# A COMPUTATIONAL STUDY ON BACTERIAL TRIGGER FACTOR AND RIBOSOME DYNAMICS

by

Gökçe Ezeroğlu

B.S., Chemical Engineering, Middle East Technical University, 2011

Submitted to the Institute for Graduate Studies in

Science and Engineering in partial fulfillment of

the requirements for the degree of

Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2014

# ACKNOWLEDGEMENTS

# ABSTRACT

## A COMPUTATIONAL STUDY ON BACTERIAL TRIGGER FACTOR AND RIBOSOME DYNAMICS

Trigger factor (TF), found in bacterial cells and chloroplasts, is the first chaperone that welcomes the newly synthesized nascent polypeptide chains to acquire their native fold. In this thesis, four independent 200 ns long molecular dynamics (MD) simulations were performed in explicit solvent at different ionic strengths (29 mM, 50 mM and two simulations for 150 mM). Root mean square deviations from initial structure and residue mean square fluctuations from the average structure indicated high conformational flexibility of TF. Large oscillations in gyration radius were also observed, but a fully collapsed state reported recently was not detected. A unique coil to helix formation in the binding loop took place during 29 mM run. Principal component analysis of MD trajectories and anisotropic network modelling (ANM) calculations were carried out to reveal the collective dynamics of apo TF. Domain decompositions within TF obtained from both of these methods were consistent. In the first collective mode, there is generally an anti-correlated motion between the three major domains, namely binding (BD), core (CD) and head (HD) domains. HD motion with respect to CD is more dominant at 150 mM ionic strength, however opening/closing between BD-CD and CD-HD appear at the same time at 29 mM and 50 mM runs. All snapshots from four simulations were clustered after performing overall alignment and domain-wise alignment. Resulting clusters further elucidated the inter-domain and intra-domain conformational changes. Dynamics of TF-ribosome complex were investigated after docking centroids from major clusters of apo TF runs on to a fragment of BD that was co-crystallized with ribosome 50S subunit. ANM was applied on eight different ribosome 50S-docked TF complexes. The collective dynamics were modified upon complex formation depending on the conformation and position of TF docked on 50S. For most docked conformers, HD was larger compared to apo TF, moving again in an anti-correlated manner with respect to BD and CD. In two of the docked conformers, TF moved as a fully correlated entity on 50S. Distribution of total solvent accessiable surface values calculated for all structure was similar to the distribution of gyration radii.

# ÖZET

## BAKTERİLERDE BULUNAN TRİGGER FACTOR PROTEİNİ VE RİBOZOM DİNAMİKLERİ ÜZERİNE HESAPSAL ÇALIŞMA

Bakteri ve kloroplast hücrelerinde bulunan 'Trigger factor' (TF) proteini, yeni sentezlenen protein zincirleriyle ile ilk etkileşen ve onların özgün katlanmış yapılarına ulaşmalarına yardımcı olan bir proteindir. Bu tezde 200 ns uzunluğunda dört bağımsız moleküler dinamik (MD) simülasyonu su ortamında farklı iyonik güçlerde gerçekleştirilmiştir (29 mM, 50 mM ve 150 mM için iki simülasyon). Arayapıların başlangıç yapısından kök ortalama kare sapmaları ve rezidülerin ortalama yapıya göre ortalama kare dalgalanmaları TF'teki yüksek esnekliği göstermiştir. Jirasyon yarıçapında yüksek salınımlar gözlenmiş olmasına rağmen literatürde TF için belirtilmiş tam katlanmış hal tespit edilmemiştir. 29 mM simülasyonunda bağlanma bölgesinde ilmekten heliks oluşumu gerçekleşmiştir. Simülasyonların temel bileşen analizleri ve anizotropik ağyapı modeli (ANM) hesaplamaları ile serbest TF proteininin kollektif dinamikleri incelenmiştir. İki metot ile elde edilen bölge dağılımları birbirleriyle uyumludur. İlk kollektif modda, üç ana bölge olan bağlanma bölgesi (BD), merkez bölgesi (BD) ve kafa bölgesi (HD) birbirbirleriyle zıt hareket etmektedir. 150 mM simülasyonu için kafa bölgesinin merkez üzerine olan hareketi daha baskınken, 29 mM ve 50 mM simülasyonlarında bağlanma-merkez ve merkez-kafa bölgelerindeki açılıp kapanma aynı andadır. Dört simülasyonu oluşturan tüm arayapılar proteinin tümü ve bölgeler bazında hizalanap, kümelenmiştir. Elde edilen kümeler, bölge içi ve bölgeler arası yapısal değişimleri izah etmiştir. Büyük kümelerdeki yapılar, TF ile birlikte kristalize edilmiş olan 50S ribozom altbiriminin bağlanma bölgesine yanaştırılmış ve ribozome-TF kompleksinin dinamikleri incelenmiştir. ANM sekiz farklı 50S-TF yanaştırılmış kompleksine uygulanmıştır. TF'nin 50S üzerindeki pozisyonu ve konformasyonuna bağlı olarak kollektif dinamikleri değişmiştir. Yanaştırılmış komplekslerin çoğunda, kafa bölgesi serbest haldeki TF'e göre daha geniştir, ve serbest yapıdaki gibi bağlanma ve merkez bölgeleriyle zıt olrak hareket etmektedir. İki yanaştırılmış yapıda ise kafa bölgesi 50S ile tamamen ilişkili olarak hareket etmektedir. Toplam çözücü ulaşabilir yüzey alanı ve jirasyon yarıçapının dağılımları uyumludur.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS

| | |
|---|---|
| Å | Angstrom |
| $C$ | Covariance Matrix |
| $Cij$ | Normalize cross correlations |
| Cl | Chlorine |
| $dt$ | Time step |
| $F_i$ | Force on atom i |
| fs | femtosecond |
| $h$ | Heaviside constant |
| **H** | Hessian Matrix |
| $i$ | Any Residue |
| °K | Degree in Kelvin |
| kDa | Kilodalton |
| $m_i$ | Mass of particle i |
| mM | mili molar |
| $N$ | Number of residues |
| Na | Sodium |
| nm | Nanometer |
| ns | nanosecond |
| $O_{ij}$ | Overlap between i and j |
| ps | picosecond |
| $r_c$ | Cutoff distance |
| $r_{com}$ | Center of mass of the chain |
| $R_g$ | Gyration radius |
| $r_i$ | Position of particle i |
| $r_i^{ave}$ | Average position |
| $R_{ij}$ | Distance between node pair i and j. |
| $r_i^{ref}$ | Position of initial Structure |
| $t$ | Time |
| T | Total number of recorded snaphots |

| | |
|---|---|
| **T** | Transpose |
| **u** | Eigen vector |
| **v** | Eigen vector |
| $V$ | Potential energy |
| $v_i$ | Velocitiy of particle i |
| | |
| $\alpha_i$ | Acceleration of particle i |
| $\gamma$ | Force constant |
| $\Delta\boldsymbol{R}$ | Fluctuation vector |
| ~ | Almost or equal to |
| ° | Degree |

# LIST OF ACRONYMS/ABBREVIATIONS

| | |
|---|---|
| AMBER | Assisted Model Building with Energy Refinement |
| ANM | Anisotropic Network Model |
| Atm | Atmosphere |
| BD | Binding Domain |
| CD | Core Domain |
| CHARMM | Chemistry at Harvard Macromolecular Mechanics |
| COM | Center of Mass |
| D. radiodurans | Deinococcus Radiodurans |
| DNA | Deoxyribonucleic Acid |
| E. coli | Escherichia Coli |
| H | Hessian Matrix |
| H. marismortui | Haloarcula Marismortui |
| HD | Head Domain |
| Hsp | Heat Shock Protein |
| MD | Molecular Dynamics |
| MMTSB | Multiscale Modelling Tools for Structural Biology |
| NAMD | Nanoscale Molecular Dynamics Program |
| PCA | Principal Component Analysis |
| PDB | Protein Data Bank |
| PME | The Particle-mesh Ewald |
| PPIase | Peptidylprolyl Isomerase |
| RMSD | Root Mean Square Deviation |
| RMSF | Root mean Square Fluctuation |
| RNA | Ribonucleic Acid |
| SASA | Solvent Accessible Surface Area |
| sHsp | Small Heat Shock Protein |
| TF | Trigger Factor |
| VMD | Visual Molecular Dynamics |

# 1. INTRODUCTION

Ribosome that is the basic unit of any cell, works as a macromolecular machine taking major part in protein synthesis (Sauter, 2011). Proteins are synthesized from ribosome as polypeptide chains. After synthesis, they take a well-defined three-dimensional shape, which is essential for their functions. For a polypeptide chain, taking this well-defined structure, or so-called native structure, is defined as protein folding (Bruce *et al*., 2002).

Folding is vital for a protein to do its function; however, it is a highly error prone process (Herczenik *et al*., 2008). During folding of newly synthesized polypeptide chains, misfolding or partial folding can occur, which leads to loss of protein function and its subsequent degradation. Fatal disorders can also be observed upon misfolding, such as Parkinson's and Alzheimer's diseases (Cohen and Kelly, 2003). Therefore, proteins need the assistance of other proteins, named as chaperones, for folding into the correct native structure.

For the last 35 years or so, experimental studies on chaperones have been carried out to understand its function in the cell (Quan and Bardwell, 2012). Trigger factor (TF), which is the subject of this thesis, is a ribosome associated molecular chaperone, and found in bacteria and chloroplasts (Hoffman *et al.*, 2010). TF was first discovered in 1988, and suggested to take a role in the early stages of protein synthesis (Lill *et al*., 1988).

Up to now, the best characterized TF structure available in protein data bank (PDB) (Berman *et al*., 2000) is present for *Escherichia coli* (*E. coli*) with PDB ID: 1W26 (Ferbitz *et al*., 2004). 48 kDa and 432 residue long TF, is the first chaperone that welcomes the newly synthesized nascent chains, coming out of the ribosomal exit tunnel. It provides a cage-like structure for polypeptide chains, and helps nascent chains to acquire their native fold.

TF attaches to the L23 ribosomal protein which is located on the exit of the ribosomal tunnel (Ferbitz *et al*., 2004). To date, a 113 residue long *E. coli* TF was co-crystallized

with the 50S subunit of *Deinococcus Radiodurans* (*D. radiodurans)* (PDB ID: 2AAR) (Baram et. al., 2005) in full atomistic resolution. Secondly, a 35 residue long *E. coli* TF was co-crystallized with 50S subunit of *Haloarcula Marismortui* (*H. marismortui* ) (PDB ID: 1W2B) (Ferbitz et. al., 2004) in alpha-carbon resolution

In this thesis, the structure-function relationship of *E. coli* TF was studied by utilizing molecular dynamics (MD) simulations and anisotropic network model (ANM) (Atilgan *et al*., 2001). The collective dynamics, such as conformational flexibility and domain decompositions, are analyzed for both apo TF and several TF conformers docked on to the ribosome 50S subunit using the available TF fragment co-crystallized with 50S. Four independent 200 ns long MD simulations were performed on apo TF at different ionic strengths (one run at 29 mM, another run at 50 mM and two runs at 150 mM) for this purpose. The trajectories were analyzed in detail and compared with two recent MD studies on apo TF (PDB ID: 1W26) (Singhal *et al*., 2013; Thomas *et al*., 2013). So far, there is no reported study on dynamics of intact TF complexed with ribosome. This thesis provides a first ANM-based approach for understanding the collective dynamics features of such a huge complex.

# 2.  THEORETICAL BACKGROUND

## 2.1.  Protein Folding and Molecular Chaperones

Proteins, which participate in almost every biological process, are the most versatile biological structures. Each protein is composed of at least one polypeptide chain, and characterized by the sequence and the number of amino acids in the chain. A polypeptide chain consists of up to 20 different amino acids. The number and the sequence of the amino acids for a polypeptide chain are determined by the information encoded in DNA.

After being synthesized by the ribosome, each polypeptide chain adopts a unique three-dimensional structure, which determines its function. In molecular biology, folding refers to the process, by which a newly synthesized nascent polypeptide chain adopts its stable, compact three-dimensional shape, called as the native structure. Folding process that is essential for a polypeptide chain to perform its function is driven by two dominant forces, namely hydrophobic interactions (Kauzmann, 1959) and conformational entropy (Harano and Kinoshita, 2005).

Folding is a complex process due to the many weak, non-covalent interactions. A polypeptide chain take large number of possible conformations. Although principles of folding is studied for many years, how the final fold is determined by aminoacid sequence is still unclear (Fersht, 2008).

Chaperones are proteins that interact, stabilize or help another protein to take its native structure, without being present in the final structure (Hartl 1996; Hartl and Hayer-Hartl, 2009). Different classes of structurally unrelated chaperones have been described. Their members are often known as stress proteins or heat shock proteins (Hsp), small heat shock proteins (sHsp), ribosome associated proteins and chaperonins. Heat shock proteins are grouped based on their molecular weight on the order of kilodalton, namely Hsp40, Hsp60, Hsp70, Hsp90, Hsp100 (Chang *et al*., 2007; Tang *et al*., 2007). Hsp keeps the integrity and function of other cellular proteins in stressful conditions they are up-regulated by cells and prevents protein deformation (Ellis, 1987). Chaperonins consist of two groups,

as group I and group II. (Ellis 1990, Gupta, 1990; Trent *et al.*, 1991; Kubota *et al*., 1994; Valpuesta *et al*., 2005; Horwich *et al*., 2007 ) Group I is found in all eubacteria and endosymbiotic organelles like mitochondria, chloroplasts, and related organelles like hydrogenosomes and mitosomes whereas group II chaperonins are present in archaebacteria and in the cytosol of all eukaryotes (Bukau *et al.*, 1998; Gutsche *et al.*, 1999; Ellis and Hartl, 1999; Bui *et al*., 1996) The GroEL/GroES complex found in *E.coli* is the best characterized chaperonin complex. This complex prevents aggregate formation (Fenton *et al*., 2003) Ribosome-associated chaperones prevent misfolding and aggregation of newly synthesized proteins. In eubacteria, trigger factor (TF) is the first ribosome – associated chaperone to bind to the nascent chain (Blaha *et al.,* 2003).

## 2.2. Bacterial Trigger Factor

*E coli* Trigger factor (TF), which is so far the best understood ribosome-associated chaperone, is found in all bacteria and chloroplasts. It is the first chaperone that encounters the nascent polypeptide chain, coming from the ribosome tunnel ( Hoffmann *et al.*, 2010).

*E.coli* TF, which is the topic of this thesis, is a 48 kDa and 432 residue long dimer chaperone. It has a dragon-shaped structure, composed of three domains, namely N-terminal domain, peptidylprolyl isomerase (PPIase) domain, and C-terminal domain. N-terminal domain (residuesA1-A110), also called the "Binding Domain-BD" forms the tail of this dragon shaped *E. Coli* TF structure. PPIase (residues A150-A245) and C-terminal domain (A111-A149 and A 246-A432) form the ''Head Domain-HD'' and ''Core Domain-CD'' of this molecule, respectively (Hoffman et. al, 2010). In Figure 2.1, the domains of TF is represented.

N-terminal domain carrying a conserved signiture motif binds to the 50S subunit of the ribosome. It specifically binds to the L23 ribosomal protein that is located at the tunnel exit. N-terminal and C-terminal domains form an open cavity with two arms that are parts of C-terminal domain PPIase domain, located at the distal end, is proposed to prolong the residence time of the protein within the cavity, and support the co-translational (folding during the synthesis) folding of polypeptide chain (Ferbitz *et al*., 2004).
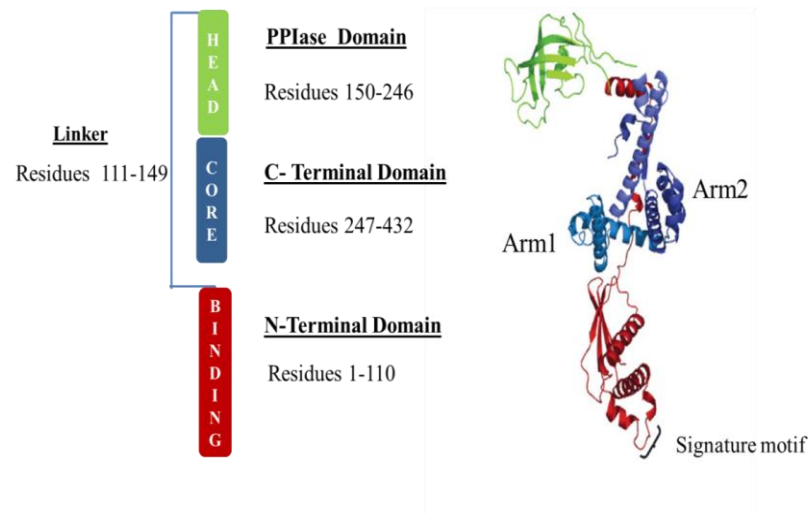
Figure 2.1. Domain regions of *E. coli* TF (Hoffman *et al*., 2010).

TF binds to unfolded nascent polypeptides to delay folding until sufficient information is present to allow stable folding of the newly synthesized protein (Agasha *et al*., 2004). In Figure 2.2, the cage-like structure of TF together with L23 (blue) and L29 (red) ribosomal binding domains are depicted. The same coloring scheme is used, as in Figure 2.1.
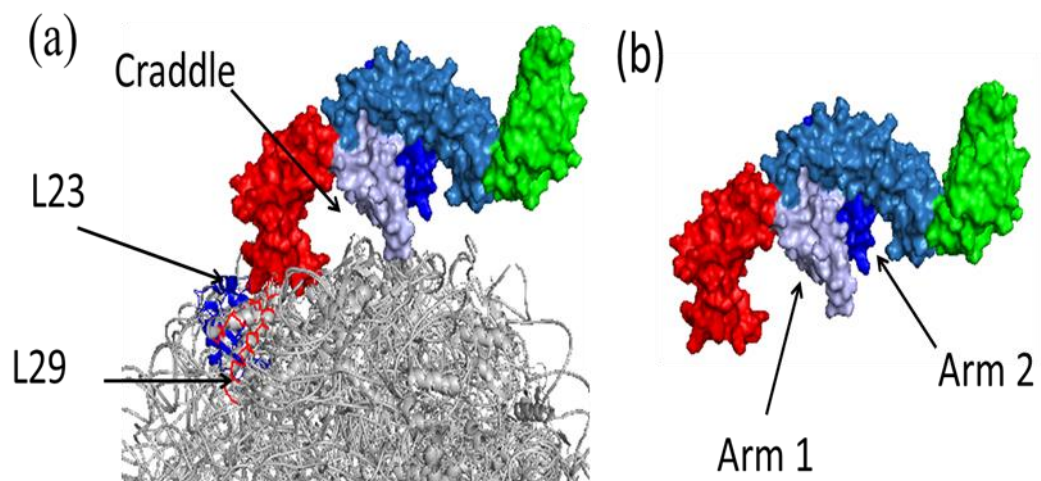


Figure 2.2 Structure of TF (a) cage like structure of TF aligned on to the exit of ribosomal tunnel with L23 and L29 ribosomal proteins (b) apo-TF.

## 2.3.    TF-ribosome Interaction

In nature, prokaryotes carry 70S ribosome composed of two main subunits: a small (30S) and a large (50S) subunit. The large subunit carrying the ribosomal exit tunnel is composed of 5S RNA subunit (120 nucleotides), 23S RNA subunit (2900 nucleotides) and 31 proteins (Bruce et. al., 2002). Synthesized nascent chains emerge from the ribosomal exit tunnel. In bacterial ribosome, the tunnel formed predominantly by 23S rRNA and segments of ribosomal proteins L4, L22 and L23. There is a huge constriction in the tunnel around 30 Å from the peptidyl transferase center. At this point, a conserved β-hairpin loop of L22 comes into close proximity to an extended loop of L4. At its distal end, the ribosomal exit tunnel widens up. The rim of the exit point is composed of RNA, a ring of four ubiquitously conserved ribosomal proteins (L22, L23, L24 and L29) and additional specific proteins (Kramer, 2009). In Figure 2.3, TF conformation aligned on 50S is depicted with L29 (red), L23 (blue), L22 (green) and L24 (orange) ribosomal proteins.
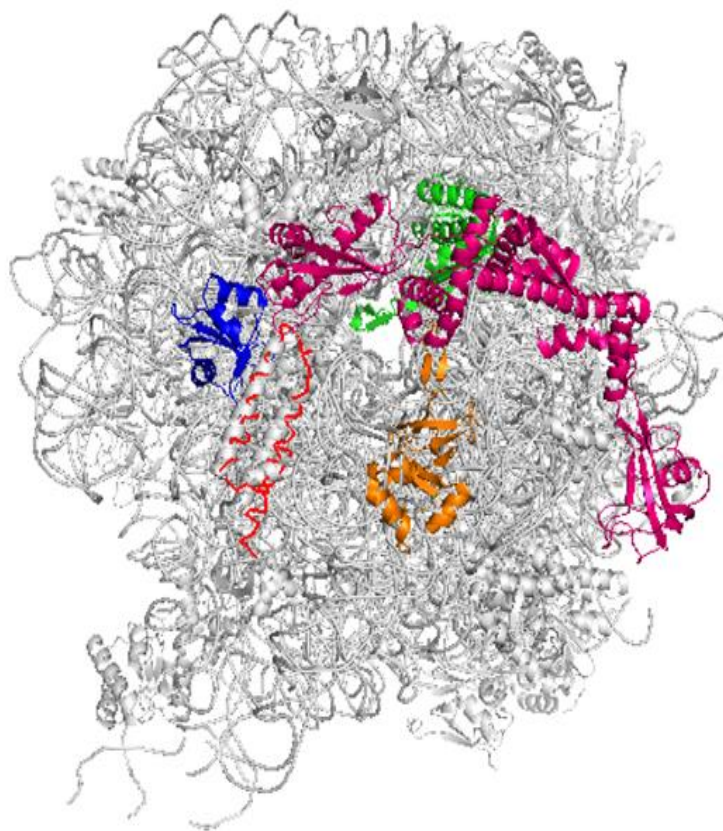


Figure 2.3. TF conformation aligned on 50S.

## 2.4. Computational Studies on TF

Recently, two MD studies have been published that discuss the dynamics of TF in explicit solvent (Singhal *et al*., 2013; Thomas *et al*., 2013). In the first study (Singhal et. al, 2013), 12 independent 250 ns long, atomistic MD simulations were performed at 50 mM ionic strength and around room temperature (~295 °K). Eight simulations were performed with AMBER03 forcefield parameter (Duan *et al*., 2003) and OPLS-AA forcefield (Kaminski *et al*., 2001) parameter was utilized for the remainings.

Conformational states of TF were categorized by looking at the center of mass (COM) distances between the domain pairs, namely BD-arm1 and HD-arm2. Observed TF conformations were categorized as fully collapsed (FC), semi-collapsed (SC), and collapsed-deformed (CD). FC state was observed when both of the domain pairs were stacked together. However, SC state was observed when only the HD-arm1 collapsed. On the other hand, CD state was observed in the case of collapsed and further deformation in BD-arm2 domain pair. FC state was observed in six of the trajectories performed with AMBER03 forcefield, remaining one simulation at the same force field showed CD state.

GROMACS 4.5.1 package (Hess *et al*., 2008) was utilized for the preparation of the systems in the reported study. AMBER03 structures were solvated in dodecahedron periodic box with a diameter of 13.2 nm. Same box with a diameter of 14 nm was used as a solvation box for the other structures.

In the second study (Thomas *et al*., 2013) two 1.5 μs long MD runs are performed with OPLS-AA/L and AMBER ff99SB-ILDN (Hornak *et al*., 2006; Lindorff-Larsen *et al*., 2010) force field parameteres. GROMACS 4.5 (Hess *et al*., 2008; Berendsen, 1995) package is utilized. Around 500 ns later, OPLS simulation predicts a more compact state by exhibiting a shorter distance for head domain- arm1 pair.

# 3.   METHODS

## 3.1.  Molecular Dynamics (MD) Simulation

Molecular dynamics simulation is a computational method that enables us to investigate the equilibrium and dynamic properties of biological systems. MD provides detailed information about the time dependent motions of atoms, which cannot be studied experimentally. It enables us to investigate the conformational dynamics of biological structures.

MD method is based on the Newton's second law, and defined as;

$$F_i = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r_i}{dt^2} \tag{3.1}$$

where $t$ is time and, $F_i$, $m_i$, $a_i$, $v_i$, $r_i$, are the force, mass, acceleration, velocity, and position for atom $i$, respectively.

Force on atom $i$ can be defined as the gradient of the potential energy, $V$, which is a function of the coordinates of the $N$ atoms in the system.

$$-\frac{\partial V(r_i, \dots, r_N)}{\partial r_i} = m_i \frac{d^2 r_i}{dt^2} \tag{3.2}$$

$V$ is an empirical function, which includes terms for bonded (bond length, angle, torsion angle) and non-bonded (van der Waals and electrostatic) interactions between atoms.

If the force on each atom is known, it is possible to integrate the equations of motions, which yields a trajectory that describes the positions, velocities, and acceleration of the particles with respect to time.

Solution of equation of motion for all the atoms within the system is complicated, and necessitates numerical integration. The Velocity Verlet algorithm (Verlet, 1967), is commonly employed for integration. This algorithm finds positions, velocities and accelerations at time $(t + dt)$ based on coordinates at time $t$ with little compromise on precision.

$$r_i(t + dt) = r_i(t) + v_i(t)dt + \frac{1}{2} a_i(t)dt^2 \tag{3.3}$$

$$v_i(t + dt) = v_i(t) + \frac{1}{2} [a_i(t) + a_i(t + dt)]dt \tag{3.4}$$

### 3.1.1. NAMD Software

Nanoscale Molecular Dynamics program (NAMD) (Philips *et al*., 2005) is a simulation package designed for analyzing large biomolecular systems. NAMD uses the molecular graphics program VMD (visual molecular dynamics) (Humphrey *et al*., 1996) for simulation setup and trajectory analysis. NAMD is file-compatible with several commonly used potential energy functions and file parameters. In this thesis, simulations were performed using NAMD Software Version 2.8. VMD 1.9 is used for analyzing trajectories in terms of equilibrium and dynamic properties of the system.

### 3.1.2. Simulation Parameters

Four independent 200 ns long MD simulations of *E.coli* TF were performed at three different ionic strengths, namely 29mM, 50mM and 150mM. The physiologic pH of the cell falls in the range of 100-200 mM, and two of these simulations, called as 150mM and 150mM_2, have 150 mM ionic strength. The initial coordinates of apo TF was extracted from the crystal structure. VMD 1.9 simulation package was used for the preparation of input files. *NPT* ensemble was generated by keeping total number of atoms/moles (*N*), pressure (*P*), and temperature (*T*) constant throughout the simulations. The initial protein structure was solvated by creating a 12 Å water layer thickness in each direction from the largest coordinate in the corresponding direction, by adding TIP3P type water molecules (Jorgensen *et al*., 1983). The resulting periodic box dimensions are given in Table 3.1 for

independent runs. Later the structures were ionized by adding the necessary number of $Na^+$ and $Cl^-$ ions to reach the specified ionic strength.

CHARMM22 force field was used with the par_all27_prot_lipid.inp parameter file and top_all27_prot_lipid.inp topology file (MacKerell et.al., 1998). Energy minimization was performed to remove any steric clashes between the atoms and any unfavorable geometries. 10000 minimization steps were used to minimize the potential energy. Langevin dynamics was used to keep the system at a constant temperature of 310°K, and constant pressure of 1 atm. RATTLE (Andersen, 1983) algorithm was used for satisfying the constraints. The particle-mesh Ewald (PME) method (Darden *et al.*, 1993) was employed for electrostatics. Data was recorded in every 2.0 fs and the resulting coordinates are recorded at each 4 ps.

Table 3.1. Simulation systems details.

| Simulation name | Periodic box dimensions (Å) | Added $Na^+$ ions | Added $Cl^-$ ions |
|---|---|---|---|
| 29 mM | 100.64 x 117.57 x 110.92 | 23 | - |
| 50mM | 100.64 x 117.57 x 110.92 | 60 | 37 |
| 150mM | 100.64 x 117.57 x 110.92 | 133 | 110 |
| 150mM_2 | 100.64 x 117.57 x 110.92 | 133 | 110 |

### 3.2. Principal Component Analysis (PCA)

Principal component analysis is a method to reduce the number of parameters in a data set. When there are many parameters within the system, PCA is submitted for reducing the dimensionality. Principal components are the directions where most of spread out observed.

In MD trajectories, identifying essential motions that may be relevant-function is important. Throughout the simulations, random fluctuations and collective motions occur at the same time. Therefore, it is hard to distinguish these motions. PCA can help filter

collective motions from local, fast motions, and facilitate the study of long term protein dynamics.

The approach of PCA comes from diagonalizing the covariance matrix (Amadei et.al. , 1993). The covariance matrix, *C*, of size $3N \times 3N$ ($1296 \times 1296$) is created from the deviation of alpha-carbon atoms from the average structure. In this way, a set of eigenvectors and corresponding eigenvalues can be obtained.

### 3.3.   Anisotropic Network Model (ANM)

Anisotropic network model (ANM) falls under the category of the elastic network models, which are used for examining structure-dynamics-function relationship of biomolecular structures via extracting the biologically relevant, collective modes of motion. ANM is a coarse-grained normal mode analysis performed for a protein structure represented as inter-connected springs. In this model, the alpha-carbons of residues are generally considered as the nodes of the network, and close-neighboring node pairs are connected by identical harmonic-springs bonds (Atilgan *et al*., 2001). In ANM, the potential energy of *N* interacting residues is expressed as,

$$V = \gamma/2 \sum_i \sum_j h(r_c - R_{ij})\left(\Delta \mathbf{R}_j - \Delta \mathbf{R}_i\right)^2 \tag{3.5}$$

where $\gamma$ is the force constant, $r_c$ is the cutoff distance, $R_{ij}$ is the distance between the node pair i,j. $h(r_c - R_{ij})$, which is defined as Heaviside step function, equals to 1 if the distance between the $(i,j)$ pair falls within the cutoff distance, but 0 otherwise. $\Delta R_i$ is a $3N$ dimensional vector representing the fluctuations of the position vector $R_i$ of residue $i$, $(i = 1 \dots, N)$ for all interactions within the system.

The overall potential energy of a structure with *N* interaction sites can be simplified as,

$$V = (\gamma/2)\Delta \mathbf{R}^{\mathrm{T}}\mathbf{H}\Delta \mathbf{R} \tag{3.6}$$

where $\Delta\mathbf{R}$ is the fluctuation in position vector matrix, $^{\mathbf{T}}$ is the transpose, and $\mathbf{H}$ is the Hessian matrix. Hessian matrix composed of the second derivatives of the potential for over all interactions sites.After the diagonalization of $\mathbf{H}$, the calculation of $3N$-6 normal modes and the respective eigenvalues leads to the collective dynamics of the elastic network.

In this thesis, phospate atoms for nucleotides and alpha-carbon atoms for residues are selected as nodes for the docked TF-50S complex. Elastic network model is created with 13 Å cutoff distance between node pairs.

# 4. RESULTS

## 4.1. RMSD Calculations

The aim of this analysis is to observe the equilibration period for 200 ns long simulations. Root mean square distance or deviation (RMSD) calculations were based on the N-Cα-C atoms of the backbone skeleton of TF.

Firstly, all snapshots were aligned onto the initial snapshot using overall structural alignment. Then, the RMSD between each snapshot and the initial structure was calculated using the RMSD trajectory tool of VMD. RMSD of a specific structure or snapshot with respect to a a reference structure, is defined as,

$$\text{RMSD(t)} = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} |(r_i(t)-(r_i)^{ref}|^2}{N_{atoms}}} \tag{4.1}$$

where $N_{atoms}$ gives the number of atoms, whose positions are being compared. $r_i(t)$ is the position of atom $i$ in the structure at time t, and $(r_i)^{ref}$ is its position in the reference or initial structure.

In Figure 4.1, RMSD results for four independent simulations are shown. The first 20 ns of the simulations, indicated by the dashed vertical line, are considered as the equilibration period, and discarded from further analyses due to major readjustments taking place during this period. RMSDs observed in Figure 4.1, span quite large ranges, possibly due to the high flexibility of the three domains forming the TF molecule. The RMSDs with respect to the initial structure exceeds 10 Å and even 16 Å in 150 mM_2 run. At this point, it needs to be determined whether these large RMSDs are mostly resulting mostly from inter-domain flexibility or intra-domain conformational rearrangements.
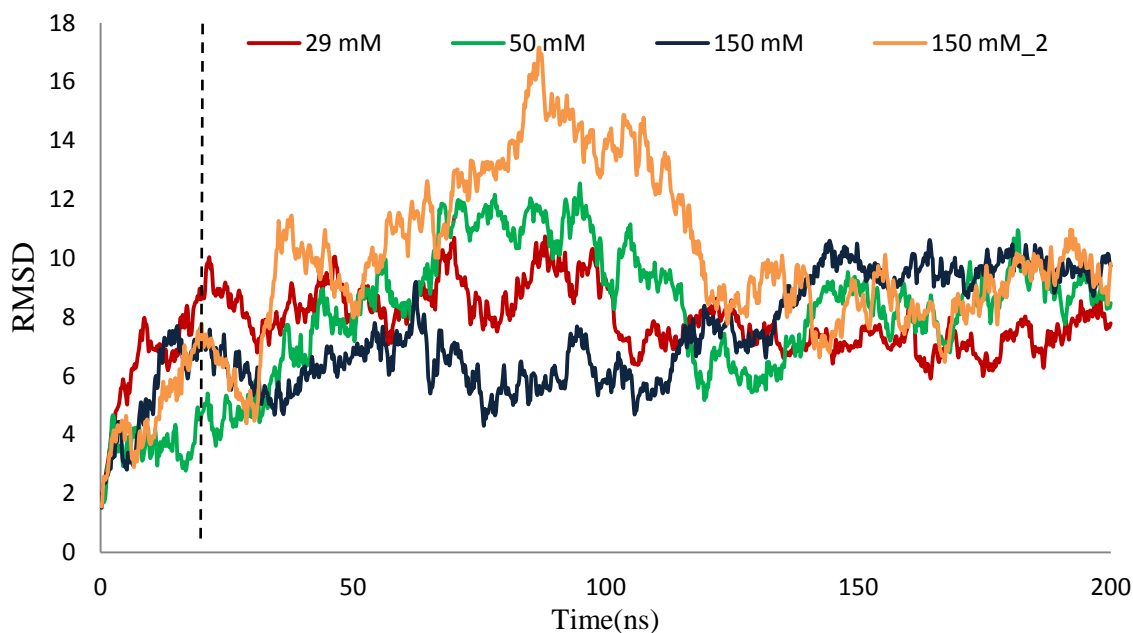
Figure 4.1. RMSD values based on overall alignment.

In order to determine the conformational flexibility of each domain separately, domain-wise alignments were done, and RMSDs of snaphots were calculated accordingly. In other words, alignments were repeated based on binding domain (BD: residues 1 to 110), core domain (CD: residues 111 to 149 and 247 to 432) and head domain (HD: residues 150 to 246) independently, and resulting RMSD profiles were plotted for the corresponding domains. In Figure 4.2, domain-wise RMSD profiles are plotted for 29 mM, 50 mM, 150 mM and 150 mM_2 runs, respectively.

In Figure 4.2a, fluctuations for the binding domain (BD) are higher compared to the other two domains. Such a trend was not observed in the previous study, which presented an independent 500 ns long run at the same ionic strength (Can, 2012). RMSDs of the BD in other simulations at different ionic strength are not as high as the 29 mM one. This increase in RMSD of BD is in fact related to a significant structural rearrangement from coil to helix, which will be discussed later in Section 4.6. In Figure 4.2, CD and HD fluctuate approximately within the same range at all ionic strengths, however fluctuations for CD are slightly higher than HD. In general, the intra-domain RMSD values, plotted in Figure 4.2, are lower than the ones based on overall structural alignment, in Figure 4.1, indicating significant inter-domain flexibility.
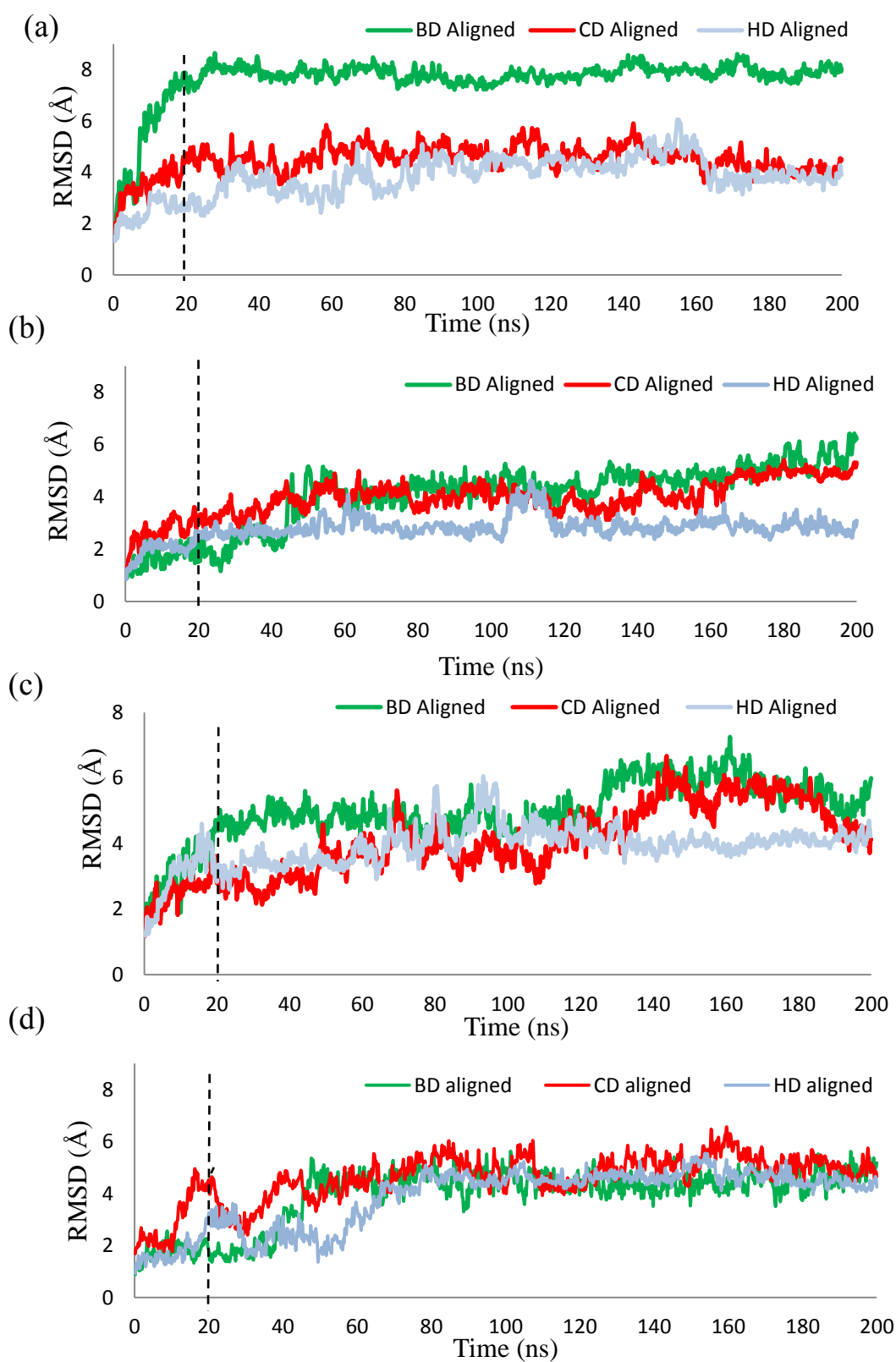
Figure 4.2. Intra-domain RMSDs for (a) 29 mM (b) 50 mM (c) 150 mM (d) 150 mM_2 runs.

Domain-wise alignments for the different ionic strengths are also compared in Figure 4.3. Again 29 mM run shows higher RMSD trend for BD, which indicates that BD of 29 mM run exhibits more internal rearrangements compared to the other runs. RMSDs for CD and HD increase up to 5 to 6 Å.
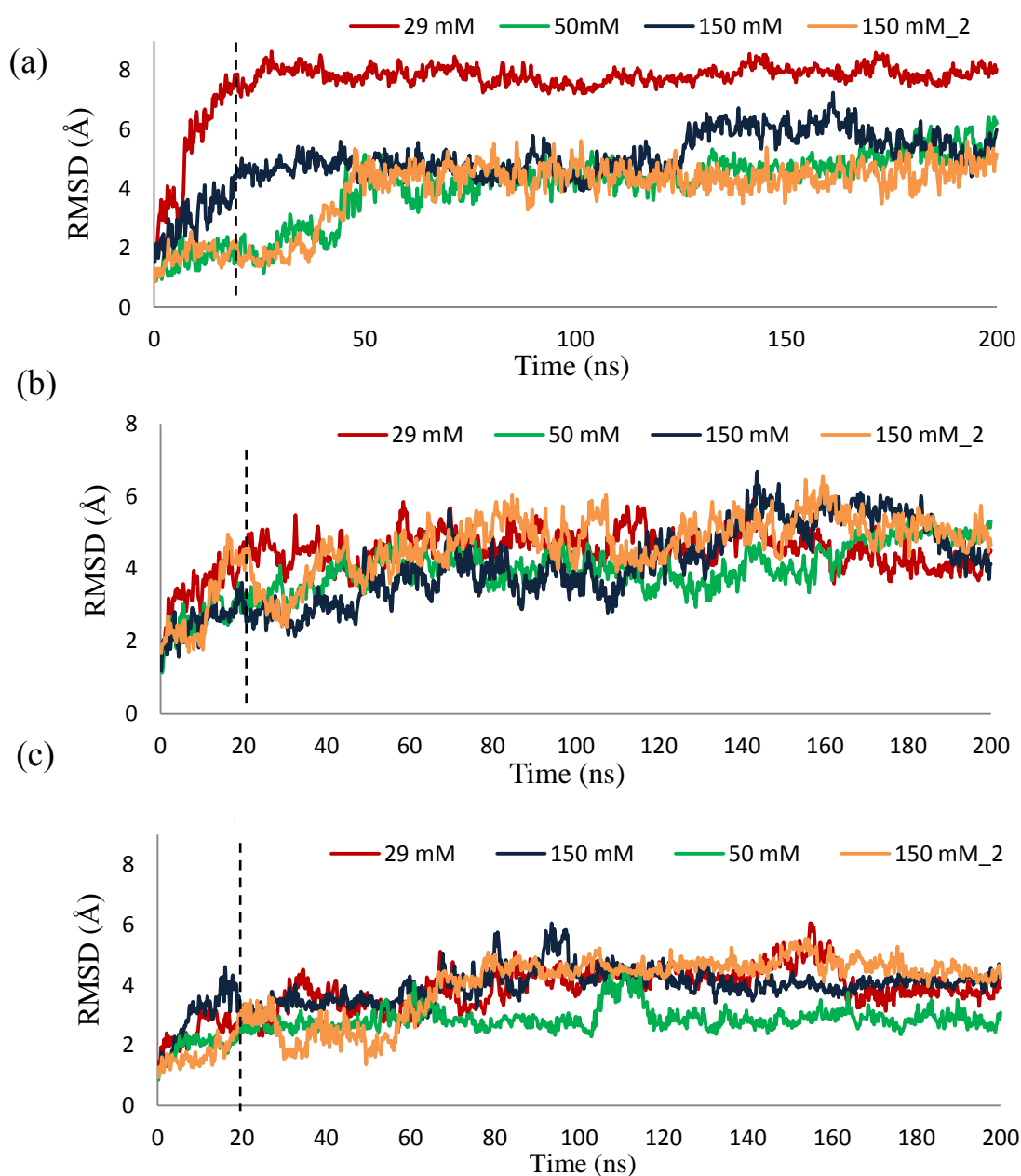


Figure 4.3. Domain-wise alignments for four simulations (a) BD wise alignment (b) CD wise alignment (c) HD wise alignment.

## 4.2. Residue Mobility Through RMSF

Residue mobility was examined by root mean square fluctuation (RMSF) calculations. RMSF is a measure of the deviation between the instantaneous position of particle i and its average position, defined as:

$$RMSF_{(i)} = \sqrt{\frac{\sum_1^T |(r_i(t) - (r_i)^{ave}|^2}{T}} \qquad (4.2)$$

where T is the total number of recorded snapshots, $r_i$ is the instantaneous position of particle, and $r_i{}^{ave}$ is its average position based on all snaphots.

RMSF calculations were done based on the equilibrated region of the simulations (between 20 ns-200 ns). Both overall alignment and domain-wise alignments were performed as in the previous section.

In Figure 4.4, the results for the 29 mM, 50 mM, 150 mM, 150 mM_2 runs, can be observed in panels a to d, respectively. Dark blue curve represents the overall alignment, and domain-wise alignments are presented by red (CD), green (BD), and light blue (HD) curves. The peaks for overall alignment are higher than the domain-wise alignments as expected. Each domain is indicated on the figures by dashed lines. CD is composed of two non-contiguous regions.



Figure 4.4. RMSFs for (a) 29 mM (b) 50 mM (c) 150 mM (d) 150mM_2 runs.
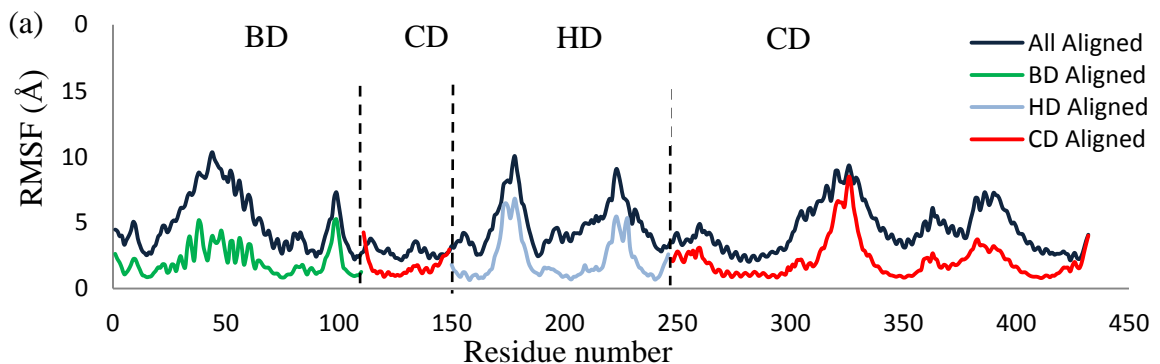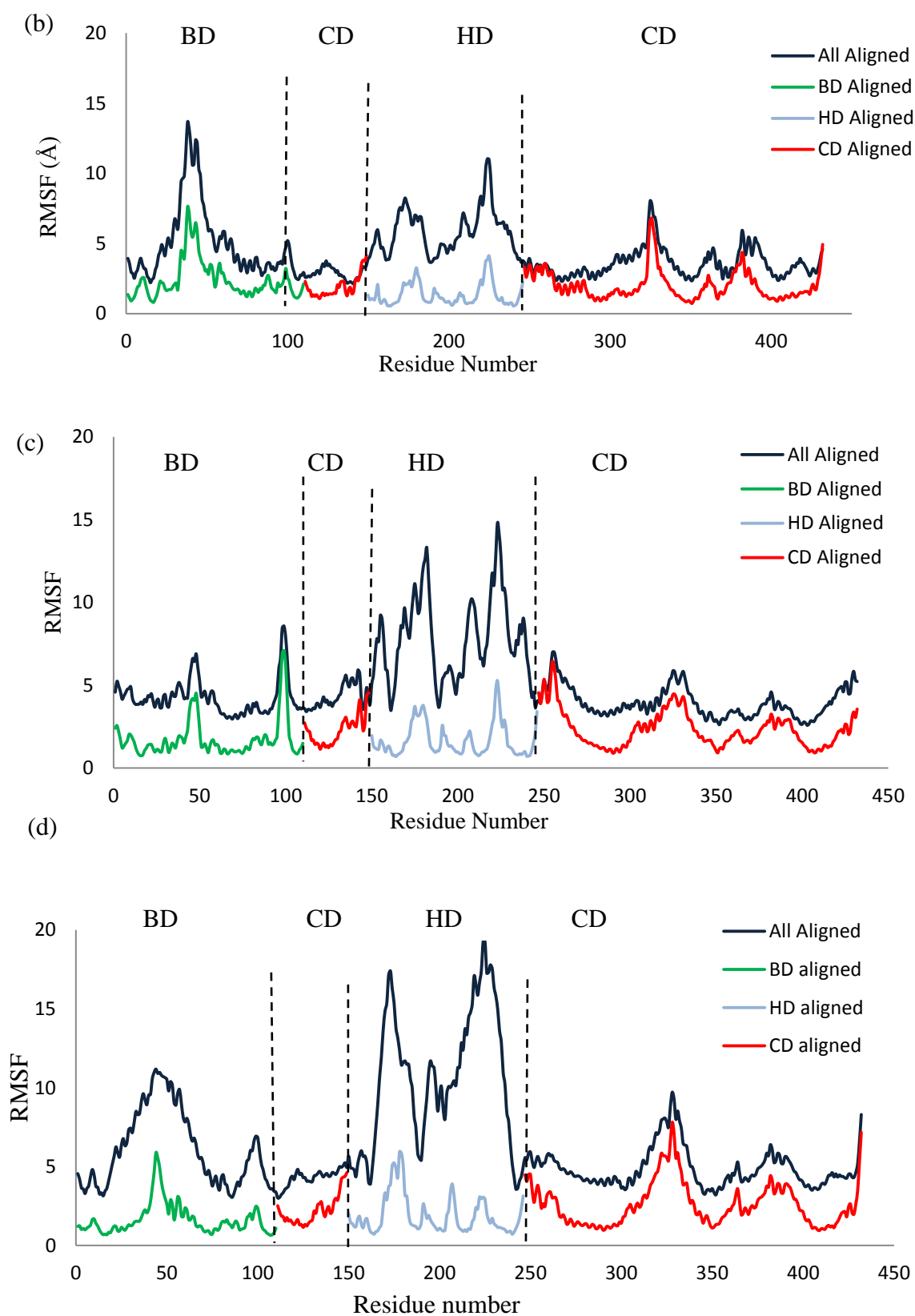
Figure 4.4. RMSFs for (a) 29 mM (b) 50 mM (c) 150 mM (d) 150mM_2 runs cont.

For a comparison of different runs, RMSF values based on overall and intra-domain alignments are plotted in Figures 4.5 and Figure 4.6, respectively. Peaks are generally observed at the same regions for four independent simulations. Arm 1 and arm 2 are the highly mobile region corresponding to the peaks in the core. Signature motif that TF binds to ribosome gives the highest peak in BD region (residues 42-49). However, the magnitudes of the peaks vary for the different ionic strengths.

In Figure 4.7a, average conformations of each run is color coded based on corresponding RMSF values from overall alignment. The code runs form dark blue (relatively immobile) to red (highly mobile) with respect to increasing RMSF values, i.e residue mobility. All runs are colored using the same RMSF range in Figure 4.5, without any normalization. HD of 150 mM_2 run has the highest RMSF values; therefore, colour of this domain is close to red. In Figure 4.7b, the residues corresponding to major peak regions in 150 mM_2 run, which have RMSF values above 3 Å, are colored red on the average conformation.
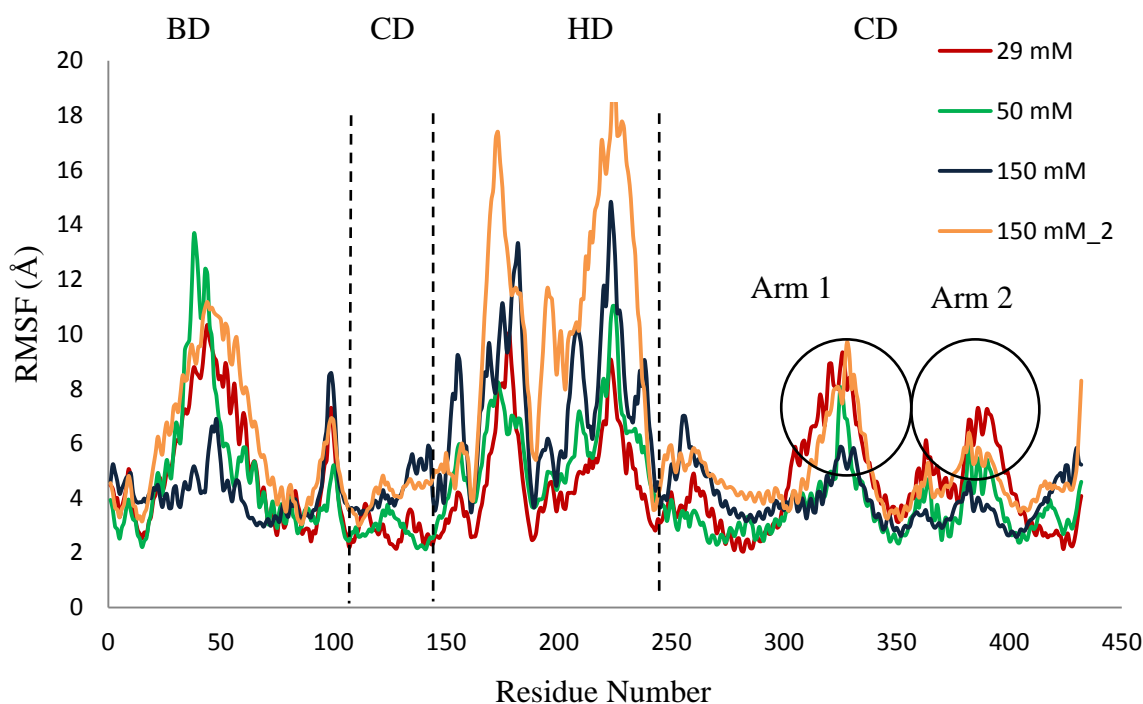


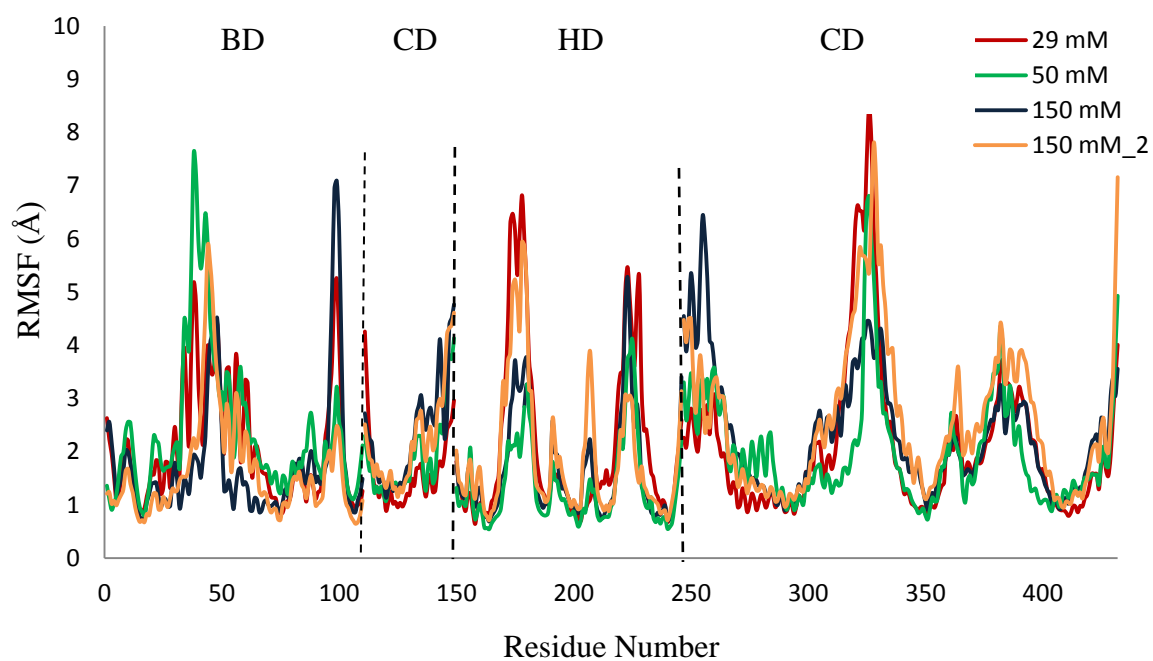Figure 4.5. RMSF values for different runs based on overall alignment.

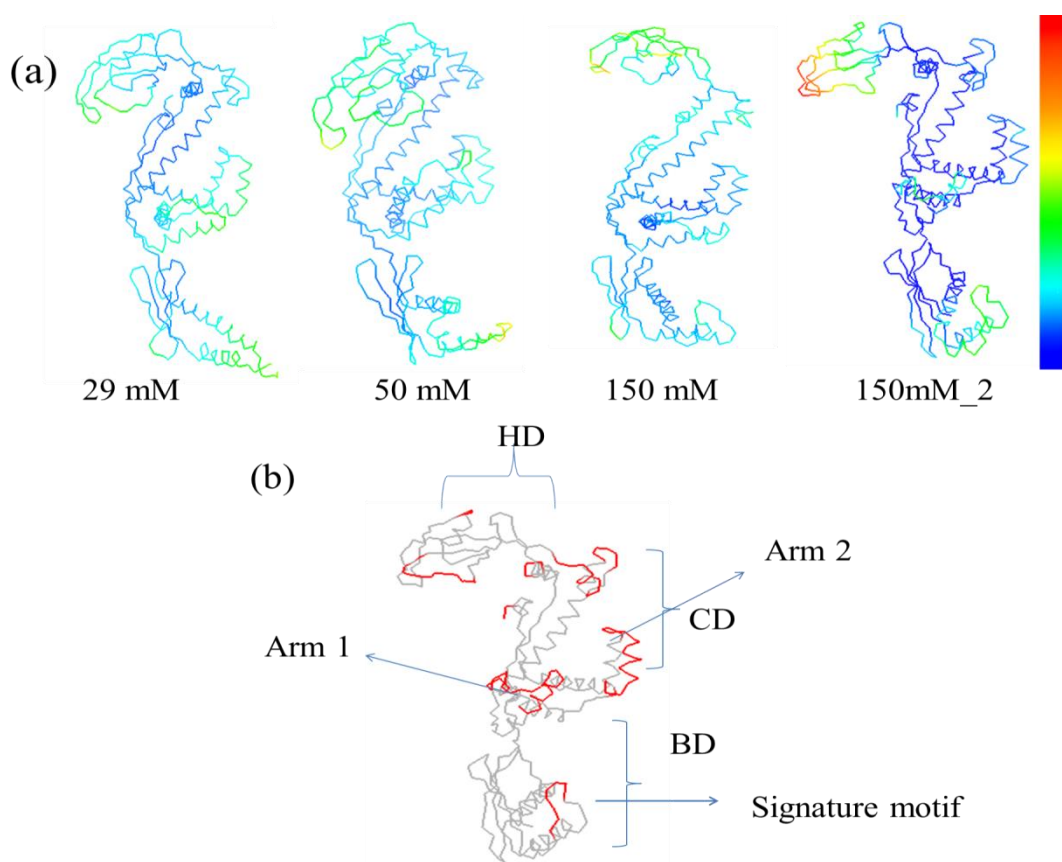Figure 4.6. RMSF values based on domain-wise alignment.



Figure 4.7. Colored conformations based on overall aligned RMSF values (a) for different ionic strengths (b) showing peak regions in 150 mM_2 run.

### 4.3. Opening and Closing Motion of TF

### 4.3.1. Radius of Gyration

Radius of gyration ($R_g$) of a protein chain is an indicator of its overall size and compactness. It is calculated as root mean square distance of atoms or residues from the center of mass of a chain. $R_g$ for each snapshot is calculated as,

$$R_g^2 = \frac{1}{N} \sum_{k=1}^{N} (r_k - r_{com})^2 \tag{4.3}$$

where $N$ is the number of atoms, and $r_{com}$ is the center of mass of the chain.

The $R_g$ values are plotted as a function of time in Figure 4.8 for the different runs. The ranges of $R_g$ values spanned during 200 ns runs are quite large, indicating the flexibility of TF. Large oscillatory behavior is observed in most runs, possibly related with an opening/closing type of motion. At around 20 ns, all runs have similar $R_g$, whereas the largest difference in $R_g$ among the runs is observed at around 100 ns.
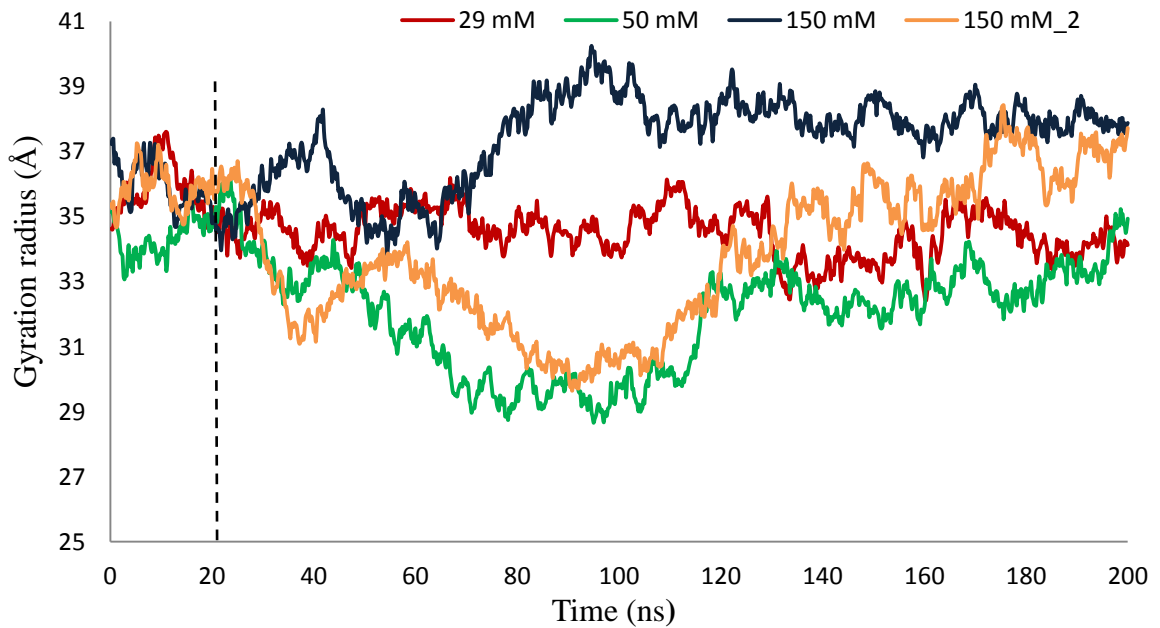


Figure 4.8. Gyration radius for 200 ns long simulations.

To observe the conformational changes during independent runs, conformations at 20, 100, and 200 ns are shown in Figure 4.9 (for 29 mM, 29mM_L and 50 mM runs) and Figure 4.10 (150 mM and 150 mm_2 runs). Cyan, green and magenta in the first three panels stand for the conformations at 20, 100 and 200 ns, respectively. The last panel indicates all conformations aligned on the 20 ns conformation of the corresponding simulation.



| (a) | | | |
| Rg=35.0 Å | Rg=34.0 Å | Rg=34.1 Å | |
| (b) | | | |
| Rg= 40.4 Å | Rg=34.9 Å | 32.8 Å | |
| (c) | | | |
| Rg=34.7 Å | Rg=29.7 Å | Rg=34.9 Å | |

Figure 4.9. Conformations at 20, 100 and 200 ns, and aligned structures for (a) 29 mM (b) 29mM_L (c) 50mM runs.

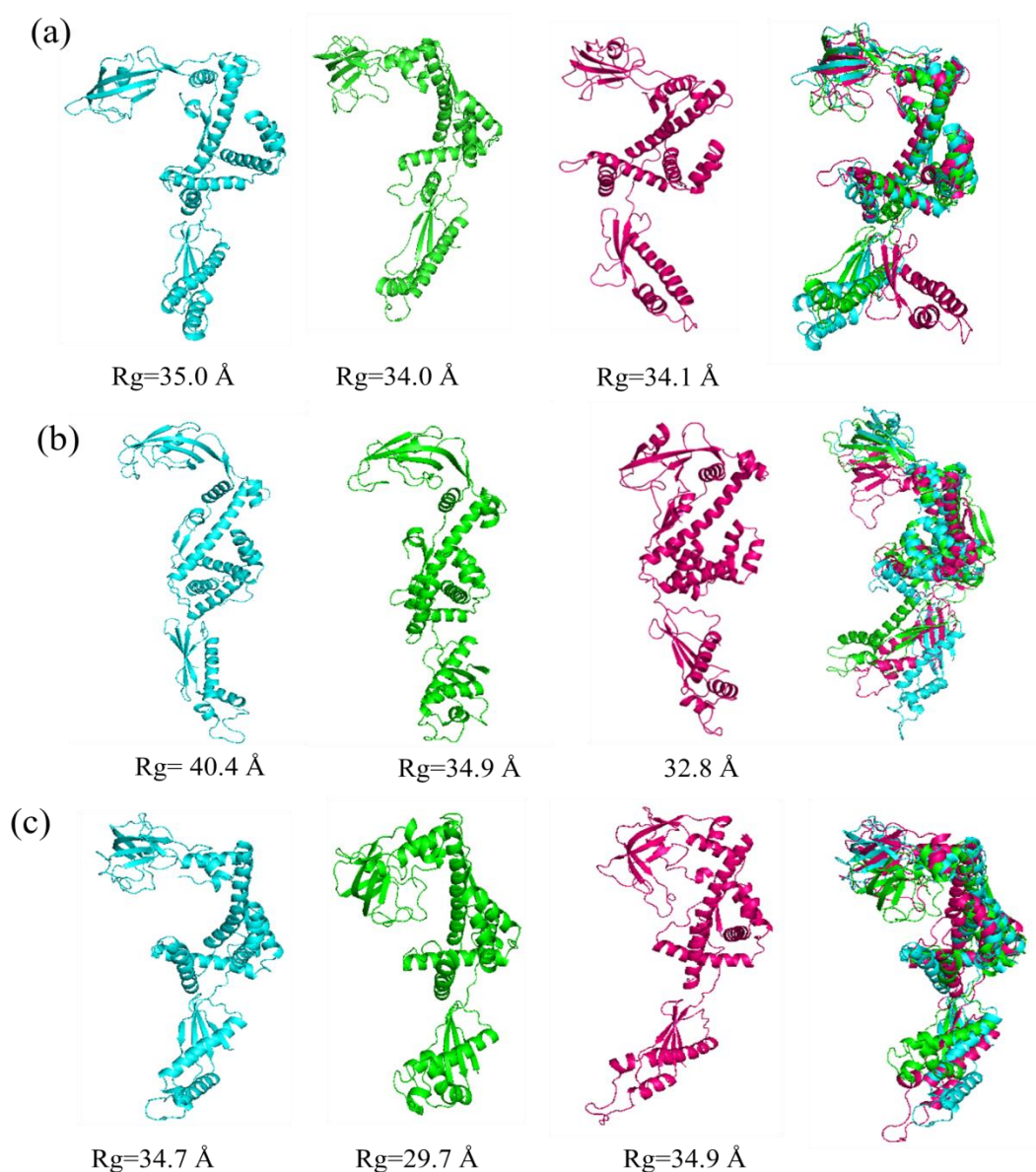In Figure 4.9, motion of the BD for 29 mM run is quite dominant compared to the other runs. Conformational change from open to closed form, and then back to open form can be clearly observed for 50 mM run. In Figure 4.10, conformational changes for 150 mM and 150mM_2 simulations are compared. Different from Figure 4.9, motion of the HD is dominant for 150 mM and 150 mM_2 runs.
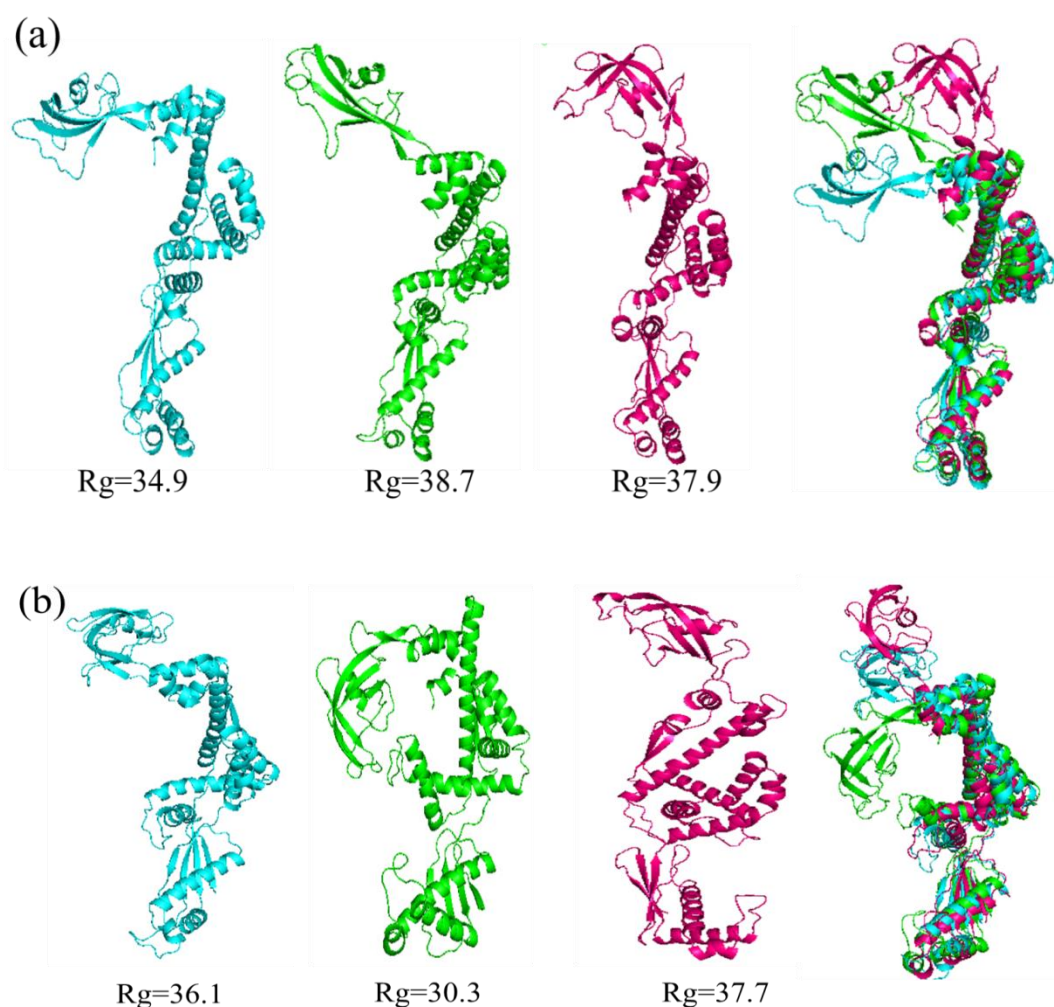


Figure 4.10. Conformations at 20, 100 and 200 ns and aligned structures for (a) 150mM, (b) 150mM_2 runs.

TF is known to act as a cage like protective tent for the newly synthesized nascent chains (Hoffmann and Bukau, 2009). Therefore, by looking at the $R_g$ profiles in Figure 4.8, we can predict the opening and the closing motions of the structure during 200 ns long

simulations. The alternating trends of the $R_g$ profiles can be linked to the opening and the closing motion of TF.

In Figure 4.8, concentrating after the 20 ns, trends of 50 mM and 150 mM_2 structures are similar to each other. In the first half of the simulations, their $R_g$ profiles decrease, and then increase. In other words these two simulations firstly show closing and then opening motion. 150 mM stucture behaves in the opposite way. On the other hand, oscillatory behavior for the 29 mM structure is not as distinct as in the previous three cases. Its $R_g$ values at 20, 100, and 200 ns are nearly the same. As a result of this $R_g$ fluctuations, the distance between domains also changes, which will be discussed in Section 4.3.2.

In Figure 4.11, relative frequencies of $R_g$ values are presented for the different ionic strengths. The result of an independent 500 ns MD run (Can, 2012), which has 29 mM ionic strength was also used in this analysis. The highly populated $R_g$ values fall in the range of 33-35 Å for 29 mM ionic strength. 50 mM ionic strength run visits lower values in comparison with the maximum at 33 Å. In contrast, 38 Å and 39 Å are the mostly observed $R_g$ values for 150 mM ionic strength structures, indicating that it adopts more extended conformations compared to the lower ionic strengths. In Figure 4.12, number of snaphots versus $R_g$ values are plotted in a histogram diagram for each independent run separately.
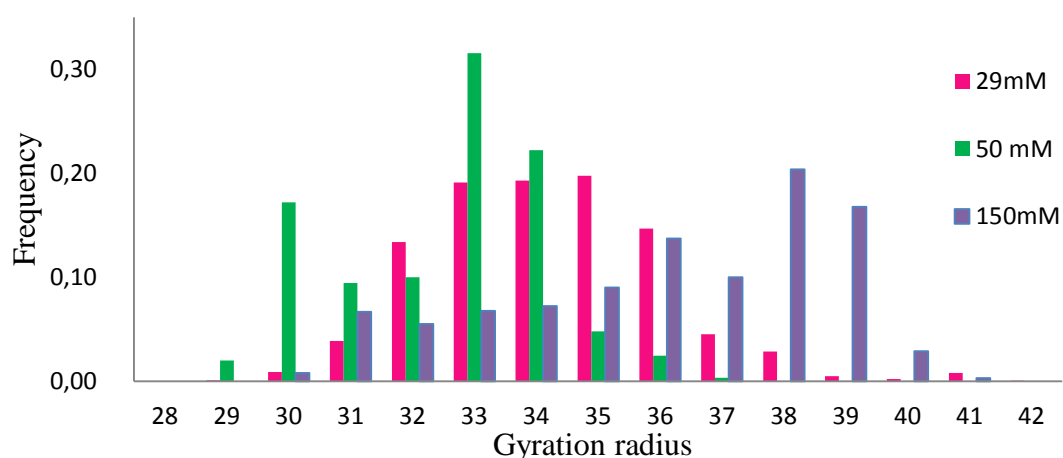


Figure 4.11. Normalized distributions of gyration radius for different ionic strengths.
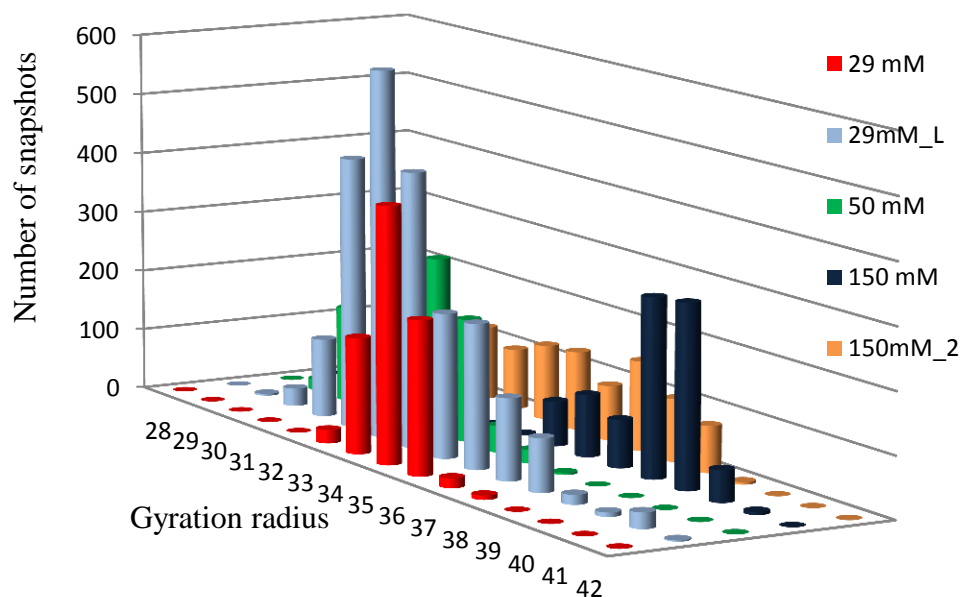
Figure 4.12. Gyration radius histograms for independent simulations.

### 4.3.2. Inter-domain Distances and Angles

Dynamics of free TF in water indicate that the inter-domain distance changes due to flexibility. The HD-BD inter-domain distance was calculated between the center of masses (COM) of head and binding domains, and plotted as function of time in Figure 4.13. Moreover, inter-domain angle defined between the BD − CD and CD − HD vectors was determined based on the COMs of these domains, and plotted in Figure 4.14. Some kind of an opening/closing behavior of the domains observed in both figures are consistent with gyration radius.
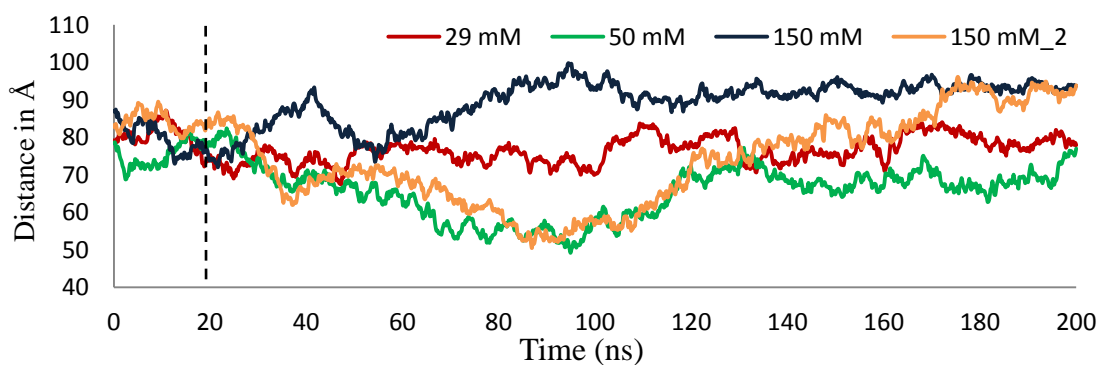


Figure 4.13. Inter-domain distances between BD and HD for 200 ns runs.
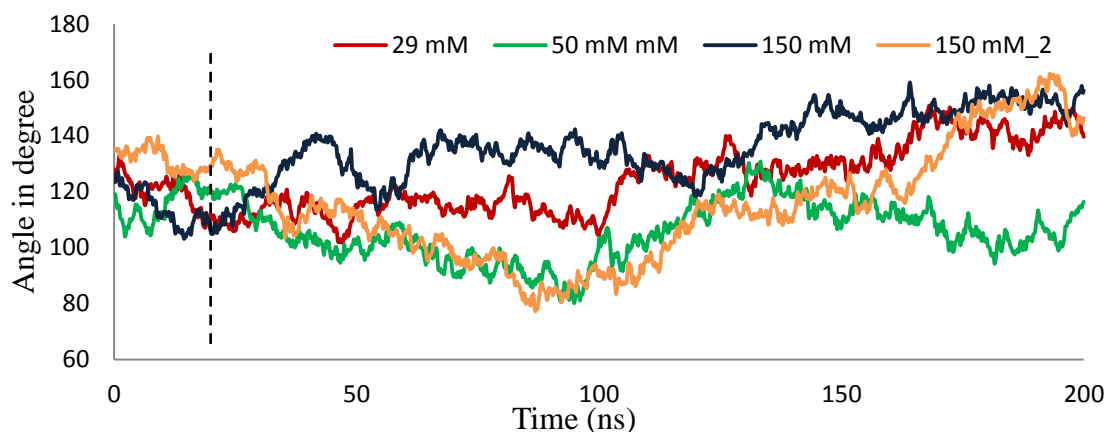
Figure 4.14. Inter-domain angles between BD and HD for independent 200 ns runs.

In Figure 4.15, the inter-domain distance versus angle plot is given for the equilibrated part of the simulations. High correlation coefficients, based on Pearson's method, are obtained between two parameters, namely 0.69, 0.86, 0.84, 0.71, 0.95 for 29 mM, 29mM_L, 50 mM, and 150 mM and 150 mM_2 simulations, respectively.



Figure 4.15. The correlation between inter domain distances and angles for all runs.

In Figure 4.16, positive correlations are observed between $R_g$ and inter-domain distance (HD and BD) are plotted for each run. Correlation coefficients are calculated based on Pearson's method, namely 0.68, 0.98, 0.97, 0.95, and 0.90 for 29 mM, 29mM_L, 50 mM, 150mM, and 150 mM_2 runs, respectively.

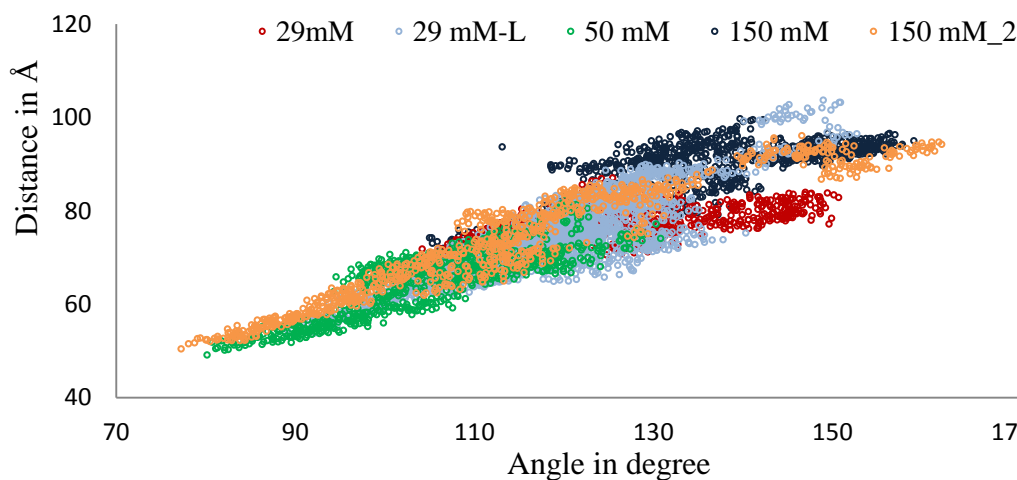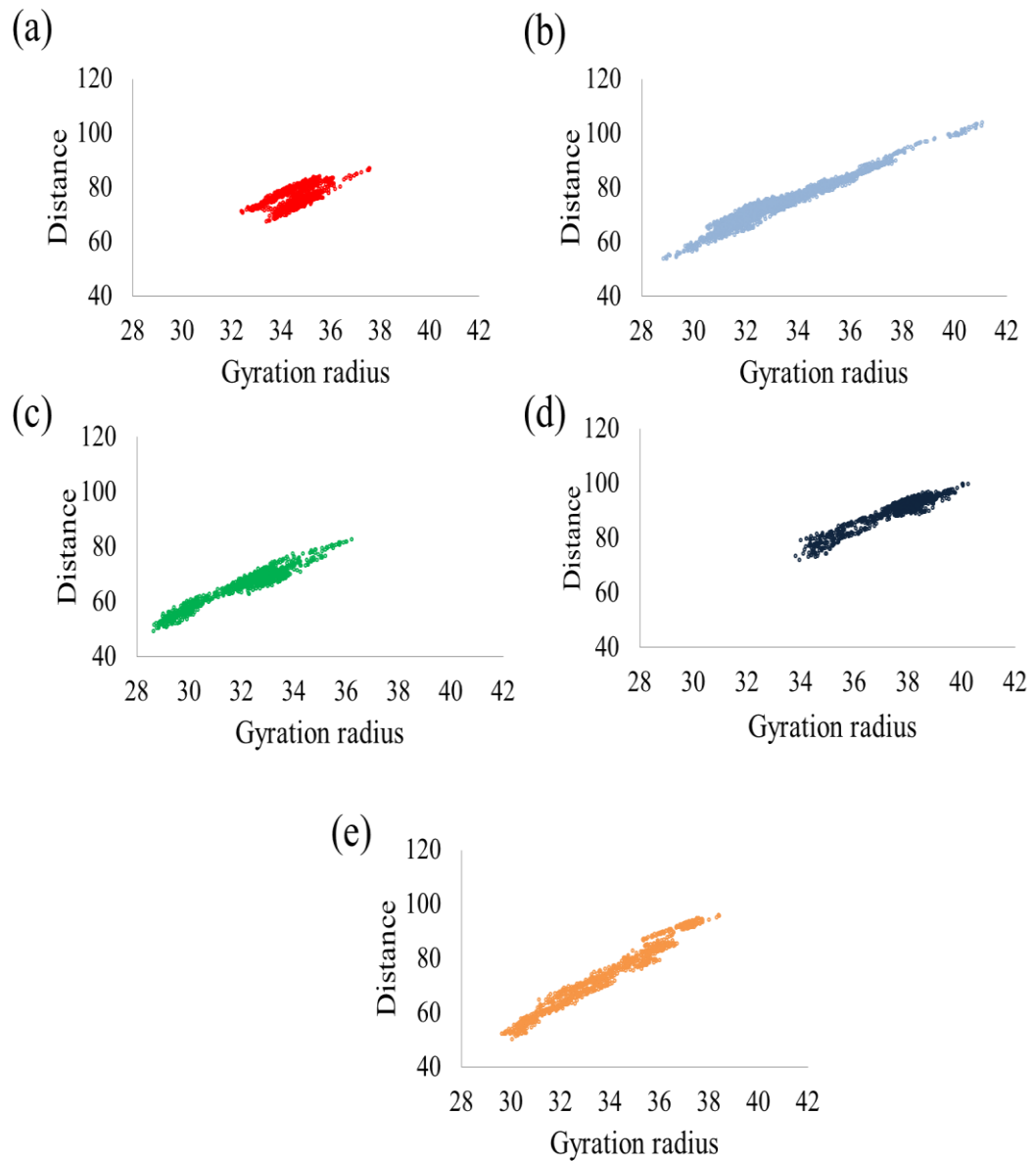Figure 4.16. The correlation between inter-domain distances gyration radius for (a) 29 mM (b) 29mm_L (c) 50mM (d) 150 mM (e) 150 mM_2 runs.

### 4.3.3. Fully Collapsed State Not Visited

In literature, collapsed study of TF for 50 mM simulation is reported by looking at its COM distance between BD-Arm2 and HD-Arm1 (Singhal *et al*., 2013). In this thesis, five simulations are investigated in terms of collapse behavior by looking at their COM distances between three domains. While making the analysis, residues for the domains are used as specified in Table 4.2 (Singhal *et al*., 2013). Collapsed behavior of independent

simulations are investigated in Figure 4.17. In this figure, two black squares represent the regions that are stated as semi-collapsed and the fully-collapsed regions in the reported article. In that study, region I is observed when the distance between head-arm 1 and N BD–arm 2 is around 2.6 nm and 4.6 nm, respectively. On the other hand, region II corresponds to 2.6 nm distance between head and arm1, and 2 nm distance between N terminal and arm 2.

In Figure 4.17, the collapsed state cannot be observed for equilibrated part of the simulations. The distance between the head and arm 1 is wider compared to the reported histogram (Singhal *et al*., 2013). Behavior of the simulations are far away from the collapsed state.
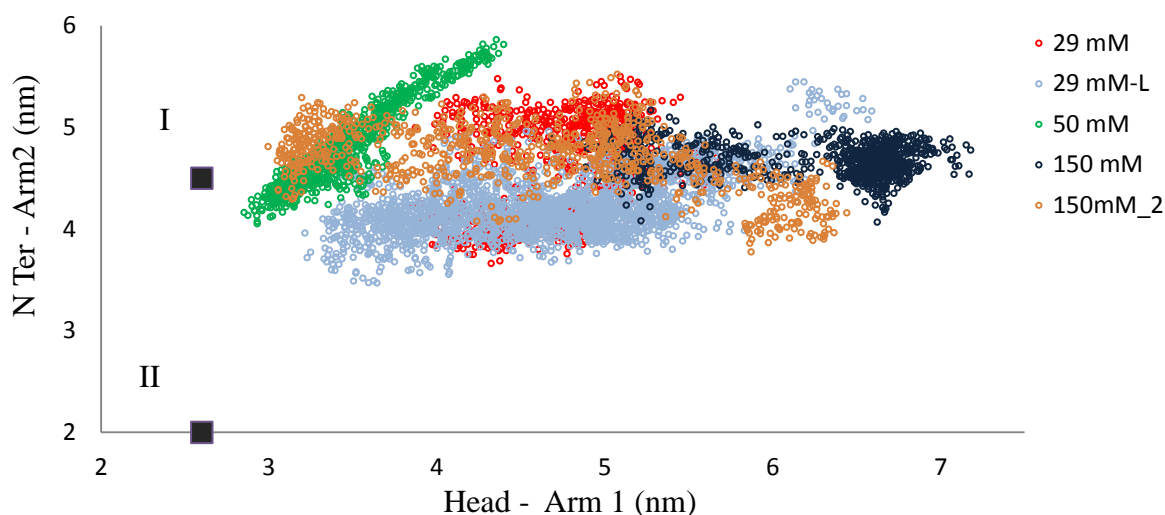


Figure 4.17. Scatter plot domain-arm distances to observe collapsed behavior.

### 4.4.  Essential Dynamics of Apo TF

Essential motions, also named as anharmonic modes, of TF can be viewed by applying the principal component analysis (PCA) on each MD trajectory using alpha-carbon atom coordinates only. PCA was applied to the equilibrated part of each trajectory, which is aligned on to its initial snapshot. In general, the essential dynamics can be expressed by the lowest-indexed modes that bear the highest variance or contribution to total motion. In Figure 4.18, the cumulative percent explanation for the first 20 modes are depicted. For 150 mM_2 run, the contribution of the first mode to the overall motion is

about 70%, which is larger than other simulations that fall in the range of 40-50%. About 80-90% of the total motion is expressed by the cumulative 10 modes for all trajectories.
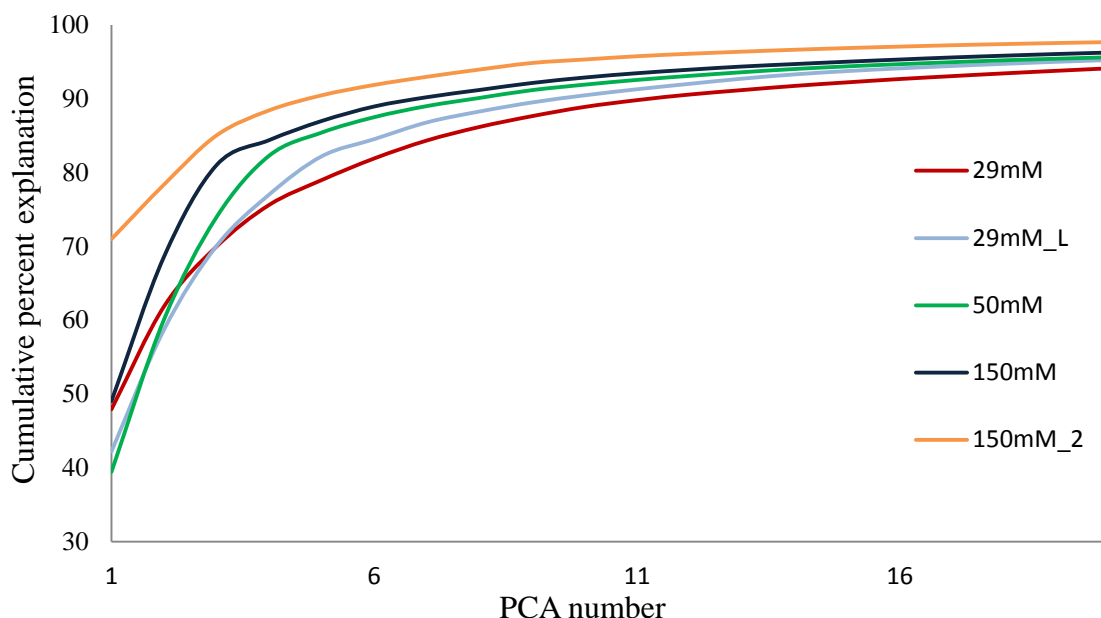


Figure 4.18. Cumulative percent explanation of the first 20 PCA modes.

Residue cross-correlation maps for the first mode, and based on cumulative 10 modes are shown for independent runs in Figure 4.19 and Figure 4.21, respectively. These maps were created based on α-carbon atoms of the structures. Normalized cross-correlations, $C(i,j)$, between residue fluctuations defined in Equation 4.4, where ($\Delta \boldsymbol{R}_i$) is the fluctuation in the position vector $R_i$ of site i.

$$C(i,j) = \frac{(\Delta \boldsymbol{R}_i . \Delta \boldsymbol{R}_j)}{\left[(\Delta \boldsymbol{R}_i . \Delta \boldsymbol{R}_i)(\Delta \boldsymbol{R}_j . \Delta \boldsymbol{R}_j)\right]^{1/2}} \qquad (4.4)$$

As indicated in the color bar in Figure 4.19, red refers to fully correlated motion, while blue indicates anti-correlated motion. The positively correlated blocks along the diagonal correspond to the dynamic domains of the protein, which are tentatively indicated within black squares on each map, and listed in Table 4.1. In the same table, domain regions obtained after application of ANM on average structure of each MD run is also reported, and will be discussed later.
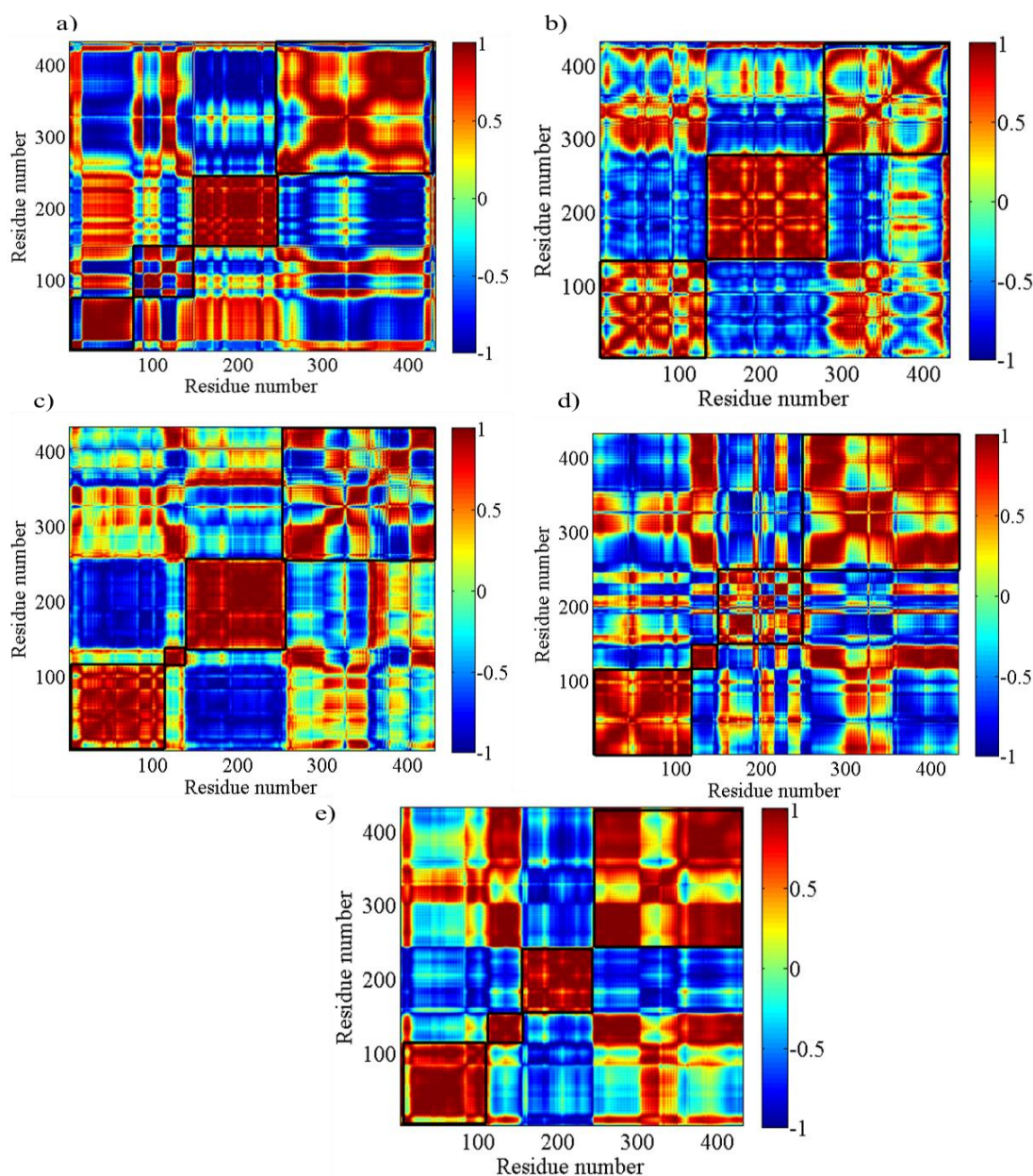
Figure 4.19. Residue cross-correlation maps based on first mode for (a) 29 mM (b) 29mM_L (c) 50 mM, (d) 150 mM (e)m150 mM_2 runs.

The dynamic domain decomposition can also be compared with the ones reported in literature, which are listed in Table 4.2. The small linker region cannot be observed for 29mM_L simulation. For 50 mM, 150mM and 150 mM_2 runs, the domain regions are in parallel to the reported domains. BD region is smaller for 29 mM run compared to the other runs and reported values. In some of the runs, the reported domains do not appear as a single red block but can be rather decomposed into smaller entities.

Table 4.1.  Domain decomposition based on first mode from PCA or ANM

| Run / Method | Binding Domain | Linker | Head | Core |
|---|---|---|---|---|
| 29 mM PCA | 1-75 | 76 -149 | 150-246 | 247-432 |
| 29 mM-L PCA | 1-132 | 133-280 | | 281-432 |
| 50 mM PCA | 1-114 | 115-142 | 143-257 | 258-432 |
| 150 mM PCA | 1-117 | 118-151 | 152-246 | 247-432 |
| 150 mM_2 PCA | 1-114 | 115-154 | 155-243 | 244-432 |
| 29mM ANM | 1-110 | 111-150 | 151-245 | 246-432 |
| 29mM_L ANM | 1-114 | 115-154 | 155-242 | 243-432 |
| 50  mM ANM | 1-106 | 107-142 | 143-253 | 254-432 |
| 150 mM ANM | 1-110 | 111-154 | 155-241 | 242-432 |
| 150mM_2 ANM | 1-111 | 112-146 | 147-245 | 246-432 |

Table 4.2. TF domain regions reported in literature.

| | Binding Domain | Linker | Head | Core |
|---|---|---|---|---|
| Hoffmann *et al*., 2010 | 1-110 | 111-149 | 150-245 | 246-432 |
| Ludlam *et al*., 2004 | 1-110 | - | 150-246 | 247-388 |
| Singhal *et al., 2013* | 1-149 | 111-149 | 150-245 | 246-432 |

In Figure 4.20, the four domains reported in literature (Hoffman *et al*., 2010), are colored for independent simulations. In these structures, BD, linker, HD and core domains are coloured as red, blue, hot pink and green, respectively. In fact, the linker and the core domain are not separate domains since in most runs they are positively correlated with each other.
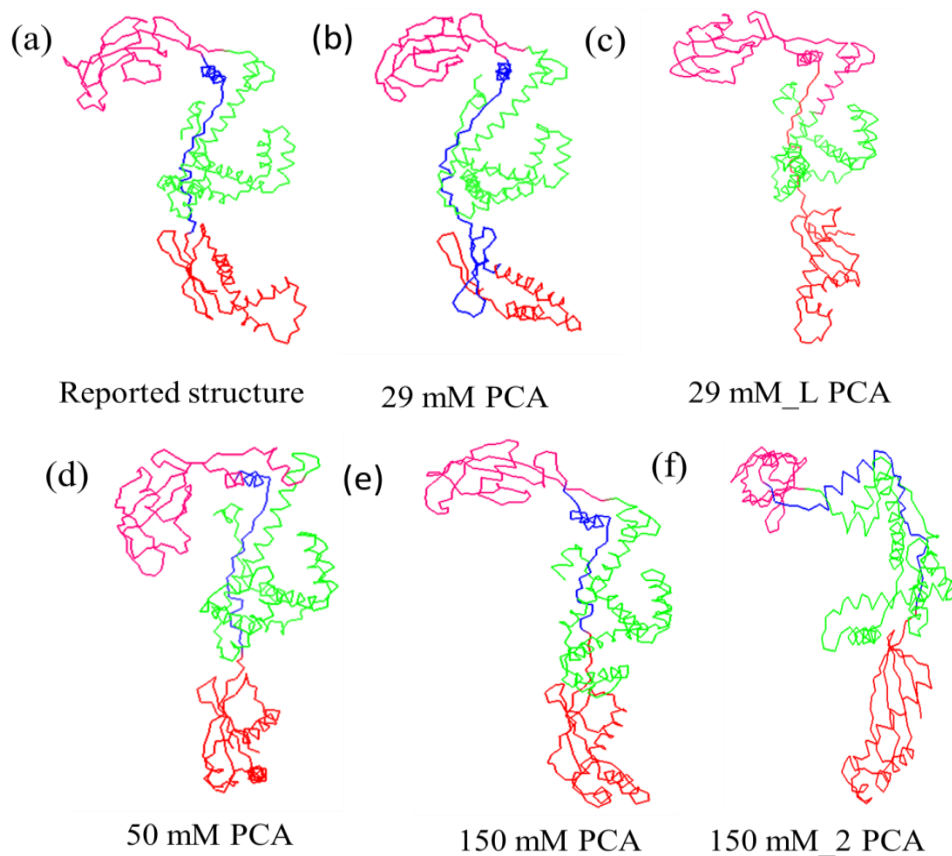


Figure 4.20. Coloured conformations based on decompositon of domains (a) reported structure ( Hoffman *et al*., 2010) (b) 29 mM PCA (c) 29 mM_L PCA (d) 50mM PCA (e) 150mM PCA (f) 150mM_2 PCA.

In Figure 4.21, residue cross-correlation maps are presented based on the cumulative contribution of the first ten modes. Similar features in terms of domain decomposition and inter-domain correlations can be seen in these figures, due to the dominance of the first mode features.
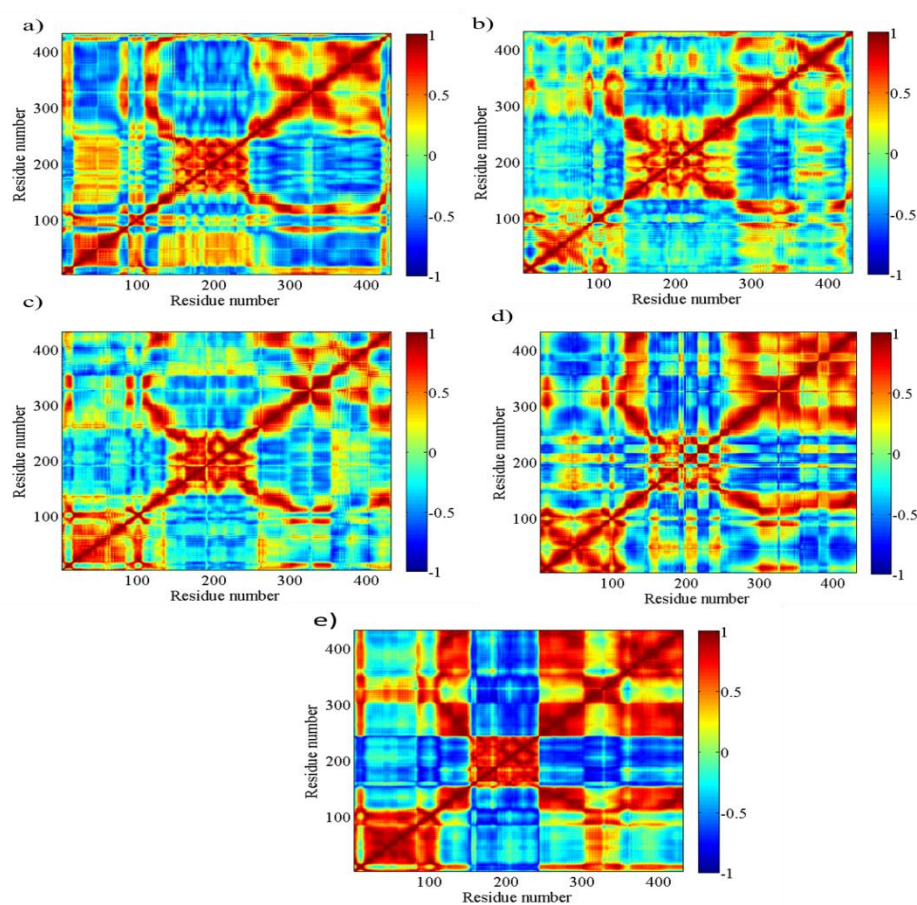
Figure 4.21. Residue cross-correlation maps based on the cumulative action of first ten modes for (a) 29 mM (b) 29mM _L (c) 50 mM (d) 150 mM (e)150mM_2 runs.
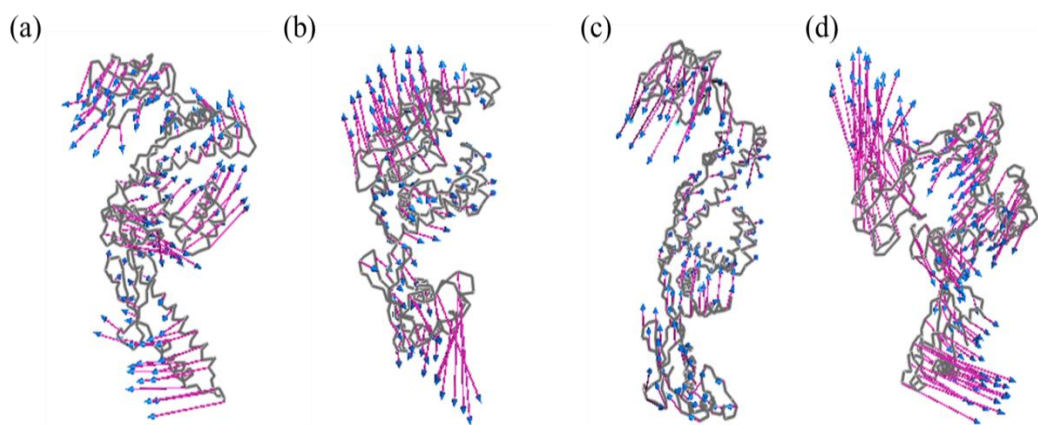


Figure 4.22. Deformation vectors based on $1^{st}$ mode of PCA for (a) 29 mM (b) 50mM (c) 150mM (d) 150 mM_2 runs.

In Figure 4.22, deformation vectors of the first mode that has the largest contribution to the overall motion is presented on the average structure of each run.

One-to-one correspondance between the first 10 anharmonic modes from two independent runs can be investigated by their mode overlap matrix. Overlap, $O_{ij}$ between 3$N$-dimensional eigenvectors $\mathbf{u}_i$ and $\mathbf{v}_j$ corresponding to the subspaces of two different trajectories is the inner product or dot product between $\mathbf{u}_i$ and $\mathbf{v}_j$. To determine $O_{ij}$ between different runs, the average structure from the two runs need to be aligned first. Overlap matrixes between 150 mM run and all other runs are given in Figure 4.23.
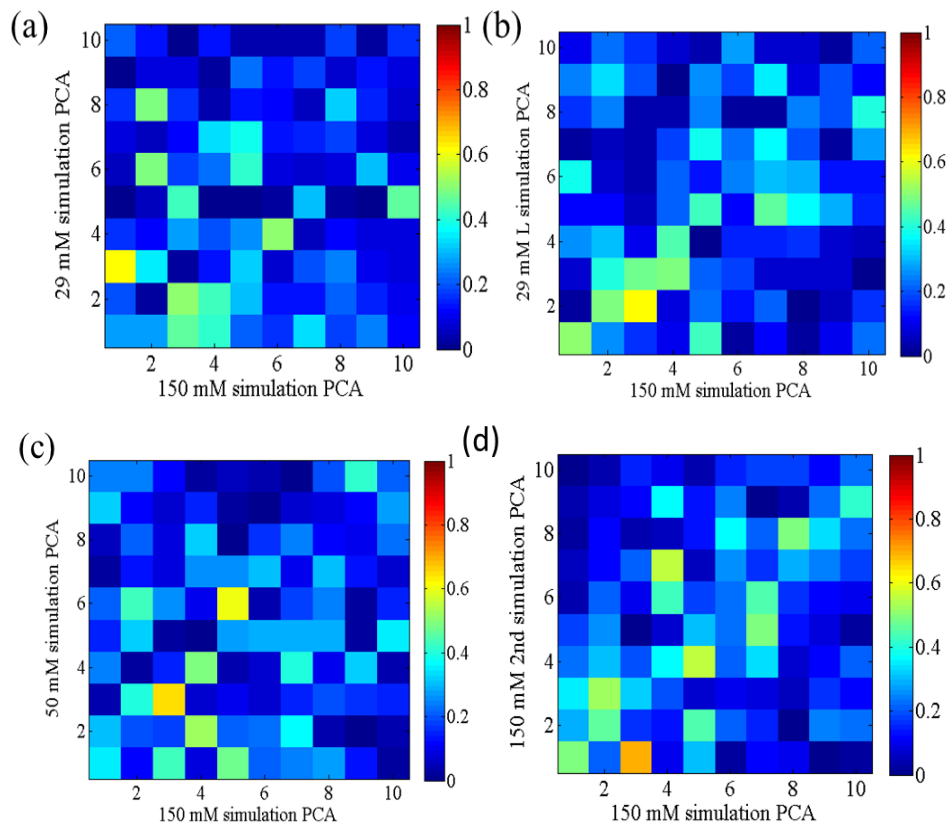


Figure 4.23. Mode overlap matrices between 150 mM run and (a) 29 mM, (b) 29 mM_L (c)50 mM (d) 150 mM_2 runs.

Closeness of $O_{ij}$ value to one (red color) indicates high correlation between the motions depicted by the two modes from different runs. There are several modes that present a fair overlap value above ~0.5 (green and yellow squares) on each plot. The extent

of overlap between the independent runs are reasonable for MD simulations. The correlation between the low indexed modes from independent 150 mM runs can be noted here, as we are going to concentrate on these two runs in the upcoming sections due its physiological relevance.

### 4.5. Collective Dynamics of Apo TF from ANM

Collective vibrational modes of TF are also determined by applying ANM on the average structure of each run. Cross-correlation maps based on the action of the first mode only and cumulative 10 modes are created in Figure 4.24 and Figure 4.25, respectively. Apparent domain regions are squared in Figure 4.24.
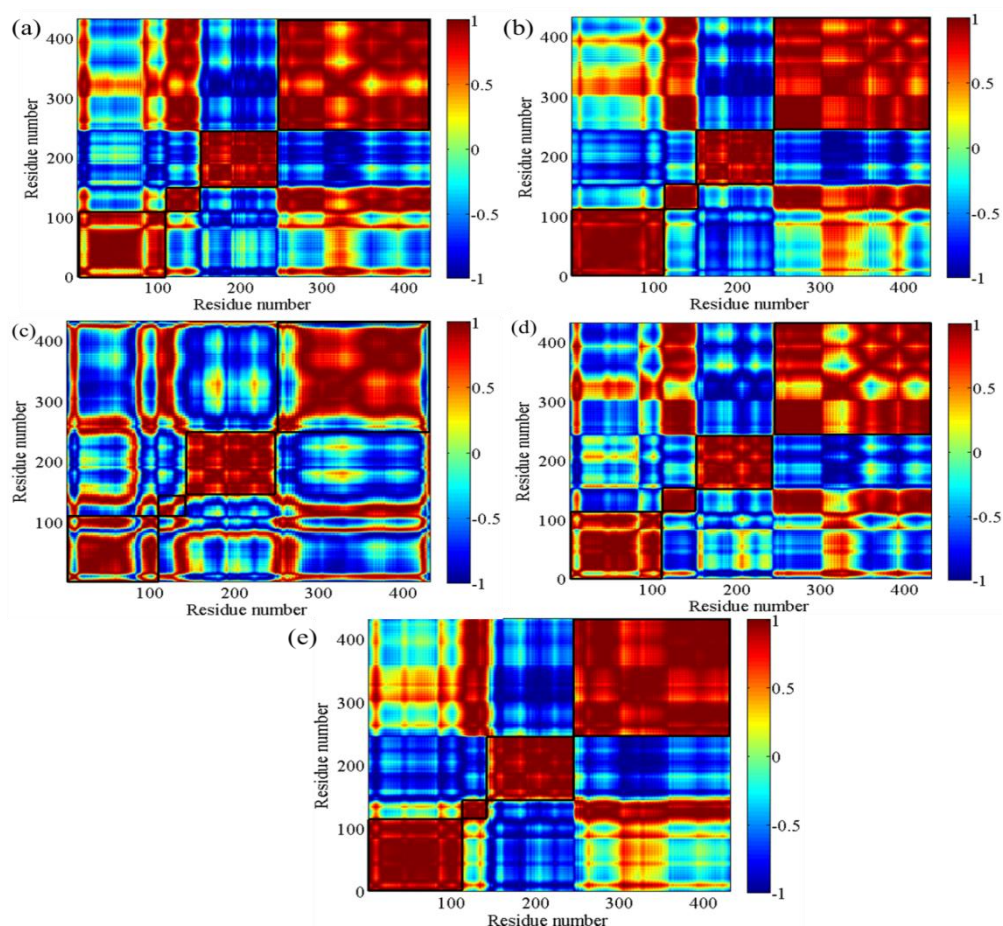


Figure 4.24. Cross-correlation map using first mode of ANM applied on average structures from (a) 29mM (b) 29mM_L (c) 50 mM (d) 150 mM (e) 150 mM_2 runs.
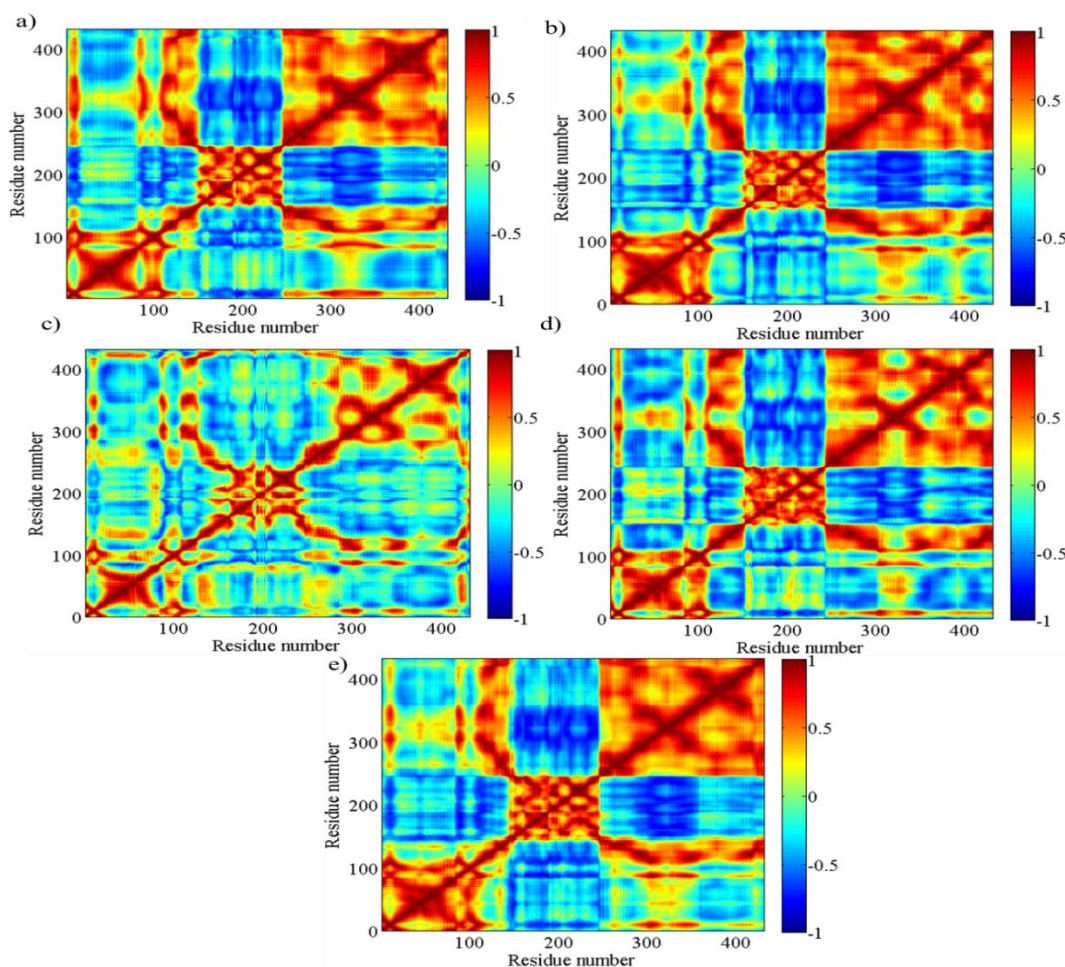
Figure 4.25. Cross-correlation for cumulative 10 ANM modes applied on average structures from (a) 29mM (b) 29mM_L (c) 50 mM (d) 150 mM (e) 150 mM_2 runs.

In the case of ANM, the domain features seem more distinct, and present a common picture for all runs. Moreover, the domain decomposition in ANM given in Table 4.1 highly resembles the reported residue ranges in Table 4.2.

In Figure 4.26, the overlap matrixes for ANM applied on average structures of different MD runs are compared. The ANM modes are much more corrrelated with each other than the PCA modes, previously shown in Figure 4.23. Specifically, the highly correlated first, second and third modes are shown in Figure 4.27 for 150 mM and 150mM_2 runs in vector notation.
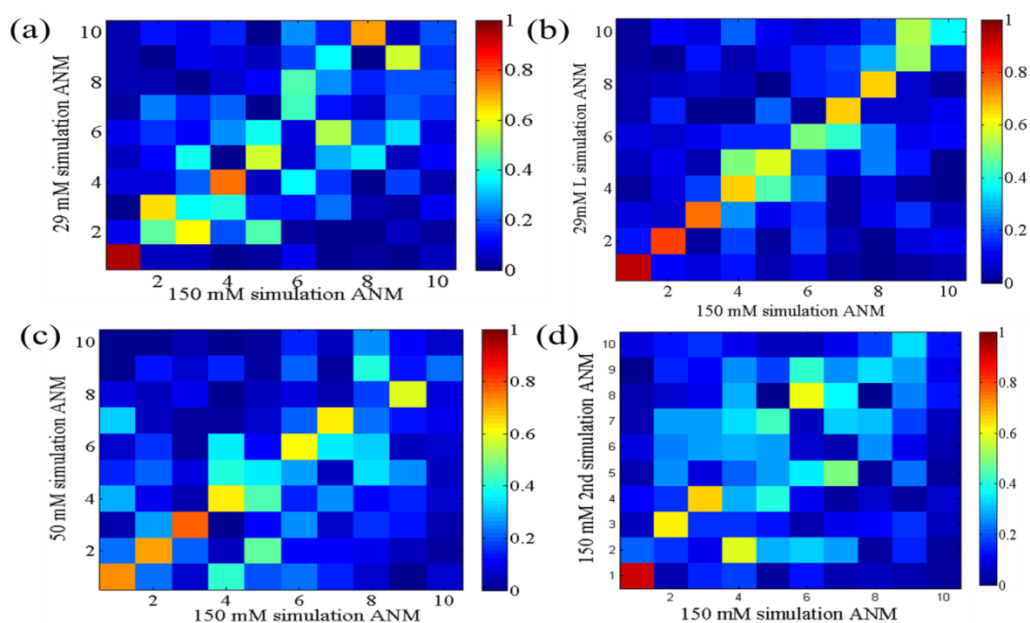
Figure 4.26. Overlap matrixes for ANM modes between 150 mM and (a) 29mM (b)29mM_L rn (c) 50mM and (d)150 mM runs.
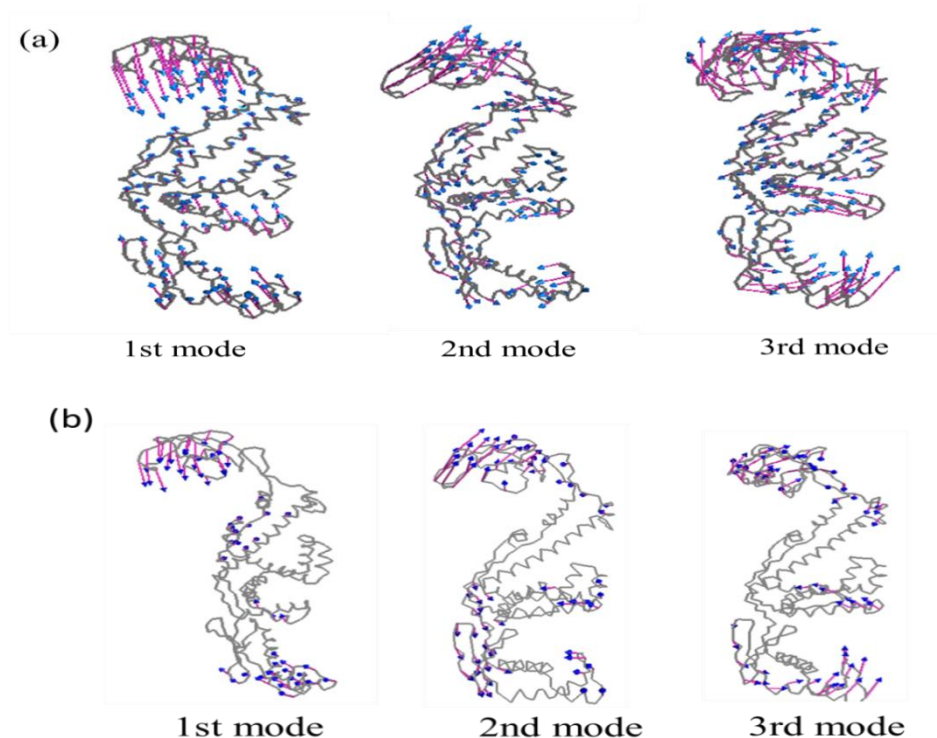


Figure 4.27. Vector notation for ANM applied on the average structure (a) 150 mM (b) 150 mM_2 runs.

### 4.6. Clustering of MD Snapshots

Snaphots of the all simulations (29 mM_L, 29 mM, 150 mM, 50 mM and 150 mM_2) were clustered together by using MMTSB toolit (Multiscale modelling tools for structural biology) (Feig et. al., 2004). Before clustering, all snaphots were superimposed by overall alignment on to the first frame of the 29mM_L run. K-means clustering method was used based on heavy atoms, with an RMSD cutoff of 6 Å. For 6 Å cutoff, 29 distinct clusters were obtained for five independent runs. In Figure 4.28, centroid numbers versus frame numbers are depicted. Snapshots of the same trajectories are located in the consecutive centroids with a few exceptions.

Clusters obtained after BD alignment are shown in Figure 4.29. In this figure, concatenated frames, were again superimposed on to the first snaphot of the 29 mM_L simulation, based on BD only. 5 centroids were obtained, with 6 Å cut off distance. To have a more detailed distribution, 4 Å cut off distance was selected, resulting 15 centroids.

In Figure 4.30 and Figure 4.31 simulations are clustered based on HD and CD alignments, respectively. Both for HD and CD aligned clustering, 4 Å RMSD cut off distance is selected. Accordingly, 13 and 17 centroids are obtained for HD and CD aligned clusters, respectively.
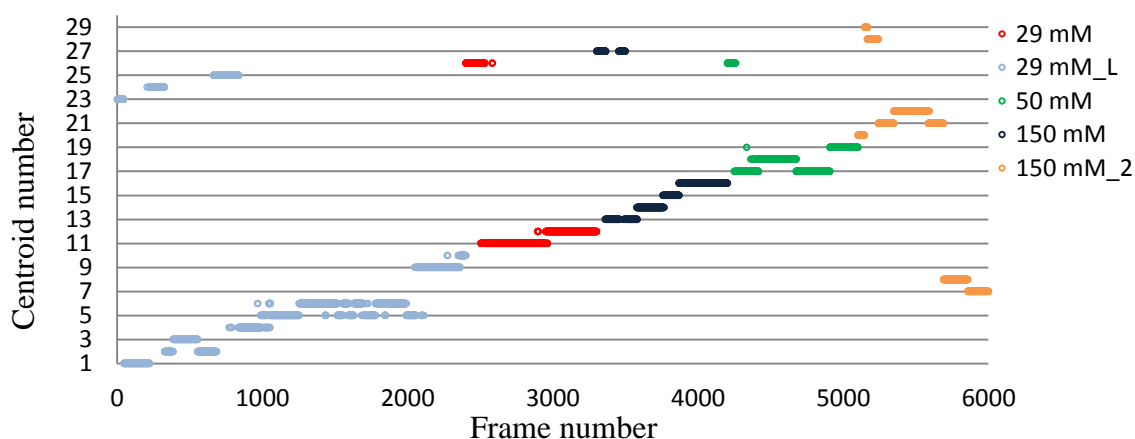


Figure 4.28. Centroids for concatenated frames based on overall alignment.
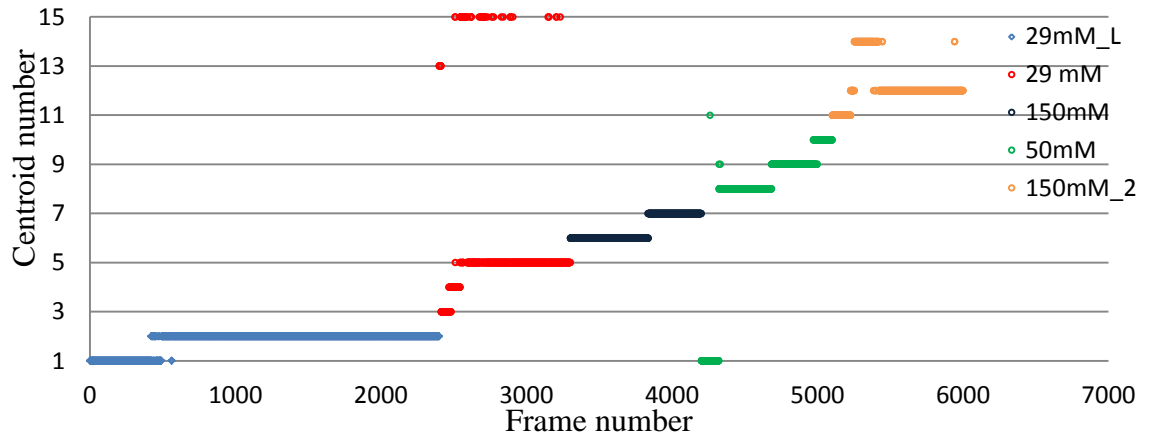
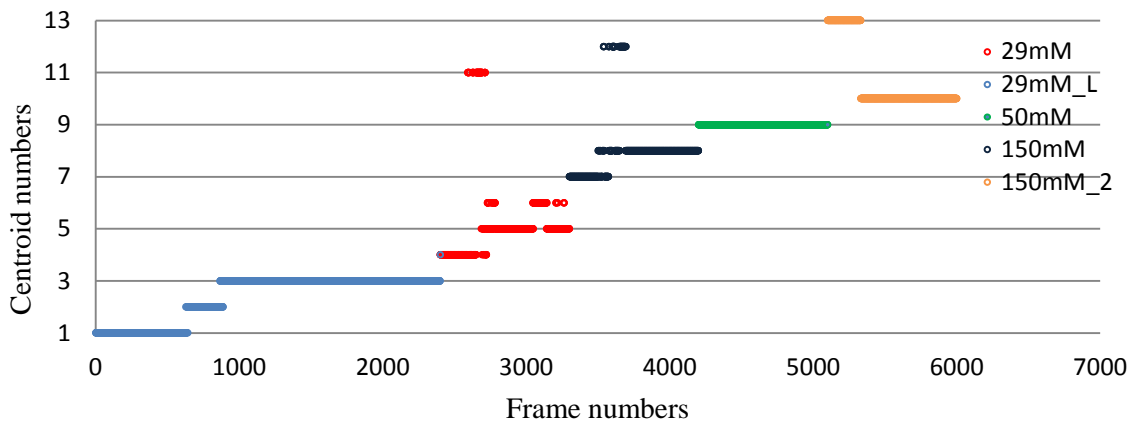Figure 4.29. Centroids for concatenated frames based on BD alignment.



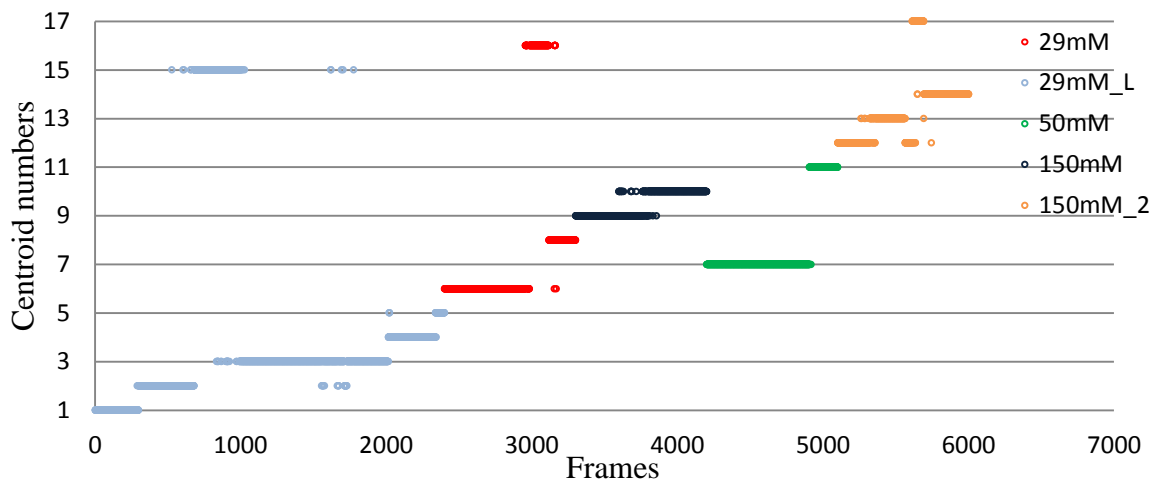Figure 4.30. Centroids for concatenated frames based on HD alignment.



Figure 4.31. Centroids for concatenated frames based on CD alignment.

Conformational changes based on overall, BD, CD and HD alignment are presented for 29 mM, 29mM_L, 150mM and 150mM_2 runs; therefore, two most populated centroids are selected for each run, and given in APPENDIX A, Table A.1. The best representative conformations of the regarding centroids are compared based overall and domain-wise alignment.

In Figure 4.32, conformational rearrangements of 29 mM and 29 mM_L runs are compared based on overall, BD, CD, HD aligned clusters. Looking at the overall aligned cluster centroids in Figure 4.32a, the dominant deformations of the 29 mM run results from the movement of BD.

In Figure 4.32b, one of the helices next to the binding loop is observed to lengthen due to the loss of the link region for 29 mM run. The formation of this longer helical region during the initial stages of 29 mM run is shown by the indicated snapshots in Figure 4.33. In the first 20 ns, rearrangement from coil to helix is observed, and this form stands throughout the simulation.

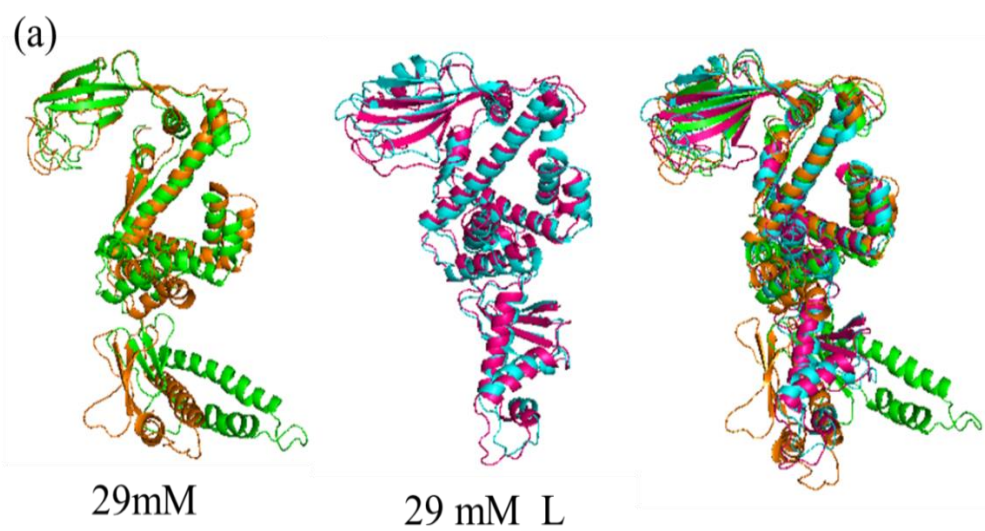

Figure 4.32. Alignment of conformations from the regarding centroids for 29 mM, 29mM_L and both runs for (a) overall (b) BD (c) HD (d) CD alignments.

Figure 4.32. Alignment of conformations from the regarding centroids for 29 mM, 29mM_L and both runs for (a) overall (b) BD (c) HD (d) CD alignments cont.

(a) initial position

(b) 4 ns

(c) 8 ns

(d) 21 ns

(e) 20 ns
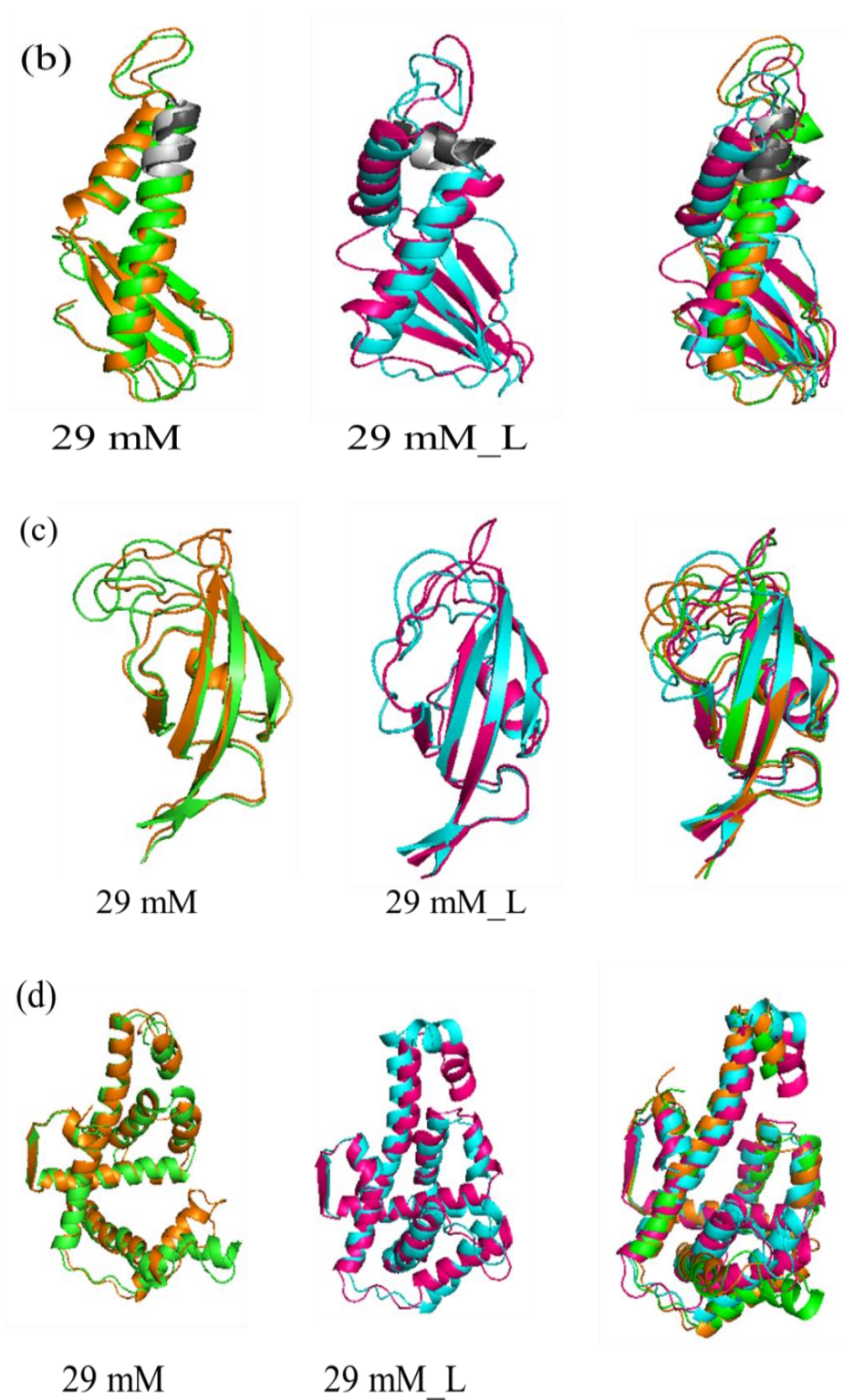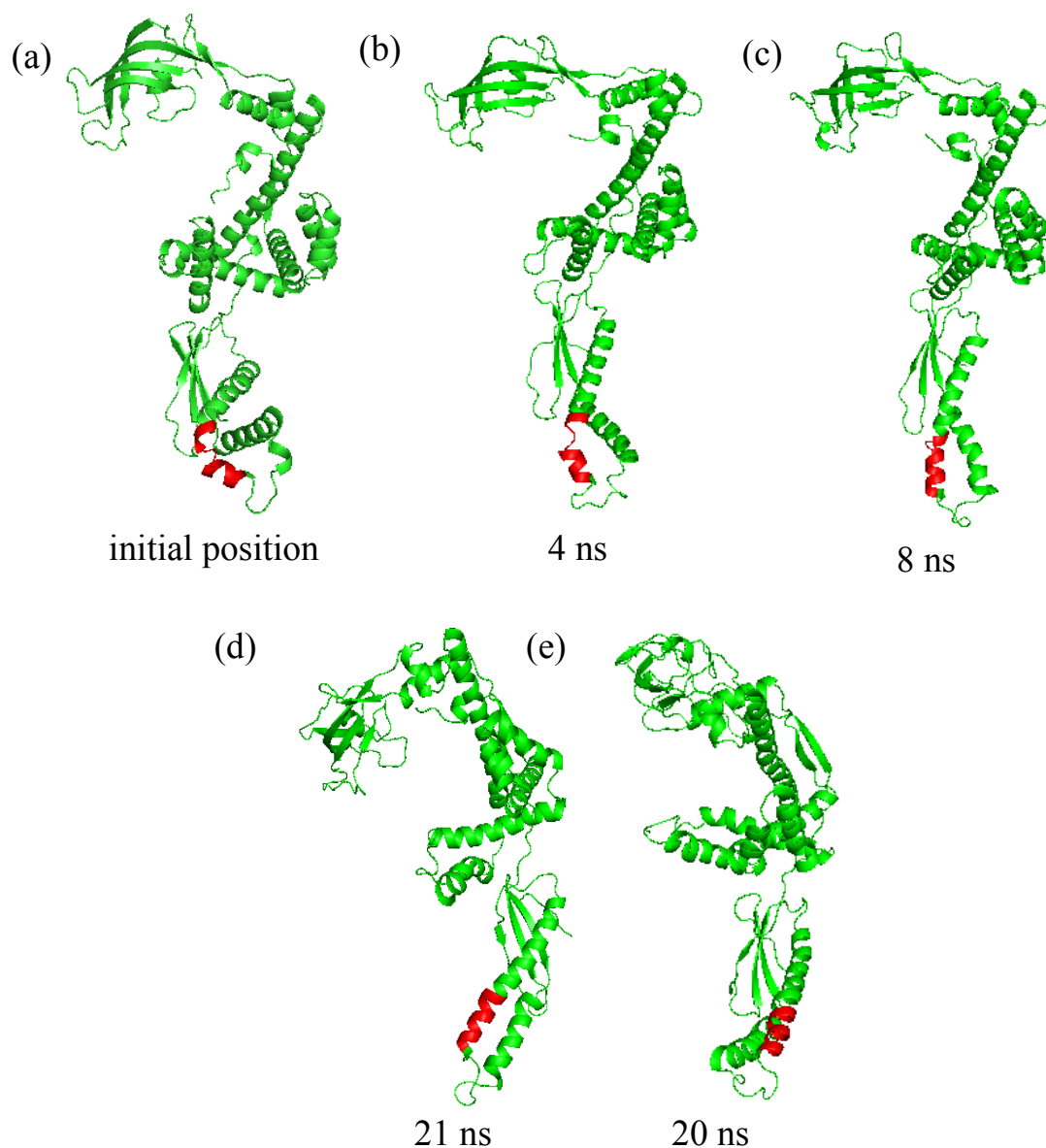
Figure 4.33. Extension of helix in binding region during initial stages of 29 mM run.

In Figure 4.34, the dominant clusters for 150 mM and 150 mM_2 runs (overall alignment) indicate major movements of HD. In line with this observation, $R_g$ values excluding HD, i.e. including BD and CD residues, are plotted together with normalized distribution values in Figure 4.35 and 4.36 respectively. The major oscillations in gyration radius without HD is lower than that observed by the whole structure, given in Figure 4.8. HD motion with respect to CD is more dominant at 150 mM ionic strength, however opening/closing between HD-CD and CD-HD appear at the same time at 29 mM and 50 mM runs.

(a)

150 mM           150 mM  2

(b)

150 mM           150 mM_2

(c)

150 mM           150 mM  2

Figure 4.34. Alignment of conformations from the regarding centroids for 150 mM and 150 mM_2 runs (a) overall alignment (b) BD alignment (c) HD alignment (d) CD alignment.

(d)



150 mM                    150 mM_2

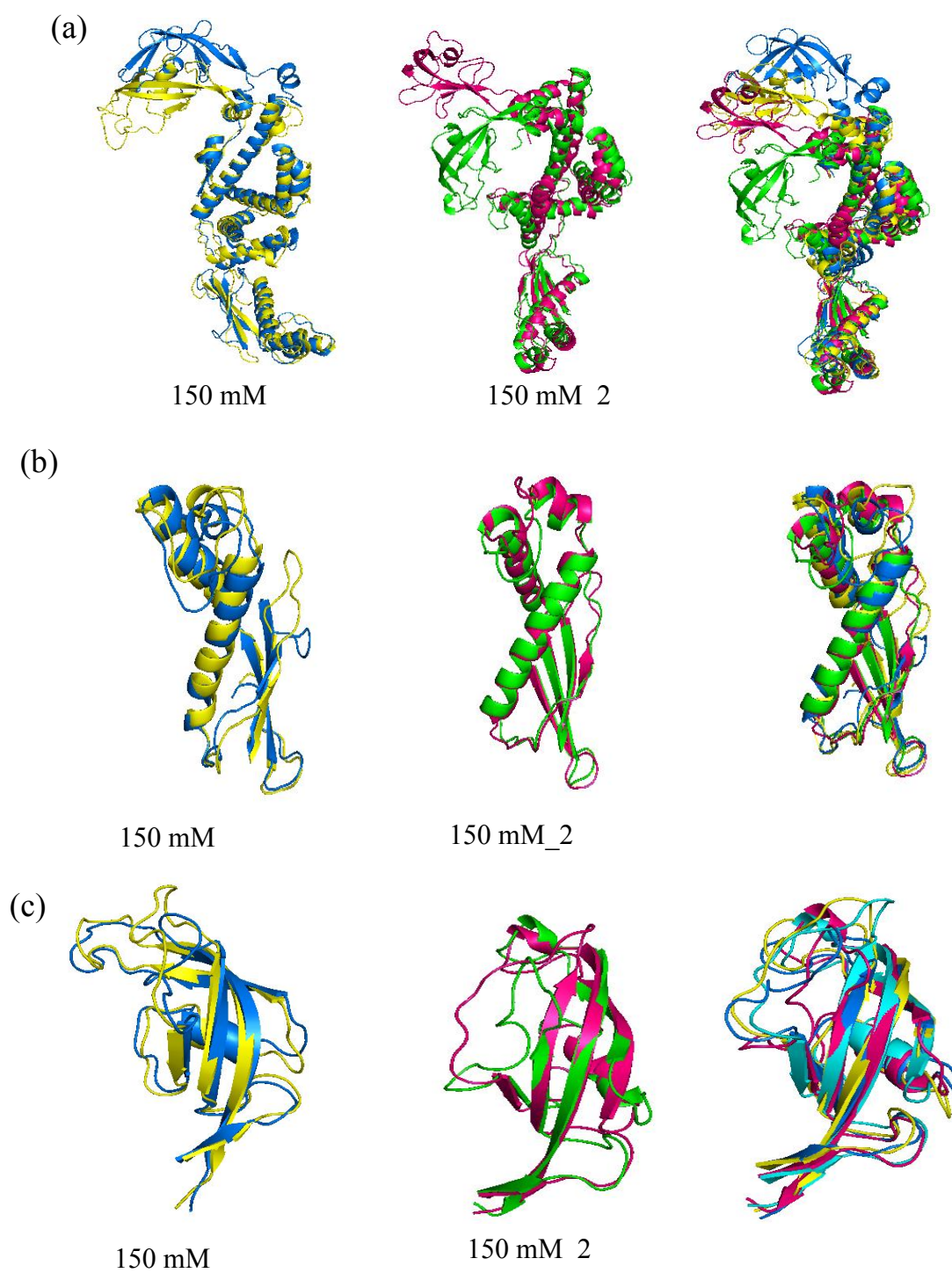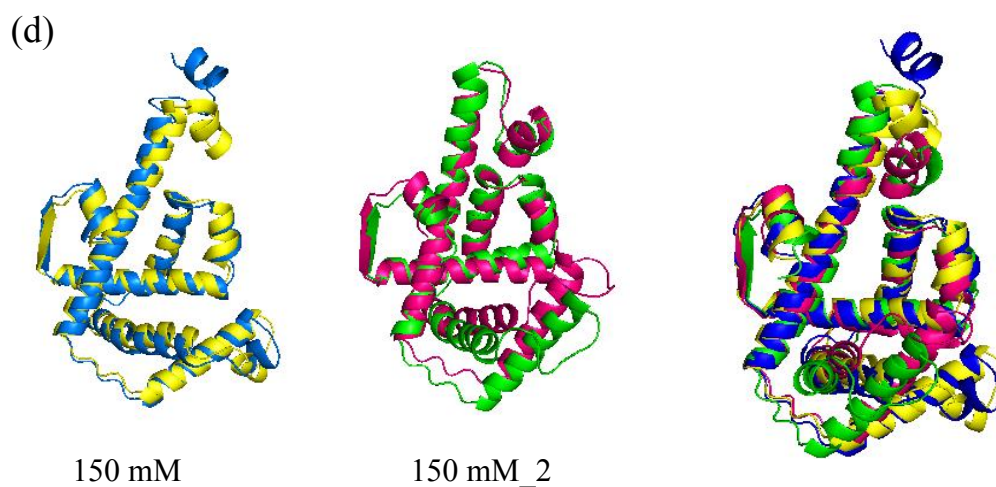Figure 4.34. Alignment of conformations from the regarding centroids for 150 mM and 150 mM_2 runs (a) overall alignment (b) BD alignment (c) HD alignment (d) CD alignment cont.
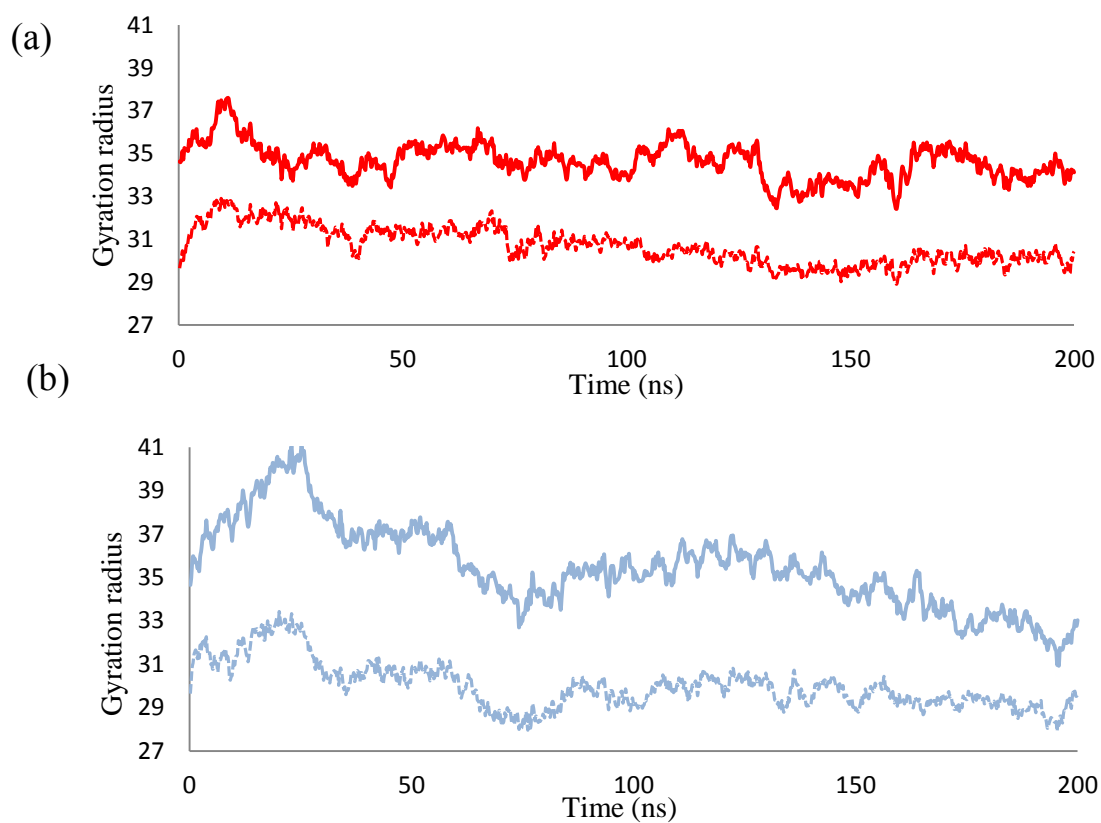


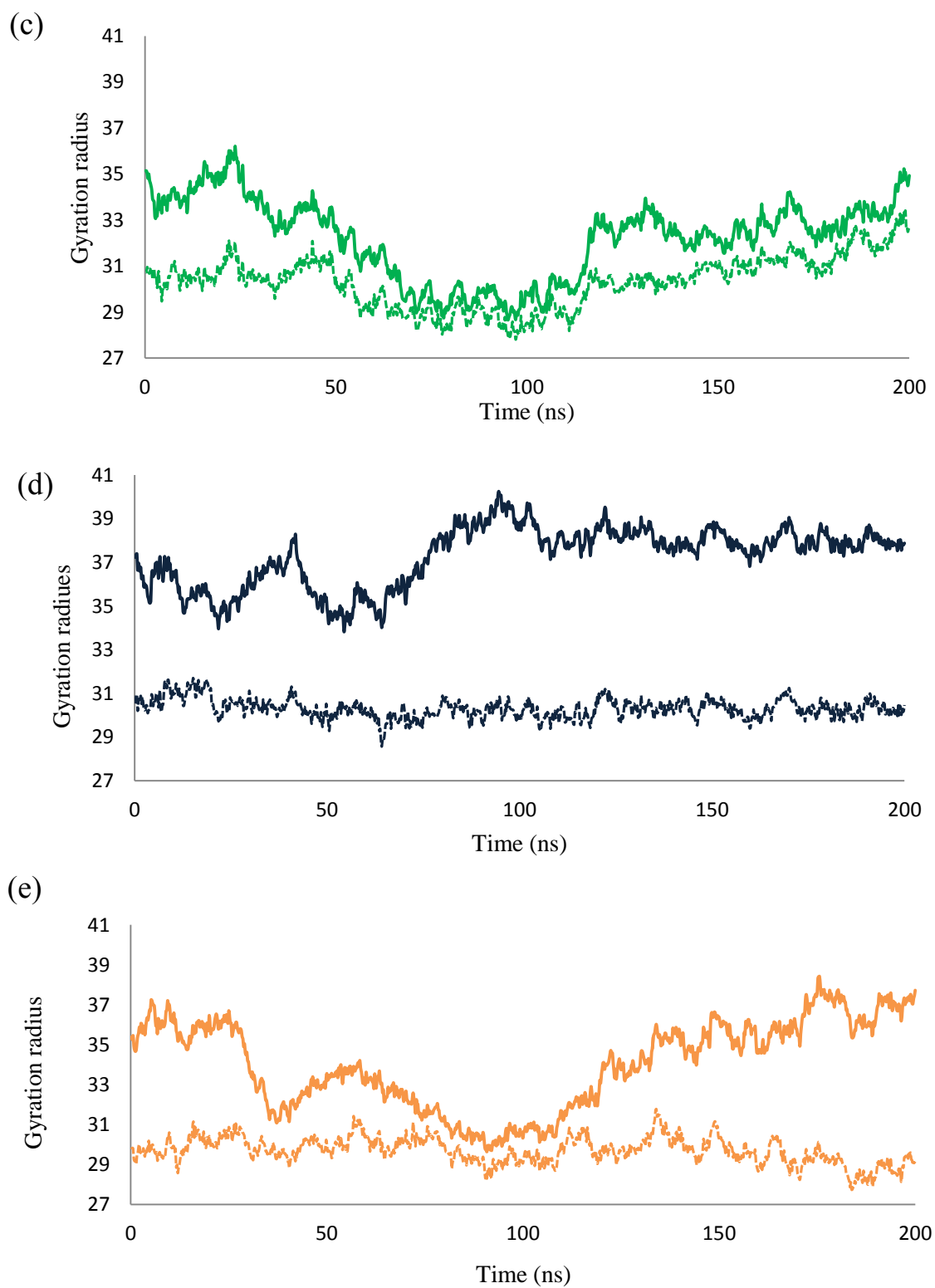Figure 4.35. Gyration radius excluding HD for (a) 29mM (b) 29mM_L (c) 50 mM (d) 150 mM (e) 150 mM_2.

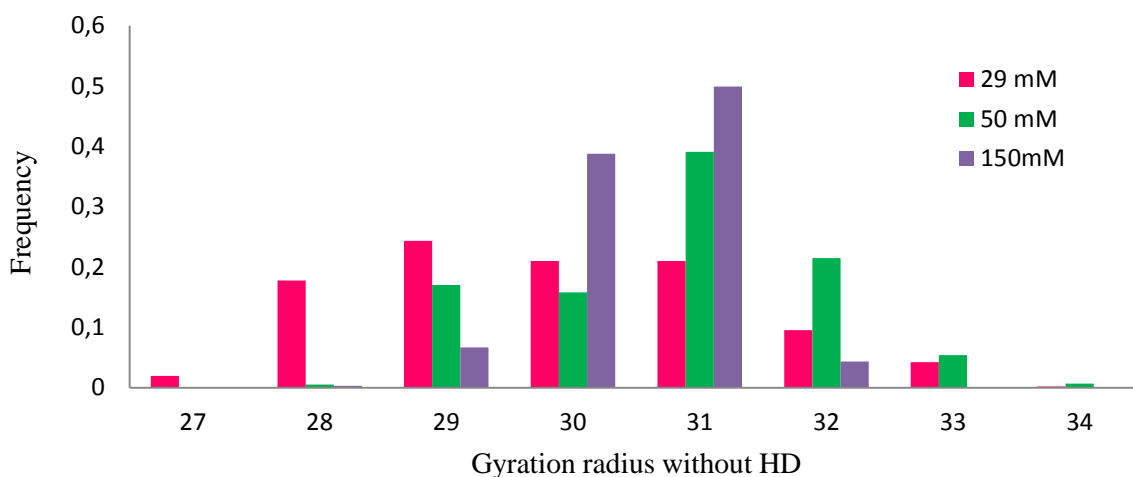Figure 4.35. Gyration radius excluding HD for (a) 29mM (b) 29mM_L (c) 50 mM (d) 150 mM (e) 150 mM_2 cont.

Figure 4.36. Normalized distribution of gyration excluding HD.

## 4.7.   Collective Dynamics of TF Docked on Ribosome 50S Subunit

Short segments from TF binding domain has been co-crystallized with the 50S subunit of ribosome, which are reported in two different structures (PDB ID:1W2B and PDB ID: 2AAR). In 1W2B, coordinates of a 35-residue long *E. coli* TF binding domain (residues 25-59) are resolved in atomistic detail, while a 113-residue long *H. marismortui* TF binding domain (residues 1-113) is reported as a coarse-grained chain with alpha-carbon coordinates in 2AAR. The atomistic ribosome structure in 1W2B contains 50S ribosomal proteins; L2P, L3P, L4P, L5P, L6P, L7AE, L10E, L13P, L14P, L15P, L15E, L18P, L18E, L19E, L21E, L22P, L23P, L24P, L24E, L30P, L31E, L32E, L37AE, L39E, L44E, whereas the coarse-grained structure contains only L23 and L29.

In this section, a composite structure is obtained by alignment of these 50S subunits from 1W2B and 2AAR. The main reason for this alignment is to have the intact 50S (atomistic detail) with the relatively longer BD segment (113-residue long segment) bound to it. Then, apo TF conformations obtained from MD simulations were aligned or 'docked' on this coarse-grained BD segment in the composite structure, which resulted in a complex structure of intact ribosome 50S with intact TF. Complexes formed using this specific alignment procedure will be called 'docked TF' in the following text.

Figure 4.37 shows such a docked conformation from 150 mM MD simulation at 78 ns from two different perspectives. The energy of this 50S-TF complex was minimized to

release any unfavorable contacts. Energy minimization was performed in implicit solvent using AMBER11 as the force field. Later, the collective dynamics of the complex structures before and after energy minimization were each examined by ANM. The results were used to create the cross correlation maps based on the first and cumulative 10 ANM modes for TF only in Figures 4.38 and 4.39. The trends of the two maps are similar to each other, indicating that energy minimization did not affect the domain dissections.

In Figure 4.40, ANM was applied on the apo form of the docked TF conformation. The domain decomposition is consistent with the previously reported ones in Table 4.2 and Figure 4.24. Comparison of docked and apo TF results indicates that the linker region (residues 111- 149) cannot be observed as a separate entity in the docked TF. Instead, it is moving together with a part of HD rather than the CD as in apo TF. In docked TF, the CD is much smaller than in apo TF. In docked TF, another region that actually moves together with BD appears, i.e. the BD is enlarged.

Selected conformers from overall and BD aligned clusters are similarly docked to the ribosome, and given in APPENDIX A. The position of 50S docked conformations for all aligned and BD aligned clusters can be found in Figure A.1 and Figure A.2, respectively.
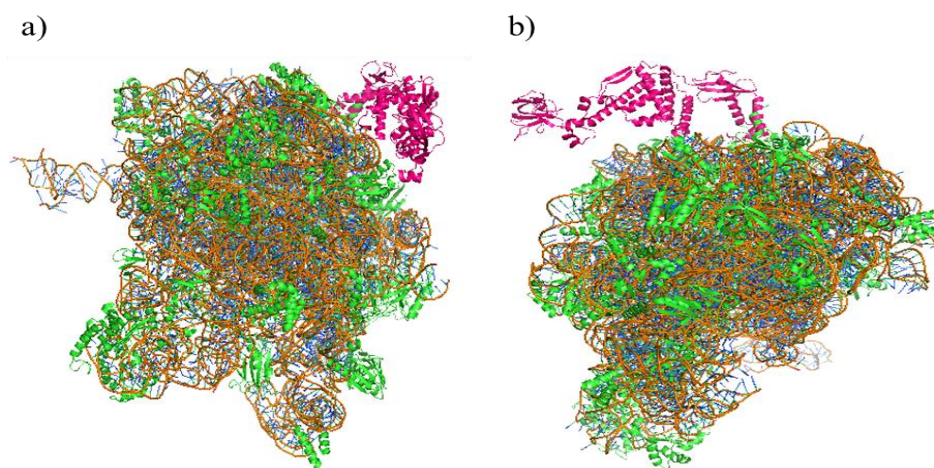


Figure 4.37. 78$^{th}$ ns conformation from 150 mM MD run (at centroid14) aligned on to the 50S ribosomal subunit (a) right view (b) left view.
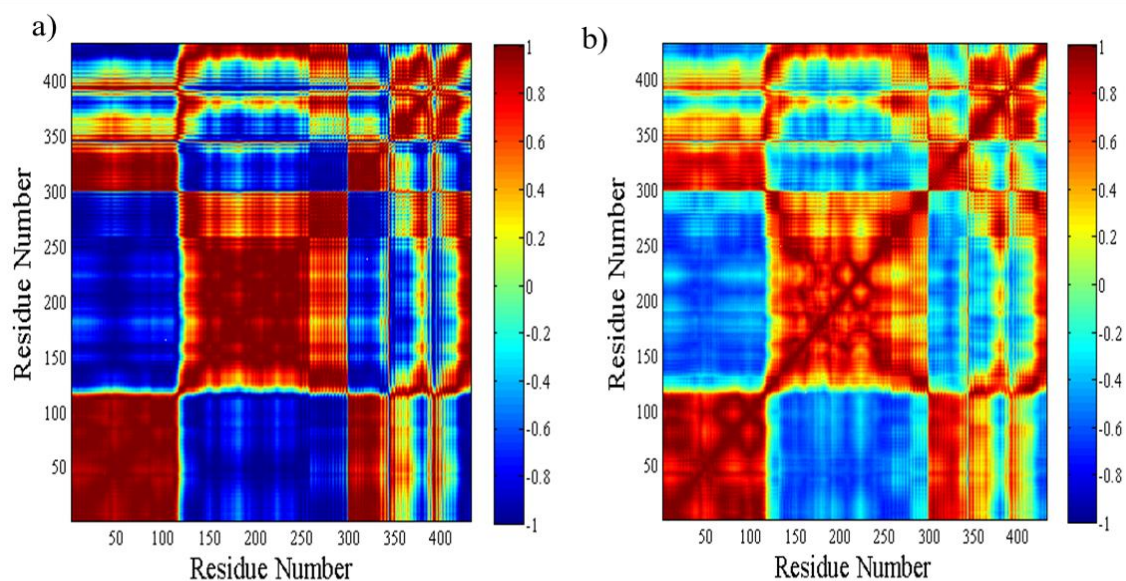
Figure 4.38. ANM cross-correlation maps for TF part of the docked structure (at 78 ns) after energy minimization, based on (a) 1$^{st}$ mode (b) cumulative 10 modes.
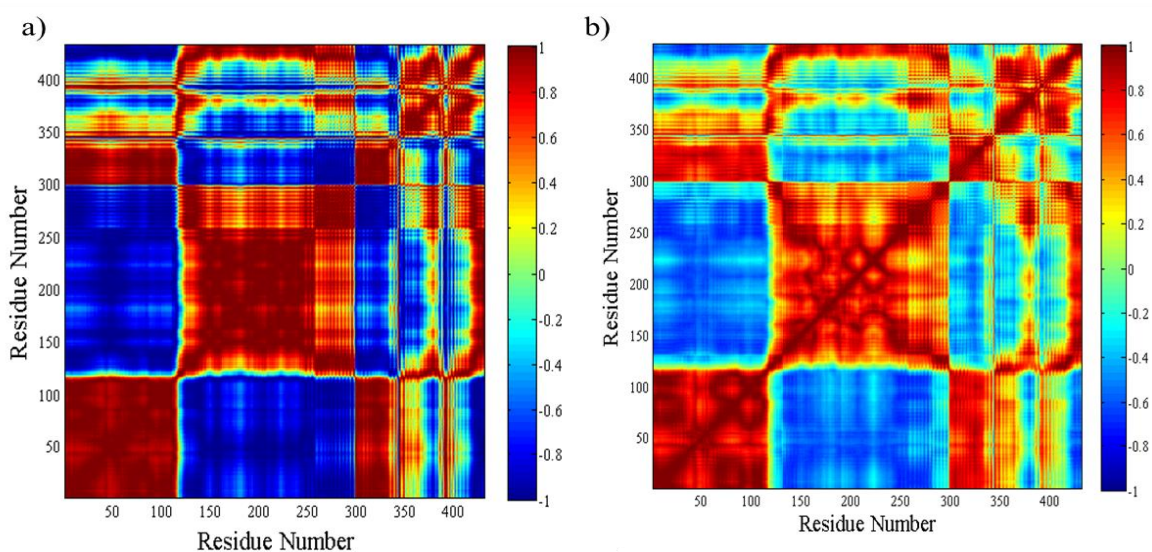


Figure 4.39. ANM cross correlation maps for TF part of docked structure (at 78 ns) without perfoming energy minimization, based on (a) first mode ( b) cumulative 10 modes.
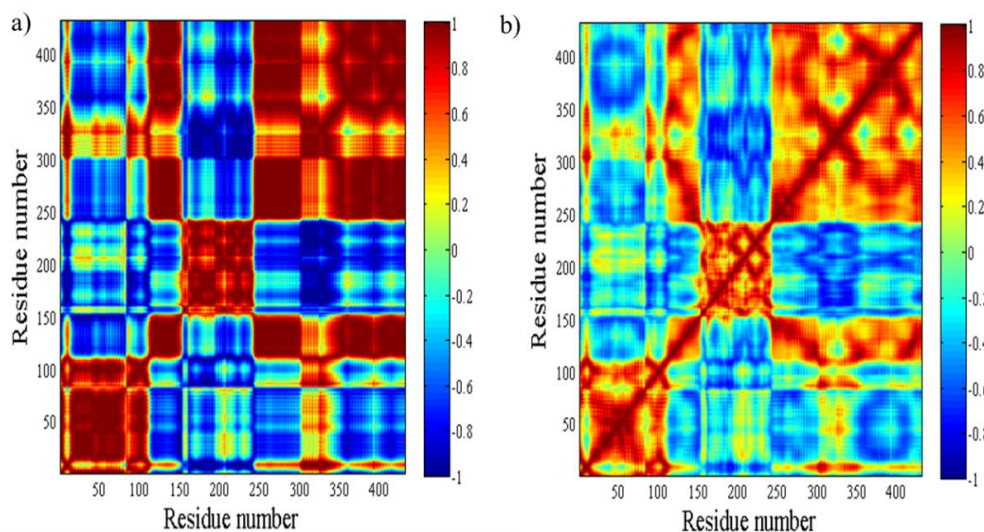
Figure 4.40. ANM cross-correlation maps for apo structure (at 78 ns), based on (a) 1st mode (b) cumulative 10 modes.

Since the ionic strength of a cell medium is close to 150 mM, other TF conformations at this ionic strength were selected for further ANM analyses. Eight conformations were chosen from the highly populated clusters in Figure 4.28 for this purpose. Four conformations from 150 mM run (centroids 13, 14, 15, 16) and four from 150 mM_2 run (centroids 7, 8, 21, 22) were used for ANM analyses. Cross-correlation maps resulting from the apo structures for the first mode are presented in Figures 4.41 and 4.42 for the respective runs. The alignments of TF on 50S and the corresponding cross-correlation maps for the first mode are shown in Figures 4.43 and 4.44 for 150 mM and 150mM _2 runs, respectively.

For 150 mM run, all conformers, apo or docked forms, exhibit similar cross-correlation maps and domain decomposition as the 78 ns snapshot reported above. In docked clusters given in Figure 4.43, the linker is correlated with the HD, so there is a larger positively correlated HD block in the middle of figure. The CD is a smaller block and part of it is now correlated with the BD. The domain decompositions for these docked TF centroids are given in Table 4.2.

Figure 4.41. Cross-correlation maps based on first ANM mode of apo TF from 150 mM run.
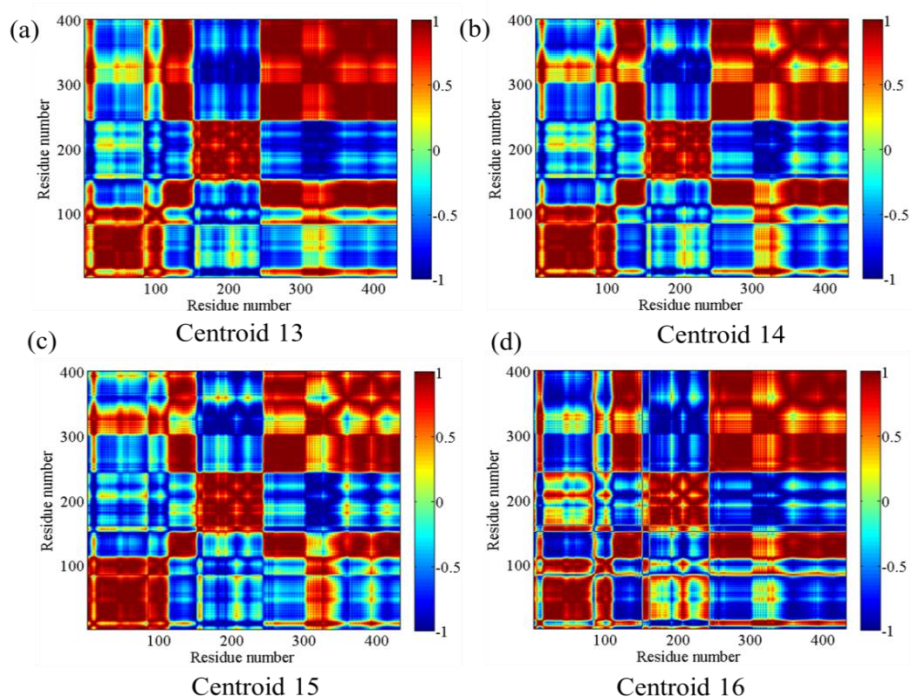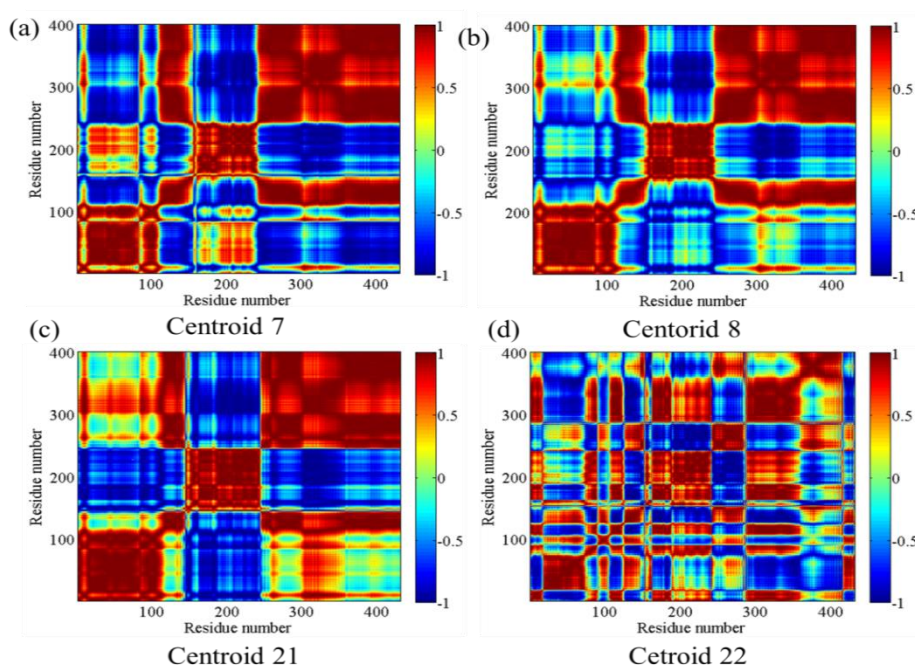


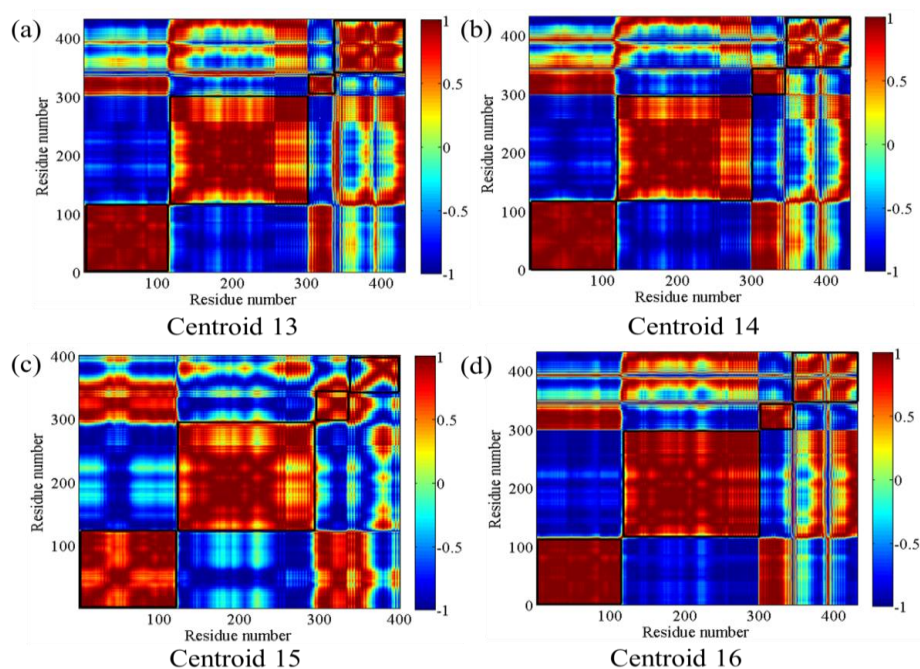Figure 4.42. Cross-correlation maps based on first ANM mode of apo TF from 150 mM_2 run.

Figure 4.43. Cross-correlation map based on first ANM mode for 150 mM centroids docked on 50S.
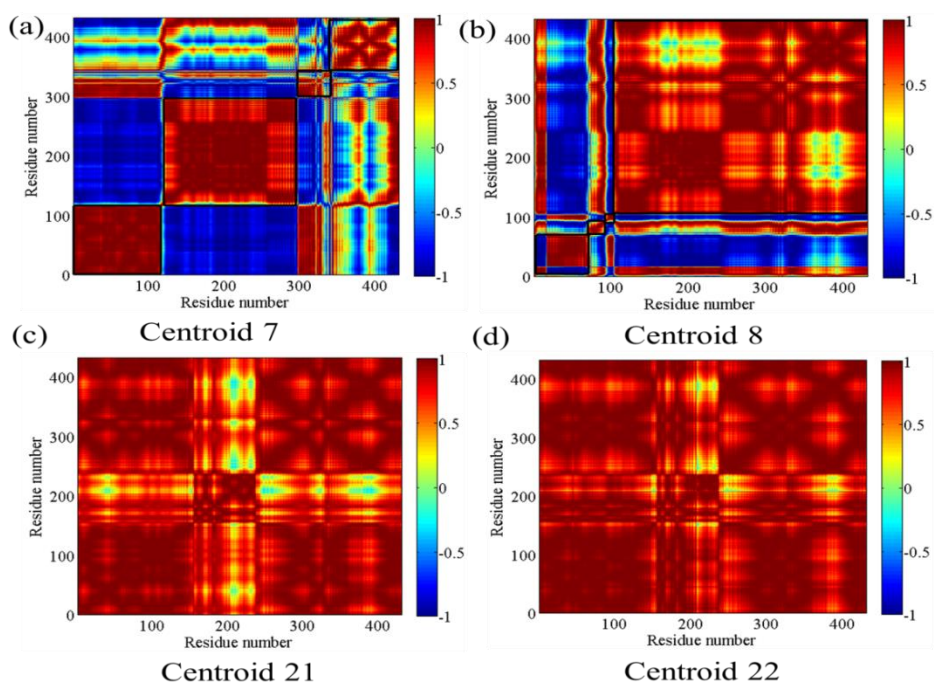


Figure 4.44. Cross-correlation map based on first ANM mode for 150 mM_2 centroids docked on 50S.

Table 4.3. Domain dissection for 150 mM centroids docked on 50S

| 150 mM | Binding Domain | Linker | Head Domain | Core Domain |
|---|---|---|---|---|
| Centroid 13 | 1-115, 301-346 | - | 116-300 | 347-432 |
| Centroid 14 | 1-117,298-342 | - | 118-297 | 343-432 |
| Centroid 15 | 1-121, 294-339 | - | 122-293 | 340-432 |
| Centroid 16 | 1-116, 298-345 | - | 117- 297 | 346-432 |

Table 4.4. Domain dissection for 150_2 mM centroids docked on 50S

| 150 mM_2 | Binding Domain | Linker | Head Domain | Core Domain |
|---|---|---|---|---|
| Centroid 7 | 1-117, 298-341 | - | 118-297 | 342-432 |
| Centroid 8 | 1-71, 94-105 | - | 72-93, 106-432 | |
| Centroid 21 | | | | |
| Centroid 22 | | | | |

On the other hand, some of the centroids from 150 mM_2 run display quite different domain decompositions in Figure 4.44. For centroid 8 (panel b), the BD and the rest of TF form two main domains. For centroids 21 and 22 (panels c and d), the whole TF acts as a correlated entity, i.e. there is no inter-domain flexibility. The correlations in the apo form of these last two centroids display some different features from the common picture that has emerged for all other apo structures shown previously.

For comparing first mode motions in more detail, mode vectors for 150 mM (centroid 14) and 150 mM_2 (centroid 22) runs are created in Figure 4.45 and Figure 4.46, respectively.
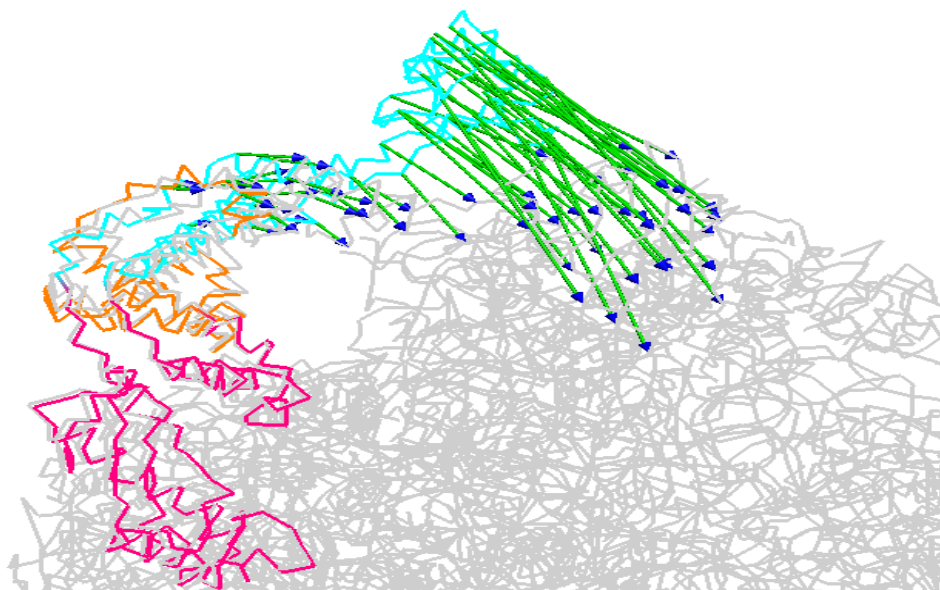
Figure 4.45. First mode deformations of 50S docked TF (centroid 14).

In Figure 4.45, mode vectors for the first mode of the motion is represented. BD (pink), CD (orange) and HD (blue) are colored based on the resulting domain decompositions, given in Table 4.3. Mode vectors shows the direction of motion, which is just seen for HD region.
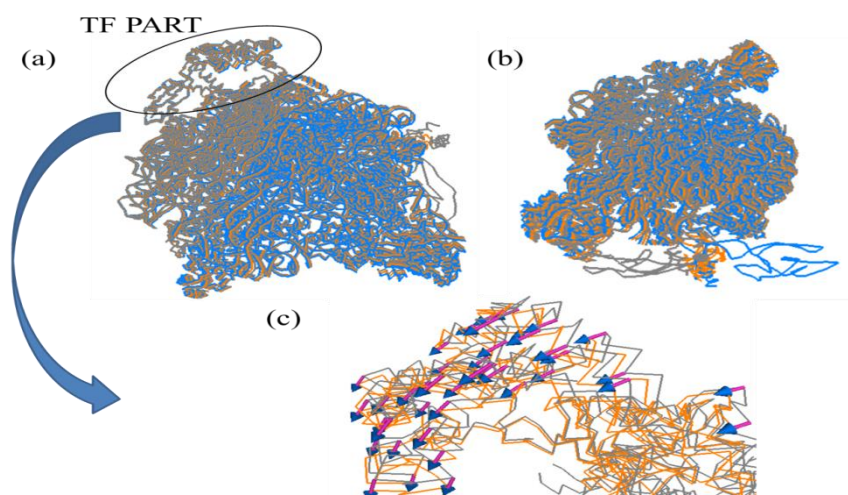


Figure 4.46. First mode deformations of 50S docked TF (centroid 22) (a) position of TF on 50S (b) position of TF from different perspective (c) mode vectors for TF part.

In Figure 4.46, 50S docked TF conformes is depicted from different perspectives in panel a and b. In Figure 4.45, (centroid 14) the major displacements are observed in TF, whereas ribosome 50S is relatively immobile. In contrast, in Figure 4.46c (centroid 22) TF shows little displacement, while 23S RNA part of ribosome moves. As observed in Figure 4.46, HD of TF is quite close to ribosome, therefore, it cannot move independently. Since its domains move collectively, domain dissections are not observed in Figure 4.44d.

Resulting domain dissections are coloured in Figure 4.47 and Figure 4.48 for 150 mM and 150 mM_2 simulations, respectively. Positions of BD (red), CD (orange) and HD (cyan) with ribosome are depicted. L23 (blue) and L29 (red) ribosomal proteins can also be seen in the same figures.
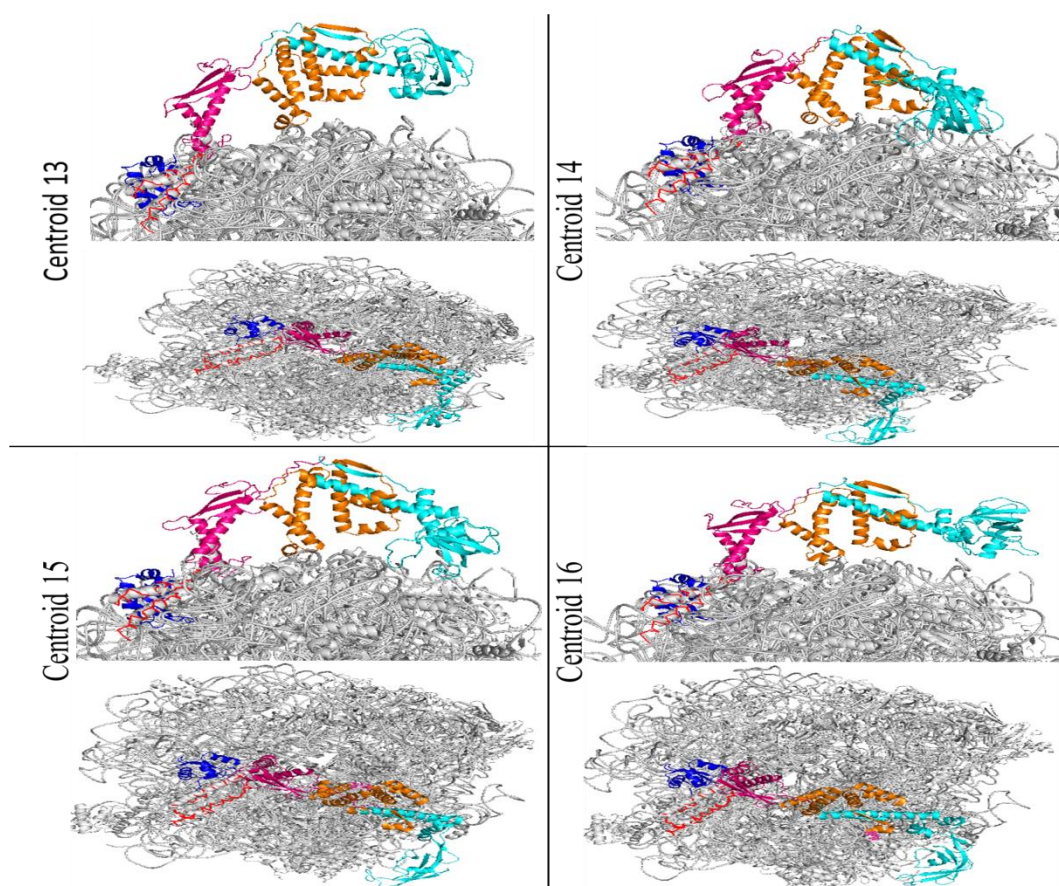


Figure 4.47. Positions of TF with 50S ribosomal protein for 150 mM centroids.
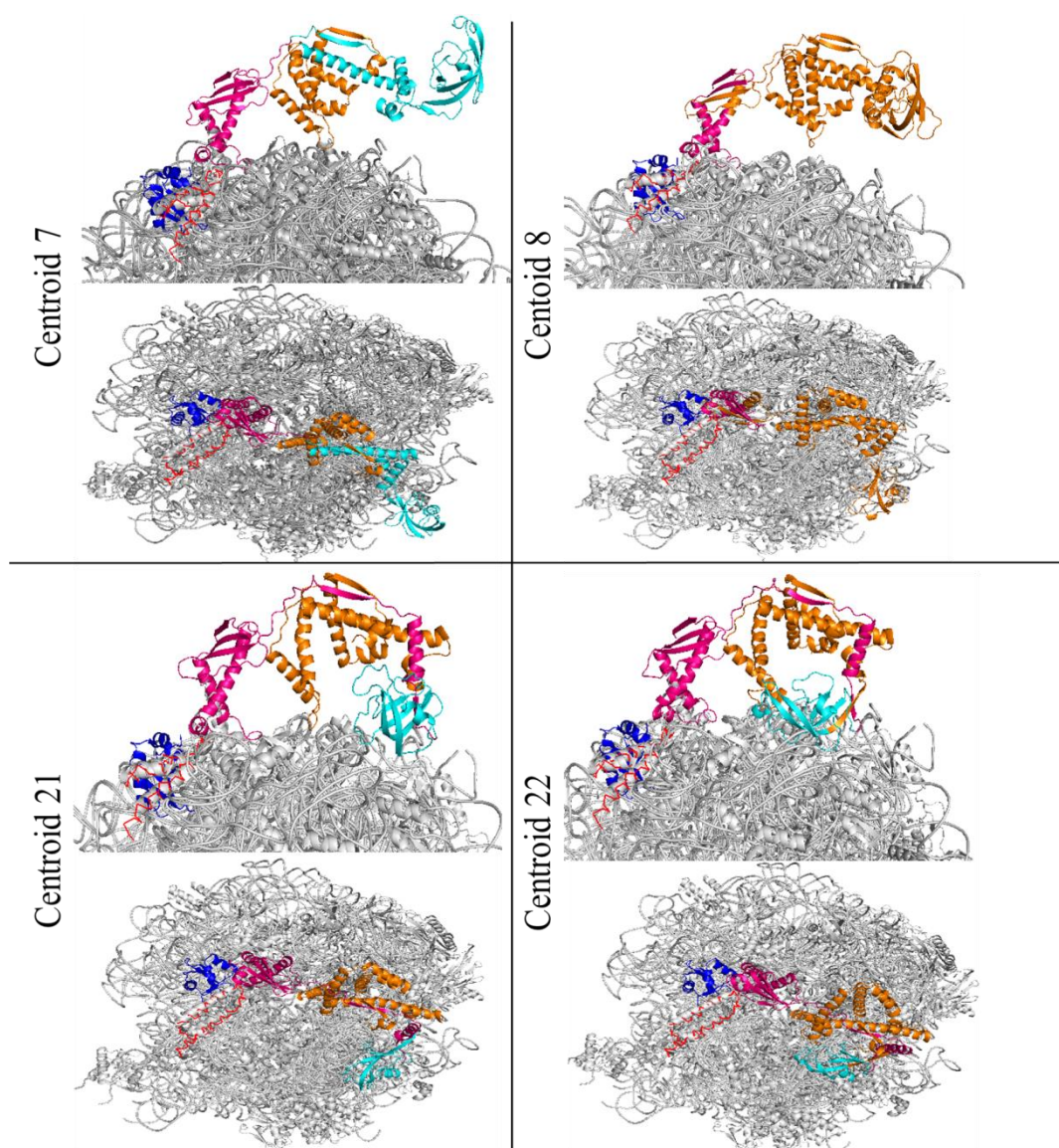
Figure 4.48. Positions of TF with 50S ribosomal protein for 150 mM_2 centroids.

Overlap matrix is also created between apo-TF and ribosome- bound TF for the selected centroids for 150 mM and 150 mM_2 runs, in Figure 4.49 and 4.50 respectively. In Figure 4.49, the first three modes of apo-TF correlates especially with the first 6 modes of ribosome-bound TF. This correlation trend resembles to the ones in 4.49a. However, in Figure 4.50, correlation trend is not as high as in 150 mM run. The lowest correlation matrix is seen for Figure 4.50d.
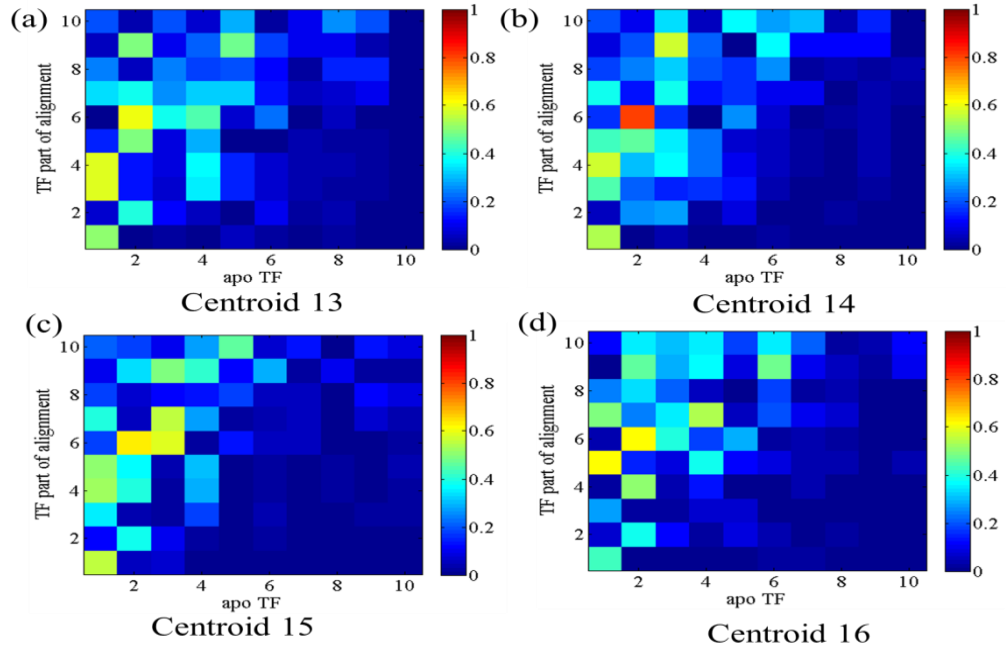
Figure 4.49. Overlap matrix for ANM applied apo TF and TF part of aligned complex for 150 mM runs .
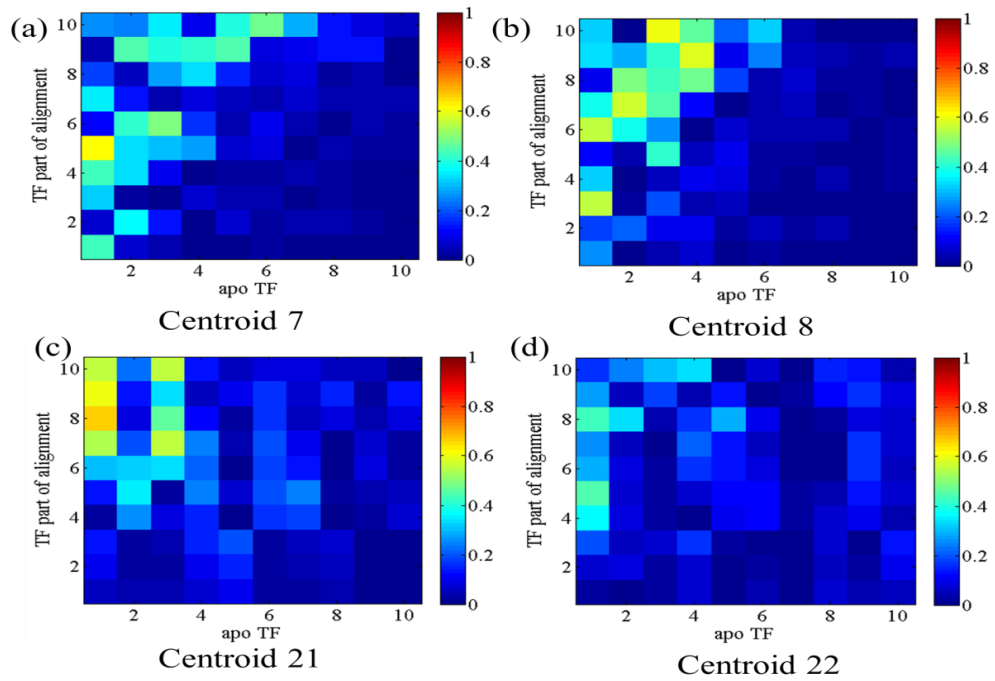


Figure 4.50. Overlap matrix for ANM applied apo TF and TF part of aligned complex for 150 mM_2 runs .

Moreover, positions of TF on 50S ribosome subunit is also depicted for 29 mM, 29mM_Land 50 mM simulations, based overall alignment cluster results. Different TF positions on 50S can be seen.
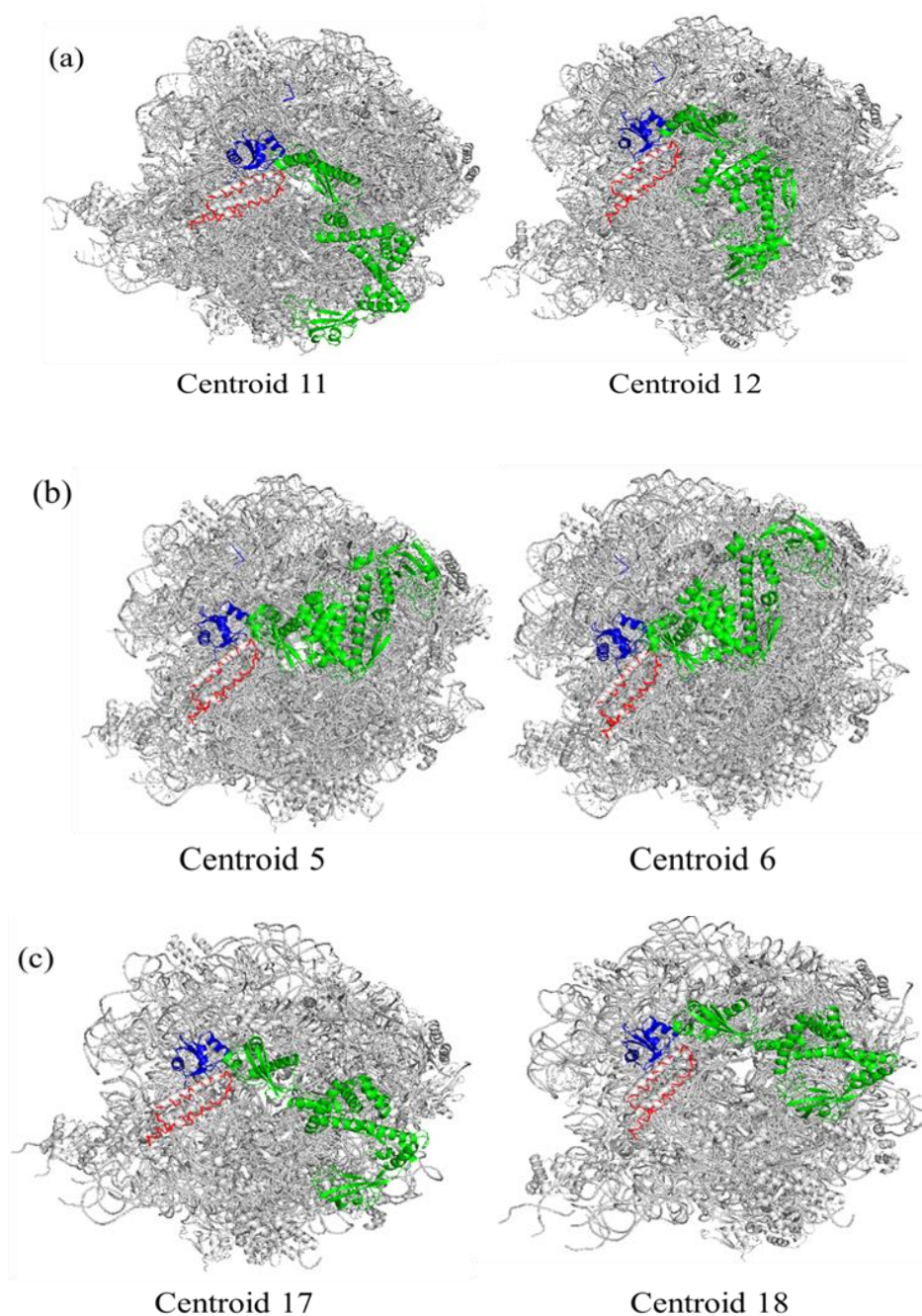


Figure 4.51. Positions of L23 and L29 proteins on 50S ribosome aligned complex for (a) 29 mM (b) 29 mM_L (c) 50 mM runs.

## 4.8.    Solvent Accessible Surface Area (SASA) Calculation

Total SASA of whole protein for the independent simulations, and normalized distributions are plotted in Figure 4.52. In this figure, normalized distributions are in a similar fashion with the ditribution of gyration values, given in Figure 4.11. Total hydrophobic SASA value plot again with normalized distributions for whole protein is depicted as well in Figure 4.53. Distribuiton manner of total SASA and total hyrophobic ssasa values are found as parallel.
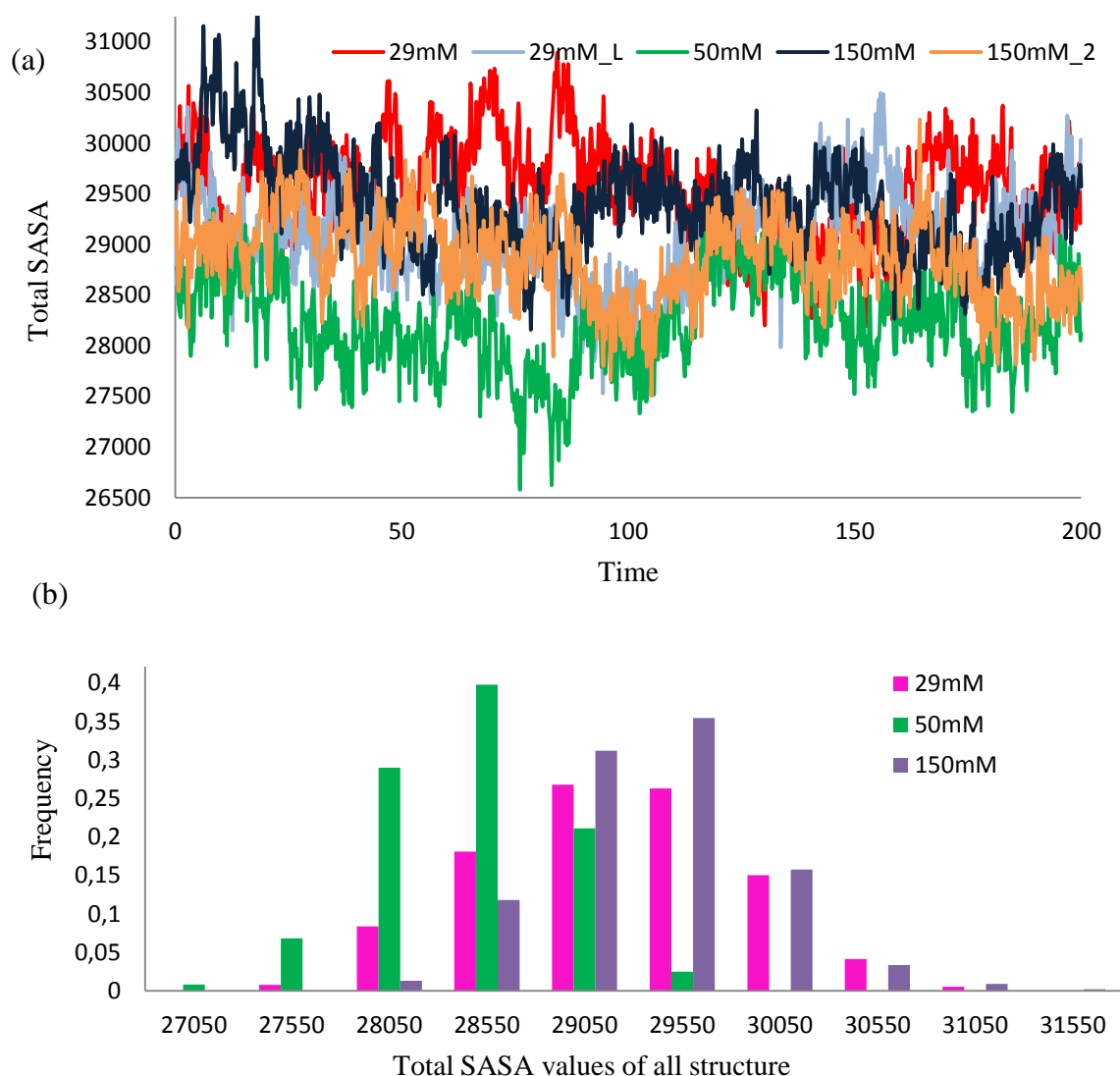


Figure 4.52. Total SASA values for whole protein at different ionic strengths (a) change in time (b) normalized distributions.
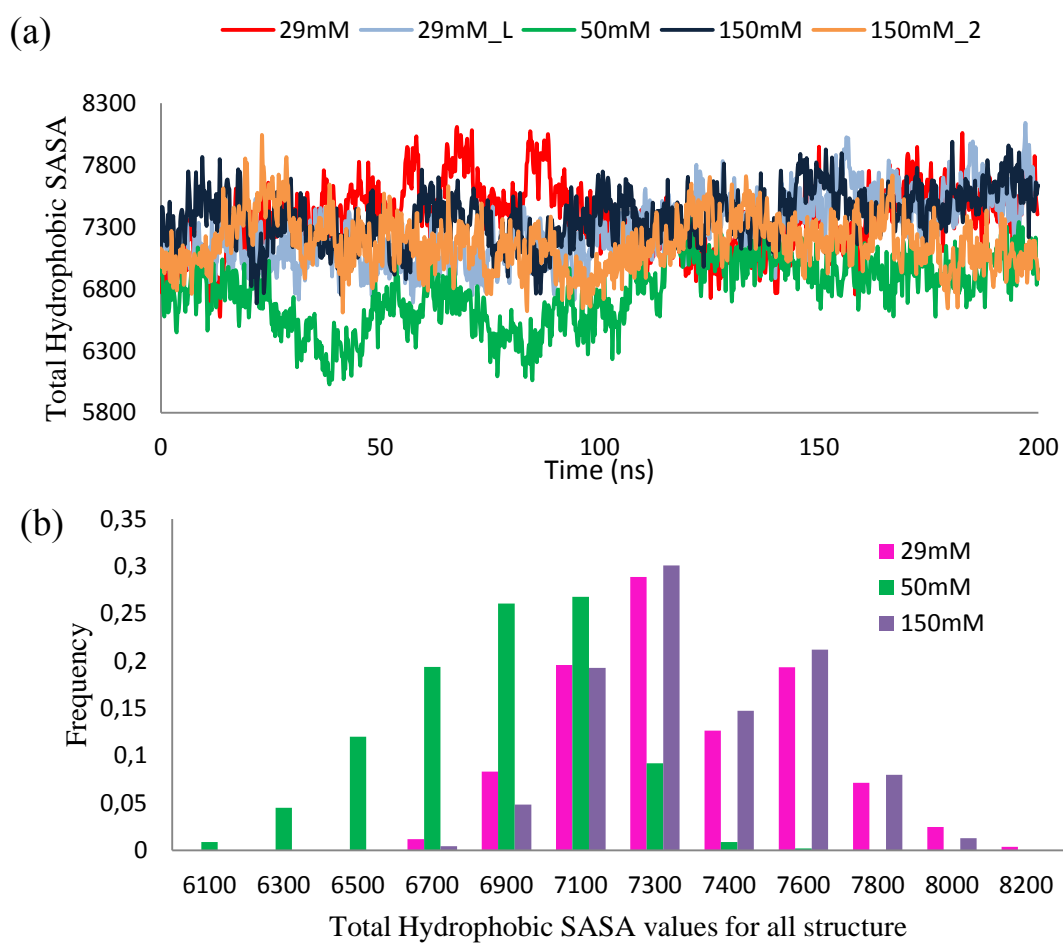
Figure 4.53. Total hydrophobic SASA values for whole protein at different ionic strengths (a) change in time (b) normalized distributions.

# 5.  CONCLUSION

Four independent 200 ns long MD simulations were performed at different ionic strengths (29 mM, 50 mM and two simulations for 150 mM) in explicit solvent to investigate conformational dynamics of *E. coli* TF (PDB ID:1W26). RMSD and RMSF analyses indicated significant inter-domain and intra-domain conformational flexibility. The binding loop in BD, the two arms (arm1 and arm2) in CD and two unstructured loops in HD are observed as the most mobile regions. Large oscillations observed in $R_g$ values were linked to the opening closing motion of TF between HD-CD and CD-BD pairs. While the HD motion with respect to CD is more dominant at 150 mM ionic strength, opening closing between HD-CD and CD-BD appear at the same time at 29 and 50 mM runs. Histograms of SASA and $R_g$ distributions at different ionic strengths indicate a similar trend where 50 mM run exhibits relatively lower values compared to others. In the 29 mM simulation, a unique structural rearrangement in BD, namely extension of a helix next to the binding loop was detected.

Conformational states of TF were also investigated based on the COM distances between HD-arm1 and BD-arm2. Collapsed stated of TF, which was reported based on six 250 ns long MD runs with AMBER03 forcefield (Singhal *et al*., 2013) and a 1.5 μs long MD run with OPLS-AA/L forcefield (Thomas *et al*., 2013), could not be observed in current simulations performed with CHARMM22 forcefield. This disparity may either originate from the different force fields used or the duration of the simulations. Specifically, the collapsed state was detected at the early stage of 250 ns simulations, whereas in 1.5 μs simulation it was detected after 500 ns.

PCA and ANM calculations were carried out to reveal the collective dynamics of TF. Domain decompositions obtained from residue cross-correlations by both methods were consistent, indicating an anti-correlated motion between three major domains, namely BD, CD and HD. Overlap matrices between different runs indicate higher correlations for ANM compared to PCA modes, name first 10 modes.

All simulations were clustered based on overall and domain-wise alignments. Inter-domain and intra-domain conformational changes were also observed based on these

clusters. Dynamics of TF-ribosome complex were investigated after docking centroids from major clusters of apo TF runs on to a fragment of BD that was co-crystallized with ribosome 50S subunit. Performing energy minimization on a 50S-docked TF structure did not alter the results of ANM calculations. Four conformers were selected from each of the 150 mM and 150 mM _2 runs, and ANM was applied on these eight different 50S-docked TF complexes. Domain decomposition was performed based on collective dynamics of 50S-TF complexes. HD was found larger than for apo TF in six of the conformers. However, for two docked conformers TF showed a fully correlated dynamics acting as a single domain. This observation was linked to the relatively compact TF and its close association with 50S. Thus, the docked positions of conformers on 50S affect the dynamics of the complex to great extent.

# APPENDIX A: POSITIONS OF CONFORMERS

## A.1. Selected Conformations Aligned on Ribosome

Conformations for each independent simulations are selected based on all and BD aligned clustered centroids. Later, these conformations are aligned on to the ribosome and their positions are observed.

Table A.1. Selected centroids for each run based on all aligned clustered results.

| Simulation name | Selected Centroids | |
|---|---|---|
| 29 mM | Centroid 11 | Centroid 12 |
| 29_L mM | Centroid 5 | Centroid 6 |
| 50 mM | Centroid 17 | Centroid 18 |
| 150 mM | Centroid 14 | Centroid 16 |
| 150 mM_L | Centroid 8 | Centroid 22 |

Table A.2. Selected centroids for each run based on BD aligned clustered results.

| 29mM | Centroid 5 | Centroid 15 |
|---|---|---|
| 29mM_L | Centroid 1 | Centroid 2 |
| 150mM | Centroid 6 | Centroid 7 |
| 50mM | Centroid 8 | Centroid 9 |
| 150 mM_2 | Centroid 12 | |

Ribosome aligned conformations, results from all aligned and BD aligned clusters, are depicted in Figure A.1, and Figure A.2, respectively.
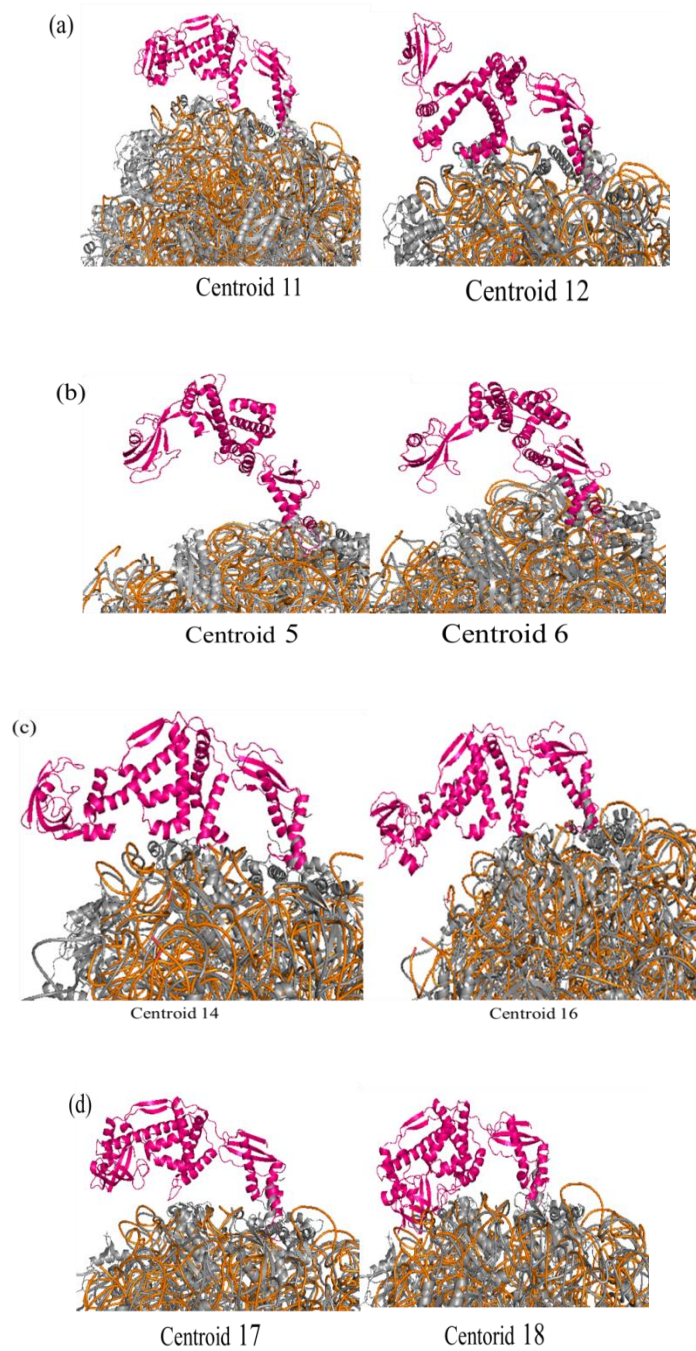


Figure A.1. Representative conformations, results from all aligned clusters, aligned on ribosome for (a) 29 mM (b) 29 mM_L (c) 150 mM (d) 50 mM (e) 150mM_2 runs.

Figure A.1. All alignment conformations aligned on ribosome for (a) 29 mM (b) 29 mM_L (c) 150 mM (d) 50 mM (e) 150mM_2 runs cont.



Figure A.2. BD aligned conformations aligned on ribosome for (a) 29 mM (b) 29 mM_L (c) 150 mM (d) 50 mM (e) 150mM_2 runs.

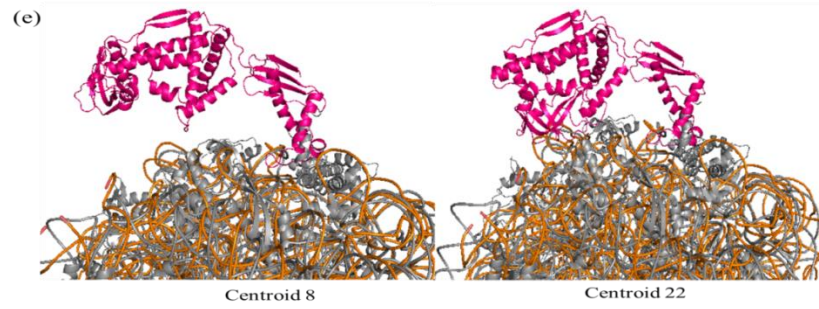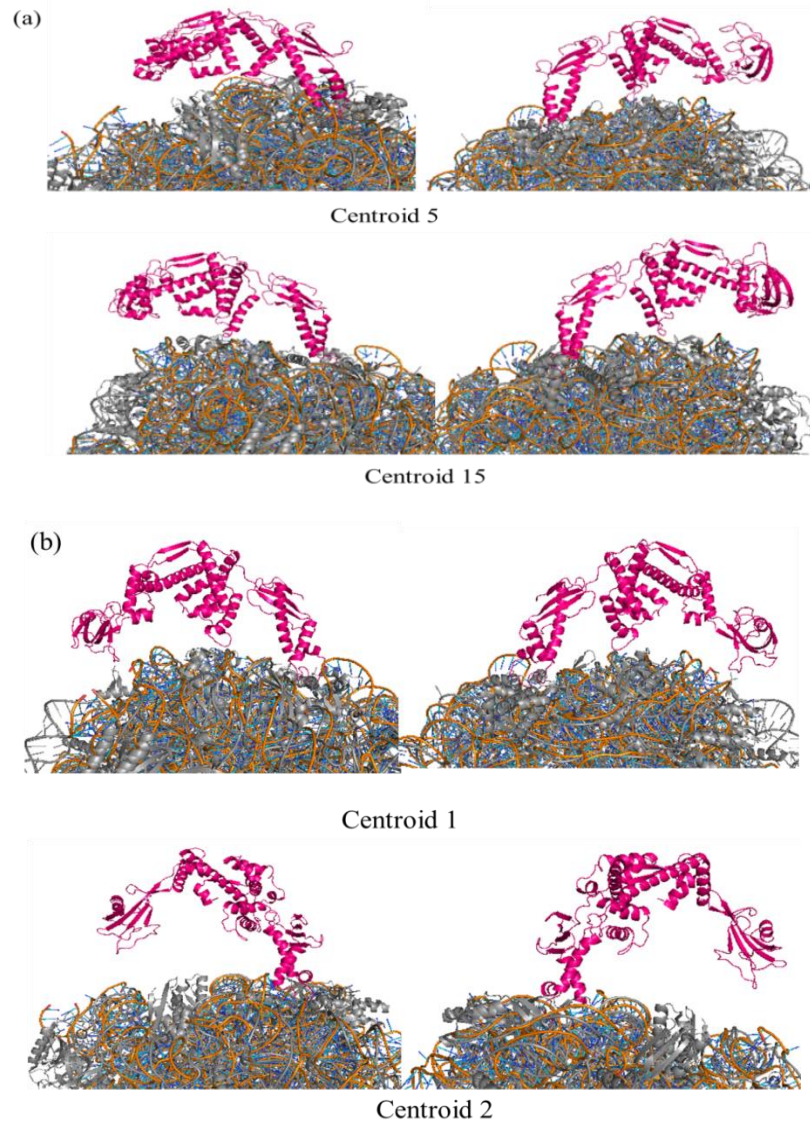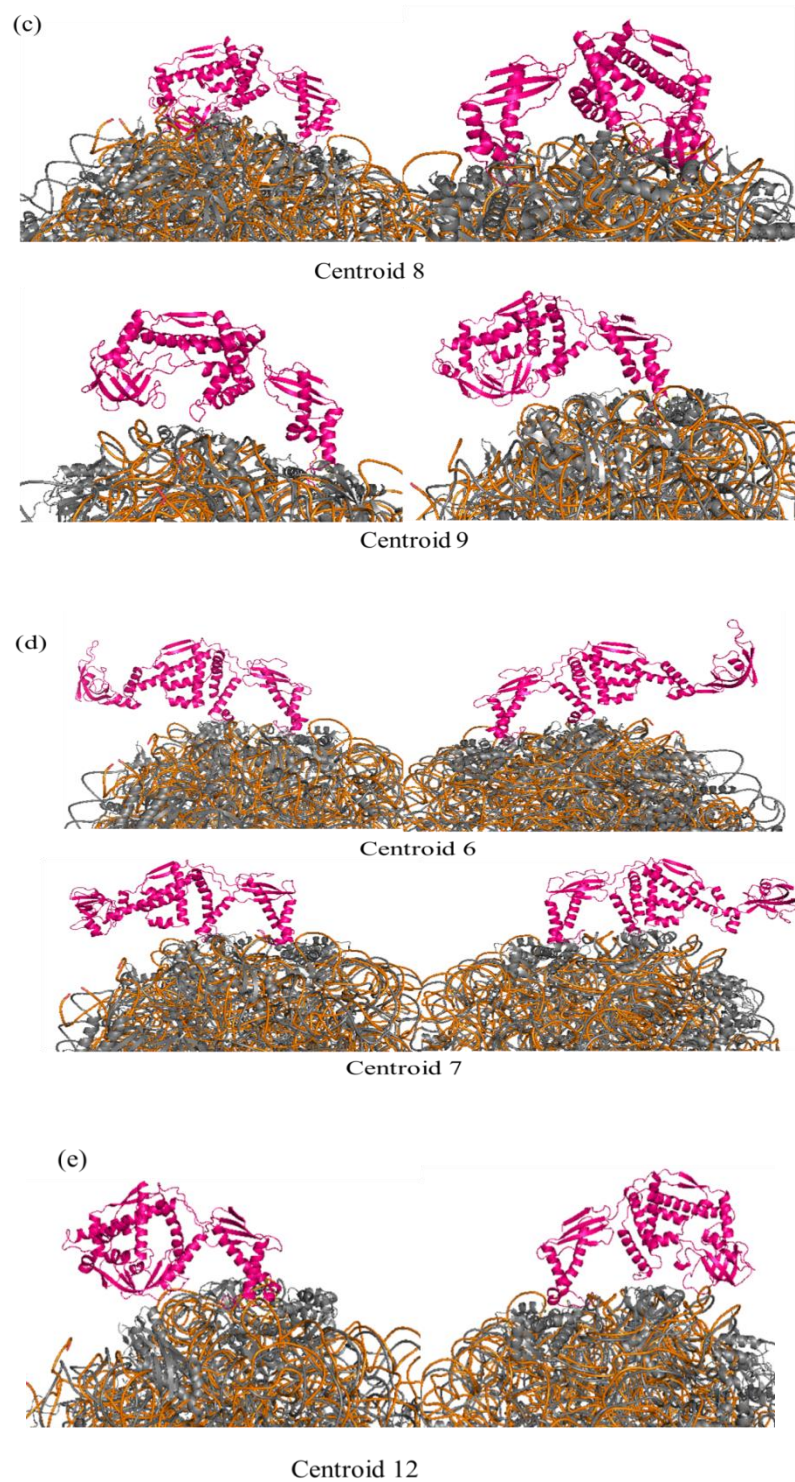Figure A.2. BD alignment conformations aligned on ribosome for (a) 29 mM (b) 29 mM_L (c) 150 mM (d) 50 mM (e) 150mM_2 runs cont.

# REFERENCES

Agashe, V. R., S. Guha, H. C. Chang, P. Genevaux, M. Hayer-Hartl, M. Stemp, C. Georgopoulos, F. U. Hartl and J. M. Barral, 2004, "Function of Trigger Factor and DnaK in Multidomain Protein Folding: Increase in Yield at the Expense of Folding Speed", *Cell*, Vol. 117, No. 2, pp. 199-209.

Andersen, H. C., 1983, "Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations", *Journal of Computational Physics*, Vol. 52, No. 1, pp. 24-34.

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, 2001, "Anisotropy of Fluctuation Dynamics of Proteins with An Elastic Network Model", *Biophysical Journal*, Vol. 80, No. 1, pp. 505-515.

Baram, D., E. Pyetan, A. Sittner, T. Auerbach-Nevo, A. Bashan and A. Yonath, 2005, "Structure of Trigger Factor Binding Domain in Biologically Homologous Complex with Eubacterial Ribosome Reveals Its Chaperone Action", *Proceedings of the National Academy of Sciences of the United States of America*, U S A, Vol. 102, No. 34, pp. 12017-12022.

Berendsen, H. J. C., D. Vanderspoel and R. Vandrunen, 1995, "Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation", *Computer Physics Communications*, Vol. 91, No. 1-3, pp. 43-56.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, 2000, "The Protein Data Bank", *Nucleic Acids Research*, Vol. 28, No. 1, pp. 235-242.

Blaha, G., D. N. Wilson, G. Stoller, G. Fischer, R. Willumeit and K. H. Nierhaus, 2003, "Localization of the Trigger Factor Binding Site on the Ribosomal 50 S Subunit", *Journal of Molecular Biology*, Vol. 326, No. 3, pp. 887-897.

Bui, E. T. N., P. J. Bradley and P. J. Johnson, 1996, "A Common Evolutionary Origin for Mitochondria and Hydrogenosomes", *Proceedings of the National Academy of Sciences of the United States of America,* U S A, Vol. 93, No. 18, pp. 9651-9656.

Bukau, B. and A. L. Horwich, 1998, "The Hsp70 and Hsp60 Chaperone Machines", *Cell*, Vol. 92, No. 3, pp. 351-366.

Chang, H. C., Y. C. Tang, M. Hayer-Hartl and F. U. Hartl, 2007, "SnapShot: Molecular Chaperones, Part I", *Cell*, Vol. 128, No. 1, pp. 212.

Cohen, F. E. and J. W. Kelly, 2003, "Therapeutic Approaches to Protein-Misfolding Diseases", *Nature*, Vol. 426, No. 6968, pp. 905-909.

Darden, T., D. York and L. Pedersen, 1993,"Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems", *Journal of Chemical Physics*, Vol. 98, No. 12, pp. 10089-10092.

Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, 2003, "A point-charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-phase Quantum Mechanical Calculations", *Journal of Computational Chemisty*, Vol. 24, No. 16, pp. 1999-2012.

Ellis, J., 1987, "Proteins as Molecular Chaperones", *Nature*, Vol. 328, No. 6129, pp. 378-379.

Ellis, R. J. and F. U. Hartl, 1999, "Principles of Protein Folding in the Cellular Environment", *Current Opinion in Structural Biololgy*, Vol. 9, No. 1, pp. 102-110.

Ellis, R. J., 1990, "Molecular Chaperones: the Plant Connection", *Science*, Vol. 250, No. 4983, pp. 954-959.

Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling and N. Ban, 2004,"Trigger Factor in Complex With the Ribosome Forms A Molecular Cradle for Nascent Proteins", *Nature*, Vol. 431, No. 7008, pp. 590-596.

Fersht, A. R., 2008, "From the First Protein Structures to Our Current Knowledge of Protein Folding: Delights and Scepticisms", *Nature Reviews Molecular Cell Biology*, Vol. 9, No. 8, pp. 650-654.

Gupta, R. S., 1990, "Sequence and Structural Homology between a Mouse T-Complex Protein Tcp-1 and the Chaperonin Family of Bacterial (Groel, 60-65-Kda Heat-Shock Antigen) and Eukaryotic Proteins", *Biochemistry International*, Vol. 20, No. 4, pp. 833-841.

Gutsche, I., L. O. Essen and F. Baumeister, 1999, "Group II Chaperonins: New TRiC(k)s and Turns of a Protein Folding Machine", *Journal of Molecular Biology*, Vol. 293, No. 2, pp. 295-312.

Harano, Y. and M. Kinoshita, 2005, "Translational-entropy Gain of Solvent Upon Protein Folding", *Biophysical Journal*, Vol. 89, No. 4, pp. 2701-2710.

Hartl, F. U., 1996, "Molecular Chaperones in Cellular Protein Folding", *Nature*, Vol. 381, No. 6583, pp. 571-580.

Hayer-Hartl, M. and A. P. Minton, 2006, "A Simple Semiempirical Model for the Effect of Molecular Confinement Upon the Rate of Protein Folding", *Biochemistry*, Vol. 45, No. 44, pp. 13356-13360.

Hendrick, J. P. and F. U. Hartl, 1993, "Molecular Chaperone Functions of Heat-Shock Proteins", *Annual Review of Biochemisty*, Vol. 62, No. 349-384.

Herczenik, E. and M. F. Gebbink, 2008, "Molecular and Cellular Aspects of Protein Misfolding and Disease", *Fedaration of American Societies for Experimental Biology Journal*, Vol. 22, No. 7, pp. 2115-2133.

Hess, B., C. Kutzner, D. van der Spoel and E. Lindahl, 2008, "GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation", *Journal of Chemical Theory and Computation*, Vol. 4, No. 3, pp. 435-447.

Hoffmann, A., B. Bukau and G. Kramer, 2010, "Structure and Function of the Molecular chaperone Trigger Factor", *Biochimica et Biophysica Acta* (*BBA*) *- Molecular Cell Research*, Vol. 1803, No. 6, pp. 650-661.

Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, 2006, "Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters", *Proteins-Structure Function and Bioinformatics*, Vol. 65, No. 3, pp. 712-725.

Horwich, A. L., W. A. Fenton, E. Chapman and G. W. Farr, 2007, "Two Families of Chaperonin: Physiology and Mechanism", *Annual Review of Cell and Development Biology*, Vol. 23, No. 115-145.

Humphrey, W., A. Dalke and K. Schulten, 1996, "VMD: Visual Molecular Dynamics", *Journal of Molecular Graphics & Modelling*, Vol. 14, No. 1, pp. 33-38.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, 1983, "Comparison of Simple Potential Functions for Simulating Liquid Water", *Journal of Chemical Physics*, Vol. 79, No. 2, pp. 926-935.

Kaminski, G. A., R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, 2001, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides", *Journal of Physical Chemistry B*, Vol. 105, No. 28, pp. 6474-6487.

Kauzmann, W., 1959, "Some Factors in the Interpretation of Protein Denaturation", *Advances in Protein Chemistry*, Vol. 14, No. 1-63.

Kubota, H., G. Hynes, A. Carne, A. Ashworth and K. Willison, 1994,"Identification of Six Tcp-1-related Genes Encoding Divergent Subunits of the TCP-1-containing Chaperonin", *Current Biology*, Vol. 4, No. 2, pp. 89-99.

Lill, R., E. Crooke, B. Guthrie and W. Wickner, 1988, "The "Trigger Factor Cycle" Includes Ribosomes, Presecretory Proteins, and the Plasma Membrane", *Cell*, Vol. 54, No. 7, pp. 1013-1018.

Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, 2010, "Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field", *Proteins-Structure Function and Bioinformatics*, Vol. 78, No. 8, pp. 1950-1958.

MacKerell, A.D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M.,1998, *The Encyclopedia of Computational Chemistry,* John Wiley & Sons. pp. 271–277.

Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, 2005, "Scalable Molecular Dynamics with NAMD", *Journal of Computational Chemistry*, Vol. 26, No. 16, pp. 1781-1802.

Quan, S. and J. C. Bardwell, 2012, "Chaperone Discovery", *Bioessays*, Vol. 34, No. 11, pp. 973-981.

Singhal, K., J. Vreede, A. Mashaghi, S. J. Tans and P. G. Bolhuis, 2013, "Hydrophobic Collapse of Trigger Factor Monomer in Solution", *PLOS ONE*, Vol. 8, No. 4, pp. e59683.

Tang, Y. C., H. C. Chang, M. Hayer-Hartl and F. U. Hartl, 2007,"SnapShot: Molecular Chaperones, Part II", *Cell,* Vol. 128, No. 2, pp. 412.

Trent, J. D., E. Nimmesgern, J. S. Wall, F. U. Hartl and A. L. Horwich, 1991, "A Molecular Chaperone from a Thermophilic Archaebacterium Is Related to the Eukaryotic Protein T-Complex Polypeptide-1" *Nature*, Vol. 354, No. 6353, pp. 490-493.

Valpuesta, J. M., J. Martin-Benito, P. Gomez-Puertas, J. L. Carrascosa and K. R. Willison, 2002, "Structure and Function of a Protein Folding Machine: the Eukaryotic Cytosolic chaperonin CCT", *The Federation of European Biochemical Societies Letters*, Vol. 529, No. 1, pp. 11-16.

Verlet, L., 1967,. "Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules", *Physical Review*, Vol. 159, No. 1, p. 98.