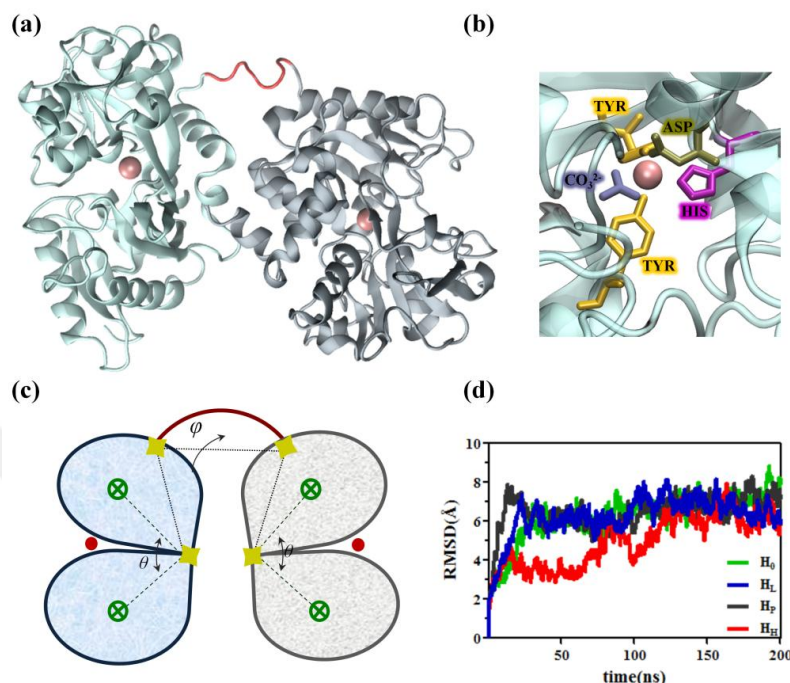


Figure 4.16 (a) Three-dimensional structure of holo hTf based on PDB code 3v83. The N-lobe (residues 1-331), linker (residues 332-338) and the C-lobe (residue 339-677) are displayed in blue, red and gray, respectively. This color coding of the regions is used throughout the text. The irons are displayed as pink spheres. (b) The identical ferric ion binding site consist of an Asp, two Tyr, and a His residues along with the synergistic carbonate anion. (c) Schematic of hTf displaying the color code used in (a), the points used to define the reduced degrees of freedom are $\otimes$ to show the centers of mass of the N- and C-lobes and $\blacklozenge$ display the beginning and end points of the linker, and the hinge residues between the subdomains of each domain, Y95 and Y426. We define the lobe bending angles θ_N and θ_C , measured by the angle between the two centers of mass and the hinge residue between each subdomain of the respective N- and C-lobes. The lobe torsional angle ϕ , is defined by the four points shown. These are the major degrees of freedom on which we project all trajectories to monitor the sampled conformations. (d) Heavy atom RMSD curves for the full structure of hTf in H group at various IS values. Systems at zero, 50, 130 and 210 mM are displayed in green, blue, black and red, respectively. The same color coding is used throughout the text for the various IS.

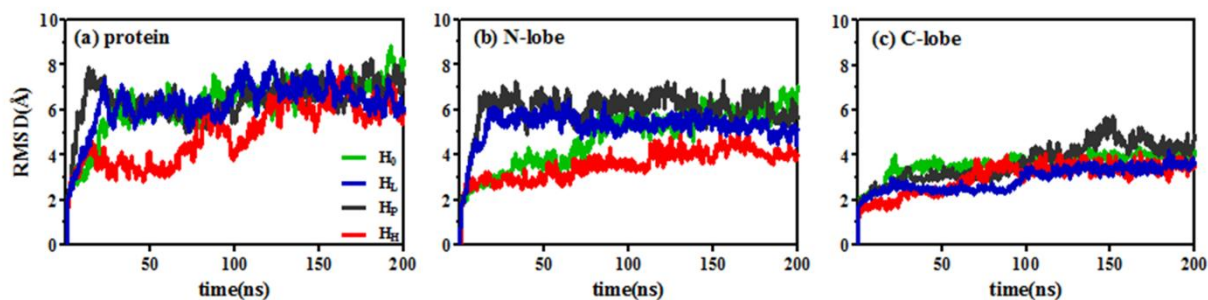


Figure 4.17 RMSD curves for H systems are displayed. The overall RMSD (a) as well as that of the N-lobe (b) and the C-lobe (c) are reported (green, blue, black and red, respectively for H_0 , H_L , H_P and H_H).

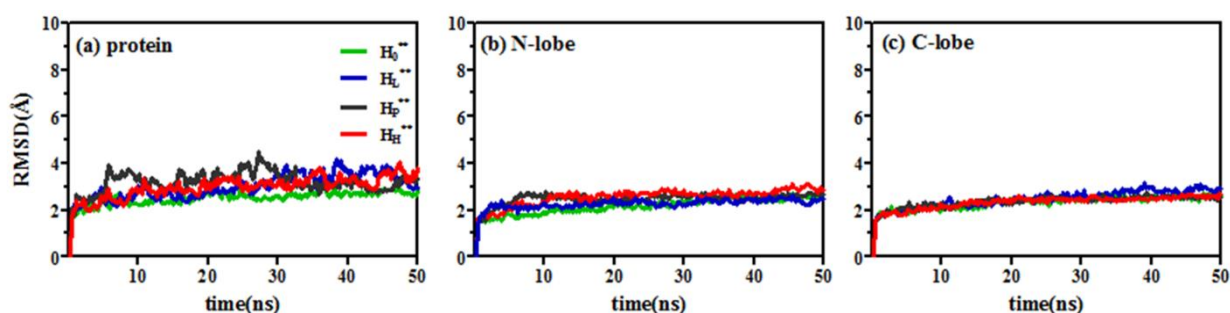


Figure 4.18 RMSD curves for H^{**} systems are displayed. The overall RMSD (a) as well as that of the N-lobe (b) and the C-lobe (c) are reported (green, blue, black and red, respectively for H_0^{**} , H_L^{**} , H_P^{**} and H_H^{**}).

Meanwhile, H^{**} systems where all four tyrosines in the binding site are negatively charged display relatively compact structures whereby fluctuations get stabilized at ~ 2 Å in the two lobes, while the overall RMSD is ~ 3 Å (Figure 4.18a). We observe that different IS environments do not affect the overall tight structure of the protein in the individual lobes of the H^{**} systems. Thus, we do not detect any large conformational rearrangements if the coordinating tyrosines are negatively charged. In fact, the motions of the linker are constrained allosterically by the tyrosines so that the relative motions of the two lobes are also restricted. Since the H systems represent the dynamics of the protein after lobe opening is triggered by tyrosine protonation, we hypothesize that the specific effect of the non-synergistic anions will be observed in the conformational rearrangements following an initial event. We therefore first study in detail the effect of amount of salts on the initial hTf structures in the next subsection.

4.2.1.2. Addition of salts initiates a balanced electrostatic potential distribution on the protein surface

Kinetic studies of iron release from hTf point out that iron release will not be observed from either lobe when the ionic strength of the reaction mixture is zero, independent of temperature (24). Since the initial event in iron release is identified as the protonation of one or both iron coordinating tyrosines, possibly from a nearby second shell residue (55, 56), the presence of ions in the solvent is expected to have a major effect. We first study the accessibility of the tyrosines by the solvent in the H^{**} systems to understand if there are any local arrangements that would allow for proton transfer from the solvent. The distances from the iron coordinating residues to the iron in the final equilibrated structures in each trajectory is given in Table 4.3 along with the solvent accessible surface area (SASA) of the region. For the Asp and Tyr residues, the distance remains at ~ 2 Å. While the distances to the histidines are greater, it is not large enough to accommodate water as verified by the SASA of the four tyrosines, which remain to be 0 throughout the trajectories in all the H^{**} systems. Furthermore, by detailed inspection of the trajectories, we find that there are no rare events that lead to instantaneous opening of the iron binding region.

Table 4.3 Average ferric ion distances from interacting side chain of coordinating residues (all values in Å). Distances are measured between OD1, OH and NE2 atoms of ASP, TYR and HIS residues, respectively and ferric ion. Errors are ~ 0.01 Å. SASA has been calculated for residues surrounding ferric ion (D63, Y95, Y188 and H249 for the N-lobe; D392, Y426, Y517 and H585 for the C-lobe). We did not include water molecules in the calculations.

FE -	D63	Y95	Y188	H249	SASA	D392	Y426	Y517	H585	SASA
H_0^{**}	2.17	1.95	1.97	3.88	19 $\pm$ 5	2.06	1.95	1.96	5.43	53 $\pm$ 12
H_L^{**}	2.05	1.96	1.97	3.85	32 $\pm$ 11	2.03	1.96	1.98	4.85	57 $\pm$ 14
H_P^{**}	2.05	1.97	1.97	3.92	14 $\pm$ 5	2.05	1.95	1.96	5.30	41 $\pm$ 10
H_H^{**}	2.12	1.96	1.97	4.30	16 $\pm$ 7	2.10	1.95	1.96	5.39	46 $\pm$ 10
H_0	2.07	5.86	6.85	5.31	180 $\pm$ 52	2.09	3.97	4.50	5.47	169 $\pm$ 28
H_L	2.10	10.17	9.63	5.05	146 $\pm$ 27	2.08	3.60	5.24	5.39	152 $\pm$ 28
H_P	2.09	5.70	7.47	5.56	177 $\pm$ 34	2.06	5.90	6.35	2.31	158 $\pm$ 33
H_H	2.05	6.28	6.28	2.32	130 $\pm$ 24	1.99	4.37	5.40	2.78	166 $\pm$ 32

It has been proposed that protonation of the tyrosines is facilitated by proton transfer from neighboring residues such as those in the dilysine trigger on the N-lobe and a triad of residues consisting of K534, R632 and D634 on the C-lobe (53). Thus, the effect of salt in the surrounding water may be indirect, by creating the suitable electrostatic environment that may trigger such proton transfer. Using a similar idea, we have previously shown that enlarged positively charged patches on the protein surface lead to iron release at the lower pH (16). In Figure 4.19, we display the electrostatic potential distributions around the protein conformations at 50 ns time points. The calculations are carried out by the APBS package (57) with default settings and are visualized by the VMD package (58). We find that in the absence of ionized water, the protein surface is negatively charged, while in the presence of ions, including the 50 mM case, there is a balance of negative and positive patches. Considering the long range effects, we conclude that the protons available for transfer will be attracted to the surface in the 0 mM case, while in the rest of the situations where the surface interactions are balanced, they will have a higher probability of transferring to specific locations inside the protein. Therefore, we propose that the presence of salt affects the holo protein by electrostatic screening at the surface, thus channeling direct electron transfer between the side chains of the second and first shell residues. These observations are consistent with the lack of iron release in the absence of ions where the protein surface potential is remains predominantly negative.

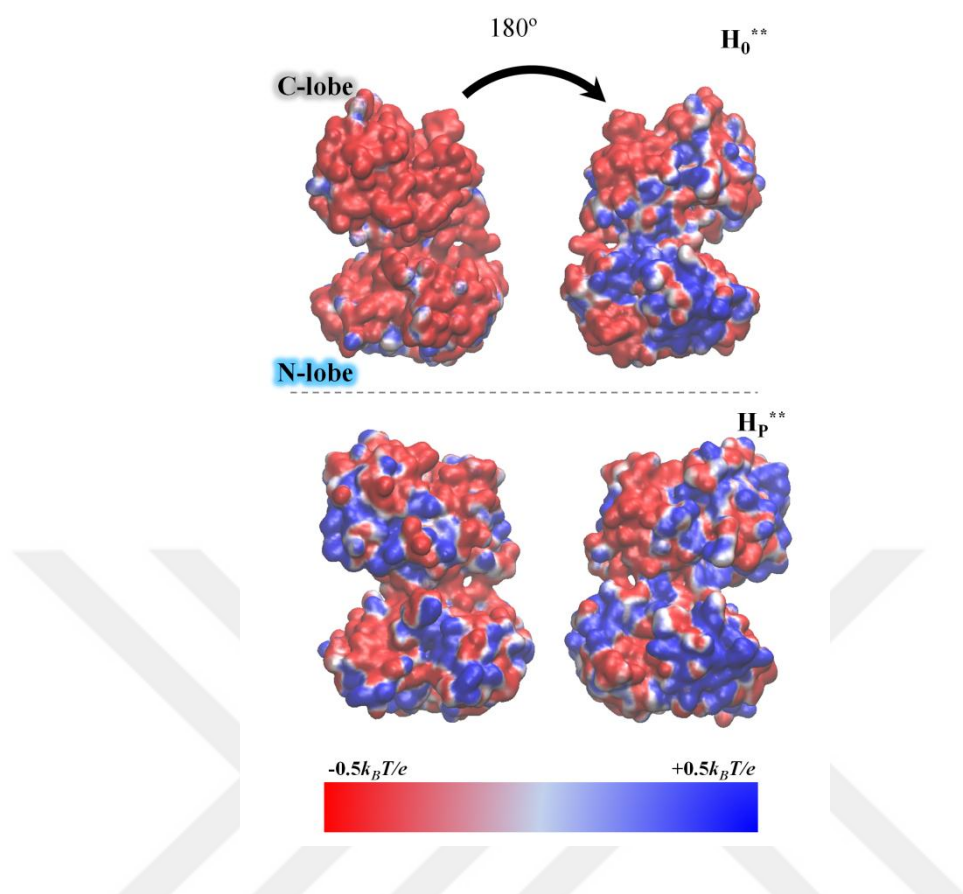


Figure 4.19 Electrostatic isocontours of H^{**} group (at 0 mM top row and 130 mM bottom row). 50 ns time point snapshots of electrostatic isocontours of hTf are drawn at $0.5 k_B T/e$ are displayed. Blue is positive and red is negative. For each case, hTf is shown in two orientations.

We note that, we have also checked the distribution of ions around the protein. While similar potential distributions at the three IS values exist, the radial distribution functions (RDFs) of Na^+ ion around negative side chains are dependent on the IS (Figure 4.20). We have previously made a similar observation on calmodulin whereby the population of Na^+ ions near the protein decreases with increasing IS (59). At physiological IS, Na^+ ions prefer to interact more with the free Cl^- ions than the negatively charged surface residues due to the entropic gain caused by moving in the solvent instead of being restricted to the protein surface.

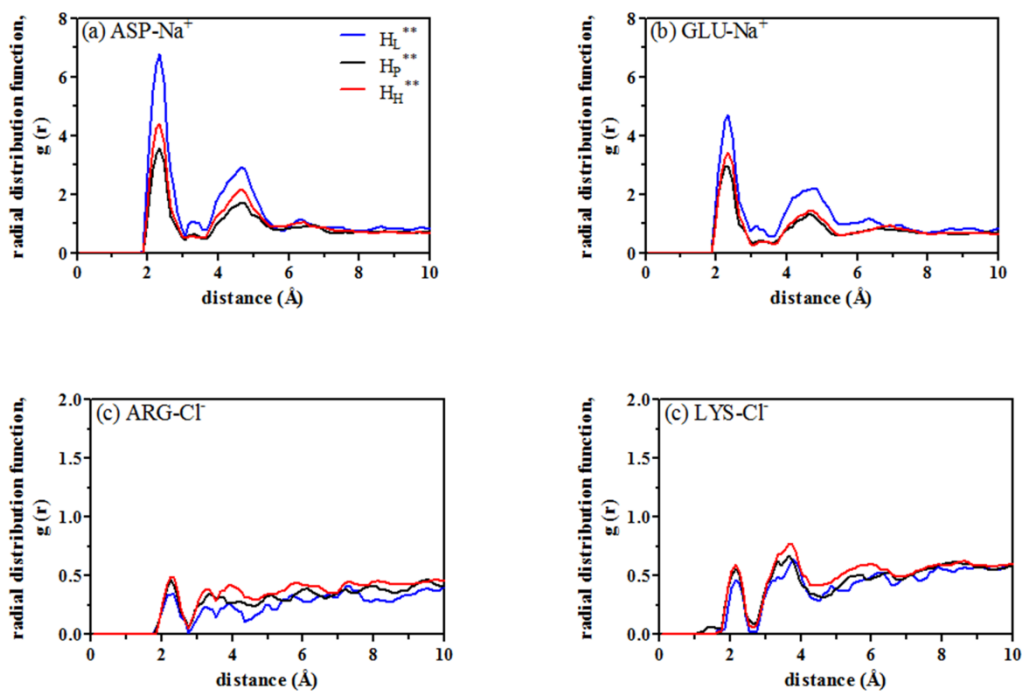


Figure 4.20 Radial distribution functions (RDF, $g(r)$) between side-chain atoms and nonspecific ions in the solvent (a), (b) Comparing distribution of Na^+ around side chain atoms of Asp and Glu residues of the protein in H_L^{**} (blue), H_P^{**} (black), H_H^{**} (red) systems. (c), (d) Comparing distribution of Cl^- ions around side chain atoms of Arg and Lys residues. Na^+ ions are accumulated around protein surface while Cl^- are more prone to be scattered in the solvent.

4.2.1.3. Upon triggering lobe opening, conformational preferences of the two lobes become IS dependent

To find the origins of the large RMSD values in H systems we explore the global and local changes in protein motions and fluctuation patterns. To this end, in Figure 4.21, we report the root mean square fluctuations (RMSF) profiles averaged over six pieces of 30 ns long MD trajectories in the H systems; we discard the first 20 ns of each simulation as the time required for equilibration. We find that residues 31, 87, 103, 140, 165, 180, 197, 215, 229, 255, 277, 288 from the N-lobe, residue 334 from the linker and residues 360, 415, 435, 472, 475, 502, 525, 544, 565, 612 and 642 are locally mobile in all systems as marked by lines in Figure 4.21. In fact, despite the relatively large RMSD profiles of these systems, the overall fluctuation patterns in various IS are very similar, with Pearson correlation coefficient between any pair of the RMSF series being larger than 0.8. However, the linker region is less mobile in the absence of ions (H_0) (shaded region in Figure 4.21). Interpreted together with the RMSD profiles of Figure 4.17, we do not find any distinct feature in residue fluctuations

that would lead to the observed differences in iron release behavior of hTf under different IS conditions.

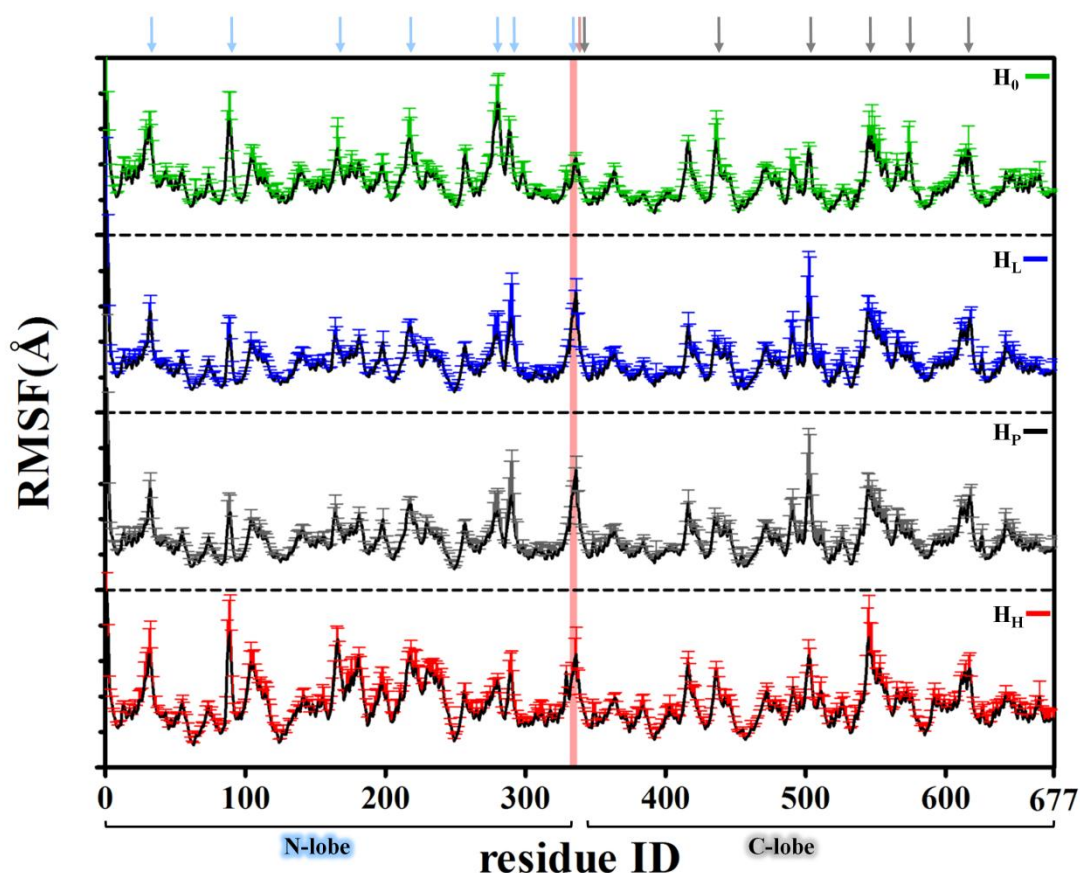


Figure 4.21 C_{α} root mean square fluctuations averaged over 30 ns long trajectories for **H** group systems. Error bars reflect the uncertainty obtained from averaging six sets of 30 ns length after removing the first 20 ns of the simulations for equilibration. Residues that display more than 50% change in standard deviation at different IS are indicated with arrows. The linker region (residues 332-338, shaded in transparent red) is less mobile in zero IS as compared to systems with larger IS.

Since the RMSF measures the local mobilities of residues, we then monitor the dynamic global conformational changes in hTf by calculating the reduced degrees of freedom in both lobes via the angle changes in the jaws of the two lobes (Figure 4.22). As described in the methods section and schematically shown in

Figure 4.16c, for θ_N (θ_C) measurements, we calculate the center of mass coordinates of N_1 (C_1), C_{α} coordinates of residue 95 (426) and center of mass coordinates of N_2 (C_2).

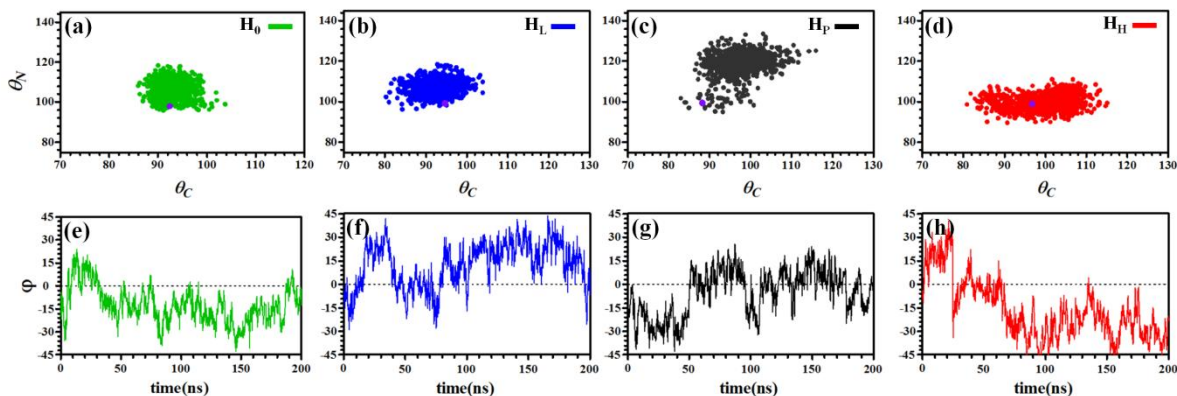


Figure 4.22 Conformations sampled by (a) H_0 , (b) H_L , (c) H_P and (d) H_H systems calculated by projecting on the reduced degrees of freedom, defined in Figure 4.16c. The initial position following energy minimization and 200 ps MD simulation is shown by the purple asterisk. The N-lobe samples an open state projected on a wide range of domain opening angle (θ_N). The largest conformational change in the N-lobe is observed in H_P system. The domain opening in the C-lobe increases monotonically with increasing IS and displays the widest range for H_H system. (e-h) Trajectories of the torsional angle (ϕ) quantifying the relative positioning of the two lobes. Initially they are all at $\sim -4^\circ$. The H_P system is more labile and samples wider range of angles in frequently alternating twisting motions.

To form a baseline for the range of values sampled in the iron-bound case, we first consider the H^{**} systems that are restricted in their motion, since they monitor the system prior to triggering domain opening. The θ_N values have average values of $103 \pm 2^\circ$, $101 \pm 2^\circ$, $103.0 \pm 2^\circ$ and $104 \pm 2^\circ$, while for the θ_C they are $97 \pm 2^\circ$, $96 \pm 2^\circ$, $96 \pm 1^\circ$ and $96 \pm 2^\circ$ for H_0^{**} , H_L^{**} , H_P^{**} and H_H^{**} , respectively. Thus, the mean and the standard deviations are the same irrespective of the IS in the H_0^{**} systems. Furthermore, the fluctuations around the average value are in the $10\text{-}12^\circ$ range for either lobe (Figure 4.23).

In comparison, the average θ_N angle is calculated as $107 \pm 4.3^\circ$, $107 \pm 4^\circ$, $119 \pm 6^\circ$ and $100 \pm 5^\circ$ for H_0 , H_L , H_P and H_H . Monitoring the trajectory of the changes in θ_N and θ_C angles in Figure 4.22a-d shows that all the systems fluctuate around an equilibrium θ_N except for the H_P system where its θ_N increases to achieve a more open pose in the first 50 ns and then equilibrates (the N-lobe opens to the same angle region within 10 ns in the second, independent run of 150 ns length). To be more precise, the θ_N changes within the ranges $[96^\circ, 118^\circ]$, $[96^\circ, 118^\circ]$, $[95^\circ, 134^\circ]$ and $[90^\circ, 111^\circ]$ for H_0 , H_L , H_P and H_H , respectively. Thus, θ_N of H_P sweeps a wider range of $\sim 39^\circ$, and is centered on a more open conformer, while its motions are dampened by further increasing IS.

The average θ_C angles in the C-lobe of H_0 , H_L , H_P and H_H fluctuate less and stay near the range sampled by the H_0^{**} systems, with values of $93\pm3^\circ$, $93\pm4^\circ$, $97\pm5^\circ$ and $100\pm6^\circ$, respectively. The θ_C changes within the respective ranges of $[86, 104^\circ]$, $[80^\circ, 104^\circ]$, $[83^\circ, 116^\circ]$ and $[81^\circ, 115^\circ]$. For the C-lobe, as the IS is increased, the range sampled by this angle also increases from $18 \rightarrow 24 \rightarrow 33 \rightarrow 34^\circ$ with a jump in the $50 \rightarrow 130$ mM range.

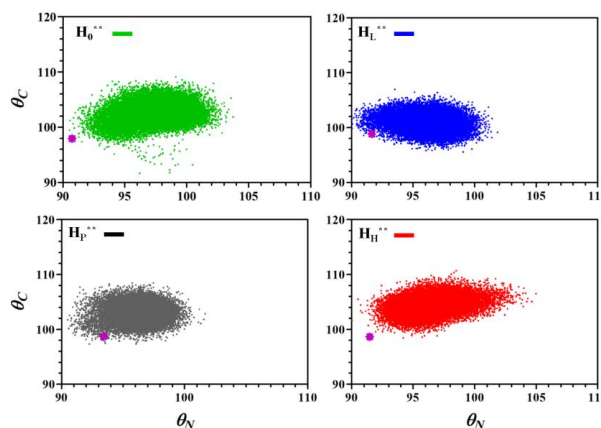


Figure 4.23 Conformations sampled by (a) H_0^{**} , (b) H_L^{**} , (c) H_P^{**} and (d) H_H^{**} systems is calculated by projecting on the reduced degrees of freedom, defined in Figure 4.16c.

Thus, while H^{**} group displays no domain opening motions in either the N-lobe or the C-lobe, proteins of H group move in a wide distribution of angles in its lobes. Furthermore, protonating the coordinating tyrosines will trigger domain opening in both lobes irrespective of IS. We find that the conformational change in the N-lobe is enhanced by increasing the IS and reaches a maximum at physiological IS where further increase in the IS results in the reduced motions of the protein both in terms of domain opening in the N-lobe and the fluctuations around the average value. The average value of θ_C angle characterizing domain opening in the C-lobe stays near that of the H^{**} systems at all IS studied, but increasing IS enhances the fluctuations slightly in domain opening motions. We predict that more flexible motions of the C-lobe are not favored physiologically, possibly due to its stability necessary for recognition by TfR that occurs on this lobe (60).

Operation of hTf also requires proper positioning of the two lobes with respect to each other, as controlled by the linker. Distribution of φ angles defined as the torsional angle (

Figure 4.16c) between the hinge residue in the N-lobe (Y95), two end points of the linker (C331 and C339) and the hinge in the C-lobe (Y426) are given in Figure 4.22e-h. In

this definition, the major axes of the two lobes are in the same plane for $\varphi = 0^\circ$. We record torsional angle values as $3 \pm 9^\circ$, $-17 \pm 9^\circ$, $-26 \pm 11^\circ$ and $-17 \pm 9^\circ$ for H_O^{**} , H_L^{**} , H_P^{**} and H_H^{**} , respectively. The H^{**} group systems display a single peaked distribution in conformity with the low RMSD values recorded for the whole protein as displayed in Figure 4.18. Thus, the negative charges on the coordinating tyrosines leading to the tight iron binding site as a local property restricts the global sliding motions of lobes relative to each other, both inhibiting domain hinge opening necessary for iron release and lobe-lobe twisting necessary for receptor binding.

The φ angle distribution of H systems span a wider range of values covering the range $[-45^\circ, 45^\circ]$ (Figure 4.22e-h). Recording the trajectory of φ angle indicates that all systems except for H_P move toward a favorite φ angle and stay there, whereas H_P tends to jump between two states centered on ~ 30 and $\sim 0^\circ$ (Figure 4.22g). Therefore, we observe that while the relative positioning of the lobes is locked at definitive values in the absence of ions in the solvent (-30°) as well as at low (30°) and relatively high IS (-30°), at the optimal IS the protein displays large, alternating twist motions in the lobe-lobe torsion space, with most conformations keeping the major axes of the two lobes in the same plane (0°) favorable for binding to the TfR.

4.2.1.4. Physiological salt concentration is optimal for iron release from the N-lobe

Focusing on the iron binding site topology, in Table 4.3, we list the distance between iron and interacting side chain of every coordinating residue in the two lobes averaged over 200 ns of MD simulations for H systems. In the distorted octahedral coordination of the iron binding site, even before tyrosines are protonated to trigger domain opening, histidines are naturally further away from iron than the other coordinating residues (at ~ 4 Å in the N-lobe, and ~ 5 Å in the C-lobe). Since the H^{**} systems remain compact with minimal domain opening angles, we conclude that the proximity of the histidines are not crucial for preventing iron release. On the other hand, in all the systems in both groups, Asp residues (D63 and D392) are quite stable and do not display any change in their interaction distance from iron. This behavior is due to the strong electrostatic interactions between aspartic acid and ferric ion. D63 and D392 are located at the edge of N_1 and C_1 subdomains, respectively, and disruption of their interaction with iron will result in complete domain opening, an event we do not

expect to observe under the current simulation conditions (this point will be further discussed with the iron binding strengths listed in Table 4.4). Iron coordinating tyrosines (Y95, Y188, Y426 and Y517) display large movements in the H group, while they have a fixed distance to Fe^{3+} in the H^{**} systems due to the negative charge on the tyrosines in this simulation set. H249 of H_H is closer to Fe^{3+} than in any other system in H group because of the relatively low amount of domain opening of the N-lobe in H_H (Figure 4.22d). H585 is quite mobile in H_0 and H_L and its distance from iron increases up to ~ 5.5 Å while it keeps its coordination to iron in H_P and H_H . The distance change pattern in the N-lobe shows that stability of D63 and D392 from N_1 and N_2 side of the iron binding cleft prevent complete domain opening in the H group. However, near physiological IS, the N-lobe is in an open state while the C-lobe is closed, in conformity with experimental observations whereby in the absence of TfR, iron release rates are faster from the N-lobe by a factor of 1.4 at 37°C (54.7×10^{-3} and $40.4 \times 10^{-3} \text{ min}^{-1}$, respectively; the difference is 3.5 fold at 25 °C) (53).

Thus, the physiological IS is optimized for hTf function in that, (i) it renders flexibility to the N-lobe for maximal lobe opening leading to the faster release rates in the absence of TfR, (ii) it restricts motions in the C-lobe for receptor recognition, and (iii) maintains a conformation suitable for TfR binding (the major axes of the two lobes of hTf are in the same plane in TfR form; PDB code 1suv (60)).

4.2.1.5. R124 is the major KISAB site in the N-lobe

To understand how the presence of ions affects the protein conformation and iron binding properties, we study the distribution of the solvated non-synergistic ions (Na^+ and Cl^-) duplicate MD runs of hTf in H group proteins. He et al. studied chloride effect on iron release in detail (61), and showed that the direct effect of chloride on iron release rates is very small. However, it has a synergistic effect pointing to a rate determining step. This is corroborated by their D63 mutants where the loose structure is more affected by the presence of chloride.

In the H group where lobe opening has been triggered, we calculate the radial distribution functions of Na^+ ions around negative side chain of Asp and Glu residues and Cl^- ions around positive side chains of Arg and Lys residues. Here, we do not consider H_0 system in our calculations since it contains only several Na^+ cations for system neutralization purposes (Table 2.3). We note that on the N-lobe there are 24, 16, 12, and 27 Asp, Glu, Arg and Lys residues, respectively. Their respective numbers on the C-lobe are 20, 24, 12, and 31.

Furthermore, all Asp, Glu, Arg and Lys residues reside on or near the protein surface which we have verified through individual SASA calculations.

Na^+ ions display greater preference for the protein surface relative to their counter ions based on their large RDFs around the protein surface (Figure 4.24). The interactions between Na^+ ions and Asp residues are roughly similar in all systems ($g(r) \sim 7-8$; Figure 4.24a). However, their propensity towards Glu residues slightly depend on IS in the order $H_H < H_L \approx H_P$ (Figure 4.24b). The number densities of Cl^- around Arg (Figure 4.24c) or Lys (Figure 4.24d) are about eight times lower than that of Na^+ around Asp and Glu. Notably, Cl^- ions interact strongly with Arg residues in the H_P system while increasing or decreasing the IS restores these interactions to the same level as those with lysines.

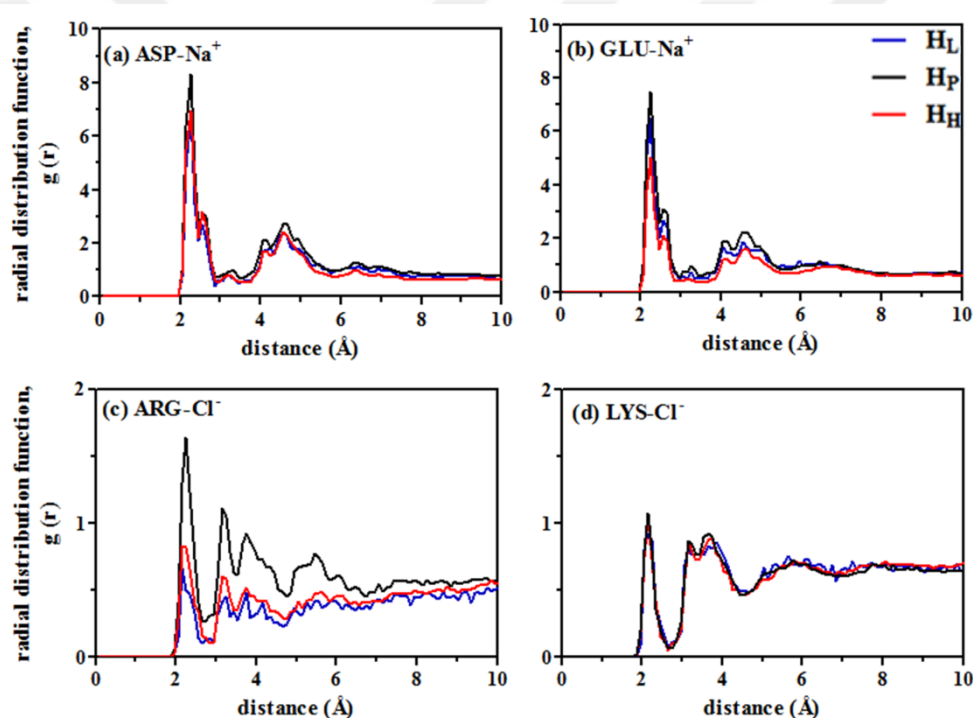


Figure 4.24 Radial distribution functions (RDF, $g(r)$) between side-chain atoms and nonspecific ions in the solvent (a), (b) Comparing distribution of Na^+ around side chain atoms of Asp and Glu residues of the protein in H_L (blue), H_P (black), H_H (red) systems. (c), (d) Comparing distribution of Cl^- ions around side chain atoms of Arg and Lys residues. Na^+ ions are accumulated around protein surface while Cl^- are more prone to be scattered in the solvent.

Anion binding to KISAB sites is said to boost extracellular iron release to ligands such as siderophores or therapeutic chelating agents used to treat iron overload (20). He et al. report that anions help maintain conformational change from closed to open state by binding to residues inside the cleft (61). The binding of anions to hTf may occur at places near ferric ion binding site, which are positively charged (2, 62). Particularly, it has been predicted that anion binding to Arg residues in the N₂ subdomain induces a conformational change in the N-lobe before the conformational change in the C-lobe (23, 63). To corroborate these experimental findings, we have individually monitored the different regions on the protein structure at various IS, and we find that the most significant interaction of Cl⁻ ions is around R124 of the N₂ subdomain. We find that interactions of Cl⁻ ion with R124 displays significant differences relative to the average distribution of Cl⁻ ions around all Arg residues on the protein surface.

The protonation of critical residues and domain opening in the N-lobe of H group facilitates the Cl⁻ ion entry to the iron-binding cleft where Cl⁻ ion is free to contact R124. We find that a Cl⁻ ion is present in the iron-binding cleft of the N-lobe in five out of the six runs (we do not find specific interactions between R124 and a Cl⁻ only in the second 150 ns simulation at 130 mM IS). Sample interactions are displayed in Figure 4.25 where the Cl⁻ binding is long lived; in the H_p systems Cl⁻ binding is less long lived and exchanges between different Cl⁻ ions is observed with an average lifetime of 7 ns. The interaction between Cl⁻ and R124 is either tight with a distance of 2.1 Å (Figure 4.25a), or looser, with Cl⁻ occupying the region between R124, K206 and K296 and having an average distance of 6.5 Å to R124 (Figure 4.25b).

In hTf structure, R124 anchors carbonate anion in the ferric binding cleft. Carbonate bonds bidentately to R124 and the N-terminus of an α -helix through hydrogen bonding (64). The combination of ferric binding residues, carbonate and carbonate anchoring residues maintains the optimum size and shape for ferric exit pathway. The R124-Cl⁻ interaction might change carbonate and iron binding mode, leading to carbonate followed by ferric ion mobility in the cleft. The perturbation of carbonate-R124 interaction via a Cl⁻ ion and hence loose binding of carbonate may readily facilitate its protonation and help facilitate iron release. At some time points, the same Cl⁻ ion coordinates K206 and K296 (Figure 4.25b), the two residues forming the so-called dilysine trigger (65). EPR spectra of K206Q and K296Q

indicate altered iron binding properties and similar Cl^- binding curves compared to the WT, suggesting that Cl^- binding to these two residues are not effective in inducing conformational

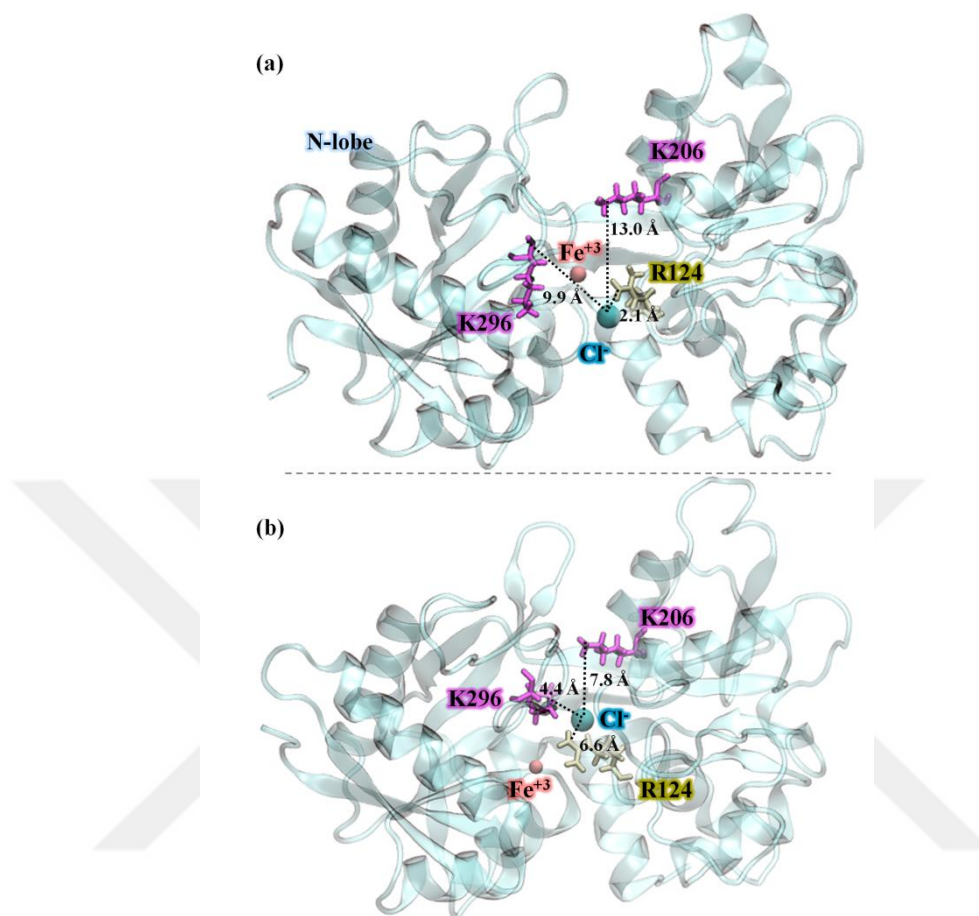


Figure 4.25 A Cl^- ion (cyan sphere) enters into the N-lobe binding cleft and establishes a persistent electrostatic interaction with residues R124 (tan stick), K206 and K296 (purple sticks). The distances between amino group of each residue and the Cl^- ion are labeled on the figure. Only N-lobe (transparent cartoon) is shown for clarity.

changes at physiological pH and arguing against a role for K206 and K296 as a KISAB site (22). However, we find that the contact between R124/K206/K296 and Cl^- ion might nevertheless support domain opening. This contact prevents the hydrogen bonding and pairing between the two lysine residues and holds the N-lobe in an open conformation.

We note that we also monitor the ferric ion coordination by Cl^- ion in the C-lobe and observe that Cl^- interaction with K534 occurs in one H_P and one H_H system run where the C-lobe is relatively more open than H_L system so that a Cl^- ion may enter the cleft and interact with ferric ion. The interaction between Cl^- is stronger in H_P system where it also coordinates

R632 in its second coordination shell. Thus, the role of KISAB sites on the C-lobe might also be examined in future studies.

4.2.1.6. Exposed iron cannot dissociate without chelators

Once we identify that there is IS dependent conformational change in the iron binding clefts following tyrosine protonation, the question arises as to whether iron is expected to dissociate in this conformation. We therefore calculate the binding affinities of ferric ions in various systems using SMD simulations (equations 2.8-2.11) to determine affinities. First, we perform SMD for the H^{**} systems to determine the quantitative range of accuracy of our SMD procedure, since there is experimental data for this system at neutral pH and in the absence of the TfR (66). For this purpose, the 50, 100 and 200 ns equilibrated snapshots are chosen as the initial structures for the SMD experiment.

The standard free energy estimations for ligand dissociation from the N-lobe and the C-lobe are reported in Table 4.4. We caution against over-interpretation of the exact free energy values since hydration free energies of multivalent metal ions are known to be underestimated by Lennard-Jones 12-6 type of potentials (see refs (29) and (67) for details). In addition, there will also be small errors to correct for the relaxation of the harmonically restrained C_α coordinates in the binding site upon complete dissociation of iron (see subsection SMD Simulations under Methods). However, the interpretation of the free energy differences between the different systems is more reliable as all systems studied have these errors.

The results show that ferric ion extraction requires substantially more work in H^{**} relative to the H group, for both the N-lobe and the C-lobe. This is a result of the strong interaction of iron coordinating residues, particularly the deprotonated tyrosines, with the ferric ion and closed domains of H^{**} systems. In this case, iron dissociation from both lobes are on the same order of magnitude within error bounds, except for the case which lacks ions in solution, where the N-lobe is tighter than the C-lobe by 7.5 ± 2.2 kcal/mol. The experimentally evaluated ferric ion binding constants at physiological pH and in various salts are on the order of $K_d \sim 10^{-22}$ M (66, 68, 69). Using equation 2.11, this corresponds to standard dissociation free energies of ~ 31 kcal/mol at 310 K, on the order of the energies displayed in Table 4.4.

Table 4.4 Estimated free energy differences upon ferric ion binding in H and H^{**} systems in the N- and the C-lobes (all values in kcal/mol.)

IS/ ΔG°	N ^{**}	C ^{**}	N	C
0	-36.4±0.9	-28.9±2.0	-16.4±0.6	-21.9±3.0
50	-35.1±0.5	-33.1±0.8	-17.2±0.1	-25.1±1.1
130	-34.9±1.8	-30.3±0.7	-19.4±0.4	-23.6±2.0
210	-33.2±1.0	-33.1±0.8	-17.7±0.6	-19.2±1.0

Once iron coordination is disrupted by the protonated tyrosines, iron dissociation is easier to achieve (corroborated by lower ΔG° in H compared to H^{**}), more so on the N-lobe than the C-lobe. The overall considerably larger amount of work necessary to remove ferric ion from the C-lobe relative to the N-lobe originates from the C-lobe structure. While the hinge region in the N-lobe (residues 94-100) is slightly shorter than that in the C-lobe (residues 424-433) that the latter is more rigid due to the presence of additional disulphide bridges. The control of the binding cleft is therefore accomplished by the internal linker connecting the rigidified C₁ and C₂ subdomains (70). Furthermore, kinetic iron release rate from the C-lobe is slower than the release rate from the N-lobe in the absence of TfR (14), in conformity with the tighter C-lobe domains and stronger iron binding affinities following triggered lobe opening by tyrosine protonation at IS up to and including the physiological value. We note however, that even in the situations with the lowest standard free energies of dissociation, the iron is still in the tight binding regime of pico Molar range. Thus, while tyrosine protonation triggers the conformational changes leading to lobe opening and presents the iron to the solvent, D63 on the N-lobe and D392 on the C-lobe remain in direct contact with the iron (Table 4.3), preventing spontaneous dissociation. The conformation, however, provides enough solvent accessibility to the iron for chelators to pick it up from the binding site.

CHAPTER 5. INCREASED SUBSTRATE AFFINITY IN THE ESCHERICHIA COLI L28R DIHYDROFOLATE REDUCTASE MUTANT CAUSES TRIMETHOPRIM RESISTANCE

We analyze the structural changes in the WT (WT^{DHF} , WT^{TMP^+} and WT^{TMP}) and L28R mutants ($L28R^{DHF}$, $L28R^{TMP^+}$ and $L28R^{TMP}$) of *E. coli* DHFR that are induced by DHF and TMP binding via 210 ns long MD simulations for each system. The simulations for the WT systems are duplicated to assess the reproducibility of our observations. The root mean square deviation (RMSD) profiles of the six systems studied with respect to the 1rx2 structure are displayed in Figure 5.1. The RMSD from the initial, closed x-ray structure (protein data bank, PDB, code 1rx2¹¹), for WT^{DHF} , WT^{TMP} and WT^{TMP^+} are 3.8 ± 0.6 , 3.5 ± 0.6 and 2.2 ± 0.3 Å, respectively, averaged over the whole trajectory for the duplicate MD simulations. For the L28R mutants, these are 2.2 ± 0.5 , 3.3 ± 0.1 and 2.7 ± 0.4 Å, respectively, for $L28R^{DHF}$, $L28R^{TMP}$ and $L28R^{TMP^+}$. Besides, we closely monitored the M20 loop, which controls the catalytic activity of DHFR in all runs by computing the average RMSD from the closed, open and occluded structures (PDB codes 1rx2, 1ra1 and 1rx7¹¹, respectively). We verified that it remained in the closed conformation in all runs.

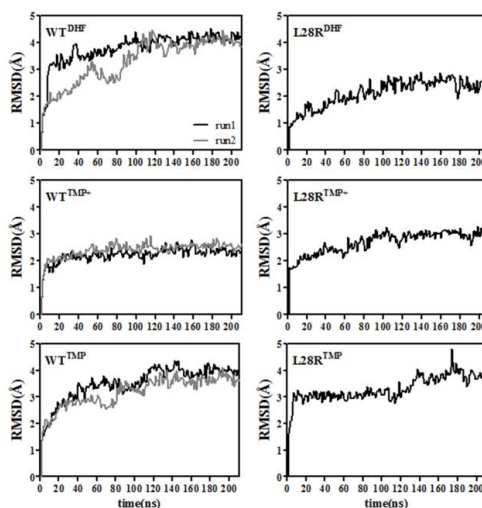


Figure 5.1. RMSD profiles for the 210 ns long MD simulations. On the left column, DHF, TMP^+ and TMP bound systems in WT are reported in black and gray, respectively for the first and second runs. On the right column, DHF, TMP^+ and TMP bound systems in the L28R mutant are displayed.

While studying the enzyme's conformational behavior during the MD simulations by analyzing the auto and cross correlations (see below), we discard the first 10 ns of the simulations accounting for the time required for equilibration (Figure 5.1). We then divide the rest of the constant pressure MD simulation trajectory (10-210 ns) of each system into equally spaced and equilibrated pieces of 40 ns to evaluate the characteristic motions averaged over all chunks. As such, the results will reproduce the motions on the time scale up to tens of nanoseconds.

In Figure 5.2a, the DHFR-DHF-NADPH ternary (1rx2) structure is displayed with loops color coded. In Figure 5.2b, we display the root mean square fluctuations (RMSF) of all the complexes averaged over the five equally spaced chunks of 40 ns along with those from experimental B-factors for the 1rx2 PDB structure. The most significant changes in RMSF are observed in the M20, CD, FG and GH loops (Figure 5.2b) which are highlighted in orange, purple, gray and red, respectively. The overall pattern reproduced by the RMSFs for all six systems are also in accord with the experimental B-factors despite shifts in relative magnitudes. The importance of these loops in *E. coli* DHFR has been addressed in the literature. The closed to occluded transition of the M20 loop that occurs on the micro to milliseconds time scale involves breaking M20 hydrogen bonds with the FG loop and establishing new ones with the GH loop.¹¹ A triad of interactions is known to exist between these loops. In addition, residues 67-69 on the CD loop have been shown to exhibit a consistent increase in rigidity upon trimethoprim binding on the ps-ns time scale motions of the backbone dynamics.¹⁰ They are coupled in stability and function to G121 on the FG loop, despite separation exceeding 25 Å.⁵⁴

5.1. Protonated state of trimethoprim in WT enzyme and L28R mutant

Previous studies utilizing ¹³C NMR spectra suggest that in WT DHFR structure, when bound to DHFR, ¹³C labeled trimethoprim is protonated (TMP⁺) on position N1 (Figure 2.1).⁵⁵ Therein, it is reported that trimethoprim molecules bound to D27N mutant lose a proton at this position at all pH values and is in deprotonated form (TMP). Interestingly, the inhibitor bound to D27S mutant displays pH-dependent protonation state. On the other hand, force field parameters for trimethoprim⁵⁶ molecule in the literature utilize united atom models for CH₂ and CH₃ groups and thus are not adequate for the level of detail required in this study. Therefore, we have parameterized trimethoprim in its neutral and protonated states (TMP and

TMP⁺). We have described the parameterization protocol used for trimethoprim force field under the Materials and Methods section; the resulting parameters are listed in appendix A.

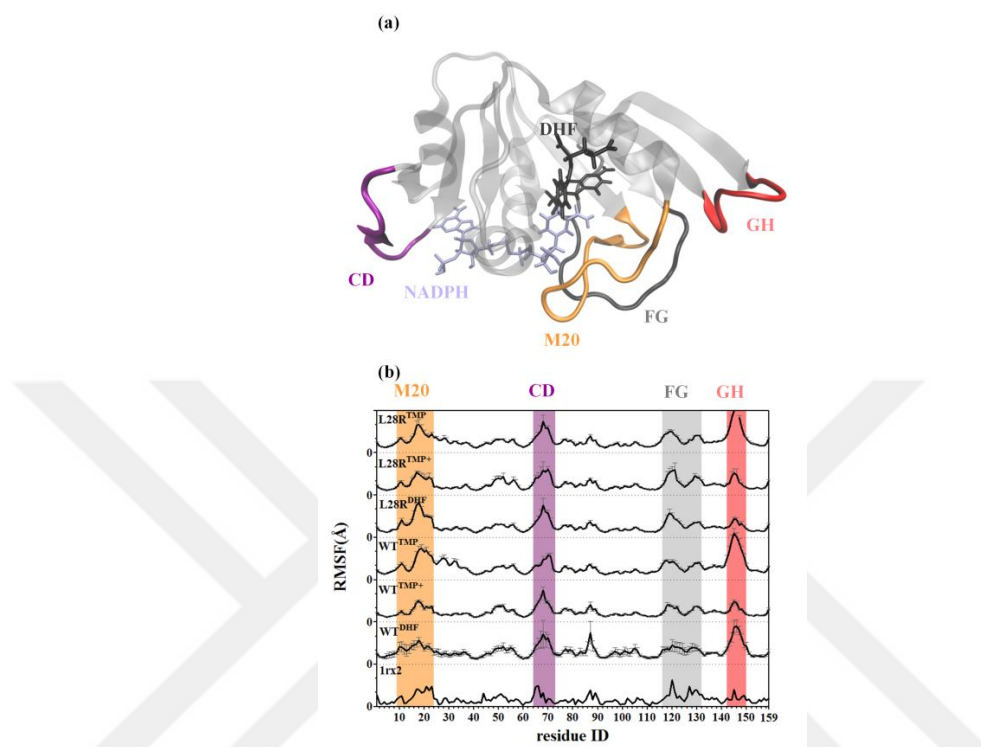


Figure 5.2 (a) Cartoon representation of *E.coli* DHFR for the PDB structure 1rx2 is displayed in transparent gray. The characteristic loops of the enzyme are shaded in orange, purple, dark gray and red for the M20, CD, FG and GH loops, respectively. DHFR is complexed with DHF (black) and NADPH (lavender) in its native form. (b) RMSF profiles of WT^{DHF}, WT^{TMP+}, WT^{TMP}, L28R^{DHF}, L28R^{TMP+} and L28R^{TMP} averaged over all equally spaced chunks of 40 ns are displayed; error bars for each residue are shown in gray. The average fluctuations obtained from crystallographic B-factors are displayed in bottom panel. The regions of the characteristic loops of the enzyme are shaded in orange, purple, gray and red indicating M20, CD, FG and GH loops, respectively.

We verified that the WT enzyme in its inhibitor bound state reproduces structures that are compatible with a protonated TMP⁺ which has been shown by NMR to be the state found when bound to *E. coli* DHFR.⁵⁵ After docking trimethoprim into DHFR's substrate binding site, residues I5, D27, I94, and Y100 are in close proximity of the ligand and form hydrogen bonds with TMP/TMP⁺ (Figure 5.3a-b). I5 and I94 make the hydrogen bonds through their backbone oxygen atoms, while D27 and Y100 form hydrogen bonds with the ligand via their side chains.

We monitor the average hydrogen bond distances throughout the WT^{TMP+} trajectory (Figure 5.3c). All the initial hydrogen bonds between TMP⁺ and DHFR fluctuate around the initial distances reducing the motions of the inhibitor in the binding site. The initial distances between hydrogen bond donor and acceptor for I5, D27 (two hydrogen bonds), I94, and Y100 are 1.8, 1.6, 1.6, 2.0 and 3.5 Å and average values of 2.8 ± 0.6 , 2.7 ± 0.9 , 1.9 ± 0.3 , 3.2 ± 0.4 and 3.1 ± 0.4 Å (mean $\pm$ standard deviation), respectively, are recorded to validate that all the hydrogen bonds are preserved. The C4-C5-C7-C1P dihedral (see Figure 2.1 for the atom labels) characterizing the trimethoxyphenyl methyl tail motions sweep the $180\pm30^\circ$ range. The C5-C7-C1P-C2P dihedral fluctuates around $-91.0\pm20.0^\circ$ throughout the 210 ns. Lack of librational motions in WT^{TMP+} is quantified by the contribution fraction of the rotation along the three principal axes of the vector connecting the NA2 and C7 atoms as described and visually demonstrated under Models and Methods (see also equation 2.7) where the major axis of motion has a fraction of 0.80, with the two minor axes contributing only 0.11 and 0.09. A large contribution to one axis and insignificant contributions to the remaining two represent a stiff molecule.⁵⁷

Although TMP is not in its neutral form when bound to WT DHFR, WT^{TMP} structure is studied as a necessary step for evaluating the changes due to the deprotonation. In fact, while TMP⁺ is prevented from going through free movements in the binding cavity due to strong hydrogen bonding of position N1 and the carboxylate group of D27, the opposite is observed for TMP due to the loss of the aforementioned hydrogen bonds (Figure 5.3d). We observe that all the initial hydrogen bonds between TMP and DHFR are broken in WT^{TMP} complex due to the tumbling motions of TMP inside the active site without exiting on the time scale of the simulations. In alignment with this observation, we find that the contribution fraction of the three principal axes of the vector connecting the NA2 and C7 atoms to TMP movements in this case is 0.42, 0.31 and 0.27. The close values along the three axes of TMP bound to WT^{TMP} represent a free rotation of the molecule inside the binding cavity.⁵⁷ Upon such free rotational movements of TMP inside the binding site, the initial distances between hydrogen bond donor and acceptor for I5, D27 (two hydrogen bonds), I94, and Y100 increase to the average values of 10.8 ± 2.5 , 9.6 ± 2.7 , 9.7 ± 2.7 , 10.7 ± 2.7 and 10.3 ± 2.3 Å, respectively (Figure 5.3d). Furthermore, the internal movements of TMP that involve the dihedrals connecting the two rings present in the antibiotic structure change in time. The C4-C5-C7-C1P dihedral characterizing the trimethoxyphenyl methyl also moves in a wider range.

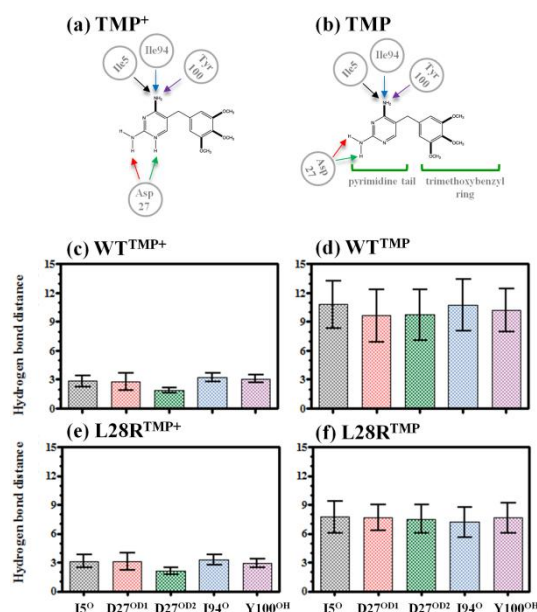


Figure 5.3 (a), (b) Schematic representation of hydrogen bonds in TMP^+ and TMP binding site in E.coli DHFR. Hydrogen bonds to the indicated amino acid on DHFR (gray circles) are shown by the colored arrows. (c-f) Average hydrogen bond distances formed between trimethoprim and binding site for WT^{TMP^+} , WT^{TMP} , $\text{L28R}^{\text{TMP}^+}$, L28R^{TMP} , respectively. Standard deviations are shown by vertical lines. Color coding follows the colors used in the arrows in (a) and (b).

Consequently, the pyrimidine ring of TMP changes its configuration, occasionally returning to the vicinity of the initial binding pose. While adopting multiple configurations, the very mobile TMP can establish a plethora of short-lived hydrogen bonds with other residues in the pocket, namely A6, I14, M16, E17, N18, M20, W22, N23, L24, S49 and W113. These favorable interactions prevent the drug from leaving the binding site despite the broken initial contacts, and the protein can accommodate the highly mobile TMP inside the active site, occasionally flipping into and out of its original orientation. Thus, the extra positive charge ensures tight binding of trimethoprim (TMP^+).

It should be noted that increased TMP mobility in the binding pocket has experimentally been observed in the complex of *Lactobacillus casei* (*L. casei*) DHFR with trimethoprim and NADP^+ .⁵⁸ Ring-flipping is achieved upon substantial conformational changes in the protein around the trimethoxybenzyl ring.⁵⁹ NMR studies of binary and ternary complexes of *L. casei* DHFR reveals that trimethoxybenzyl ring conformational rearrangements occur on the millisecond range and they require transient breaking of the hydrogen bonds to disrupt the interactions of trimethoprim which requires conformational changes in the protein to satisfy the predicted barrier heights.⁶⁰ Meanwhile, NMR

measurements on *E. coli* DHFR complex that point to two conformations for trimethoprim detects librational motions in trimethoprim which are restricted upon coenzyme binding.¹⁷ The librations are consistent with the binding modes of TMP^+ observed on the tens of ns time scales observed in the current MD simulations.

We also study the L28R mutant in the presence of both TMP (L28R^{TMP}) and TMP^+ ($\text{L28R}^{\text{TMP}^+}$) to determine the protonation state of the inhibitor in this mutant under the prevailing conditions. As in the WT protein, the hydrogen bond donor–acceptor pair distances in $\text{L28R}^{\text{TMP}^+}$ stay near their initial values (Figure 5.3e). The hydrogen bond distances between I5, D27 (two hydrogen bonds), I94, and Y100 that are originally at 1.8, 1.6, 1.6, 2.0 and 3.6 Å have the average values of 3.1 ± 0.7 , 3.1 ± 0.9 , 2.1 ± 0.3 , 3.3 ± 0.5 and 2.9 ± 0.4 Å, respectively, over the course of the 210 ns trajectory. Thus, while the hydrogen bonds are maintained as in the WT, the distances are elongated by an average of 0.2 Å in the mutant. C4-C5-C7-C1P dihedral rotates from its initial value of 156° and equilibrates around $-100\pm9^\circ$. Meanwhile, the C5-C7-C1P-C2P dihedral remains near its initial value of -108° with the average value of $-93\pm12^\circ$. Overall, the motions of TMP^+ in the pocket are as constrained as in WT, with the contribution fractions of the major axes being 0.83, 0.09, and 0.07. Because of the lengthened average hydrogen bond distances and the distorted C4-C5-C7-C1P dihedral angle, we conclude that TMP^+ binding is weakened by the L28R mutation, consistent with the increased K_i value (24.63 ± 1.34 nM) with respect to the WT (2.39 ± 1.06 nM).

When TMP is neutralized in the presence of the L28R mutation, the distances between the donor - acceptor pair distances of I5, D27 (two hydrogen bonds), I94, and Y100 in L28R^{TMP} have the average values of 7.7 ± 1.6 , 7.6 ± 1.3 , 7.5 ± 1.4 , 7.2 ± 1.5 and 7.6 ± 1.6 Å, respectively (Figure 5.3f). The initial hydrogen bonds between TMP and DHFR are no longer permanent. Nevertheless, compared to WT^{TMP} , the donor–acceptor pair distances shorten. While the oscillations around the dihedrals C4-C5-C7-C1P and C5-C7-C1P-C2P are maintained in L28R^{TMP} . The most significant difference from its WT counterpart is that the rigid body rotational motions of TMP inside the binding cavity are much more restricted despite the lack of hydrogen bonds. The latter is quantified by the contribution fractions of the major axes of 0.69, 0.23, and 0.08 (Figure 2.3).

We have also carried out 50 ns long MD simulations on TMP^+ and TMP bound D27N and D27S mutants to validate the approach we use here to determine the stable form of the

drug by monitoring the five hydrogen bonds. The former mutant is known to bind trimethoprim in its neutral state, while the latter carries significant populations of both forms of the drug, as shown by ^{13}C NMR resonance spectra.⁵⁵ Thus, in our simulations, the D27N^{TMP} should maintain significant amounts of the characteristic hydrogen bonds compared to $\text{D27N}^{\text{TMP}^+}$. Furthermore, D27S^{TMP} to $\text{D27S}^{\text{TMP}^+}$ should both maintain the hydrogen bonds. The relevant hydrogen bond distances are displayed in Figure 5.4 and are in excellent agreement with these expectations. In summary, the hydrogen bond distribution patterns summarized in this section delineate that the TMP^+ form is the more stable binding partner for the WT and the L28R enzyme. We will therefore confine our discussion to the TMP^+ bound forms of both the WT and mutant enzyme in the rest of the manuscript.

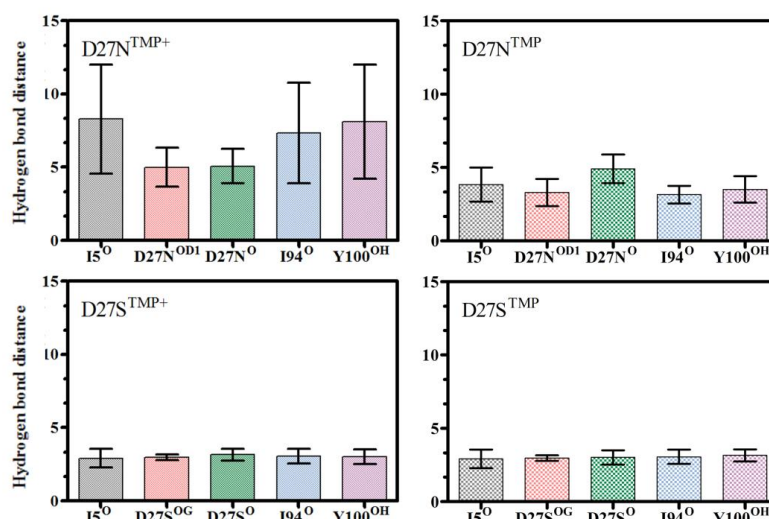


Figure 5.4 Average distances of characteristic binding pocket hydrogen bonds recorded in 50 ns trajectories of the D27N and D27S mutants for TMP^+ and TMP bound cases. Standard deviations are shown by vertical lines; color coding is the same as in figure 2. The more stable form of the drug is the one which maintains the characteristic hydrogen bonds. For D27N, this is the TMP bound form in contrast to the WT which accommodates TMP^+ (Figure 5.3); the D27S mutant binds both forms.

5.2. Conformational preferences of DHF-bound DHFR

In Table 5.1, local structural changes of the systems are reported in terms of individual RMSD of major loops (color coded in Figure 5.2a). The average structures WT^{DHF} and L28R^{DHF} are displayed in Figure 5.5a. The RMSD between them is 2.7 Å and we trace the difference mainly to the mobility of the GH loop. While M20 and FG loop conformations of WT^{DHF} are close to the crystal structure, the GH loop displays a major conformational change

from its initial position (Figure 5.5a, grey). Conversely, in the L28R^{DHF}, the GH loop maintains its initial shape (Figure 5.5a, red).

We attribute the lower RMSD of L28R^{DHF} as compared to WT^{DHF} mainly to its sluggish GH loop with an RMSD of 1.5 ± 0.2 Å in the mutant as opposed to 4.2 ± 0.6 Å in the WT. The enhanced mobility of the GH loop in WT^{DHF} is quantified by the RMSFs (Figure 5.2b). While the fluctuations in this region are relatively higher than the B-factors recorded in the 1rx2 coded x-ray structure, we note that the state simulated in this work is not in any of the x-ray structures. Instead, the cofactor is in reduced form as opposed to the oxidized form in the 1rx2 structure, and DHF is represented as 5-protonated 7,8-dihydrofolate for the reactant in the hydride transfer step as opposed to folate in the experimental structure.¹¹ Meanwhile, the total ion current abundance % per residue plots in recent ultraviolet photo dissociation mass spectroscopy measurements confirm enhanced motions in the ternary complex in solution, corroborating our observation of the mobile GH loop under some conditions.⁶¹

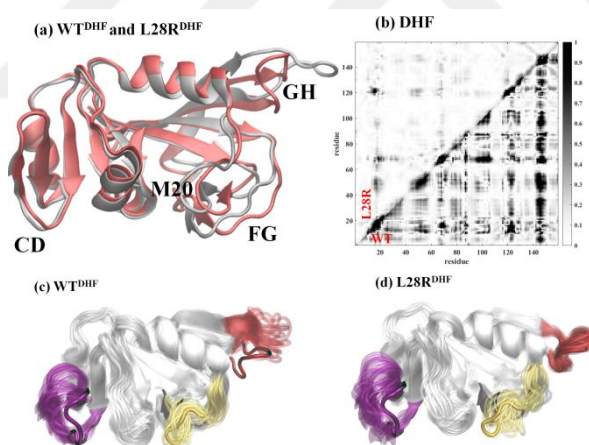


Figure 5.5 (a) Average structure of WT^{DHF} (gray) is superposed on that of L28R^{DHF} (red). The main difference is observed in M20, FG and GH loops. (b) The residue-residue cross correlation maps between WT^{DHF} (lower diagonal) and L28R^{DHF} (upper diagonal), averaged over chunks of 40 ns. Correlations are much stronger, extending over the whole protein in the WT compared to the L28R mutant. (c,d) Superposed structures of WT^{DHF} and L28R^{DHF}, clearly display the mobility of the loops. 20 equally spaced snapshots are shown.

The overall fluctuation patterns of DHF bound complexes of DHFR represent the role of loop dynamics in global motions of the enzyme. However, to gain insight on how the loop motions are coupled to each other and the rest of the residues, we construct the covariance

matrix between the fluctuations of the residue networks.⁶² The residue-residue cross correlation maps averaged over all the chunks are displayed in Figure 5.5b providing information if the processes are coupled over the ten nanoseconds time scale; the lower triangular region is for WT^{DHF}, while the upper part is for L28R^{DHF}. Equally spaced, superposed snapshots from this trajectory are displayed in Figure 5.5c.

Table 5.1 Loop and overall RMSD (Å) in the systems with respect to initial structure, averaged over the 210 ns trajectories.

	M20	CD	FG	GH	overall
WT^{DHF}	2.7±0.7	2.9±0.8	2.3±0.5	4.2±0.6	3.8±0.6
L28R^{DHF}	2.6±0.7	2.8±0.7	2.8±0.5	1.5±0.2	2.2±0.5
WT^{TMP+}	2.4±0.2	1.8±0.4	2.9±0.5	1.5±0.2	2.2±0.3
L28R^{TMP+}	2.6±0.4	3.2±0.6	2.7±0.3	1.5±0.2	2.7±0.4

In WT^{DHF}, there are strong correlations between almost all parts of the protein (Figure 5.5b, lower triangle). In particular, the M20 loop is correlated to residues 47-59, CD loop, residues 85-89, 95-107, FG and GH loops. Meanwhile, residues 119-126 of FG loop and residues 142-149 located on GH loop are highly correlated with the entire structure of the system. A strong coupling is also detected between residues 95-107 and the CD loop.

In Figure 5.5b, upper triangle, we display the effect of the L28R mutation on the residue cross correlations. The mutation suppresses the overall cross correlations between the different regions of the enzyme in the presence of DHF compared to the WT systems. In particular, the concerted movements of the GH loop with the rest of the protein disappear. However, we observe a remaining correlation of the M20 loop (Figure 5.5d, yellow loop) with some CD and FG loop residues as well as residues 42-55. The enhanced motions of the FG loop and the overall more ordered structure of the L28R^{DHF} system are apparent in Figure 5.5d.

Comparing the local and global dynamics of the WT and L28R mutant enzyme, we next monitor the local interactions of the native ligand through the hydrogen bonds established with the enzyme. We assess the changes in hydrogen bond distributions of the active site in all systems (Figure 5.6). The residues in the DHF binding site of DHFR are I5, A6, A7, M20, D27, L28, F31, R52, R57, I94 and T113. Among them, backbone of residue I5

and side chains of residues D27, R52 and R57 can form hydrogen bonds with DHF atoms of NA2, N3 and N8 as the hydrogen bond donors and O, O1 and O2 as the hydrogen bond acceptors (Figure 5.6a; see Figure 2.2 for atom labeling scheme). Average acceptor – donor distances for each hydrogen bond formed between DHF and the binding site in WTDHF and L28RDHF are displayed in Figure 5.6b-c. In both enzymes, I5^O-DHF^{H8} (2.5 ± 0.6 Å), D27^{OD1}-DHF^{NA2} (2.1 ± 0.4 Å), and D27^{OD2}-DHF^{N3} (1.9 ± 0.3 Å) hydrogen bonds are maintained throughout the MD simulations, keeping I5 and D27 in constant contact with DHF. I5 and D27 are located on a rigid and stable α -helix and β -strand, respectively. In the WT system, the remaining monitored hydrogen bonds on R52 and R57 begin to break by 10 ns of MD simulations with occasional reformation and breaking events throughout the trajectory due to the mobility of the C6-C9-N10-C14 dihedral angle connecting the pterin ring and p-aminobenzoyl glutamate tail of DHF. The dihedral angle changes in the range $[-154^\circ, 113^\circ]$ while the hydrogen bond occurs only for the dihedral angle range of $[-109^\circ, 113^\circ]$. More negative angles result in twisting the p-aminobenzoyl glutamate tail, turning the hydrogen bond acceptor site away from R57, losing the R57^{HH11}-DHF^{O1} hydrogen bond. R52 and R57 reside on a highly mobile and flexible loop (residues 51-58) which displays large fluctuations and its RMSD increases up to 2.7 Å. Consequently, p-aminobenzoyl glutamate tail of DHF is provided with enough freedom to move in large swinging motions while the pterin ring is kept in position by I5 and D27. Because of a freely moving tail, the distance of the R52^{HH21}-DHF^O hydrogen bond increases up to 16.4 Å by 10 ns of MD simulations, with an average value of 6.2 ± 2.1 Å. The R57^{HH11}-DHF^{O1} and R57^{HH21}-DHF^{O2} hydrogen bonds are kept intact until 30 ns, after which the distances increase up to 8.3 ± 3.1 and 9.0 ± 3.5 Å, respectively. We find that the movements of the pterin tail increases the R57^{HH11}-DHF^{O1} and R57^{HH21}-DHF^{O2} distances. We quantify the mobility of DHF by calculating the principle axis of p-aminobenzoyl glutamate tail in the WT system. The principal axis calculations were carried out for a vector connecting the N10 position to CD position as described in the Models and Methods section (also see equation 2.7). The contribution fraction of the three principal axes to DHF p-aminobenzoyl glutamate tail movements in WTDHF is 0.45, 0.33 and 0.22 (Figure 2.3a). Comparable contributions along the three axes of DHF bound to WT system represent free rotation of the tail.⁵⁷

We note that the strengthening the hydrogen bonds in the binding pocket only explains the enthalpic contribution to the Gibbs free energy difference. We have used pulling

experiments to predict the relative binding affinity of DHF to DHFR in the WT and the mutant. To facilitate ligand dissociation in each system, we have used snapshots at the 10, 50, 90, 130, 170 and 210 ns time points of the trajectories as described in the methods section. The total free energy differences, ΔG , between ligand free and ligand bound states are 17.0 ± 1.0 , and 15.8 ± 1.0 kcal/mol, for WT^{DHF} and L28R^{DHF}, respectively, which imply that the systems have indistinguishable stability, within error bounds ($\Delta\Delta G = 1.2 \pm 1.4$ kcal/mol), assuming the L28R mutation has similar free energetic cost in the DHF-bound and DHF-free enzyme. Thus, despite the hydrogen bond stabilization of the ligand in the L28R mutant, the entropy loss due to the lack of fluctuations in the p-aminobenzoyl glutamate tail and the long side chain of residue 28, along with the suppressed mobility of the GH loop (Figure 5.2b and Figure 5.5, Table 5.2), compensate for the energetic gain and the two systems have comparable overall stability. On the other hand, the extra hydrogen bonds established between the ligand and the enzyme in the mutant (Figure 5.6d-g) are expected to lead to a larger energetic cost to reach the transition state ($\Delta H_{L28R}^{\ddagger} > \Delta H_{WT}^{\ddagger}$). Therefore, the barrier heights for ligand dissociation from the L28R mutant is expected to be higher. In fact, using the relation $K_m = (k_f + k_{cat})/k_f$ where k_f and k_r are the forward and reverse reaction rates for substrate binding, k_{cat} is the reaction rate for product formation, and substituting the experimentally measured values for K_m and k_f listed in Table 5.2 as well as an $\Delta\Delta G$ of 1.2 kcal/mol disfavoring DHF binding in the mutant, we find that the ratio of reaction rate for DHF dissociation, k_r^{WT}/k_r^{L28R} , is expected to increase in the range 2-20 depending on the actual value of k_f^{WT} . Furthermore, since THF is only a single hydrogen different than the 5-protonated 7,8-dihydrofolate in the current simulations, one would expect similar hydrogen bonding patterns for THF, and a similar increase in barrier heights.

This is in conformity with k_{cat} values we measure which are 8.16 ± 3.27 and 1.30 ± 0.01 s⁻¹, respectively for the WT and the L28R mutant enzyme (Table 5.2). We note that THF release is the rate limiting step in the catalytic cycle. Nevertheless, the thermodynamics of the inhibitor bound states must also be assessed to evaluate the impact of the mutation in the competitive inhibition of DHFR which we address next.

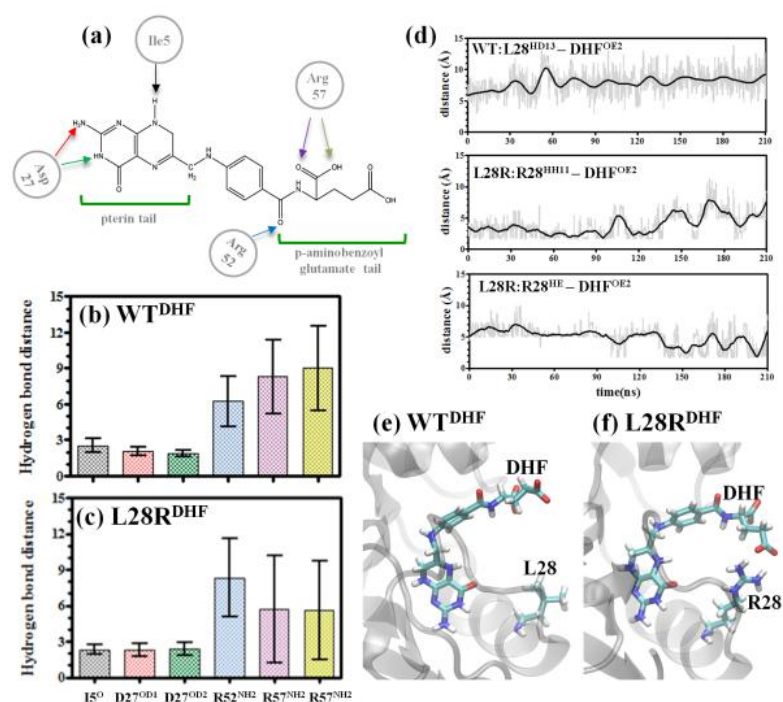


Figure 5.6 (a) Schematic representation of hydrogen bonds in DHF binding site in *E.coli* DHFR. Hydrogen bonds to the indicated amino acid on DHFR (gray circles) are shown by the colored arrows. (b), (c) Average hydrogen bond distances formed between DHF and binding site for WT^{DHF} and L28R^{DHF}, respectively. Standard deviations are shown by vertical lines. Color coding follows those used in the arrows in (a). (d) Sample donor-acceptor distance between the methyl group of residue 28 and DHF. For the WT, L28 makes no hydrogen bonds with the p-aminobenzoyl glutamate tail of the substrate. Sample donor-acceptor distances between methyl group of R28 and DHF indicates there are sustained hydrogen bonds between the α -amino group of R28 and p-aminobenzoyl glutamate tail of DHF in the L28R mutant. (e), (f) sample pose of DHF relative to residue 28 in WT^{DHF} and L28R^{DHF}, respectively. R28 interacts with DHF p-aminobenzoyl glutamate tail via its side chain and further stabilizes the ligand.

Table 5.2 Experimental results for the competitive inhibition of DHFR.

ITC measurements					Competitive inhibition measurements		
	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/mol/K)	K_d (nM)	k_{cat} (s ⁻¹)	K_m (μ M)	K_i (nM)
WT	-11.2 $\pm$ 0.5	-13.1 $\pm$ 1.4	-6.5 $\pm$ 3.0	4.5 $\pm$ 0.9	8.16 $\pm$ 3.27	1.49 $\pm$ 0.14	2.39 $\pm$ 1.06
L28R	-10.8 $\pm$ 0.2	-6.8 $\pm$ 0.5	13.4 $\pm$ 2.6	13.1 $\pm$ 1.5	1.30 $\pm$ 0.01	0.62 $\pm$ 0.01	24.63 $\pm$ 1.34
Difference	+0.4 $\pm$ 0.5	+6.3 $\pm$ 1.5	+19.9 $\pm$ 4.0				
	(+0.6 $\pm$ 1.1)*						

*free energy difference calculated via pulling experiments

5.3. L28R mutation disrupts water organization at the trimethoprim – enzyme interface.

The overall structure of trimethoprim-bound DHFR is nearly unaltered upon the inclusion of the L28R mutation, as evident from the average structures displayed in Figure 5.7a. The RMSD between WT^{TMP⁺} and L28R^{TMP⁺} average structures is only 1.1 Å. The largest change with respect to the x-ray structure occurs in the CD loop (Table 5.1). Introducing TMP⁺ into the DHFR structure leads to a system where the motions of the loops are weakly dependent on each other in both the WT and the L28R mutant (Figure 5.7b). In the WT, M20 loop is correlated with residues 45-55, CD and FG loop. CD loop also displays small, but detectable correlations with residues 76-105. These results are in conformity with NMR spectroscopy findings on *E. coli* DHFR in complex with trimethoprim where the closed conformation was found to predominate the ensemble of structures upon drug binding.⁸ It was proposed therein that, while filling the active site sends a signal for conformational change on the μ s–ms time scale, the drugged-complex remains locked in the closed conformation. Moreover, NMR measurements on backbone and side chain methyl groups showed that the motional landscape was also simplified on the ps–ns time scale that is also sampled in the current work. In particular, side-chain methyl order parameters measuring the ps–ns dynamics in the closed, trimethoprim bound protein indicates a general rigidification. This is in accordance with the lack of strong correlations in protein fluctuations of WT^{TMP⁺} system (Figure 5.7b) as well as the lower RMSF values (Figure 5.2b).

Equally spaced snapshots from the WT^{TMP⁺} and L28R^{TMP⁺} simulations are superimposed in Figure 5.7c,d for comparison. The covariance matrix of L28R^{TMP⁺} in the upper triangle of Figure 5.7b, shows that the drug binding leads to very similar correlations to those of WT^{TMP⁺} system in corresponding residue positions. Together with the fact that the hydrogen bonds in the binding pocket are nearly unaltered in the WT and mutant enzyme (Figure 5.3c,e), we conclude that the L28R mutation has a minute effect on the trimethoprim binding propensity of the enzyme.

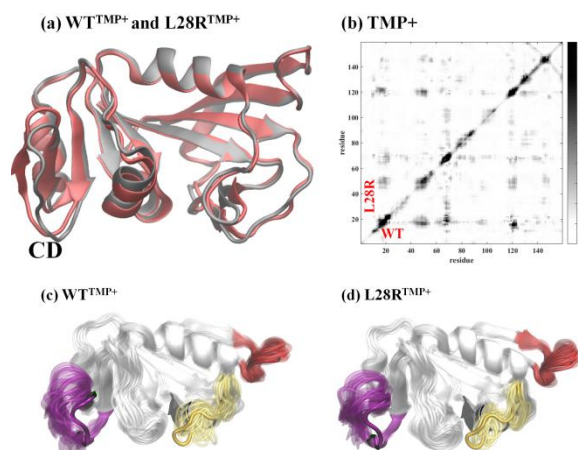


Figure 5.7 (a) Average structure of WT^{TMP+} (gray) is superposed on that of L28R^{TMP+} (red). The main difference is observed in the CD loop. (b) The residue-residue cross correlation maps between WT^{TMP+} (lower diagonal) and L28R^{TMP+} (upper diagonal), averaged over chunks of 40 ns. Correlations are subsided in TMP⁺ bound forms, slightly more so in the L28R mutant than the WT. (c,d) Superposed structures of WT^{TMP+} and L28R^{TMP+} clearly display loop motions. 20 equally spaced snapshots are shown.

To measure the relative stability of TMP⁺ in the binding pocket, we have again resorted to free energy of binding calculations. Using snapshots at the 10, 50, 90, 130, 170 and 210 ns time points for both systems as described in the Materials and Methods section, ΔG between ligand free and ligand bound states are calculated as 11.6 ± 0.9 and 11.0 ± 0.7 kcal/mol for WT^{TMP} and L28R^{TMP+} systems, respectively. We note that since TMP is found in neutral form in the water environment, these values cannot be used to calculate the exact K_d . However, the difference between them may be assessed to determine the relative stability of TMP in the binding pocket, which is implied to be very similar, within error bounds ($\Delta\Delta G = 0.6 \pm 1.1$ kcal/mol). In fact, we have carried out ITC experiments for the binding of trimethoprim to the WT and L28R mutant. The experimental data are presented in Figure 5.8, and the calculated ΔG , ΔH and ΔS values are listed in Table 5.2.⁶³ The experimental free energy difference between the mutant and the WT (0.4 ± 0.5 kcal/mol) are in conformity with the calculated value. ITC experiments find K_d of 4.5 and 13.1 nM respectively for WT and L28R systems in the presence of trimethoprim. The nM range calculated dissociation constants are in agreement with experimental K_d values reported for trimethoprim bound WT system (9.1 nM).¹⁶

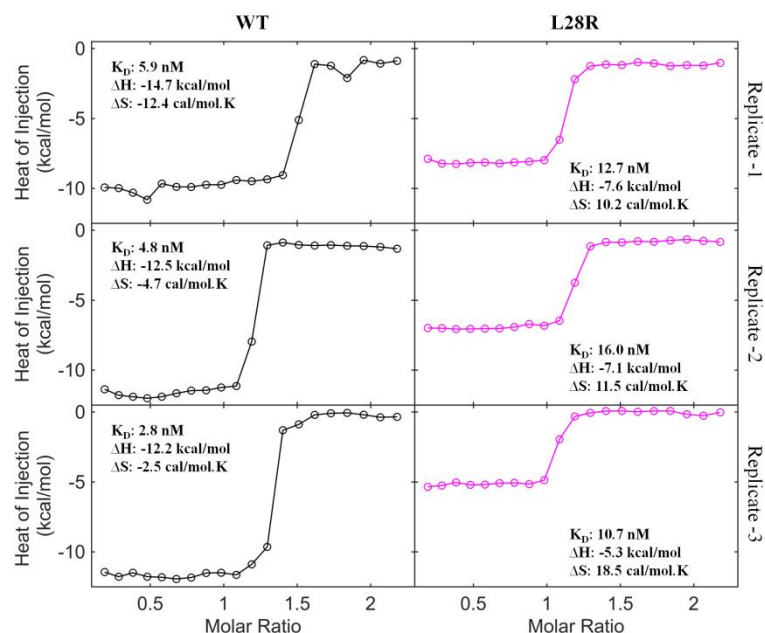


Figure 5.8 Isothermal Titration Calorimetry results for TMP binding are shown for WT and L28R DHFR. Each experiment is repeated in triplicate. All K_D , ΔH , ΔS and ΔG values are significantly different between the WT and the L28R mutant.

While the free energy differences for trimethoprim dissociation are similar for the WT and the mutant, we find that there is energy/entropy compensation. The enthalpy of binding of the drug is less favorable in the mutant, which is not directly apparent from a mere consideration of the average 0.2 Å increase in the hydrogen bond distances within the binding pocket (Figure 5.3c,e). This energetic loss is compensated by the increased entropy change (Table 5.2) from -6.5 ± 3.0 cal/mol/K to 13.4 ± 2.6 cal/mol/K.

The entropy change due to binding is associated with the following contributions: The largest is due to the solvent entropy increase arising from the release of solvent molecules upon binding. Conformational entropy change accounts for the variations in the degrees of freedom of the protein (e.g. restricted side chain mobility) and the ligand; it may make a positive or a negative contribution. Furthermore, rotational and translational degrees of freedom loss in both the protein and the ligand upon binding will have a negative effect. Finally, ordering of water at the complex interface may contribute unfavorably to entropy change and favorably to enthalpy change.⁶⁴

We find that the WT protein in fact harbors an environment for ordered water when bound to trimethoprim, corroborating the measured entropy and enthalpy changes listed in

Table 5.2. The average solvent accessible surface area of the trimethoprim in the WT protein is 112 Å² in the WT while it is reduced to 99 Å² (standard deviation is 20 Å² in both cases), leading to a tightening of ~10 %. When we monitor the trajectories, we find that a structural water molecule is located next to the NA2 atom of the TMP⁺ molecule that is stable for the last 50 ns of the first trajectory and the last 100 ns of the second trajectory. No such water molecules are found in the mutant. The environment of the coordinating water molecule in the WT is displayed and compared with that of the mutant in Figure 5.8. Residues 5, 111 and 113 from the β sheet beneath the TMP binding site accommodate the interfacial water molecule that makes an additional hydrogen bond with trimethoprim while not disrupting the one with D27. In the L28R mutant, the binding site is slightly tightened so that a water molecule may no longer squeeze between the β sheet, D27 and trimethoprim.



CHAPTER 6. CONCLUSIONS AND FUTURE OUTLINE

6.1. Conclusions

Combining all atom MD schemes and coarse-grained analytical techniques, we characterize the local and global motions; conformational change propensity; distribution of conformational substates under various environmental factors; and ligand stability in protein-protein, protein-ligand and protein-inhibitor complexes. We argue that the mechanical basis of the proteins' function originates partly from the ligand and loop motions locally and long-range allosterically driven motions globally.

We explore Hessian matrix of protein complexes to identify functional information relevant to long-term dynamics. We employ PRS to extract essential components that contribute to a pre-selected conformational change. Since large-scale motions may require key residues to mediate correlated motions between different regions of the protein complexes, we use PRS to predict those involved in the conformational transitions between the bound and free proteins in protein-protein complexes. Later, we study hTf within selected conformational states and various environmental factors. The simulations have implications for iron-binding/release mechanisms of hTf, as well as its recognition at the cell surface in the iron loaded form and release by the cell in the iron free form. We note that all results pertain to the dynamics of hTf in the absence of TfR and the important role the receptor plays in iron release from hTf is not addressed in this thesis. Finally, we study DHFR in complex with DHF and trimethoprim drug in its WT and L28R mutant forms. We explain the coupling between enzyme fluctuation and catalytic function by reporting the details of the molecular motions of DHFR binding to its substrate (DHF) and its inhibitor (trimethoprim) in two protonation states with sampling of equilibrium conformations on the time scales of 200 ns. The role of ligand dynamics and its local and allosteric effects on the enzyme is helpful in understanding potential inhibitory activity of trimethoprim on DHFR in particular and the antibiotic resistance phenomena in general.

The major findings are itemized as following:

1. *Protein-protein complex interface does not always control dissociation.* We have applied the computationally simple and efficient PRS method in combination with ENMs to

extract residues that might exhibit unique contributions to dissociation process of protein-protein complexes. We have collected a set of 25 non-redundant, functionally diverse protein complexes that feature an excellent pool to study the critical residues triggering the conformational transition propensity between bound and unbound forms of the proteins constructing the pair. PRS reveals that while for a group of protein pairs, the residues involved in recovering the conformational shifts are mainly confined to regions with large RMSD, there are others where these residues from parts of the protein unaffected structurally during the binding/unbinding processes. Strikingly, only one of the protein complexes displays hot residues from the interface. Other cases include residues scattered all around the protein sections. 16 of the protein pairs display remote control of dissociation reaction in which electrostatic potential distribution along protein surfaces play the major role. In five of the protein pairs, large local conformational changes display auto-controlled motions. In the remaining protein pairs, we observe a complex pattern consist of a combination of various characteristics. A thorough statistical analysis of the residues involved in PRS analysis reveals that the method picks 59% of hot residues from the ones residing on unstructured loops, which account for 45% of the all residues present in the analysis pool. This observation is explained by the dichotomy of flexibility – stiffness of loops present in each secondary-loop-secondary motifs. We validate the approach by illustrating the biological significance of top residues selected by PRS. We find that the residues whose perturbation leads to the observed conformational changes usually correspond to either functionally important residues or highly conserved residues.

2. *Coupling between reaction kinetics and protein motions may mechanically stress the enzyme/substrate binding or disrupt stabilization.* We have applied PRS as a powerful method to extract residues that might exhibit unique contributions in understanding enzyme inhibition. Application of PRS to subtilisin in its apo and inhibitor bound conformation provides us with a range of spatial locations on the protein structure. These residues are predominantly clustered and they display biologically important roles in the catalytic mechanism and stability of the enzyme. PRS is effective in that the communication in the protein is deduced only from the fluctuation patterns of the residues. The specificity of residues is indirectly incorporated in the analysis, insofar as their effect on the fluctuations. On the other hand, many of the sites affecting stability/enzymatic activity do so due to their specific roles, e.g. due to its hydrogen bond formation. It remains to be further studied how

nature places specific residues on sites that are also mechanically coupled to the rest of the structure.

3. *Domain opening at serum pH is coupled to protonation states of iron coordinating tyrosines in hTf.* We show that the holo conformation may be destabilized by manipulating protonation states of some distinct residues. Large conformational changes may be delimited or induced by targeting perturbation-sensitive residues selected based-on their predicted pK_a values (Table 2.2). By mimicking two different pH environments and monitoring reduced degrees of freedom for intra and inter-lobe motions, we distinguish two possible pathways for conformational changes occurring prior to iron release. At serum pH, we observe the closed $\rightarrow$ PO transformation within 200 ns because of charging two tyrosine residues in the iron-binding cleft; this structure remains stable up to 0.5 μ s. We expect PO $\rightarrow$ open transformation to follow, but on a longer time scale, since the two lobes cannot make the necessary step of moving past each other by $\varphi = 50^\circ$ under these conditions (Figure 4.6, center column). Yet, the distance of iron to its coordinating residues within 200 ns in the PO structure implies that the dissociation event towards the apo conformation has already begun (Table 4.1). Conversely, at the lower pH of the endosome, a direct closed $\rightarrow$ open switch is observed within 200 ns (Figure 4.2); the two lobes sliding past each other is thus triggered by lowering the pH.

4. *In the absence of suitable iron chelators and TfR, mild acidification of the endosome is not sufficient to stimulate iron dissociation from the cleft.* In systems where domain separation occurs (H, PO, H^{Fe+N} , H^{Fe+C} , H_N , $H^\dagger$), iron is observed to progress toward the outside of the binding cleft, but never gets released. In both lobes, carbonate anions remain within the iron binding cleft while having a degree of mobility correlated with the motions of the iron binding residues. This is consistent with experiments where carbonate has alternative conformations (175) and remains bound to the protein even after iron is released from both lobes (73). The observed behavior in the MD simulations on individual lobes indicates that the motions of the two are concerted. The N-lobe acts as a pulling machine that stimulates the C-lobe to open in the full hTf structure. Therefore, the bilobal nature of hTf is a biological advantage, promoting iron release as a result of interlobe communication. At endosomal pH, we observe a simultaneous conformational switch of the two lobes as opposed to the delayed switch at serum pH. In fact, for the holo form, PRS can explain the lack of iron release propensity at

serum pH and its ease of release at endosomal pH. In the former case, there are no single residues whose perturbation is expected to lead to the conformational change while in the latter, a plethora of residues readily lead to the conformational change. We find that electrostatic surface distributions drastically shift at the two pH values, providing a physics-based explanation of the observed phenomena.

5. *Residues located on mechanically adept regions trigger local dynamical contributions that lead to global rearrangements of the protein structure towards pre-selected conformations.* Via PRS, we have been able to determine sites on the apo form whose perturbation is expected to shift the conformation distribution towards holo-like structures. We then tested this idea with conventional MD simulations where these locations are perturbed via synergistic anion docking, which is the natural ligand expected to enter the iron binding cavity and prepare it for iron uptake. The simulations show that such a local perturbation leads to global motions on both lobes (Figure 4.11).

6. *Allosteric control and mechanical manipulation of rapidly evolving residues regulate hTf dissociation from the bacterial pathogen.* We offer a possible mechanism on how single residue substitutions may be drastically effective under the positive selection in the evolutionary path of hTf (170), despite the conjecture that single-point mutations would not be resistant toward TbpA binding due to the large buried interface ($\sim 2500 \text{ \AA}^2$ (65)). We have shown that residues that undergo positive selection to avoid iron sequestration by pathogens are in direct contact with two patches of residues disclosed by PRS to be mechanically capable of steering hTf away from binding the pathogenic protein. Therefore, we propose target locations on the surface of hTf for which ligands may be designed to modify the hTf-TbpA interface and prevent the iron scavenging process without altering the functioning of the protein.

7. *The extent of conformational changes under various IS has effects at two different stages of iron dissociation.* In the first stage that requires protonation of tyrosines, the salt in the solution has a mean field effect by altering the electrostatic potential distribution on the protein surface, harboring the balanced charge environment necessary for directed protonation events in the protein. In the second stage that involves iron dissociation from the partially open lobes, we show that the Cl^- ions enter the binding site and affect the dynamics locally.

8. *The extent of the coupling between different loops of DHFR enzyme in WT and mutant forms significantly changes under drug binding relative to the native substrate.* Changes in the correlated motions of loops with each other and with the rest of the protein structure suggest that the dynamics of DHFR is altered upon drug binding and the mutation. M20 and GH loops motions prove to be crucial to catalysis reaction of DHFR in WT system. Their lost correlation in the entire enzyme structure in WT^{TMP⁺} contributes to the inhibition process. However, not being able to retrieve M20 and GH long-range correlations in L28R^{TMP⁺} shows that the L28R substitution may not be fully effective in developing a resistance mechanism in the presence of TMP⁺. In combination with less mobility of GH loop, long-range cross correlation of the L28R^{DHF} system is lost. The long-range connection within the enzyme structure is abandoned in the WT^{TMP⁺} system where loop motions are subsided. L28R mutation reduces the correlations across the enzyme and presents a structure that is ready to get rid of the ligand

6.2. Future outlook

Throughout this thesis, we have made efforts to address molecular details of hTf protein structure in the various stages of its cycle. While hTf process is complicated, we have demonstrated molecular basis of the iron release mechanism from both the N- and the C-lobes of hTf. However, many problems remain to be answered; one of which is the conformational changes associated with lobe-lobe communication. the results of classical MD simulations and time evolution of the C _{α} RMSF plots (Figure 4.9) shows that there is a time delay between the structural variations of the N-lobe and C-lobe. Corroborated by cross-correlation maps of H and H[†] systems (Figure 4.12) in which the two lobes move in a cooperative manner in the former while they move independently in the latter, we predict a cross-talk between the two lobes at physiological pH which is lost at endosomal pH. Presence or absence of such a communication between two domains results in their different transitional dynamics under various pH values. We aim to answer the question on how the different molecular motions that operate in different time scales affect each other. To this end, we will extract the timescales of motions in each lobe from MD simulations of full and isolated structures of hTf at two different pH values via 5-200 ns-long MD simulation and predict response time (τ) of local motions distributed over different time scales. Based on the collected data and through

combination of multiple viscoelastic descriptors, we will propose a viscoelastic model of this protein, which describes the time dependency of the protein's response in different environments.

Yet another extension of the present thesis will discuss the dynamics of DHFR in the presence of twelve set of mutations consist of single, double, triple, quadruple and quintuple combination of P21L, A26T, L28R, W30G, W30R and I94L mutations. We will study various systems via 210 ns long classical MD simulation in the presence of DHF and trimethoprim (in two protonation state). We compare the effect of mutations by constructing cross-correlation maps of residue fluctuations. The dynamics of binding site is further studied via monitoring the hydrogen bond distribution. The conformational changes of functional loops in response to various mutations as contributing factors to catalytic activity will also be studied. We will also calculate the free energy of substrate/inhibitor binding by means of alchemical transformation for the above-mentioned mutations except for the ones containing proline mutations. The free energy changes will help explain the stability and binding affinity of the ligands in each mutation scenario while correlating thermodynamic properties of the system to kinetic parameters within the competitive binding scheme.

APPENDIX A

For a proper representation of trimethoprim interactions with the protein, we have parameterized trimethoprim in two states (TMP and TMP⁺). We described the parameterization protocol used for trimethoprim force field under the Materials and Methods section.

A1. Topology files

```

RESI TMP 0.0

ATOM NA2 NN1T -0.164000 ! HA41 HA42
ATOM HA21 HN1T 0.148000 ! \ /
ATOM HA22 HN1T 0.148000 ! NA4 H2P H81
ATOM C2 CAT -0.096000 ! | |
ATOM N3 NC2T -0.144000 ! C4 C7 C2P O3P-C8-H81
ATOM N1 NC2T -0.039000 ! / \ / \ / \ /
ATOM C6 CAT -0.026000 ! N3 C5 C1P C3P H81
ATOM H6 HPT 0.090000 ! || | || |
ATOM C4 CAT -0.046000 ! HA21 C2 C6 C6P C4P H91
ATOM NA4 NN1T -0.185000 ! \ / \ // \ // \ |
ATOM HA41 HN1T 0.130000 ! NA2 N1 C5P O4P-C9-H92
ATOM HA42 HN1T 0.130000 ! / / / |
ATOM C5 CAT -0.010000 ! HA22 O5P H93
ATOM C7 CT2T -0.075000 ! |
ATOM H71 HAT 0.090000 ! H101-C10-H102
ATOM H72 HAT 0.090000 ! |
ATOM C1P CAT 0.137000 ! H103
ATOM C2P CAT -0.095000 !
ATOM H2P HPT 0.090000 !
ATOM C3P CAT -0.054000 !
ATOM O3P OST -0.141000 !
ATOM C8 CT3T -0.137000 !
ATOM H81 HAT 0.090000 !
ATOM H82 HAT 0.090000 !
ATOM H83 HAT 0.090000 !
ATOM C4P CAT -0.066000 !
ATOM O4P OST -0.055000 !
ATOM C9 CT3T -0.193000 !
ATOM H91 HAT 0.090000 !
ATOM H92 HAT 0.090000 !
ATOM H93 HAT 0.090000 !
ATOM C5P CAT -0.054000 !
ATOM O5P OST -0.141000 !
ATOM C10 CT3T -0.137000 !
ATOM H101 HAT 0.090000 !
ATOM H102 HAT 0.090000 !
ATOM H103 HAT 0.090000 !
ATOM C6P CAT -0.095000 !
ATOM H6P HPT 0.090000 !

BOND NA2 HA21 NA2 HA22 NA2 C2 C2 N1 N3 C4 C4 NA4 NA4 HA41 NA4 HA42
BOND C5 C6 C6 H6 C5 C7 C7 H71 C7 H72 C7 C1P C1P C2P C2P H2P
BOND C6P H6P C3P O3P O3P C8 C8 H81 C8 H82 C8 H83 C3P C4P C4P O4P
BOND O4P C9 C5P C6P C10 H103

```

BOND C9 H91 C9 H92 C9 H93 C5P O5P O5P C10 C10 H101 C10 H102
 DOUBLE C2 N3 C4 C5 N1 C6 C2P C3P C4P C5P C1P C6P
 DONOR NA2 HA21 NA2 HA22

IC	CAT	CAT	CT2T	HAT	1.380	120.000	0.000	111.000	1.100
IC	HPT	CAT	CAT	CT2T	1.080	120.000	0.000	120.000	1.500
IC	CAT	OST	CT3T	HAT	1.340	112.000	0.000	108.500	1.500
IC	NN1T	CAT	CAT	CAT	1.400	122.000	180.0	120.000	1.380
IC	HPT	CAT	CAT	OST	1.080	120.000	180.0	120.000	1.340
IC	HPT	CAT	CAT	CAT	1.080	120.000	0.000	120.000	1.380
IC	OST	CAT	CAT	OST	1.340	120.000	0.000	120.000	1.340
IC	CAT	CAT	NN1T	HN1T	1.380	122.000	0.000	113.000	1.030
IC	CAT	CAT	CAT	CAT	1.380	120.000	180.0	120.000	1.380
IC	CAT	CAT	CT2T	CAT	1.380	120.000	180.0	111.000	1.500
IC	CT2T	CAT	CAT	CAT	1.500	120.000	180.0	120.000	1.380
IC	NC2T	CAT	NC2T	CAT	1.330	128.000	180.0	115.000	1.330
IC	NC2T	CAT	CAT	CAT	1.330	123.000	180.0	120.000	1.380
IC	NN1T	CAT	NC2T	CAT	1.400	116.000	180.0	115.000	1.330
IC	CAT	NC2T	CAT	HPT	1.330	128.000	180.0	115.500	1.080
IC	HN1T	NN1T	CAT	NC2T	1.030	113.000	0.000	116.000	1.330
IC	CAT	CAT	OST	CT3T	1.380	120.000	180.0	112.000	1.400
IC	NC2T	CAT	CAT	CT2T	1.330	123.000	180.0	120.000	1.500
IC	OST	CAT	CAT	CAT	1.340	120.000	180.0	120.000	1.380
IC	NN1T	CAT	CAT	CT2T	1.400	122.000	180.0	120.000	1.500
IC	CAT	NC2T	CAT	CAT	1.330	115.000	180.0	123.000	1.380

RESI TMPP 1.0

ATOM	NA2	NN1T	-0.213000	!	HA41	HA42				
ATOM	HA21	HN1T	0.276000	!	\	/				
ATOM	HA22	HN1T	0.276000	!	NA4		H2P	H81		
ATOM	C2	CAT	-0.085000	!						
ATOM	N3	NC2T	-0.368000	!	C4	C7	C2P	O3P-C8-H81		
ATOM	N1	NC2T	-0.004000	!	/ \ \	/ \	/ \ \			
ATOM	C6	CAT	0.264000	!	N3	C5	C1P	C3P	H81	
ATOM	H6	HPT	0.090000	!						
ATOM	C4	CAT	0.133000	!	HA21	C2	C6	C6P	C4P	H91
ATOM	NA4	NN1T	-0.367000	!	\	/ \	//	\	// \	
ATOM	HA41	HN1T	0.291000	!	NA2	N1		C5P	O4P-C9-H92	
ATOM	HA42	HN1T	0.291000	!	/		/			
ATOM	C5	CAT	0.096000	!	HA22	H11		O5P	H93	
ATOM	C7	CT2T	-0.075000	!						
ATOM	H71	HAT	0.090000	!				101-C10-H102		
ATOM	H72	HAT	0.090000	!						
ATOM	C1P	CAT	-0.276000	!				H103		
ATOM	C2P	CAT	0.119000	!						
ATOM	H2P	HPT	0.090000	!						
ATOM	C3P	CAT	-0.011000	!						
ATOM	O3P	OST	-0.319000	!						
ATOM	C8	CT3T	-0.146000	!						
ATOM	H81	HAT	0.090000	!						
ATOM	H82	HAT	0.090000	!						
ATOM	H83	HAT	0.090000	!						
ATOM	C4P	CAT	0.024000	!						
ATOM	O4P	OST	-0.067000	!						
ATOM	C9	CT3T	-0.088000	!						
ATOM	H91	HAT	0.090000	!						
ATOM	H92	HAT	0.090000	!						
ATOM	H93	HAT	0.090000	!						
ATOM	C5P	CAT	-0.011000	!						
ATOM	O5P	OST	-0.319000	!						
ATOM	C10	CT3T	-0.146000	!						
ATOM	H101	HAT	0.090000	!						

```

ATOM    H102  HAT      0.090000  !
ATOM    H103  HAT      0.090000  !
ATOM    C6P   CAT      0.119000  !
ATOM    H6P   HPT      0.090000  !
ATOM    H11   HN2      0.346000  !

```

```

BOND  NA2 HA21  NA2 HA22  NA2 C2   C2 N1   N3 C4   C4 NA4   NA4 HA41  NA4 HA42
BOND  C5 C6   C6 H6   C5 C7   C7 H71  C7 H72  C7 C1P  C1P C2P  C2P H2P
BOND  C6P H6P  C3P O3P  O3P C8   C8 H81  C8 H82  C8 H83  C3P C4P  C4P O4P
BOND  C9 H91  C9 H92  C9 H93  C5P O5P  O5P C10  C10 H101  C10 H102
BOND  C10 H103 C5P C6P  H11 N1   O4P C9
DOUBLE C2 N3  C4 C5   N1 C6   C2P C3P  C4P C5P  C1P C6P
DONOR  NA2 HA21  NA2 HA22

```

```

IC  CAT   CAT   CT2T  HAT      1.380 120.000 0.000 111.000 1.100
IC  HPT   CAT   CAT   CT2T     1.080 120.000 0.000 120.000 1.500
IC  CAT   OST   CT3T  HAT      1.340 112.000 0.000 108.500 1.500
IC  NN1T  CAT   CAT   CAT      1.400 122.000 180.0 120.000 1.380
IC  HPT   CAT   CAT   OST      1.080 120.000 180.0 120.000 1.340
IC  HPT   CAT   CAT   CAT      1.080 120.000 0.000 120.000 1.380
IC  OST   CAT   CAT   OST      1.340 120.000 0.000 120.000 1.340
IC  CAT   CAT   NN1T  HN1T     1.380 122.000 0.000 113.000 1.030
IC  CAT   CAT   CAT   CAT      1.380 120.000 180.0 120.000 1.380
IC  CAT   CAT   CT2T  CAT      1.380 120.000 180.0 111.000 1.500
IC  CT2T  CAT   CAT   CAT      1.500 120.000 180.0 120.000 1.380
IC  NC2T  CAT   NC2T  CAT      1.330 128.000 180.0 115.000 1.330
IC  NC2T  CAT   CAT   CAT      1.330 123.000 180.0 120.000 1.380
IC  NN1T  CAT   NC2T  CAT      1.400 116.000 180.0 115.000 1.330
IC  CAT   NC2T  CAT   HPT      1.330 128.000 180.0 115.500 1.080
IC  HN1T  NN1T  CAT   NC2T     1.030 113.000 0.000 116.000 1.330
IC  CAT   CAT   OST   CT3T     1.380 120.000 180.0 112.000 1.400
IC  NC2T  CAT   CAT   CT2T     1.330 123.000 180.0 120.000 1.500
IC  OST   CAT   CAT   CAT      1.340 120.000 180.0 120.000 1.380
IC  NN1T  CAT   CAT   CT2T     1.400 122.000 180.0 120.000 1.500
IC  CAT   NC2T  CAT   CAT      1.330 115.000 180.0 123.000 1.380

```

PATCHING FIRST NONE LAST NONE

A2. Parameters files

BONDS

!V(bond) = Kb(b - b0)**2

!

!Kb: kcal/mole/A**2

!b0: A

!

!atom type Kb b0

!

```

NN1T  CAT   400.000  1.400
NN1T  HN1T  465.000  1.030
CAT   NC2T  365.000  1.330
CAT   CAT   340.000  1.380
CAT   HPT   380.000  1.080
CAT   CT2T  320.000  1.500
CT2T  HAT   365.000  1.100
CAT   OST   400.000  1.340

```

```

OST   CT3T   335.000   1.400
CT3T  HAT    355.000   1.090
NC2T  HN2    471.000   1.010

```

ANGLES

```

!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!

```

!atom types			Ktheta	Theta0	Kub	S0
NC2T	CAT	CAT	56.000	123.000		
CAT	CAT	CAT	60.000	120.000		
NC2T	CAT	NC2T	96.000	128.000		
NN1T	CAT	NC2T	82.500	116.000		
NN1T	CAT	CAT	101.000	122.000		
OST	CAT	CAT	82.500	120.000		
HN1T	NN1T	CAT	87.500	113.000		
HN1T	NN1T	HN1T	90.000	115.000		
CAT	NC2T	CAT	95.000	115.000		
NC2T	CAT	HPT	89.000	115.500		
HPT	CAT	CAT	48.000	120.000		
CAT	CAT	CT2T	86.000	120.000		
CAT	CT2T	CAT	145.000	111.000		
CAT	CT2T	HAT	71.000	111.000		
HAT	CT2T	HAT	64.000	105.000		
CAT	OST	CT3T	129.000	112.000		
OST	CT3T	HAT	77.000	108.500		
HAT	CT3T	HAT	48.000	108.000		
CAT	NC2T	HN2	45.000	115.600		

DIHEDRALS

```

!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types          Kchi    n    delta
!
CAT    CAT    CT2T    HAT    0.0000    3    0.0000
HPT    CAT    CAT     CT2T    0.0000    1    0.0000

```

CAT	OST	CT3T	HAT	0.0000	3	0.0000
NN1T	CAT	CAT	CAT	0.0000	2	180.00
HPT	CAT	CAT	OST	4.2000	2	180.00
HPT	CAT	CAT	CAT	0.0000	1	0.0000
OST	CAT	CAT	OST	0.0000	1	0.0000
CAT	CAT	NN1T	HN1T	0.0000	1	0.0000
CAT	CAT	CAT	CAT	1.9000	2	180.00
CAT	CAT	CT2T	CAT	1.6500	3	180.00
CT2T	CAT	CAT	CAT	2.9000	2	180.00
NC2T	CAT	NC2T	CAT	2.8000	2	180.00
NC2T	CAT	CAT	CAT	3.0000	2	180.00
NN1T	CAT	NC2T	CAT	2.9000	2	180.00
CAT	NC2T	CAT	HPT	2.9000	2	180.00
HN1T	NN1T	CAT	NC2T	0.5000	3	0.0000
CAT	CAT	OST	CT3T	1.8500	2	180.00
NC2T	CAT	CAT	CT2T	3.0000	2	180.00
OST	CAT	CAT	CAT	2.9000	2	180.00
NN1T	CAT	CAT	CT2T	3.0000	2	180.00
CAT	NC2T	CAT	CAT	2.9000	2	180.00
NN1T	CAT	NC2T	HN2	1.2000	2	180.00
NC2T	CAT	NC2T	HN2	0.0000	2	180.00
HN2	NC2T	CAT	HPT	2.5000	1	-64.16
HN2	NC2T	CAT	CAT	2.5000	2	180.00

NONBONDED nbxmod 5 atom cdiel shift vatom vdistance vswitch -

cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 el4fac 1.0 wmin 1.5

!

!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]

!

!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)

!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j

!

!atom	ignored	epsilon	Rmin/2	ignored	eps,1-4	Rmin/2,1-4
-------	---------	---------	--------	---------	---------	------------

!

CAT	0.0	-0.070000	1.992400			
CT2T	0.0	-0.055000	2.175000	0.0	-0.010000	1.900000
CT3T	0.0	-0.080000	2.060000	0.0	-0.010000	1.900000
HAT	0.0	-0.022000	1.320000			
HN1T	0.0	-0.046000	0.224500			
HPT	0.0	-0.030000	1.358200	0.0	-0.030000	1.358200
NC2T	0.0	-0.200000	1.850000			
NN1T	0.0	-0.200000	1.850000			
OST	0.0	-0.152100	1.770000			

END

BIBLIOGRAPHY

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol.* 2002;9(9):646-52.
2. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics.* 2012;41(1):429-52.
3. Mauldin RV, Carroll MJ, Lee AL. Dynamic dysfunction in dihydrofolate reductase results from antifolate drug binding: modulation of dynamics within a structural state. *Structure (London, England : 1993).* 2009;17(3):386-94.
4. London RE, Howell EE, Warren MS, Kraut J, Blakley RL. Nuclear magnetic resonance study of the state of protonation of inhibitors bound to mutant dihydrofolate reductase lacking the active-site carboxyl. *Biochemistry.* 1986;25(22):7229-35.
5. Turano P, Lalli D, Felli IC, Theil EC, Bertini I. NMR reveals pathway for ferric mineral precursors to the central cavity of ferritin. *Proc Natl Acad Sci USA.* 2010;107.
6. James NG, Byrne SL, Steere AN, Smith VC, MacGillivray RTA, Mason AB. Inequivalent Contribution of the Five Tryptophan Residues in the C-Lobe of Human Serum Transferrin to the Fluorescence Increase when Iron is Released. *Biochemistry.* 2009;48(13):2858-67.
7. Smith J. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q Rev Biophys.* 1991;24(03):227-91.
8. Kabsch W, Pai EF, Petskoli GA, Goody RS. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature.* 1990;345:24.
9. Cammarata M, Levantino M, Schotte F, Anfinrud PA, Ewald F, Choi J, et al. Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering. *Nat Methods.* 2008;5(10):881-6.
10. Siegbahn PEM, Himo F. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science.* 2011;1(3):323-36.
11. Gao J, Truhlar DG. Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem.* 2002;53(1):467-505.
12. Atilgan C, Okan OB, Atilgan AR. Network-Based Models as Tools Hinting at Nonevident Protein Functionality. *Annual Review of Biophysics.* 2012;41:205-25.
13. Atilgan AR, Atilgan C. Local motifs in proteins combine to generate global functional moves. *Briefings in Functional Genomics.* 2012;11(6):479-88.

14. Orozco M. A theoretical view of protein dynamics. *Chem Soc Rev.* 2014;43(14):5051-66.
15. Ming D, Wall ME. Quantifying allosteric effects in proteins. *Proteins: Structure, Function, and Bioinformatics.* 2005;59(4):697-707.
16. Kim MK, Jernigan RL, Chirikjian GS. Efficient Generation of Feasible Pathways for Protein Conformational Transitions. *Biophys J.* 2002;83(3):1620-30.
17. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proceedings of the National Academy of Sciences.* 2003;100(22):12570-5.
18. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett.* 1996;77(9):1905-8.
19. Hinsen K. Analysis of domain motions by approximate normal mode calculations. *Proteins: Structure, Function, and Bioinformatics.* 1998;33(3):417-29.
20. Doruker P, Jernigan RL, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J Comput Chem.* 2002;23(1):119-27.
21. Chennubhotla C, Rader A, Yang L-W, Bahar I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys Biol.* 2005;2(4):S173.
22. Durand P, Trinquier G, Sanejouand Y-H. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers.* 1994;34(6):759-71.
23. Kurkcuoglu O, Turgut OT, Cansu S, Jernigan RL, Doruker P. Focused Functional Dynamics of Supramolecules by Use of a Mixed-Resolution Elastic Network Model. *Biophys J.* 97(4):1178-87.
24. Ahmed A, Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins: Structure, Function, and Bioinformatics.* 2006;63(4):1038-51.
25. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys J.* 2001;80(1):505-15.
26. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.* 1997;2(3):173-81.
27. Kong Y, Ming D, Wu Y, Stoops JK, Zhou ZH, Ma J. Conformational flexibility of pyruvate dehydrogenase complexes: a computational analysis by quantized elastic deformational model. *J Mol Biol.* 2003;330(1):129-35.
28. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular Mechanisms of Chaperonin GroEL–GroES Function. *Biochemistry.* 2002;41(2):491-501.

29. Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences*. 2003;100(23):13253-8.
30. Whitford PC, Miyashita O, Levy Y, Onuchic JN. Conformational Transitions of Adenylate Kinase: Switching by Cracking. *J Mol Biol*. 2007;366(5):1661-71.
31. Baysal C, Atilgan AR. Elucidating the structural mechanisms for biological activity of the chemokine family. *Proteins: Struct Funct Bioinform*. 2001;43(2):150-60.
32. Tama F, Sanejouand Y-H. Conformational change of proteins arising from normal mode calculations. *Protein Eng*. 2001;14(1):1-6.
33. Zheng W, Brooks BR, Hummer G. Protein conformational transitions explored by mixed elastic network models. *Proteins: Structure, Function, and Bioinformatics*. 2007;69(1):43-57.
34. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins: Structure, Function, and Bioinformatics*. 2002;48(4):682-95.
35. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci U S A*. 2004;101(18):6957-62.
36. Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapère J-J. Normal Mode-Based Fitting of Atomic Structure into Electron Density Maps: Application to Sarcoplasmic Reticulum Ca-ATPase. *Biophys J*. 2005;88(2):818-27.
37. Kim MK, Jernigan RL, Chirikjian GS. An elastic network model of HK97 capsid maturation. *J Struct Biol*. 2003;143(2):107-17.
38. Kim MK, Li W, Shapiro BA, Chirikjian GS. A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA. *J Biomol Struct Dyn*. 2003;21(3):395-405.
39. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An Analysis of Core Deformations in Protein Superfamilies. *Biophys J*. 2005;88(2):1291-9.
40. Han R, Leo-Macias A, Zerbino D, Bastolla U, Contreras-Moreira B, Ortiz AR. An efficient conformational sampling method for homology modeling. *Proteins: Structure, Function, and Bioinformatics*. 2008;71(1):175-88.
41. Stember JN, Wriggers W. Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *The Journal of Chemical Physics*. 2009;131(7):074112.
42. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. Principles of flexible protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*. 2008;73(2):271-89.

43. Moritsugu K, Smith JC. Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian. *Biophys J*. 2007;93(10):3460-9.
44. Lyman E, Pfaendtner J, Voth GA. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys J*. 2008;95(9):4183-92.
45. Ma J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure*. 2005;13(3):373-80.
46. Atilgan C, Atilgan AR. Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. *PLoS Comput Biol*. 2009;5(10):e1000544.
47. Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. Manipulation of Conformational Change in Proteins by Single-Residue Perturbations. *Biophys J*. 2010;99(3):933-43.
48. Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: Linear response theory. *Phys Rev Lett*. 2005;94(7).
49. Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. *Proceedings of the National Academy of Sciences*. 2006;103(34):12724-8.
50. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Mol Biol*. 2003;10(9):738-43.
51. Abdizadeh H, Guven G, Atilgan AR, Atilgan C. Perturbation Response Scanning Specifies Key Regions in Subtilisin Serine Protease for both Function and Stability. *Journal of Enzyme Inhibition and Medicinal Chemistry*, in press. 2015.
52. Atilgan AR, Aykut AO, Atilgan C. Subtle pH differences trigger single residue motions for moderating conformations of calmodulin. *J Chem Phys*. 2011;135(15):155102.
53. Aykut AO, Atilgan AR, Atilgan C. Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin. *PLoS Comput Biol*. 2013;9(12):e1003366.
54. Guven G, Atilgan AR, Atilgan C. Protonation States of Remote Residues Affect Binding–Release Dynamics of the Ligand but Not the Conformation of Apo Ferric Binding Protein. *The Journal of Physical Chemistry B* 2014;118(40):11677-87.
55. Döring K, Surrey T, Nollert P, Jähnig F. Effects of ligand binding on the internal dynamics of maltose-binding protein. *Eur J Biochem*. 1999;266(2):477-83.
56. Plunkett G, Senear DF, Zuroske G, Ryan CA. Proteinase inhibitors I and II from leaves of wounded tomato plants: Purification and properties. *Arch Biochem Biophys*. 1982;213(2):463-72.
57. Pagán M, Solá RJ, Griebenow K. On the role of protein structural dynamics in the catalytic activity and thermostability of serine protease subtilisin Carlsberg. *Biotechnol Bioeng*. 2009;103(1):77-84.

58. Sheftel AD, Mason AB, Ponka P. The long history of iron in the Universe and in health and disease. *Biochimica Et Biophysica Acta-General Subjects*. 2012;1820(3):161-87.
59. Andrews NC. Molecular control of iron metabolism. *Best Practice & Research Clinical Haematology*. 2005;18(2):159-69.
60. Hentze MW, Muckenthaler MU, Galy B, Camaschella C. Two to Tango: Regulation of Mammalian Iron Metabolism. *Cell*. 2010;142(1):24-38.
61. Cheng Y, Zak O, Alsen P, Harrison SC, Walz T. Structure of the human transferrin receptor-transferrin complex. *Cell*. 2004;116(4):565-76.
62. Lambert LA, Perri H, Halbrooks PJ, Mason AB. Evolution of the transferrin family: Conservation of residues associated with iron and anion binding. *Comp Biochem Physiol B: Biochem Mol Biol*. 2005;142(2):129-41.
63. Sun H, Li H, Sadler PJ. Transferrin as a metal ion mediator. *Chem Rev*. 1999;99(9):2817-42.
64. Wally J, Halbrooks PJ, Vornheim C, Rould MA, Everse SJ, Mason AB, et al. The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. *J Biol Chem*. 2006;281(34):24934-44.
65. Noinaj N, Easley NC, Oke M, Mizuno N, Gumbart J, Boura E, et al. Structural basis for iron piracy by pathogenic *Neisseria*. *Nature*. 2012;483(7387):53-U92.
66. Yang N, Zhang H, Wang M, Hao Q, Sun H. Iron and bismuth bound human serum transferrin reveals a partially-opened conformation in the N-lobe. *Sci Rep*. 2012;2:999.
67. Rinaldo D, Field MJ. A computational study of the open and closed forms of the N-lobe human serum transferrin apoprotein. *Biophys J*. 2003;85(6):3485-501.
68. Aisen P, Leibman A, Zweier J. Stoichiometric and site characteristics of the binding of iron to human transferrin. *J Biol Chem*. 1978;253(6):1930-7.
69. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr J Chem*. 2013;53(3-4):199-206.
70. Bairoch A, Apweiler R. The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TrEMBL. *Nucleic Acids Res*. 1996;24(1):21-5.
71. Steere AN, Byrne SL, Chasteen ND, Mason AB. Kinetics of iron release from transferrin bound to the transferrin receptor at endosomal pH. *Biochim Biophys Acta*. 2012;1820(3):326-33.
72. Amin EA, Harris WR, Welsh WJ. Identification of possible kinetically significant anion-binding sites in human serum transferrin using molecular modeling strategies. *Biopolymers*. 2004;73(2):205-15.

73. Harris WR. Anion binding properties of the transferrins. Implications for function. *Biochim Biophys Acta*. 2012;1820(3):348-61.
74. Folajtar DA, Chasteen ND. Measurement of nonsynergistic anion binding to transferrin by EPR difference spectroscopy. *J Am Chem Soc*. 1982;104(21):5775-80.
75. Grady JK, Mason AB, Woodworth RC, Chasteen ND. The effect of salt and site-directed mutations on the iron(III)-binding site of human serum transferrin as probed by EPR spectroscopy. *Biochem J*. 1995;309 (Pt 2):403-10.
76. Byrne SL, Steere AN, Chasteen ND, Mason AB. Identification of a Kinetically Significant Anion Binding (KISAB) Site in the N-Lobe of Human Serum Transferrin. *Biochemistry*. 2010;49(19):4200-7.
77. Kretchmar SA, Raymond KN. Effects of ionic strength on iron removal from the monoferric transferrins. *Inorg Chem*. 1988;27(8):1436-41.
78. Nagel ZD, Klinman JP. A 21st century revisionist's view at a turning point in enzymology. *Nat Chem Biol*. 2009;5(8):543-50.
79. Benkovic SJ, Hammes-Schiffer S. A Perspective on Enzyme Catalysis. *Science*. 2003;301(5637):1196-202.
80. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007;450(7171):913-6.
81. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, et al. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science*. 2011;332(6026):234-8.
82. Baysal C, Atilgan AR. Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. *Proteins: Struct Funct Bioinform*. 2001;45(1):62-70.
83. Baysal C, Atilgan AR. Relaxation Kinetics and the Glassiness of Proteins: The Case of Bovine Pancreatic Trypsin Inhibitor. *Biophys J*. 2002;83(2):699-705.
84. Abdizadeh H, Guven G, Atilgan AR, Atilgan C. Perturbation response scanning specifies key regions in subtilisin serine protease for both function and stability. *J Enzyme Inhib Med Chem*. 2015;30(6):867-73.
85. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nature Communications*. 2015;6:7385.
86. Schnell JR, Dyson HJ, Wright PE. Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. *Annu Rev Biophys Biomol Struct*. 2004;33(1):119-40.
87. Fierke CA, Johnson KA, Benkovic SJ. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry*. 1987;26(13):4085-92.

88. Sawaya MR, Kraut J. Loop and Subdomain Movements in the Mechanism of *Escherichia coli* Dihydrofolate Reductase: Crystallographic Evidence. *Biochemistry*. 1997;36(3):586-603.
89. Heaslet H, Harris M, Fahnoe K, Sarver R, Putz H, Chang J, et al. Structural comparison of chromosomal and exogenous dihydrofolate reductase from *Staphylococcus aureus* in complex with the potent inhibitor trimethoprim. *Proteins: Structure, Function, and Bioinformatics*. 2009;76(3):706-17.
90. Dams T, Auerbach G, Bader G, Jacob U, Ploom T, Huber R, et al. The crystal structure of dihydrofolate reductase from *Thermotoga maritima*: molecular features of thermostability1. *J Mol Biol*. 2000;297(3):659-72.
91. Miller GP, Wahn DC, Benkovic SJ. Interloop contacts modulate ligand cycling during catalysis by *Escherichia coli* dihydrofolate reductase. *Biochemistry*. 2001;40(4):867-75.
92. Huovinen P, Sundström L, Swedberg G, Sköld O. Trimethoprim and sulfonamide resistance. *Antimicrob Agents Chemother*. 1995;39(2):279.
93. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet*. 2012;44(1):101-5.
94. Alekshun MN, Levy SB. Molecular mechanisms of antibacterial multidrug resistance. *Cell*. 2007;128(6):1037-50.
95. Watson M, Liu J-W, Ollis D. Directed evolution of trimethoprim resistance in *Escherichia coli*. *FEBS J*. 2007;274(10):2661-71.
96. Yang QX, Huang FY, Lin TH, Gelbaum L, Howell EE, Huang TH. Dynamics of trimethoprim bound to dihydrofolate reductase--a deuterium NMR study. *Solid State Nucl Magn Reson*. 1996;7(3):193-201.
97. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*. 2006;313(5793):1638-42.
98. Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S. Network of coupled promoting motions in enzyme catalysis. *Proceedings of the National Academy of Sciences*. 2002;99(5):2794-9.
99. Rod TH, Radkiewicz JL, Brooks CL. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proceedings of the National Academy of Sciences*. 2003;100(12):6980-5.
100. Osborne MJ, Schnell J, Benkovic SJ, Dyson HJ, Wright PE. Backbone Dynamics in Dihydrofolate Reductase Complexes: Role of Loop Flexibility in the Catalytic Mechanism†. *Biochemistry*. 2001;40(33):9846-59.

101. Osborne MJ, Venkitakrishnan RP, Dyson HJ, Wright PE. Diagnostic chemical shift markers for loop conformation and substrate and cofactor binding in dihydrofolate reductase complexes. *Protein Science : A Publication of the Protein Society*. 2003;12(10):2230-8.
102. McElheny D, Schnell JR, Lansing JC, Dyson HJ, Wright PE. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A*. 2005;102(14):5032-7.
103. Falzone CJ, Wright PE, Benkovic SJ. Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis. *Biochemistry*. 1994;33(2):439-42.
104. Wong KF, Selzer T, Benkovic SJ, Hammes-Schiffer S. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc Natl Acad Sci U S A*. 2005;102(19):6807-12.
105. Radkiewicz JL, Brooks CL. Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. *Journal of the American Chemical Society*. 2000;122(2):225-31.
106. Thorpe IF, Brooks CL. The coupling of structural fluctuations to hydride transfer in dihydrofolate reductase. *Proteins: Structure, Function, and Bioinformatics*. 2004;57(3):444-57.
107. Liu CT, Francis K, Layfield JP, Huang X, Hammes-Schiffer S, Kohen A, et al. *Escherichia coli* dihydrofolate reductase catalyzed proton and hydride transfers: Temporal order and the roles of Asp27 and Tyr100. *Proceedings of the National Academy of Sciences*. 2014;111(51):18231-6.
108. Liu CT, Hanoian P, French JB, Pringle TH, Hammes-Schiffer S, Benkovic SJ. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proceedings of the National Academy of Sciences*. 2013;110(25):10159-64.
109. Sham YY, Ma B, Tsai CJ, Nussinov R. Thermal unfolding molecular dynamics simulation of *Escherichia coli* dihydrofolate reductase: thermal stability of protein domains and unfolding pathway. *Proteins*. 2002;46(3):308-20.
110. Binder K. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. New York 1995.
111. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-802.
112. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996;14:33-8.
113. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem*. 1983;4:187-217.
114. Raiteri P, Gale JD. Water Is the Key to Nonclassical Nucleation of Amorphous Calcium Carbonate. *J Am Chem Soc*. 2010; 132: 17623–34

115. Garcia-Viloca M, Truhlar DG, Gao J. Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase. *Biochemistry*. 2003;42(46):13558-75.
116. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-35.
117. Darden T, Perera L, Li LP, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Struc Fold Des*. 1999;7: R55–R60.
118. Andersen HC. RATTLE: A “Velocity” version of the SHAKE algorithm for molecular dynamics calculations. *J Comput Phys*. 1983;52(1):24-34.
119. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pK(a) values for protein-ligand complexes. *Proteins: Struct Funct Bioinform*. 2008;73(3):765-83.
120. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H⁺⁺: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. 2005;33:W368-W71.
121. Clarkson J, Smith DA. UV Raman evidence of a tyrosine in apo-human serum transferrin with a low pKa that is elevated upon binding of sulphate. *FEBS Lett*. 2001;503(1):30-4.
122. Sun X, Sun H, Ge R, Richter M, Woodworth RC, Mason AB, et al. The low pKa value of iron-binding ligand Tyr188 and its implication in iron release and anion binding of human transferrin. *FEBS Lett*. 2004;573(1):181-5.
123. Kubal G, Sadler PJ, Tucker A. pH-induced structural changes in human serum apotransferrin pKa values of histidine residues and N-terminal amino group determined by 1H-NMR spectroscopy. *Eur J Biochem*. 1994;220(3):781-7.
124. Dewan JC, Mikami B, Hirose M, Sacchettini JC. Structural evidence for a pH-sensitive dilysine trigger in the hen ovotransferrin N-lobe - Implications for transferrin iron release. *Biochemistry*. 1993;32(45):11963-8.
125. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*. 2009;30(16):2785-91.
126. Brooks CL, Fleischman SH. A theoretical approach to drug design. 1. Relative solvation thermodynamics for the antibacterial compound trimethoprim and ethyl derivatives substituted at the 3', 4', and 5' positions. *J Am Chem Soc*. 1990;112(9):3307-12.
127. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid parameterization of small molecules using the force field toolkit. *J Comput Chem*. 2013;34(32):2757-70.
128. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09. Wallingford, CT, USA: Gaussian, Inc.; 2009.

129. Dassault Systèmes BIOVIA. Discovery Studio Modeling Environment, Release 4.5, San Diego: Dassault Systèmes. 2015.
130. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of chemical physics*. 2003;119(6):3559-66.
131. Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett*. 1997;78(14):2690.
132. Gore J, Ritort F, Bustamante C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences*. 2003;100(22):12564-9.
133. General IJ. A Note on the Standard State's Binding Free Energy. *J Chem Theory Comput*. 2010;6(8):2520-4.
134. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem*. 2016;59(9):4035-61.
135. Atilgan C, Atilgan AR. Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. *Plos Computational Biology*. 2009;5(10).
136. Yilmaz LS, Atilgan AR. Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. *Journal of Chemical Physics*. 2000;113(10):4454-64.
137. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Structure, Function, and Bioinformatics*. 2002;47(3):281-94.
138. Haspel N, Ricklin D, Geisbrecht BV, Kavraki LE, Lambris JD. Electrostatic contributions drive the interaction between *Staphylococcus aureus* protein Efb-C and its complement target C3d. *Protein Sci*. 2008;17(11):1894-906.
139. Koepke J, Ermler U, Warkentin E, Wenzl G, Flecker P. Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 Å resolution. Structural basis of Janus-faced serine protease inhibitor specificity1. *J Mol Biol*. 2000;298(3):477-91.
140. Hwang H, Vreven T, Janin J, Weng Z. Protein-Protein Docking Benchmark Version 4.0. *Proteins*. 2010;78(15):3111-4.
141. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph Model*. 1996;14(Journal of Molecular Graphics):33-8.
142. Filipek R, Rzychon M, Oleksy A, Gruca M, Dubin A, Potempa J, et al. The Staphostatin-Staphopain Complex: A FORWARD BINDING INHIBITOR IN COMPLEX WITH ITS TARGET CYSTEINE PROTEASE. *J Biol Chem*. 2003;278(42):40959-66.

143. Bhakta T, Whitehead SJ, Snaith JS, Dafforn TR, Wilkie J, Rajesh S, et al. Structures of the diI2dIII1 Complex of Proton-Translocating Transhydrogenase with Bound, Inactive Analogues of NADH and NADPH Reveal Active Site Geometries^{†,‡}. *Biochemistry*. 2007;46(11):3304-18.
144. . !!! INVALID CITATION !!!
145. Cukuroglu E, Gursoy A, Nussinov R, Keskin O. Non-Redundant Unique Interface Structures as Templates for Modeling Protein Interactions. *PLoS One*. 2014;9(1):e86738.
146. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*. 2001;98(18):10037-41.
147. Warshel A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *J Biol Chem*. 1998;273(42):27035-8.
148. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic Basis for Enzyme Catalysis. *Chem Rev*. 2006;106(8):3210-35.
149. Pautsch A, Vogelsong M, Tränkle J, Herrmann C, Aktories K. Crystal structure of the C3bot–RalA complex reveals a novel type of action of a bacterial exoenzyme2005 2005-10-19 00:00:00. 3670-80 p.
150. Thomas PG, Russell AJ, Fersht AR. Tailoring the pH dependence of enzyme catalysis using protein engineering. 1985.
151. Gouet P, Chinardet N, Welch M, Guillet V, Cabantous S, Birck C, et al. Further insights into the mechanism of function of the response regulator CheY from crystallographic studies of the CheY-CheA124-257 complex. *Acta Crystallographica Section D*. 2001;57(1):44-51.
152. Silversmith RE, Appleby JL, Bourret RB. Catalytic Mechanism of Phosphorylation and Dephosphorylation of CheY: Kinetic Characterization of Imidazole Phosphates as Phosphodonors and the Role of Acid Catalysis. *Biochemistry*. 1997;36(48):14965-74.
153. Thomas SA, Immormino RM, Bourret RB, Silversmith RE. Nonconserved Active Site Residues Modulate CheY Autophosphorylation Kinetics and Phosphodonor Preference. *Biochemistry*. 2013;52(13):2262-73.
154. Hege T, Feltzer RE, Gray RD, Baumann U. Crystal Structure of a Complex between *Pseudomonas aeruginosa* Alkaline Protease and Its Cognate Inhibitor: INHIBITION BY A ZINC–NH₂ COORDINATIVE BOND. *J Biol Chem*. 2001;276(37):35087-92.
155. Siezen RJ, de Vos WM, Leunissen JAM, Dijkstra BW. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng*. 1991;4(7):719-37.
156. Bontemps Y, Scamuffa N, Calvo F, Khatib AM. Potential opportunity in the development of new therapeutic agents based on endogenous and exogenous inhibitors of the proprotein convertases. *Med Res Rev*. 2007;27(5):631-48.

157. Klibanov AM. Improving enzymes by using them in organic solvents. *Nature*. 2001;409(6817):241-6.
158. Neidhart DJ, Petsko GA. The refined crystal structure of subtilisin Carlsberg at 2.5 Å resolution. *Protein Eng*. 1988;2(4):271-6.
159. Braxton S, Wells JA. Incorporation of a stabilizing calcium-binding loop into subtilisin BPN'. *Biochemistry*. 1992;31(34):7796-801.
160. Hedstrom L. Serine protease mechanism and specificity. *Chem Rev*. 2002;102(12):4501-24.
161. Atilgan C, Okan OB, Atilgan AR. How orientational order governs collectivity of folded proteins. *Proteins-Structure Function and Bioinformatics*. 2010;78(16):3363-75.
162. Zhao HM, Arnold FH. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Engineering*. 1999;12(1):47-53.
163. Strausberg SL, Ruan B, Fisher KE, Alexander PA, Bryan PN. Directed coevolution of stability and catalytic activity in calcium-free subtilisin. *Biochemistry*. 2005;44(9):3272-9.
164. Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences*. 1993;90(12):5618-22.
165. Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proceedings of the National Academy of Sciences*. 1986;83(11):3743-5.
166. Rao SN, Singh UC, Bash PA, Kollman PA. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature*. 1987;328(6130):551-4.
167. Halbrooks PJ, He QY, Briggs SK, Everse SJ, Smith VC, MacGillivray RTA, et al. Investigation of the mechanism of iron release from the C-lobe of human serum transferrin: Mutational analysis of the role of a pH sensitive triad. *Biochemistry*. 2003;42(13):3701-7.
168. Eckenroth BE, Steere AN, Chasteen ND, Everse SJ, Mason AB. How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proc Natl Acad Sci U S A*. 2011;108(32):13089-94.
169. Steere AN, Byrne SL, Chasteen ND, Smith VC, MacGillivray RTA, Mason AB. Evidence that His349 acts as a pH-inducible switch to accelerate receptor-mediated iron release from the C-lobe of human transferrin. *J Biol Inorg Chem*. 2010;15(8):1341-52.
170. Barber MF, Elde NC. Escape from bacterial iron piracy through rapid evolution of transferrin. *Science*. 2014;346(6215):1362-6.
171. Byrne SL, Chasteen ND, Steere AN, Mason AB. The Unique Kinetics of Iron Release from Transferrin: The Role of Receptor, Lobe-Lobe Interactions, and Salt at Endosomal pH. *J Mol Biol*. 2010;396(1):130-40.

172. Hirose M. The structural mechanism for iron uptake and release by transferrins. *Biosci, Biotechnol, Biochem.* 2000;64(7):1328-36.
173. Mason AB, He QY, Adams TE, Gumerov DR, Kaltashov IA, Nguyen V, et al. Expression, purification, and characterization of recombinant nonglycosylated human serum transferrin containing a C-terminal hexahistidine tag. *Protein Expr Purif.* 2001;23(1):142-50.
174. Jeffrey PD, Bewley MC, MacGillivray RTA, Mason AB, Woodworth RC, Baker EN. Ligand-Induced Conformational Change in Transferrins: Crystal Structure of the Open Form of the N-Terminal Half-Molecule of Human Transferrin. *Biochemistry.* 1998;37(40):13978-86.
175. MacGillivray RTA, Moore SA, Chen J, Anderson BF, Baker H, Luo Y, et al. Two High-Resolution Crystal Structures of the Recombinant N-Lobe of Human Transferrin Reveal a Structural Change Implicated in Iron Release. *Biochemistry.* 1998;37(22):7919-28.
176. Mujika JI, Escribano B, Akhmatskaya E, Ugalde JM, Lopez X. Molecular Dynamics Simulations of Iron- and Aluminum-Loaded Serum Transferrin: Protonation of Tyr188 Is Necessary To Prompt Metal Release. *Biochemistry.* 2012;51(35):7017-27.
177. Abdizadeh H, Atilgan A, Atilgan C. Detailed molecular dynamics simulations of human transferrin provide insights into iron release dynamics at serum and endosomal pH. *JBIC Journal of Biological Inorganic Chemistry.* 2015;20(4):705-18.
178. Yang AH, MacGillivray RT, Chen J, Luo Y, Wang Y, Brayer GD, et al. Crystal structures of two mutants (K206Q, H207E) of the N-lobe of human transferrin with increased affinity for iron. *Protein Sci : A Publication of the Protein Society.* 2000;9(1):49-52.
179. MacGillivray RTA, Bewley MC, Smith CA, He Q-Y, Mason AB, Woodworth RC, et al. Mutation of the Iron Ligand His 249 to Glu in the N-Lobe of Human Transferrin Abolishes the Dilysine "Trigger" but Does Not Significantly Affect Iron Release. *Biochemistry.* 2000;39(6):1211-6.
180. Day CL, Anderson BF, Tweedie JW, Baker EN. Structure of the recombinant N-terminal lobe of human lactoferrin at 2.0-Angstrom resolution. *J Mol Biol.* 1993;232(4):1084-100.
181. He QY, Mason AB, Tam BM, MacGillivray RTA, Woodworth RC. Dual role of Lys206-Lys296 interaction in human transferrin N-lobe: Iron-release trigger and anion-binding site. *Biochemistry.* 1999;38(30):9704-11.
182. Nurizzo D, Baker HM, He Q-Y, MacGillivray RTA, Mason AB, Woodworth RC, et al. Crystal Structures and Iron Release Properties of Mutants (K206A and K296A) That Abolish the Dilysine Interaction in the N-Lobe of Human Transferrin. *Biochemistry.* 2001;40(6):1616-23.
183. Baker HM, Nurizzo D, Mason AB, Baker EN. Structures of two mutants that probe the role in iron release of the dilysine pair in the N-lobe of human transferrin. *Acta Crystallogr Sect D Biol Crystallogr.* 2007;63(3):408-14.

184. Silva André MN, Sousa Paulo RH, Coimbra João TS, Brás Natércia F, Vitorino R, Fernandes Pedro A, et al. The glycation site specificity of human serum transferrin is a determinant for transferrin's functional impairment under elevated glycaemic conditions. *Biochem J*. 2014;461(1):33-42.
185. Williams J, Moreton K. The distribution of iron between the metal-binding sites of transferrin human serum. *Biochem J*. 1980;185:483-8.
186. Vincent JB, Love S. The binding and transport of alternative metals by transferrin. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2012;1820(3):362-78.
187. Wang J, Wang J, Song W, Yang X, Zong W, Liu R. Molecular mechanism investigation of the neutralization of cadmium toxicity by transferrin. *Phys Chem Chem Phys*. 2016;18(5):3536-44.
188. Fatemi SJ, Kadir FH, Moore GR. Aluminium transport in blood serum. Binding of aluminium by human transferrin in the presence of human albumin and citrate. *Biochem J*. 1991;280 (Pt 2):527-32.
189. Khalaila I, Allardyce CS, Verma CS, Dyson PJ. A Mass Spectrometric and Molecular Modelling Study of Cisplatin Binding to Transferrin. *ChemBioChem*. 2005;6(10):1788-95.
190. He QY, Mason AB, Woodworth RC, Tam BM, MacGillivray RTA, Grady JK, et al. Inequivalence of the two tyrosine ligands in the N-lobe of human serum transferrin. *Biochemistry*. 1997;36(48):14853-60.
191. Baker HM, He Q-Y, Briggs SK, Mason AB, Baker EN. Structural and Functional Consequences of Binding Site Mutations in Transferrin: Crystal Structures of the Asp63Glu and Arg124Ala Mutants of the N-Lobe of Human Transferrin. *Biochemistry*. 2003;42(23):7084-9.
192. He Q-Y, Mason AB, Woodworth RC, Tam BM, Wadsworth T, MacGillivray RTA. Effects of Mutations of Aspartic Acid 63 on the Metal-Binding Properties of the Recombinant N-Lobe of Human Serum Transferrin. *Biochemistry*. 1997;36(18):5522-8.
193. Bobst CE, Zhang M, Kaltashov IA. Existence of a Noncanonical State of Iron-Bound Transferrin at Endosomal pH Revealed by Hydrogen Exchange and Mass Spectrometry. *J Mol Biol*. 2009;388(5):954-67.
194. Mathies G, Gast P, Chasteen ND, Luck A, Mason A, Groenen EJ. Exploring the Fe(III) binding sites of human serum transferrin with EPR at 275 GHz. *JBIC Journal of Biological Inorganic Chemistry*, in press. 2014.
195. Steinlein LM, Ligman CM, Kessler S, Ikeda RA. Iron Release Is Reduced by Mutations of Lysines 206 and 296 in Recombinant N-Terminal Half-Transferrin. *Biochemistry*. 1998;37(39):13696-703.
196. Mason AB, Byrne SL, Everse SJ, Roberts SE, Chasteen ND, Smith VC, et al. A loop in the N-lobe of human serum transferrin is critical for binding to the transferrin receptor as revealed by mutagenesis, isothermal titration calorimetry, and epitope mapping. *J Mol Recognit*. 2009;22(6):521-9.

197. Bali PK, Aisen P. Receptor-modulated iron release from transferrin-differential-effects on N-terminal and C-terminal sites. *Biochemistry*. 1991;30(41):9947-52.
198. Huebers H, Josephson B, Huebers E, Csiba E, Finch C. Uptake and release of iron from human transferrin. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*. 1981;78(4):2572-6.
199. Zak O, Aisen P. Iron Release from Transferrin, Its C-Lobe, and Their Complexes with Transferrin Receptor: Presence of N-Lobe Accelerates Release from C-Lobe at Endosomal pH†. *Biochemistry*. 2003;42(42):12330-4.
200. Pakdaman R, Abdallah FB, El Hage Chahine J-M. Transferrin, is a mixed chelate-protein ternary complex involved in the mechanism of iron uptake by serum-transferrin in Vitro. *J Mol Biol*. 1999;293(5):1273-84.
201. Hémadi M, Ha-Duong NT, El Hage Chahine JM. The Mechanism of Iron Release from the Transferrin-Receptor 1 Adduct. *J Mol Biol*. 2006;358(4):1125-36.
202. Steere AN, Miller BF, Roberts SE, Byrne SL, Chasteen ND, Smith VC, et al. Ionic Residues of Human Serum Transferrin Affect Binding to the Transferrin Receptor and Iron Release. *Biochemistry*. 2012;51(2):686-94.
203. El Hage Chahine J-M, Hémadi M, Ha-Duong N-T. Uptake and release of metal ions by transferrin and interaction with receptor 1. *Biochim Biophys Acta Biochim Biophys Acta*. 2012;1820(3):334-47.
204. Bellounis L, Pakdaman R, El Hage Chahine JM. Apotransferrin proton dissociation and interactions with hydrogencarbonate in neutral media. *J Phys Org Chem*. 1996;9(2):111-8.
205. He QY, Mason AB, Woodworth RC, Tam BM, MacGillivray RTA, Grady JK, et al. Mutations at nonliganding residues Tyr-85 and Glu-83 in the N-lobe of human serum transferrin - Functional second shell effects. *J Biol Chem*. 1998;273(27):17018-24.
206. Pakdaman R, El Hage Chahine JM. A mechanism for iron uptake by transferrin. *Eur J Biochem*. 1996;236(3):922-31.
207. Grossmann JG, Mason AB, Woodworth RC, Neu M, Lindley PF, Hasnain SS. Asp Ligand Provides the Trigger for Closure of Transferrin Molecules: Direct Evidence from X-ray Scattering Studies of Site-specific Mutants of the N-terminal Half-molecule of Human Transferrin. *J Mol Biol*. 1993;231(3):554-8.
208. Grossmann JG, Crawley JB, Strange RW, Patel KJ, Murphy LM, Neu M, et al. The nature of ligand-induced conformational change in transferrin in solution. An investigation using X-ray scattering, XAFS and site-directed mutants. *J Mol Biol*. 1998;279(2):461-72.
209. Sheff D, Pelletier L, O'Connell CB, Warren G, Mellman I. Transferrin receptor recycling in the absence of perinuclear recycling endosomes. *J Cell Biol*. 2002;156(5):797-804.

210. Dhungana S, Taboy CH, Anderson DS, Vaughan KG, Aisen P, Mietzner TA, et al. The influence of the synergistic anion on iron chelation by ferric binding protein, a bacterial transferrin. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(7):3659-64.
211. Heymann JJ, Weaver KD, Mietzner TA, Crumbliss AL. Sulfate as a Synergistic Anion Facilitating Iron Binding by the Bacterial Transferrin FbpA: The Origins and Effects of Anion Promiscuity. *J Am Chem Soc*. 2007;129(31):9704-12.
212. Siburt CJP, Mietzner TA, Crumbliss AL. FbpA - A bacterial transferrin with more to offer. *Biochimica Et Biophysica Acta (BBA)-General Subjects*. 2012;1820(3):379-92.
213. Lawrence CM, Ray S, Babyonyshev M, Galluser R, Borhani DW, Harrison SC. Crystal structure of the ectodomain of human transferrin receptor. *Science*. 1999;286(5440):779-82.
214. He QY, Mason AB, Nguyen V, MacGillivray RT, Woodworth RC. The chloride effect is related to anion binding in determining the rate of iron release from the human transferrin N-lobe. *Biochem J*. 2000;350(Pt 3):909-15.
215. Bou-Abdallah F, Terpstra TR. The thermodynamic and binding properties of the transferrins as studied by isothermal titration calorimetry. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2012;1820(3):318-25.
216. Martin R, Savory J, Brown S, Bertholf R, Wills M. Transferrin binding of Al^{3+} and Fe^{3+} . *Clin Chem*. 1987;33(3):405-7.
217. Baysal C, Erman B, Chu B. Conformational features of poly(1,1-dihydroperfluorooctyl acrylate) and poly(vinyl acetate) diblock oligomers in supercritical carbon dioxide. *The Journal of Chemical Physics*. 2001;114(12):5444-9.
218. Feeney J. NMR Studies of Ligand Binding to Dihydrofolate Reductase. *Angew Chem Int Ed*. 2000;39(2):290-312.
219. Polshakov VI, Birdsall B, Feeney J. Characterization of Rates of Ring-Flipping in Trimethoprim in Its Ternary Complexes with *Lactobacillus casei* Dihydrofolate Reductase and Coenzyme Analogues. *Biochemistry*. 1999;38(48):15962-9.
220. Searle MS, Forster MJ, Birdsall B, Roberts GC, Feeney J, Cheung HT, et al. Dynamics of trimethoprim bound to dihydrofolate reductase. *Proc Natl Acad Sci U S A*. 1988;85(11):3787-91.
221. Ohmae E, Iriyama K, Ichihara S, Gekko K. Nonadditive effects of double mutations at the flexible loops, glycine-67 and glycine-121, of *Escherichia coli* dihydrofolate reductase on its stability and function. *J Biochem*. 1998;123(1):33-41.