

MIXED-RESOLUTION ELASTIC NETWORK MODELS FOR BIOLOGICAL  
SUPRAMOLECULES

by

Ayşe Özge Kürkçüoğlu Levitas  
B.S., Ch.E, Boğaziçi University, 2001  
M.S., Ch.E, Boğaziçi University, 2003

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

Graduate Program in Chemical Engineering  
Boğaziçi University  
2008

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Pemra Doruker Turgut, for offering so much of her knowledge, encouragement and patience throughout my thesis. I have learned very much from her.

I am grateful to Prof. Canan Atılgan, Prof. Viktorya Aviyente, Assoc. Prof. Özlem Keskin and Assist. Prof. Elif Özkırımlı for their kind interest, for the time they devoted to reading and commenting on my thesis. Special thanks are due to Prof. Türkan Haliloğlu for her valuable suggestions, moral support and for her smiling face.

It was a great pleasure for me to collaborate with Prof. Robert L. Jernigan who guided me throughout of my research. I would like to thank Prof. Teoman Turgut for his kind interest, help and constructive suggestions in my study. My thanks are also due to Taner Z. Şen for his valuable comments on my study.

It was a great opportunity for me that this work has been supported by the Bogazici University B.A.P. (03A501-D, 03R104), DPT project (03K120250), TÜBİTAK project (104M247), the Turkish Academy of Sciences in the framework of the Young Scientist Award program (PD-TUBA-GEBIP/2002-1-9), COSBIOM (FP6-2004-ACC-5SA-2), and TÜBİTAK-BİDEB (2214).

I thank to all my former and present friends in Polymer Research Center for sharing life in good times and bad times.

I am grateful to my family who was always with me during my education, with love, patience and encouragement. No words would describe my gratitude to them... My sincere thanks are due to my sister Zeynep Kürkçüoğlu for her help in performing calculations and her support. Especially, I would like to give my thanks to Deyvi Levitas whose patient love and endless support enabled me to finish this thesis.

I dedicate this thesis to my family...

## ABSTRACT

### **MIXED-RESOLUTION ELASTIC NETWORK MODELS FOR BIOLOGICAL SUPRAMOLECULES**

Mixed coarse-graining approach has been recently introduced to the elastic network model to enable the modeling of a protein's native conformation with regions of high- and low-resolution. In this model, each node of the elastic network may represent a heavy atom (high-resolution), a residue or a group of residues (low-resolution), and close-neighboring nodes are connected by springs. To assign suitable cutoffs and force constants to the nodes in different resolution regions, two alternative procedures which either take the residue-based or atom-based elastic network model parameters as reference are developed. The calculation of parameters with the atom-based approach has proved to be superior due to its straightforward and realistic description of interactions and its applicability to both proteins and their complexes with RNA/DNA. The mixed coarse-graining method is validated by exploring the internal dynamics of the enzyme triosephosphate isomerase with 494 residues. The role of the dimeric enzyme's collective motions in controlling loop 6 closure and hence the catalytic activity is revealed for the first time. The supramolecular assemblage ribosome in its complex with mRNA, three tRNAs and elongation factor Tu (~11000 amino acids and nucleotides in total) is studied with residue-based and mixed coarse-grained models. As a result, large domain motions with functional importance are clearly observed, a translocation mechanism for mRNA and tRNAs is proposed, and the dynamics of the ribosomal tunnel are investigated. The decoding center with codons and anticodons in ribosome structure is modeled at atomistic level to investigate the local vibrational dynamics, difficult to attain with classical atomistic techniques. The results indicate that retaining the whole structure is critical to describe the collective dynamics of specific components in a large multi-subunit protein. The mixed resolution elastic network models have proven to be a powerful tool to study the dynamics of extremely large supramolecular assemblages at the atomistic scale with high computational efficiency.

## ÖZET

### BİYOLOJİK SÜPRAMOLEKÜLLER İÇİN KARIŞIK ÖLÇEKTE ELASTİK AĞ YAPI MODELLERİ

Elastik ağ yapı modellerinde yeni geliştirilmiş olan karışık kaba ölçek yaklaşımı, proteinlerin doğal konformasyonunun yüksek ve düşük çözünürlükte bölgeler kullanılarak modellenmesine olanak tanımaktadır. Bu modelde, elastik ağ yapının her bir düğümü, ağır bir atomu (yüksek-çözünürlük), bir rezidü veya rezidü grubunu (düşük-çözünürlük) temsil edebilir. Komşu düğümlerse birbirlerine yaylar ile bağlıdır. Farklı çözünürlükte olan bölgelerdeki düğümlere, uygun etkileşim uzaklıkları ve kuvvet sabitleri atanmalıdır. Bunun için rezidü veya atom bazlı elastik ağ yapı modellerinin değişkenlerini referans alan iki alternatif prosedür geliştirilmiştir. Atom bazlı yaklaşım ile yapılan değişken hesabı, etkileşimleri kolay ve gerçekçi tanımlaması, proteinlere ve bunların RNA/DNA komplekslerine uygulanabilmesi gibi özellikleriyle daha üstün olduğunu ortaya koymuştur. Karışık kaba ölçek metodunun geçerliliği 494 rezidü triozfosfat izomeraz enziminin iç dinamiğinin araştırılması ile onaylanmıştır. İki monomerden oluşan enzimin kolektif hareketlerinin, 'loop 6' bölgesinin kapanması ve enzimin katalitik aktivitesini kontrol etmesindeki rolü ilk kez ortaya çıkarılmıştır. Bir süpramolekül olan ribozomun mRNA, üç tane tRNA ve uzama faktörü Tu ile olan kompleksi (~11000 rezidü), rezidü bazlı ve karışık kaba ölçek modelleriyle çalışılmıştır. Sonuç olarak, işlevsel fonksiyonu olan geniş bölge hareketleri açıkça gözlemlenmiş, mRNA ve tRNA'lar için bir translokasyon mekanizması önerilmiş, ve ribozomal tüneldeki hareketler incelenmiştir. Ribozom yapısında genetik şifreyi çözen merkezle birlikte, kodon ve antikodonlar, klasik tekniklerle elde edilmesi güç olan atom seviyesindeki lokal titreşim dinamiğini incelemek amacıyla modellenmiştir. Sonuçlar, çok alt birime sahip büyük bir proteindeki belirli bileşenlerin titreşimsel dinamiğini tanımlarken, tüm yapının muhafaza edilmesinin önemli olduğunu işaret etmektedir. Karışık ölçekli elastik ağ yapı modelleri, büyük süpramoleküllerin hareketlerini, atom seviyesinde ve yüksek hesapsal verimlilikle incelemek için oldukça kuvvetli bir araçtır.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT.....	iv
ÖZET .....	v
LIST OF FIGURES .....	viii
LIST OF TABLES.....	xiii
LIST OF SYMBOLS/ABBREVIATIONS.....	xiv
1. INTRODUCTION .....	1
2. PROTEIN STRUCTURE, DYNAMICS AND FUNCTION .....	5
2.1. Protein Structure .....	5
2.1.1. Experimental Techniques to Identify Protein Structure .....	5
2.1.2. Protein Structure, Dynamics and Function.....	7
2.2. Supramolecular Assemblages .....	10
2.3. Proteins Studied .....	12
3. ELASTIC NETWORK MODEL.....	16
3.1. Anisotropic Network Model .....	16
3.2. Computational Methodology in ANM.....	17
3.3. Methodology in Mixed Coarse-Graining.....	19
3.3.1. Mixed Coarse-Graining Procedure I.....	21
3.3.2. Mixed Coarse-Grained Procedure II.....	23
4. DYNAMICS OF TRIOSEPHOSPHATE ISOMERASE .....	27
4.1. Determination of Parameters for Mixed Coarse-Graining Procedure I.....	28
4.2. Residue- and Atom-Based Models of ANM.....	30
4.3. Mixed Coarse-Grained Models of TIM .....	34
4.3.1. Mixed Coarse-Graining Procedure I.....	34
4.3.2. Mixed Coarse-Graining Procedure II .....	39
4.4. Collective Dynamics with Atomistic Detail .....	40
4.4.1. Dominant Motions as a Dimer.....	40
4.4.2. Classical Normal Mode Analysis .....	42
4.4.3. Active Site and Loop 6 Motions.....	43

4.4.4. Twisting Motion and Hinges in Slow Modes.....	47
4.4.5. Importance of Retaining Low-Resolution Regions.....	48
4.5. Effect of the Crystal Packing on Collective Dynamics of TIM.....	49
5. DYNAMICS OF BACTERIAL RIBOSOME.....	52
5.1. Structure and Function of Bacterial Ribosome.....	52
5.2. Residue-Based Models of Ribosome.....	56
5.3. Conformational Dynamics of Bacterial Ribosome.....	58
5.3.1. Ratchet-Like Rotation of Subunits.....	60
5.3.2. Motions of L1 and L7/L12 Stalks.....	61
5.3.3. Motions of Helix 44.....	65
5.3.4. mRNA and its Interactions with Ribosomal Proteins.....	66
5.3.5. Collective Motions of tRNAs.....	71
5.3.6. A Mechanistic Model Proposed for Translocation.....	73
5.3.7. Dynamics of Ribosomal Exit Tunnel.....	73
5.3.7.1. Motions from Normal Modes.....	76
5.3.7.2. Motions from Combined Normal Modes.....	79
5.4. Mixed Coarse-Grained Model of Ribosome.....	82
5.4.1. Dynamics of Decoding Center with Atomistic Model.....	83
6. CONCLUSIONS AND RECOMMENDATIONS.....	89
6.1. Conclusions.....	89
6.2. Recommendations.....	91
REFERENCES.....	93

## LIST OF FIGURES

Figure 2.1.	(a) Amino acids linked by peptide bonds (b) Primary, secondary, tertiary and quaternary structures in proteins .....	6
Figure 2.2.	(a) Normalized mean-square fluctuations of C <sup><math>\alpha</math></sup> atoms for EcoRI-DNA monomers averaged over twenty slowest and fastest modes, and (b), (c) corresponding displacement vectors for slow modes and residues for fast modes.....	9
Figure 2.3.	(a) Superimposed x-ray structures of ligand-free (gray, 8TIM.pdb) and ligand-bound (blue, 1TPH.pdb) triosephosphate isomerase (b) Active site and inhibitor PGH.....	13
Figure 2.4.	X-ray structure of bacterial ribosome complexed with mRNA and tRNAs...	14
Figure 3.1.	X-ray structures of triosephosphate isomerase with (a) residue-based model, and (b), (c) mixed coarse-graining (cg) at different levels .....	20
Figure 3.2.	Interaction parameters between nodes in mixed coarse-graining procedure I .....	22
Figure 3.3.	Interaction parameters between nodes in mixed coarse-graining procedure II .....	24
Figure 3.4.	Flow charts of the mixed coarse-graining procedures I and II.....	26
Figure 4.1.	Density of vibrational modes, $g(\omega)$ (cm), of (a) DHFR, and (b) TIM .....	29
Figure 4.2.	B-factors of one monomer of bound-TIM with residue-based cg from x-ray experiment and ANM calculations.....	31

Figure 4.3.	Alternative conformations at the first mode of (a) residue-based free TIM, (b) residue-based TIM complex and (c) atom-based TIM complex .....	32
Figure 4.4.	Alternative conformations at the second mode of (a) residue-based free TIM, (b) residue-based TIM complex and (c) atom-based TIM complex .....	33
Figure 4.5.	Overlap matrices shown for studied residue-based models .....	35
Figure 4.6.	Overlap matrices shown for studied mixed cg models.....	37
Figure 4.7.	First and second mode alternative conformations for (a) symmetric and (b) asymmetric mixed cg of TIM-PGH complex .....	38
Figure 4.8.	Slowest mode alternative mode shapes of TIM from mixed cg procedure II with low-resolution modeled at (a) $n$ and (b) $n/5$ level.....	40
Figure 4.9.	Twisting of monomers A and B, and loop 6 around rotation axes at first fourth modes. Arrows show displacement direction around axes ( $\odot$ , $\otimes$ ) following right hand rule .....	41
Figure 4.10.	Twisting motions observed from superimposed experimental x-ray structures (a) 1TRE and 1B9B (similar to second mode), (b) 1M6J and 1B9B (similar to first mode). Arrows indicate direction of deformation around axes ( $\odot$ , $\otimes$ ) following right hand rule .....	42
Figure 4.11.	Opening (yellow)/ closing (red) motion of loop 6 with active site residues, and PGH at (a) first, (b) second, (c) third and (d) fourth modes .....	45
Figure 4.12.	(a) Mean-square fluctuations at the slowest mode. Dashed lines show the threshold for hinge residues shown by spheres in (b) different perspectives .....	47

Figure 4.13. TIM shown in (a) black and (b) dashed in the generated crystal packing.....	50
Figure 5.1. Bacterial ribosome 70S and functionally important regions .....	53
Figure 5.2. Schematic representation of experimentally observed large motions of subunits 30S (blue) and 50S (gray) .....	54
Figure 5.3. Ribosomal model (a) 70S, (b) 70S+ <i>EF-Tu</i> , and (c) 70S_ no_ proteins.....	57
Figure 5.4. Comparable slow motions of three ribosome models, 70S, 70S+ <i>EF-Tu</i> and 70S_ no_ proteins .....	59
Figure 5.5. (a)-(d) Correlation maps for L1, L7/L12 stalks, and mRNA. (e)-(f) Domains on L1 and L7/L12 stalks .....	63
Figure 5.6. (a) Correlation map, and (b) domains for L1 stalk and E-tRNA.....	64
Figure 5.7. Mean-square fluctuations averaged over ten slowest modes for (a) h44, (b) H69.....	65
Figure 5.8. mRNA interacting with 16S 3' end together forming the Shine-Dalgarno (S-D) duplex and ribosomal proteins S3, S4, S5 on 30S subunit .....	66
Figure 5.9. Mean-square distance fluctuations averaged over ten slowest modes for the nucleotide A27 on mRNA and residues of S3, S4, S5 proteins .....	67
Figure 5.10. Mean-square distance fluctuations averaged over ten slowest modes for mRNA nucleotide A27 and residues of S3, S4 and S5 from three different model systems.....	68

Figure 5.11. Mean-square fluctuations averaged over ten slowest modes for (a) mRNA, and (b) 16S 3' end. Inserts display experimental B-factors for these regions .....	70
Figure 5.12. Superimposed crystal structures of mRNA with helicase active site residues of ribosomal proteins S3 and S4.....	71
Figure 5.13. Displacement vectors of tRNAs at first three modes displayed for model (a)-(c) 70S and (d)-(f) 70S+ <i>EF-Tu</i> .....	72
Figure 5.14. Ribosome complex and nascent polypeptide exit tunnel (in green).....	74
Figure 5.15. (a) Ribosomal tunnel wall with 23 rRNA and few ribosomal proteins. (b) Domains of tunnel shown on the cross-correlation map averaged over first five modes .....	75
Figure 5.16. Mean-square fluctuations averaged over first twenty modes of exit tunnel, color coded with blue to red indicating increase in mobility .....	76
Figure 5.17. Displacement vectors of tunnel wall residues for first five modes with front and top views. Alternative conformations for L22 and L4 are shown in orange and purple.....	77
Figure 5.18. Distribution of positively correlated residue pairs in (a) entrance and (b) exit domains of ribosomal tunnel from normal modes.....	78
Figure 5.19. Displacement vectors of tunnel wall residues calculated from different combinations of first five normal modes.....	80
Figure 5.20. Distribution of positively correlated residue pairs in (a) entrance and (b) exit domains of ribosomal tunnel from combined modes .....	81

- Figure 5.21. Cumulative first ten mode mean-square fluctuations are calculated for different levels of mixed cg for ribosome complex..... 84
- Figure 5.22. Displacement vectors of alternative directions (+,-) from the third mode shown in ribbon presentation of ribosome complex. A simplified scheme of motions is given on the right..... 85
- Figure 5.23. (a) Mobility of the high-resolution region with red color indicating the highest value. (b) Cross-correlation map of the high-resolution region over first ten modes ..... 86
- Figure 5.24. Cross-correlation map calculated from cumulative ten modes for the decoding center A1492, A1493 and anticodons of A- and P-tRNAs..... 87

**LIST OF TABLES**

Table 2.1.	TIM and ribosome structure details, and applied models.....	15
Table 4.1.	ANM interaction parameters for coarse-grained systems .....	30
Table 4.2.	Overlap values for residue- and atom-based model averaged over first ten modes.....	34
Table 4.3.	Overlap values for mixed cg models averaged over first ten modes.....	36
Table 4.4.	Overlap values calculated for various models of TIM .....	39
Table 4.5.	Overlap values observed between displacement vectors of residues (169-173) from different models and x-ray open-closed structures.....	44
Table 4.6.	Correlation coefficients with the atom-based elastic network model taken as the reference .....	48
Table 5.1.	Various ribosome models used in the coarse-grained model .....	58
Table 5.2.	Overlap values averaged over ten slowest modes in comparison with the reference structure 70S .....	60
Table 5.3.	Sizes of mixed coarse-grained ribosome models .....	82
Table 5.4.	Overlap values calculated for various models of 70S .....	83

## LIST OF SYMBOLS/ ABBREVIATIONS

$B_{cm,n}$	Average temperature factor over n residues
$B_i$	Debye-Waller or temperature factor of site i
$C_{ij}$	Cross-correlation coefficient between sites i and j
$g(\omega)$	Density of vibrational modes
<b>H</b>	Hessian matrix
$h(x)$	Heavyside step function
<b>I</b>	Identity matrix
$k_B$	Boltzmann's constant
L	Lagrange equation
$m_i$	Mass of site i
n	Number of residues
N	Total number of residues
Q	Coordinate
<b>Q</b>	Coordinate vector
$q_i$	Mass weighted coordinate of site i
<b>q<sub>i</sub></b>	Mass weighted coordinate vector of site i
<b>q̇</b>	First derivative of mass weighted coordinate vector
$\langle q_i^2 \rangle$	Mean-square of mass weighted coordinate of site i
$r_c$	Cutoff radius
<b>R<sub>i</sub></b>	Position vector of site i
$R_{ij}$	Distance between site i and j
$\Delta \mathbf{R}_{cm}$	Fluctuation vector of the center of mass
$\Delta \mathbf{R}_i$	Fluctuation of vector of site i
$\overline{\Delta \mathbf{R}_{cm}}$	Mean fluctuation vector of the center of mass
$\langle \Delta \mathbf{R}_i^2 \rangle$	Mean-square fluctuation of site I
$\langle \Delta \mathbf{R}_i^2 \rangle_k$	Mean-square fluctuation of site i at mode k
$\langle \Delta \mathbf{R}_{ij}^2 \rangle$	Mean-square fluctuation of distance between sites i and j
<b>S</b>	Matrix of associated eigenvectors

T	Absolute temperature
t	Time
t	Transpose
V	Potential energy
<b>v</b>	Eigenvector set
<b>w</b>	Eigenvector set
$\Delta\mathbf{x}_i$	Change in the x coordinate of the position vector of site i
$\Delta\mathbf{y}_i$	Change in the y coordinate of the position vector of site i
$\Delta\mathbf{z}_i$	Change in the z coordinate of the position vector of site i
$\alpha$	Alpha helix
$\beta$	Beta strand
$\varepsilon_i$	Frequency phase of site i
$\gamma$	Force constant
$\Lambda$	Eigenvalue matrix
$\lambda_k$	Eigenvalue of mode k
$\omega_i$	Angular frequency of site i
$\omega_k$	Frequency at mode k
A	Adenine
A	Aminoacyl
aa	Amino acid
ANM	Anisotropic Network Model
Arg	Arginine
ASL	Anticodon stem loop
cg	Coarse-graining, coarse-grained
Cryo-EM	Cryo-electron microscopy
DHAP	Dihydroxyacetone phosphate
DHFR	Dehydrofolate reductase
DNA	Deoxyribonucleic acid
E	Exit

EF	Elongation factor
G	Guanine
GAP	D-glyceraldehyde 3-phosphate
GDP	Guanosine diphosphate
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GNM	Gaussian Network Model
GTP	Guanosine triphosphate
h	Helix
His	Histidine
hr	High-resolution
lr	Low-resolution
Lys	Lysine
MD	Molecular Dynamics
mRNA	Messenger RNA
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
nuc	Nucleotide
P	Peptidyl
PDB	Protein Data Bank
PGH	Phosphoglycolohydroxamate
PTC	Peptidyl transferase center
R	Arginine
rmsd	Root-mean-square deviation
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
S-D	Shine-Dalgarno
Thr	Threonine
TIM	Triosephosphate isomerase
tRNA	Transfer RNA
Trp	Tryptophan

U

Uracil

## 1. INTRODUCTION

Large scale structural motions of proteins and nucleic acids are crucial for their cellular functions, such as enzyme catalysis, protein synthesis or chaperone assisted folding processes (Petsko and Ringe, 2004). Besides a multitude of experimental methods for identifying the structure-function relationship of macromolecules, various computational approaches are being developed to complement experimental observations *in silico*.

Three-dimensional structures and function related dynamics of large macromolecules can be determined with the use of various experimental techniques such as x-ray crystallography, cryo-electron microscopy (EM), single particle tracking, and fluorescence resonance energy transfer techniques. Recently, high-and low-resolution techniques have led to the determination of the ribosome structure and dynamics (Gabashvili *et al.*, 2000; Zhao *et al.*, 2004; Frank and Agrawal, 2000; Valle *et al.*, 2003; Jenner *et al.*, 2005; Selmer *et al.*, 2006), the moving mechanism of motor proteins (Yildiz *et al.*, 2003, 2004), sliding clamp motion of DNA polymerase on the DNA during replication (Kuriyan and O'Donnell, 1993), conformational flexibility of RNA polymerase (Darst *et al.*, 2002) and chaperone GroEL (Saibil, 2000). Although these methods provide detailed description of functional motions, they are tedious and time-consuming, for this reason efficient molecular modeling approaches need to be developed for elucidating the internal conformational dynamics of large macromolecules.

Molecular dynamics (MD) simulations and normal mode analyses (NMA) with all-atom empirical potentials developed for proteins (McCammon and Harvey, 1987; Kitao and Go, 1999) are widely used to explore molecular motions. However, limits in the computational power and in the information provided by experimental results obstruct observing functional motions of large macromolecules with these methods. In recent studies, coarse-grained protein models and simplified force fields have been very successful to describe the molecular motions of proteins and their complexes (Tirion, 1996; Bahar *et al.*, 1997; Hinsen *et al.*, 1999; Atilgan *et al.*, 2001; Tama *et al.*, 2003; Trylska *et al.*, 2005).

An efficient approach commonly used to describe the fluctuation dynamics of large proteins is effectively provided by a coarse-grained normal mode analysis, known as the elastic networks model. In the original model referred as the Gaussian network model (GNM) (Bahar *et al.*, 1997), the elastic network is constructed with  $\alpha$ -carbon sites of the amino acid residues (i.e. nodes) and they interact with other nodes that fall within a cutoff distance. Close residue pairs are linked by uniform springs with force constant  $\gamma$ . In this model, although the magnitudes of fluctuations are assumed to be isotropic, i.e. with no directional preferences, they are in accordance with X-ray crystallographic Debye-Waller factors (Bahar *et al.*, 1997; Haliloğlu *et al.*, 1997; Bahar and Jernigan, 1998; Jernigan *et al.*, 1999; Keskin *et al.*, 2000), the H/D exchange free energies (Bahar *et al.*, 1998) and the order parameters from NMR-relaxation measurements (Haliloglu and Bahar, 1999).

In general, the residue fluctuations are anisotropic (Kuriyan *et al.*, 1986), which is important in functional motions of proteins. An extension of GNM, namely the anisotropic network model (ANM) (Atilgan *et al.*, 2001), considers both the direction and magnitude of residue fluctuations. The large scale collective motions obtained from GNM, ANM, and time series analysis combined with ANM have shown to be closely related to those extracted from atomistic molecular dynamics studies for  $\alpha$ -amylase inhibitor tendamistat (Doruker *et al.*, 2000; Alakent *et al.*, 2007), restriction endonuclease *EcoRI*-DNA complex (Doruker *et al.*, 2006), ribosomal RNA 16S (Kim *et al.* 2003) and triosephosphate isomerase (Cansu and Doruker, *in press*). Moreover, GNM with Markov process has been recently developed to predict the propagation of information along low-frequency motions of proteins such as chaperone GroEL-GroES complex, and several important enzymes (Chennubhotla and Bahar, 2007a; 2007b). These coarse-grained elastic network models using various approaches have been shown to be effective and computationally efficient in the analysis of the structure-function relationship of proteins and their complexes. Recently, several large proteins and supramolecular assemblages have been modeled by ANM, such as GroEL-GroES complex (Keskin *et al.*, 2002a), tubulin (Keskin *et al.*, 2002b), ribosome complex (Wang *et al.*, 2004) and RNA polymerase (Yildirim and Doruker, 2004).

In the coarse-grained elastic networks, protein structures are treated as uniform materials undergoing mechanical motions limited by their shapes. And so far, single-node-

per-residue representations have been successful in describing the dynamics of protein complexes. In recent studies, it has been shown that as long as the overall shape of the macromolecule is kept intact (i.e. without removing or constraining any parts in the structure), the observed harmonic motion is very similar for all uniform or mixed hierarchical levels of coarse-graining (Doruker *et al.*, 2002b; Ming *et al.*, 2002; Kurkcuoglu *et al.*, 2004). Therefore, extreme levels of coarse-graining seriously reduce the computational time and make the analysis of supramolecular assemblages possible.

The mixed coarse-grained approach, recently introduced to elastic networks model, considers the native structure of a protein in high- and low-resolution regions (Kurkcuoglu 2003; Kurkcuoglu *et al.*, 2004). In this technique developed for extremely large systems, the ‘interesting’ parts of the structure are modeled at one-node-per-residue level with the rest of the structure being represented at one-node-per-multiple residues level. This model, different from previous ANM studies, considers the total mass of each residue to take into account the distribution of mass throughout the network. As a function of node size, the cutoff distance and spring constant of interactions are adjusted to maintain the flexibility of the mixed-resolution structure. With such a mixed-resolution model, the number of interaction sites is kept at a reasonable level, and the normal mode analysis can be performed with high computational efficiency for large system sizes such as the enzyme  $\beta$ -galactocidase (Kurkcuoglu *et al.*, 2004).

In this thesis, the mixed coarse-graining procedure is further developed to represent the ‘interesting’ parts of the protein, such as the active site of an enzyme, at atomistic level while keeping the rest of the system at a lower resolution. Two different approaches to incorporate atomic details to the elastic network are discussed. First approach, similar to the original model (Kurkcuoglu *et al.*, 2004) is composed of three main parts: (i) Uniform coarse-graining (i.e. one-node-per-residue) and atomic coarse-graining (i.e. full atom elastic network like in Tirion’s approach (Tirion, 1996)) of the native X-ray structures of a set of reasonably sized proteins with  $N$  residues; (ii) Establishing the relation between the force constant and the cutoff radius for different resolution regions; and (iii) Modeling the ‘interesting’ parts at atomistic level and the rest of the structure at a coarse-grained level.

In the second mixed coarse-graining approach, appropriate force constants are assigned to different sized node pairs based on atom-atom contacts of amino acids and/or nucleotides using a fixed cutoff value for atomistic interactions. In other words, the force constant is proportional to the number of atom-atom interactions between node pairs. With this approach, force constants needed for different resolution regions are not calculated explicitly, and modeling of supramolecular assemblages, hard to work with full-atom elastic network, is computationally feasible. In this study, a relatively large enzyme - triosephosphate isomerase (TIM)- and a supramolecular assemblage –the bacterial ribosome complex- are chosen due to wide experimental observations in literature in order to verify the validity of the method.

The plan of the present thesis is as follows: a brief review about structural characteristics of proteins, protein dynamics and supramolecular assemblages will be presented in the following section together with structural information about TIM and bacterial ribosome. The elastic network model and the mixed coarse-graining method that are used to elucidate the details of large systems will be presented in the third chapter. The application of the new methodology to TIM and ribosome will be discussed in detail in the fourth and fifth chapters, respectively. Finally, conclusions and recommendations about future work will be given.

## 2. PROTEIN STRUCTURE, DYNAMICS AND FUNCTION

### 2.1. Protein Structure

Proteins are complex organic macromolecules composed of twenty different amino acids joined by peptide bonds. The backbone is the same for all amino acids that build proteins: an amino group ( $\text{NH}_2$ ) and a carboxyl group ( $\text{COOH}$ ) linked by a central carbon atom ( $\text{C}^\alpha$ ). Amino acids are distinguished with their side chain (R) groups with different tendencies to interact with each other and water attributing hydrophilic, hydrophobic and amphipathic properties (Figure 2.1 (a)).

The sequence of a protein, directly decoded from genes, is its primary structure. Successive amino acids are joined together via a rigid 'peptide' bond while  $\text{N}-\text{C}^\alpha$  bond and  $\text{C}^\alpha-\text{C}$  bonds are free to rotate. This flexibility along the backbone together with its sequence allows polypeptide chains proteins to fold into a three-dimensional native form at physiological temperatures. The secondary structures of polypeptide chains can be either an alpha-helix or a beta-sheet, formed through regular hydrogen-bonding interactions between  $\text{N}-\text{H}$  and  $\text{C}=\text{O}$  groups on the backbone. Secondary structure elements together with coils or loops rearrange into a higher level called tertiary structure, which is the globular form of the protein. Many proteins are composed of more than one polypeptide chain, which constitute their quaternary structure (Figure 2.1 (b)) (Petsko and Ringe, 2004).

#### 2.1.1. Experimental Techniques to Identify Protein Structure

Recent developments in various experimental techniques reveal different and complementary information about biomolecular three-dimensional conformations. Cryo-electron microscopy (EM), electron tomography, fluorescence resonance energy transfer, single particle tracking methods give low-resolution conformations and structural transitions with no atomic details, but they can expose molecular mechanisms concerning biological functions of protein complexes in the cell (Gabashvili *et al.*, 2001; Zhao *et al.*,

2004). Usually atomic structures of macromolecules are obtained with x-ray crystallography and nuclear magnetic resonance (NMR) methods.

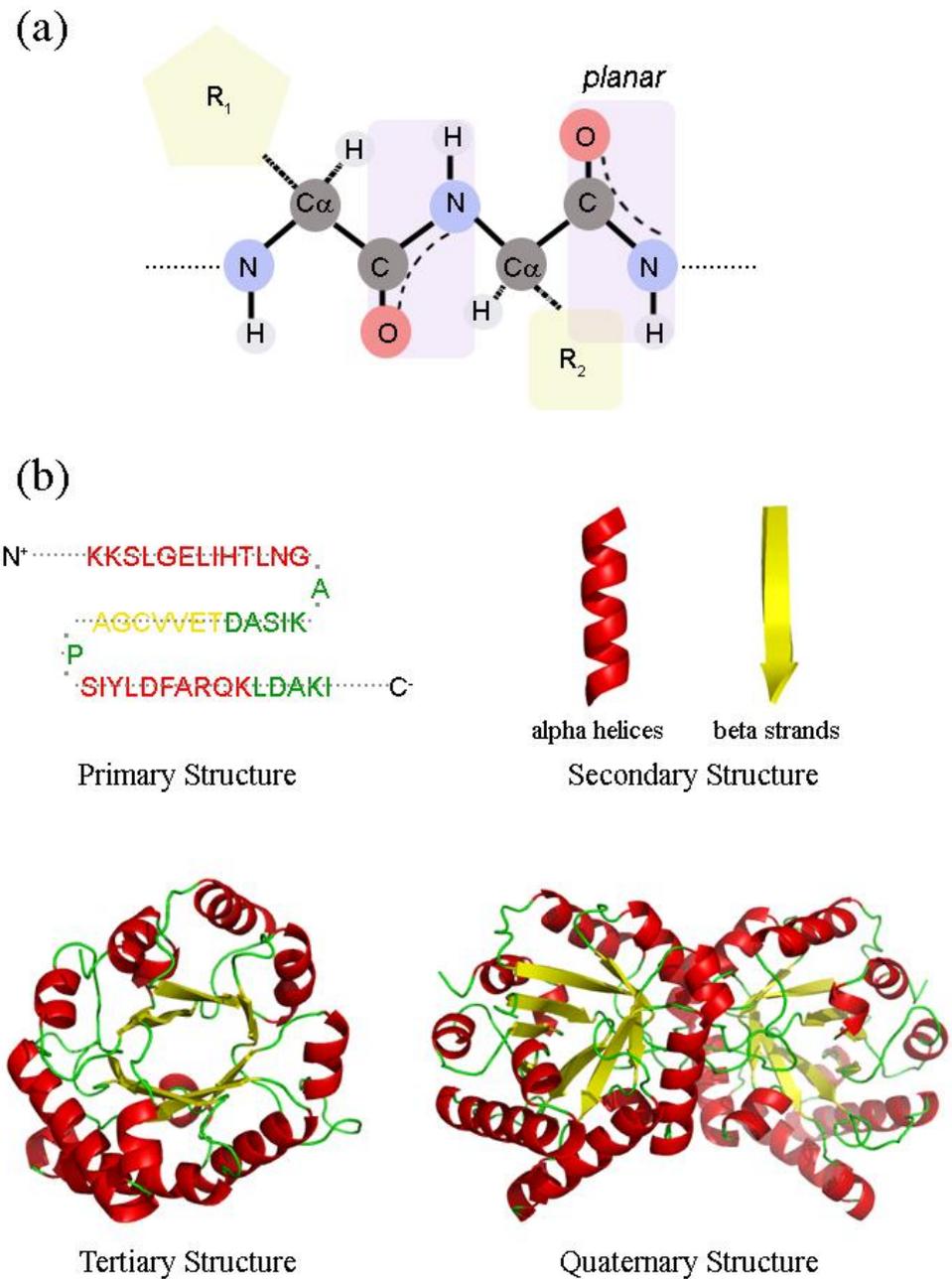


Figure 2.1. (a) Amino acids linked by peptide bonds. (b) Primary, secondary, tertiary and quaternary structures in proteins

X-ray crystallography provides the positions of non-hydrogen atoms by the use of x-rays scattering from the electron cloud of atoms. The precision of the technique depends on several factors such as experimental conditions, and resolution of the method. Crystallographic images are snapshots of frozen protein conformations, and in some cases, part of a structure may be invisible to the method due to a disorder of particular regions. This technique is suitable for all size of biomolecules that can be easily crystallized in solution.

Certain atomic nuclei, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{31}\text{P}$  have a magnetic moment or spin, and this property can be probed by nuclear magnetic resonance (NMR). This technique gives information about the relative atom distances in a molecule, and its three-dimensional model. To determine the side chain conformations of a protein, along with  $^1\text{H}$  (proton) NMR,  $^{13}\text{C}$  and  $^{15}\text{N}$  can be introduced to the macromolecule by growing it in a media enriched with these isotopes. NMR can also be used to study the flexibility and dynamics of proteins as opposed to x-ray crystallography (Branden and Tooze, 1999; Petsko and Ringe, 2004).

### **2.1.2. Protein Structure, Dynamics and Function**

In all living organisms, proteins execute and control biochemical functions defined by their structures and dynamics. As more structures are determined by experimental techniques such as x-ray crystallography, NMR, cryo-EM and other methods (see Protein Data Bank (PDB), Berman *et al.*, 2000), there is a pressing need for understanding their mechanisms. A protein or its complex can undergo different conformations in space, represented by valleys of its energy hypersurface (Frauenfelder *et al.*, 1991; 1997). And a protein experiences motions by jumping from one valley to another which defines its activity. Therefore, together with chemical diversity, collective motions of proteins are essential for functional properties, such as catalysis, binding, and operating as molecular switches (Petsko and Ringe, 2004).

There are different levels of motions in proteins. Atoms vibrate at high frequencies (i.e. fast motions,  $10^{-15}$  s) around their covalent bonds, whereas the motions of side chains are much slower ( $10^{-12}$  s). On the other hand, large domain motions changing the overall

shape of the structure occur at even lower frequencies (i.e. slow motions at  $10^{-6}$  to  $10^{-3}$  s level) and usually control the biological activity of the protein. In an earlier work (Bahar *et al.*, 1998b), it has been indicated that the residues involved with fast modes are resistant to conformational changes implying their important role in maintaining the stability of the folded state. On the other hand, residues involved in slow modes are susceptible to global motions underlying protein's biological function.

Comparison of several crystal structures of proteins expose mainly two types of collective motion near the folded state, namely hinge and shear motion that are relevant in biological activity and function (Gerstein *et al.*, 1994; Flores *et al.*, 2006). Proteins usually undergo hinge motion around one or multiple centers whose local rotational mobility determines the motion of domains as rigid bodies around these centers (Jernigan *et al.*, 1999). Hinge regions are recognized by having small displacements in the global motion of the structure at low frequencies. On the other hand, shear motions often involve sliding of domains layered one over another.

In a recent study, the collective dynamics of restriction endonuclease *EcoRI* with its DNA complex have been studied with both molecular dynamics (MD) simulations and anisotropic network model (ANM) (Doruker *et al.*, 2006). This enzyme cleaves foreign DNA from specific sites to protect prokaryotic cell. Consistent with MD results, ANM has pointed to several domains undergoing both hinge and shear motions upon DNA binding. Findings have designated hinges for the inner loops carrying DNA cleavage site residues, situated towards the middle of the enzyme. In Figure 2.2 (a), mean-square fluctuations averaged over first twenty modes are displayed with dark solid lines. Minima in this curve correspond to hinge regions and maxima describe regions of high flexibility. In Figure 2.2 (b), ribbon presentation of the enzyme-DNA complex displays displacement vectors in one alternative direction of harmonic oscillation around a hinge passing above DNA binding pocket. Shear motion is specifically observed at inner and outer loops wrapping around the DNA. Hinge and shear motions together may have significance in the sliding mechanism on DNA to find the cognate cleavage site. Moreover, residues involved with fast modes maintaining the stability of the structure have been determined from sharp peaks in Figure 2.2 (a) represented by the thin curve. These residues seem to form an inner core at the base

of the mobile loops and DNA binding region to maintain the stability of the structure (Figure 2.2 (c)).

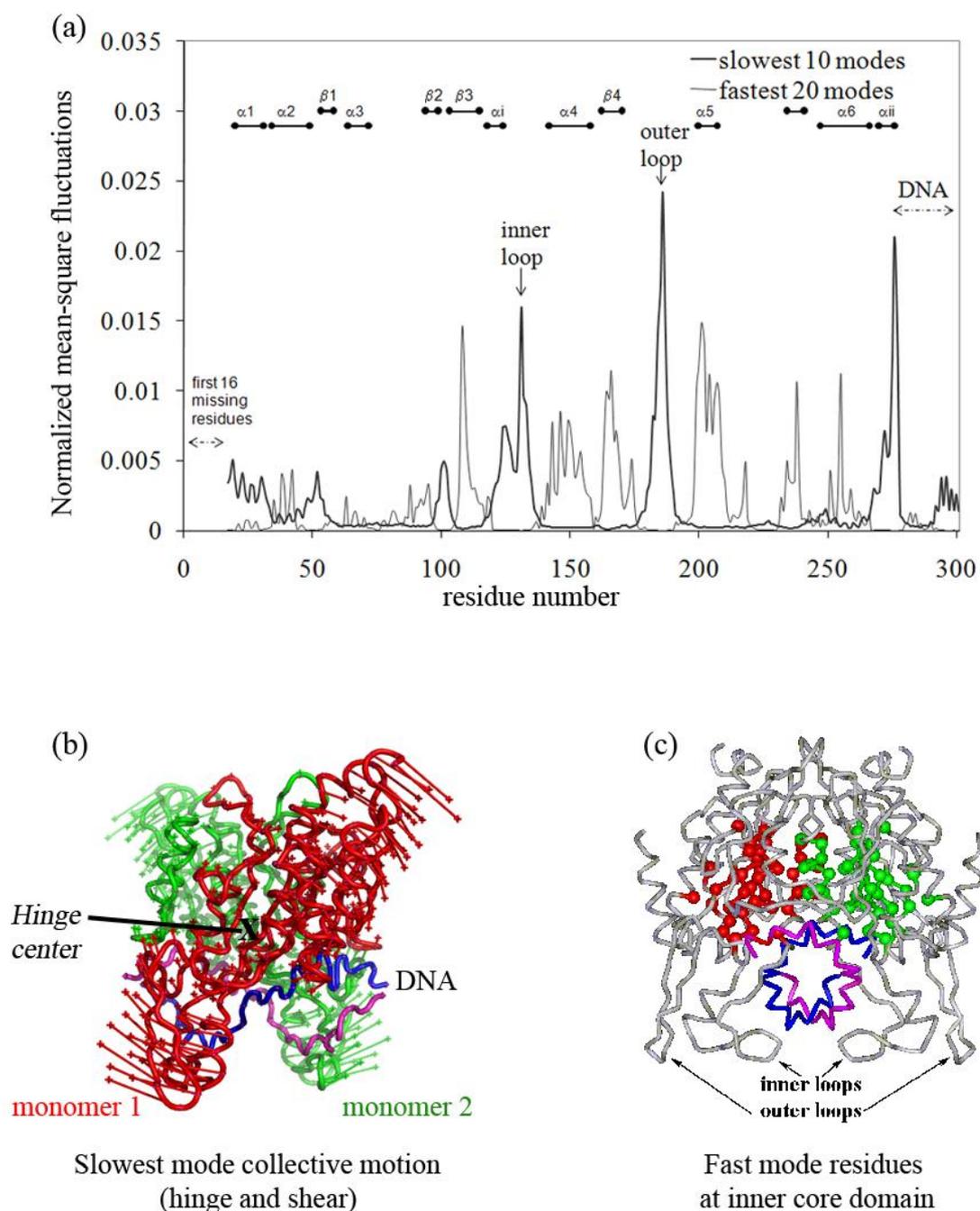


Figure 2.2. (a) Normalized mean-square fluctuations of  $C^\alpha$  atoms for *EcoRI*-DNA monomers averaged over twenty slowest and fastest modes, and (b), (c) corresponding displacement vectors for slow modes and residues for fast modes (Doruker *et al.*, 2006)

## 2.2. Supramolecular Assemblages

Few decades ago, only high-resolution structures of small proteins could be determined by x-ray crystallography or NMR. For large macromolecules electron microscopy was the reasonable method to reveal their structure and mechanisms (Cheng and Hammar, 2004). However, recent developments in technology have rapidly covered the necessity for high-resolution images of supramolecular assemblages (i.e. association of multiple biological macromolecular components) such as structures of ribosome (Ban *et al.*, 2000; Jenner *et al.*, 2005; Yusupova *et al.*, 2006), RNA polymerase (Minakhin *et al.*, 2001), large assemblies in membranes (Locher *et al.*, 1998), protein folding chaperone GroEL (Wang and Biosvert, 2003), and viruses (Fry *et al.*, 2005). Furthermore, combination of high- and low-resolution structures from various experimental techniques using computational methods such as MD and normal mode analysis (NMA) is another approach successfully applied to obtain atomistic conformations of large macromolecules (Wriggers and Chacon, 2001; Wriggers and Birmanns, 2001; Tama *et al.*, 2004).

Although high-resolution images provide necessary information about the atom rearrangements in various conformations, computational techniques are needed to determine collective motions of proteins. However, computational prediction of large-scale motions of supramolecular assemblages at atomistic detail is a daunting task (Elcock, 2002). Traditional MD or Monte Carlo simulations can describe structural fluctuations and transitions of biomolecules at the atomistic level even for ribosome complex over  $10^6$  atoms at physiological time of 22 ns (Sanbonmatsu and Tung, 2006), but such calculations are very time-consuming or even inaccessible to monitor large-scale motions at ns to  $\mu$ s level.

An interesting alternative to MD for simulating large systems is normal mode analysis (NMA). NMA can be applied to an atomic structure by describing atomic interactions by a standard molecular dynamics force field (Karplus and McCammon, 2002) or a simple Hookean potential (Tirion, 1996). As this method is based on an harmonic approximation of potential energy, collective motions are described by vibrational modes occurring around a given minimum energy. Low frequency collective motions from NMA are determined to correlate closely with experimental data related to biological function

(Tama and Sanejouand, 2001). However, in classical NMA routine, an energy minimization is required to reach the true minimum of the potential energy surface, on which the quality of modes depends. This is followed by the diagonalization of  $(3N \times 3N)$  Hessian matrix of second derivatives of potential energy, where  $N$  is the number of atoms in the structure. But, this procedure is computationally expensive for large macromolecules (Tama and Brooks, 2006). Recently, in order to reduce the size of system, coarse-grained semi-empirical potentials are introduced to describe molecular forces between residues and to simulate large systems with MD (Trylska *et al.*, 2004), and NMA (Tama *et al.*, 2003).

Suezaki and Go (1975) explored the relationship between continuous elastic bodies and low-frequency motions of biomolecules for which Young's modulus and vibrational frequencies can be derived (Suezaki and Go, 1976). This study has later inspired representing atomic structures as elastic networks, where interacting atoms are connected by uniform spring constants (Tirion, 1996). This simple Hookean potential in NMA has successfully estimated the collective fluctuations of macromolecules, noting this approach as the landmark point for extending efforts to find simpler models. Assuming that shape governs the dynamics in normal modes, macromolecular structures are further modeled in simplified coarse-grained representations (Bahar *et al.*, 1997; Hinsen, 1998). The most effective of these coarse-graining approaches are the Gaussian network model (GNM) (Bahar *et al.*, 1997) and the anisotropic network model (ANM) (Atilgan *et al.*, 2001).

One of the major drawbacks of NMA is applying an initial energy minimization to the system by using semi-empirical potentials which can result in conformation significantly different from its original structure. However, in coarse-grained elastic network models, this problem is avoided by assuming that original structure is at the minimum energy. According to GNM and ANM, a protein can be represented as an elastic network constructed by single nodes placed at  $C^\alpha$ 's of residues that are linked via harmonic springs stabilizing the folded conformation. Macromolecules are assumed to undergo Gaussian-distributed fluctuations about their mean positions, being coupled by harmonic potentials. In GNM, these fluctuations are assumed to be isotropic, with no directional preferences, whereas in ANM they are anisotropic. There is no chemical information for distinguishing amino acids, and a uniform force constant  $\gamma$  is adopted for the interaction potential between all close neighboring residue pairs (Bahar *et al.*, 1997; Atilgan *et al.*,

2001). Low-frequency motions are insensitive to the atomic details of the structure and interactions as demonstrated by Tirion (1996). Therefore, both GNM and ANM can yield a successful description of a macromolecule's internal motions and overall conformational dynamics despite their approximations even at extreme levels (Doruker *et al.*, 2002a; Kurkcuoglu *et al.*, 2004; Song and Jernigan, 2007). More importantly, hierarchical levels of coarse-grained representation allow use of low-resolution data coming from cryo-EM (Tama *et al.*, 2004).

Original GNM and ANM together with their extended versions represent a more feasible route to simulate conformational fluctuations of very large systems than conventional MD approaches. Recently, the collective motions of large macromolecules such as GroEL-GroES complex (~8000 residues) (Keskin *et al.*, 2002a; Chennubhotla and Bahar, 2006), bacterial ribosome (~ 10000 residues) (Wang *et al.*, 2004), RNA polymerase (~3000 residues) (Yildirim and Doruker, 2004) and  $\beta$ -galactocidase (~ 4000 residues) (Kurkcuoglu *et al.*, 2004) are successfully reproduced with these approaches.

### 2.3. Proteins Studied

Relatively large enzyme triosephosphate isomerase (TIM) and the supramolecular assemblage bacterial ribosome complex are studied in the thesis. These protein complexes are widely investigated both experimentally (Zhang *et al.*, 1994; Frank and Spahn, 2006) and computationally (Guallar *et al.*, 2004; Tama *et al.*, 2003).

TIM is an important enzyme in glycolysis, catalyzing the interconversion between dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP). TIM has two identical monomers of 248 residues, each forming a TIM barrel. The enzyme is fully active as a dimer (Waley, 1973). There are three active site residues for the catalytic reaction located in the middle of each TIM barrel: Glu165 and His95 participate in proton transfer and Lys13 hydrogen bonds weakly with the bridging oxygen of the ligand (Zhang *et al.*, 1994) (Figure 2.3 (b)). The functionally important flexible loop (loop 6, residues 166-176) stabilizes the ligand for the interaction with the active site residues (Knowles, 1991), protects the ligand from the solvent and inhibits the formation of a potentially toxic byproduct (Sampson and Knowles, 1992) by remaining closed during the catalytic

reaction. Loop 6 moves about 7 Å while closing over the bound ligand (Banner *et al.*, 1975); however it also opens and closes as a natural motion in the absence of ligand, which can be observed from various crystal structures (Williams and McDermott, 1995; Derreumaux and Schlick, 1998).

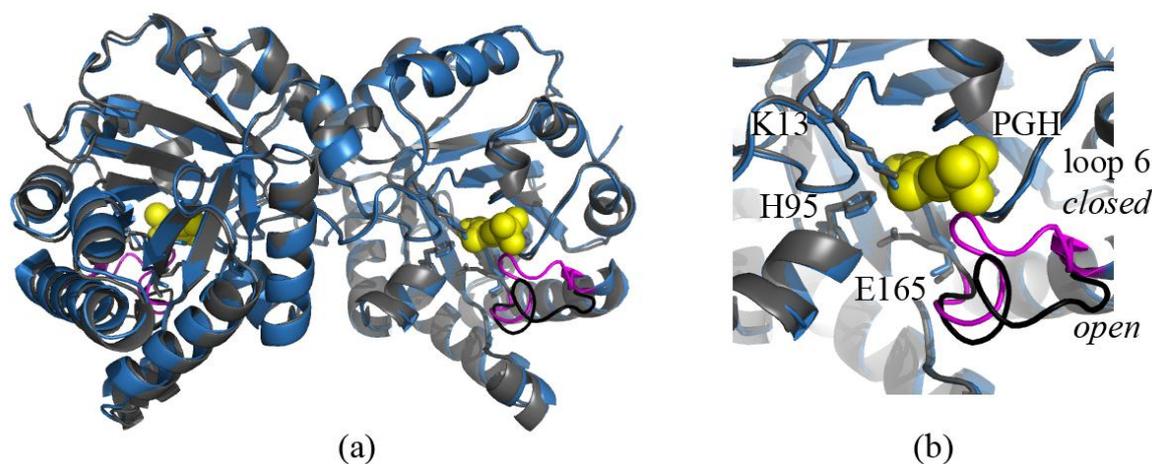


Figure 2.3. (a) Superimposed x-ray structures of ligand-free (gray, 8TIM.pdb) and ligand-bound (blue, 1TPH.pdb) triosephosphate isomerase. (b) Active site and inhibitor PGH

In this work, both free (with PDB code, 8TIM, 2.58 Å resolution) and complexed (1TPH, 1.8 Å resolution) forms of TIM (Figure 2.3 (a), (b)) are analyzed to explore the vibrational dynamics of the flexible loop 6 and the active-site residues with the new methodology. The ligand in the complex form (Zhang *et al.*, 1994) is the phosphoglycolohydroxamate (PGH), an intermediate analogue having the same structure and orientation with the substrate DHAP in the crystal structures. In Figure 2.3, the cartoon representations of the enzyme, and the other structures in all figures of the thesis, are prepared using the three-dimensional structure visualization software PyMol (DeLano, 2002).

The bacterial ribosome complexed with messenger RNA (mRNA), three transfer RNAs (tRNA) and elongation factor (EF)-Tu is a protein-RNA complex chosen to show the computational efficiency in applying the mixed coarse-grained methodology developed in this thesis to a very large biological system. Ribosome, a ‘molecular machine’ that synthesizes proteins in the cell according to the genetic information, mediates the

interaction of mRNA and tRNAs carrying the amino acids. It has two subunits namely 30S and 50S, called 70S together, in bacteria (Figure 2.4). The small subunit 30S (blue) comprises ~1500 nucleotide long 16S ribosomal RNA (rRNA) and 21 ribosomal proteins, whereas 50S (gray) is formed of ~2900 nucleotide long 23S rRNA, 120 nucleotide long 5S rRNA and more than 30 ribosomal proteins (Frank and Spahn, 2006). Low- and high-resolution ribosome structures are analyzed with residue-based and mixed coarse-grained ANM, respectively. The PDB file names for the low-resolution crystal ribosome structures studied are 1JGO (30S subunit, mRNA and tRNA) and 1GIY (50S subunit) (Yusupova *et al.*, 2001; Yusupov *et al.*, 2001). The high-resolution ribosome crystal structures are 2HGP (30S subunit, mRNA and three tRNA) and 2HGQ (50S subunit) (Yusupova *et al.*, 2006) respectively. Extended protein L9 of ribosome is excluded from the model and the missing protein L7/L12 (1DD4.pdb (Wahl *et al.*, 2000)), forming an important domain in collective dynamics (Tama *et al.*, 2003; Wang *et al.*, 2004; Trylska *et al.*, 2005) is added to the crystal structure 2HGP.

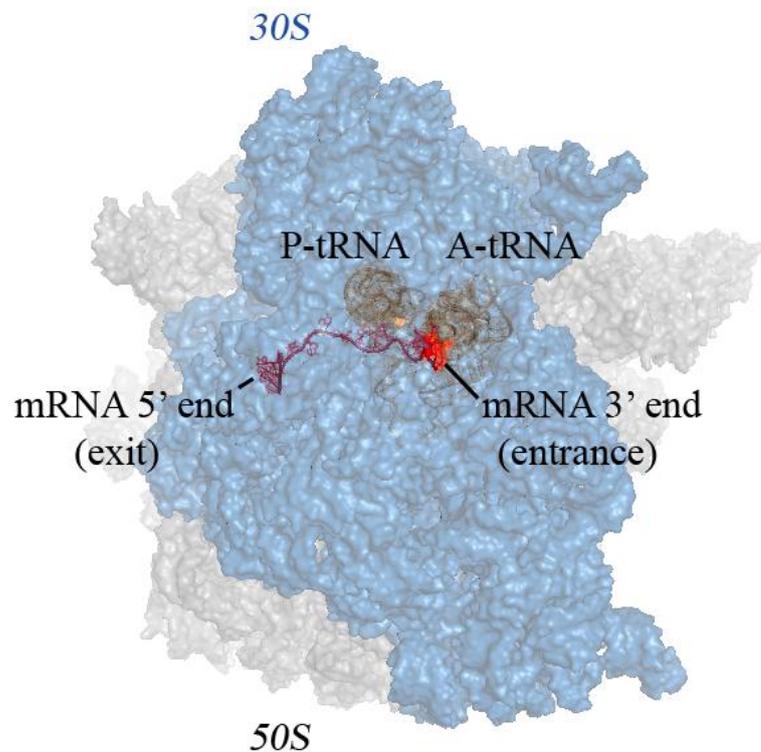


Figure 2.4. X-ray structure of bacterial ribosome complexed with mRNA and tRNAs

Details of TIM and ribosome complex structures with corresponding PDB codes are summarized in Table 2.1. Except crystal structures given in PDB files 1JGO and 1GIY for ribosome complex, all the structures are at high-resolution, i.e. Cartesian coordinates of the heavy atoms are available. Open (8TIM) and closed (1TPH) conformations of TIM are studied with residue-based, atom-based and mixed coarse-grained models. The low-resolution (lr) crystal structure of ribosome complex (1JGO, 1GIY) together with EF-Tu (1MJ1) is only investigated with residue-based model. Finally, the collective dynamics of the high-resolution (hr) structures of ribosome (2HGP, 2HGQ) with missing ribosomal protein L7/L12 (1DD4) are explored with both residue-based and mixed coarse-grained models.

Table 2.1. TIM and ribosome structure details, and applied models

Protein complex	PDB code	Resolution	Structure details	Elastic network model
TIM	8TIM	2.5 Å, hr	No ligand	Residue-based
	1TPH	1.8 Å, hr	With inhibitor PGH	Atom-based Mixed coarse-grained
Ribosome	1JGO	5.6 Å, lr	30S subunit, mRNA, tRNAs	Residue-based
	1GIY	5.5 Å, lr	50S subunit	
	1MJ1	Cryo-EM hr	Elongation factor Tu	
	2HGP	5.5 Å, hr	30S subunit, mRNA, tRNAs	Mixed coarse-grained
	2HGQ	5.5 Å, hr	50S subunit	
	1DD4	2.4 Å, hr	L7/L12 protein	

### 3. ELASTIC NETWORK MODEL

#### 3.1. Anisotropic Network Model

Elastic network models are based on the native conformations of proteins determined by high- and low-resolution techniques such as x-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (EM). Anisotropic network model (ANM) is a three-dimensional version of elastic network model incorporating anisotropy of residue fluctuations to Cartesian coordinates. It estimates the magnitude and direction vectors of residue displacements using normal mode analysis (NMA) for  $3N-6$  internal modes, for a structure of  $N$  residues.

In the coarse-grained representation of protein structures,  $C^\alpha$  atoms of amino acids and P atoms of nucleotides are accepted as interaction centers, i.e. nodes (Atilgan *et al.*, 2001; Wang *et al.*, 2004). All close-neighboring nodes are linked via springs with a universal force constant  $\gamma$ . Therefore, ANM as other elastic network models assumes that all bonded and non-bonded interactions are nonspecific in the folded structure of a protein, in contrary to classical NMA using full-atom empirical potentials to describe inter-residue interactions (Hayward, 2001). Maintaining the geometry and distribution of interactions throughout the structure, ANM provides satisfactory description of local and collective motions of molecular machines (Atilgan *et al.*, 2001).

In recent studies, several proteins from different sources, namely  $\alpha$ -amylase inhibitor tendamistat (74 residues) (Doruker *et al.*, 2000), EcoRI-DNA complex (600 residues) (Doruker *et al.*, 2006) and triosephosphate isomerase (494 residues) (Cansu and Doruker, *in press*) have been analyzed by both ANM and MD. It has been demonstrated that the collective motions at low frequencies obtained by ANM and essential motions from MD agree with high overlaps. Moreover, these motions are very robust and can be reproduced at higher and mixed levels of coarse-graining in elastic networks (Doruker *et al.*, 2000; Kurkcuoglu *et al.*, 2004). These results point that protein structures are really uniform rubbery bodies with limited repertoire of motions, determined by their shapes. As the control of these motions is very strong, it is appropriate to consider that these computed

collective motions of a protein are related to biological mechanisms such as enzyme reactions, processing steps and chaperone activities.

### 3.2. Computational Methodology in ANM

In the elastic network model composed of  $N$  identical nodes, close neighboring residues within a cutoff distance  $r_c$ , are connected by identical harmonic springs with force constant  $\gamma$ . Then, the total potential energy of the folded protein structure can be given as a summation over all harmonic interactions of  $(i, j)$  pairs,

$$V = (\gamma/2) \sum_i \sum_j h(r_c - R_{ij})(\Delta \mathbf{R}_j - \Delta \mathbf{R}_i)^2 \quad (3.1)$$

Here,  $h(x)$  is the heavy side step function [ $h(x) = 1$  if  $x \geq 0$ , and zero otherwise] and  $R_{ij}$  is the distance between sites  $i$  and  $j$  in the native structure of protein.  $\Delta \mathbf{R}_i$  is the fluctuation in the position vector  $\mathbf{R}_i$  of site  $i$  ( $1 \leq i \leq N$ ).

The potential energy can be expanded in a Taylor series at a minimum denoted by Hayward (2001)

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_0 q_i q_j + \dots \quad (3.2)$$

Here,  $q_i = \sqrt{m_j} \Delta x_j$ ,  $q_{i+1} = \sqrt{m_j} \Delta y_j$  and  $q_{i+2} = \sqrt{m_j} \Delta z_j$  are the mass-weighted coordinates, where  $j$  stands for  $N$  atoms and  $i$  for  $3N$  Cartesian coordinates.

The equations of motion in terms of any set of coordinates are given by Lagrange's equations,

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) = \left( \frac{\partial L}{\partial q_i} \right) \quad (3.3a)$$

The Lagrangian is given as the kinetic energy minus the potential energy.

$$L = \frac{1}{2} \sum_{i=1}^{3N} \dot{q}_i^2 - \frac{1}{2} \sum_{i,j=1}^{3N} \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_0 q_i q_j \quad (3.3b)$$

Lagrange in matrix form is defined as,

$$L = \frac{1}{2} \dot{\mathbf{q}}^t \dot{\mathbf{q}} - \frac{1}{2} \mathbf{q}^t \mathbf{H} \mathbf{q} \quad (3.4)$$

Here, t is used for transpose, and  $\mathbf{H}$  is the Hessian matrix calculated from the second derivatives of the potential energy with respect to mass weighted coordinates, practically equal to force constants.

The potential energy can be also expressed in matrix form as

$$\mathbf{V} = (1/2) \Delta \mathbf{R}^t \mathbf{H} \Delta \mathbf{R} \quad (3.5)$$

where,  $\Delta \mathbf{R}$  is the  $3N$ -dimensional vector of fluctuations and  $\Delta \mathbf{R}^t$  stands for its transpose. In order to calculate the normal modes of the elastic network, symmetric Hessian matrix  $\mathbf{H}$  is diagonalized into the form,

$$\mathbf{S}^t \mathbf{H} \mathbf{S} = \mathbf{\Lambda} \quad (3.6)$$

$\mathbf{\Lambda}$  is the  $(3N \times 3N)$  diagonal matrix with diagonal elements being eigenvalues or squared normal mode frequencies.  $\mathbf{S}$  is a  $(3N \times 3N)$  orthogonal matrix ( $\mathbf{S}^t \mathbf{S} = \mathbf{S} \mathbf{S}^t = \mathbf{I}$ ,  $\mathbf{I}$  being the identity matrix) where the columns are the normalized eigenvectors giving the normal mode directions of motion. In normal mode calculations, overall rotational and translational motion of the molecule are excluded corresponding to six zero eigenvalues and the overall motion is described over  $3N-6$  individual internal modes.

A new coordinate set is defined as

$$Q_i = \sum_{k=1}^{3N} S_{ik} q_i \quad i = 1, \dots, 3N \quad (3.7)$$

When Equation 3.7 is substituted into equation 3.4, the solution of the differential equation is  $Q_i = A_i \cos(\omega_k t + \varepsilon_k)$ , where each normal mode has a frequency of  $\omega_k$  with phase  $\varepsilon_k$ .

In classical dynamics, each normal mode has a time-averaged potential energy defined by  $\frac{1}{2} k_B T$ . Then, the application of equipartition law leads to the mean-square fluctuation of residue  $i$  defined in terms of mass-weighted coordinates and averaged over  $3N-6$  normal modes (Hayward, 2001; Atilgan *et al.*, 2001),

$$\langle \Delta q_i^2 \rangle = \frac{k_B T}{\gamma} \sum_{k=1}^{3N-6} \left( \frac{S_{ik}}{\omega_k} \right)^2 \quad (3.8)$$

$k_B$  is the Boltzmann's constant and  $T$  the absolute temperature and the mean-square fluctuation of  $i^{\text{th}}$  residue at  $k^{\text{th}}$  mode is given as

$$\langle \Delta R_i^2 \rangle_k = \frac{1}{m_j} \left[ \langle q_i^2 \rangle_k + \langle q_{i+1}^2 \rangle_k + \langle q_{i+2}^2 \rangle_k \right] \quad (3.9)$$

The force constant  $\gamma$  is the only adjustable parameter in the model, whose value is determined by comparing theoretical results (Equation 3.8 and 3.9) with experimental residue fluctuations given by the temperature factor or B-factor:

$$B_i = (8\pi^2 / 3) \langle \Delta R_i^2 \rangle \quad (3.10)$$

### 3.3. Methodology in Mixed Coarse-Graining

Previously (Kurkcuoglu, 2003; Kurkcuoglu *et al.*, 2004), mixed coarse-graining approach has been introduced to elastic networks to model the native structure of a protein with high- and low-resolution regions. Here, the high-resolution regions being the

‘interesting’ parts of the protein were described at one-node-per-residue level and the rest of the structure of  $n$  residues at higher levels of coarse-graining (for example by retaining  $(n/2)$ ,  $(n/5)$ ,  $(n/10)$ ,  $(n/20)$  or  $(n/40)$  residues) (Figure 3.1 (b)). As a result, a reduced number of nodes can be retained in NMA but still experimental temperature factors are reproduced successfully and meaningfully similar results are obtained compared to those obtained from detailed simulations.

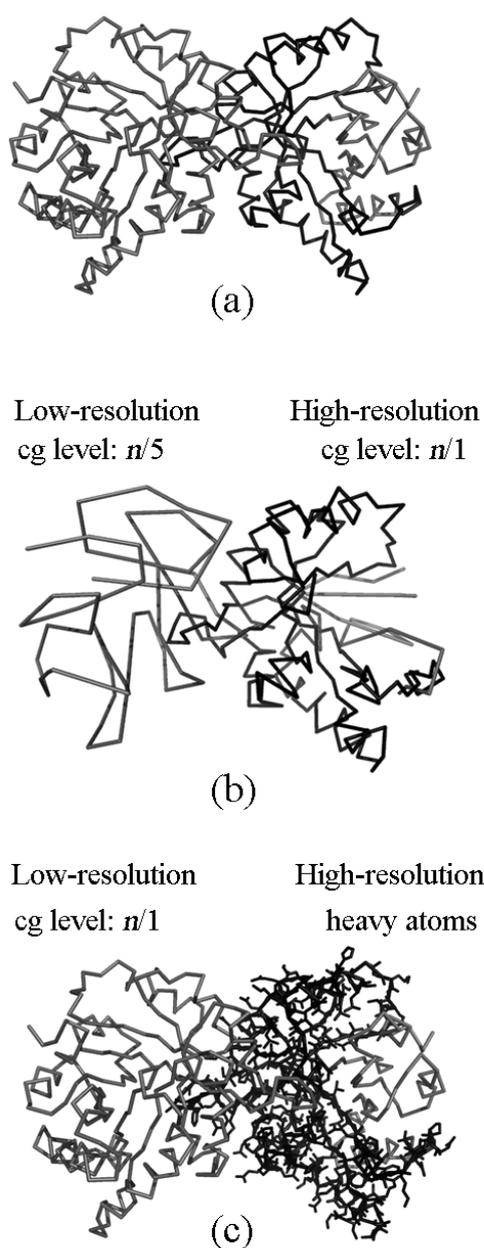


Figure 3.1. X-ray structures of triosephosphate isomerase with (a) residue-based model, and (b), (c) mixed coarse-graining (cg) at different levels

In this study, the aim is to describe the collective dynamics of extremely large systems at atomistic detail, while maintaining the computational efficiency of the elastic network models. For this purpose, the mixed coarse-graining is moved to a detailed level; each node of the elastic network represents either a single atom (high-resolution) or a residue (low-resolution) (Figure 3.1 (c)). The high-resolution regions constitute the interesting parts of a protein (e.g. active site on an enzyme) while the rest of the system is retained at coarse-grained level to maintain the shape of the structure (Kurkcuoglu *et al.*, 2005; 2006). It will be shown in results section that different conformations of large proteins can be efficiently obtained at atomistic detail, difficult to attain with full-atom MD and NMA methods.

Two different procedures are developed for mixed coarse-graining with atomistic detail. In the first approach, similar to the previous study (Kurkcuoglu *et al.*, 2004), appropriate cutoff distances and force constants are assigned for the interaction among nodes belonging to different resolution regions. These interaction parameters should be separately defined for the high-resolution, low-resolution and the interface throughout the elastic network. The cutoff distance for the low-resolution region is taken from previous works (Atilgan *et al.*, 2001; Kurkcuoglu *et al.*, 2004), and for the high-resolution region an appropriate cutoff is defined based on residue-based ANM calculations. Suitable force constants of different regions are assigned by comparing theoretical and experimental B-factors as explained above. However, for an extremely large molecule such as ribosome, this approach is computationally expensive for the calculation of the high-resolution force constant. To overcome this difficulty, the range of interactions among atoms is kept constant, and the value of force constant is assigned based on atom-atom interactions, as an alternative approach. These procedures are discussed in the following sections.

### 3.3.1. Mixed Coarse-Graining Procedure I

In the elastic network models, the cutoff distance specifying the range of interactions between nodes is scaled on the basis of radius of gyration of the polymer. In the original model (Atilgan *et al.*, 2001), a cutoff distance of  $13\text{\AA}$  was found suitable to define both short and long range interactions of coarse-grained residues represented by single points placed at  $C^\alpha$  atoms. The same value is used in this study for the low-resolution region

modeled at one-node-per-residue level. But, it is necessary to define an appropriate cutoff distance for the high-resolution region represented at atom detail. For this purpose, two sets of atom-based ANM calculations are performed at cutoff of 6Å and 9Å for several proteins. Then, their frequency distributions are compared with the results of residue-based model with 13Å cutoff, where only C<sup>α</sup> atoms are nodes of the elastic network. Finally an appropriate force constant is selected for the high-resolution (Kurkcuoglu *et al.*, 2005; 2006). Application of this methodology is discussed in detail in the following chapter.

There are two different types of nodes in the mixed resolution structure as shown in Figure 3.2. The cutoff distance determining the interaction between nodes of high- (node 1) and low-resolution (node 2) regions is defined as

$$r_{c,1,2} = \left( \frac{r_{c,1}^3 + r_{c,2}^3}{2} \right)^{1/3} \quad (3.11)$$

by considering that the masses of the nodes are proportional to the volume of spherical residues.

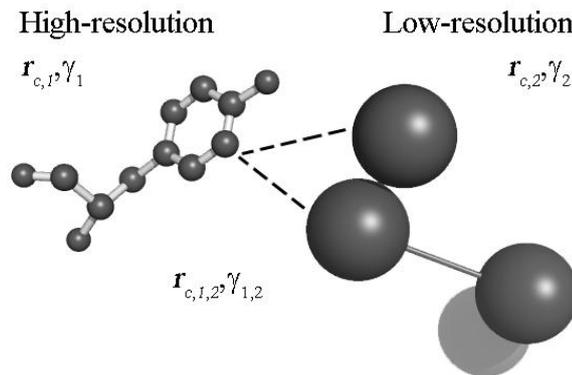


Figure 3.2. Interaction parameters between nodes in mixed coarse-graining procedure I

After fixing the cutoff radii, the effective force constant for each resolution can be calculated by matching experimental B-factors and theoretical mean-square fluctuations averaged over  $3N-6$  normal modes. The force constant between nodes 1 and 2 is found by a linear interpolation between  $(r_{c,1}, \gamma_1)$  and  $(r_{c,2}, \gamma_2)$  pairs.

A coarse-grained node is defined as the center of mass of a group of  $n$  residues with different molecular weights and coordinates. Then, the mean-square fluctuation of this node can be defined as

$$\Delta\mathbf{R}_{cm} - \Delta\bar{\mathbf{R}}_{cm} = \frac{m_1\Delta\mathbf{R}_1 + \dots + m_n\Delta\mathbf{R}_n}{\sum_{i=1}^n m_i} \quad (3.12)$$

Considering that  $\Delta\bar{\mathbf{R}}_{cm}$  is zero at equilibrium and neglecting the cross-correlations terms between residue fluctuations, it can be shown that

$$\langle \Delta R_{cm}^2 \rangle_n = \frac{\sum_{i=1}^n m_i^2 \langle \Delta R_i^2 \rangle}{(\sum_{i=1}^n m_i)^2} \quad (3.13)$$

Here summations are performed over  $n$  residues. This equation is used to estimate the B-factor for a node that is a collection of  $n$  residues, as

$$B_{cm,n} = (8\pi^2 / 3) \langle \Delta R_{cm}^2 \rangle_n \quad (3.14)$$

### 3.3.2. Mixed Coarse-Graining Procedure II

In procedure I, the effective force constant for the high-resolution is estimated by modeling protein's native conformation by considering all heavy atoms in the structure. However, this is computationally inefficient, and even impossible for very large molecules such as ribosome complex. To overcome this difficulty, another approach is developed.

All the close neighboring heavy atoms in the folded structure are connected by a harmonic spring  $\gamma$ . And, suitable force constants to link different nodes are assigned based on atom contacts. In the high-resolution region, interacting heavy atoms are linked via a spring constant whose value is set to one in this study. In the low-resolution region, one node represents an amino acid or a nucleotide with a total mass of its heavy atoms. The

force constant linking heavy nodes should be suitable to maintain the homogeneity of force in the system. So, the effective force linking two coarse-grained nodes is set equal to total number of interactions between these former residues' atoms. Therefore, mixed resolution elastic network with reasonable interactions can be modeled for extremely large molecules.

In a protein-RNA (or DNA) complex, the shape and size of an amino acid and a nucleotide differ. For this reason, the interaction distance is assigned as  $r_{c,aa}=10 \text{ \AA}$  between amino acid atoms,  $r_{c,aa,nuc}=13 \text{ \AA}$  between amino acid and nucleotide atoms and  $r_{c,nuc}=15 \text{ \AA}$  between nucleotide atoms' interactions. Then, the high- and low-resolution regions of the elastic network are determined (Figure 3.3).

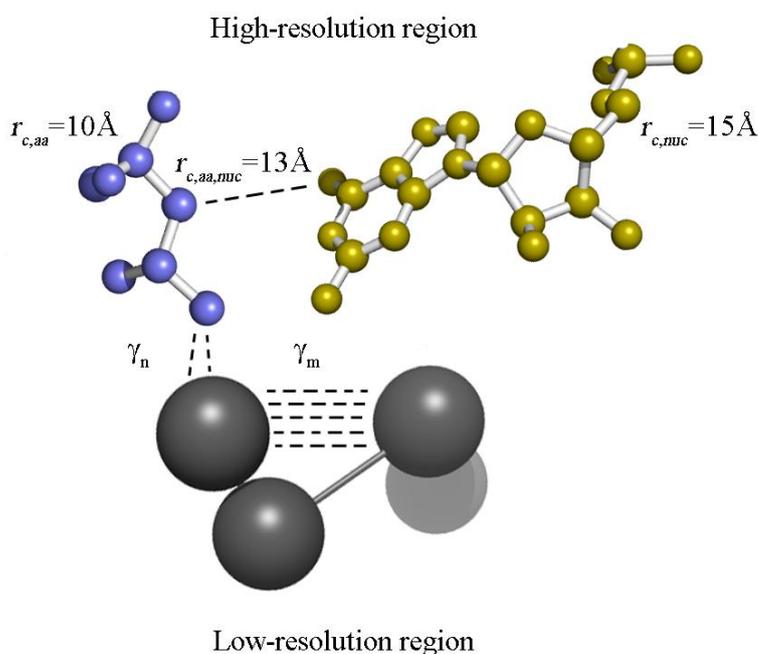


Figure 3.3. Interaction parameters between nodes in mixed coarse-graining procedure II

In the simulations, the Hessian matrix is diagonalized to estimate the displacement vectors of the residues and corresponding eigenvalues of the normal modes. Singular value decomposition can be used to calculate all  $3N$  normal modes, as in the original model (Atilgan *et al.*, 2001). However, for an elastic network model of a system containing 10000 nodes, the diagonalization of a (30000 x 30000) Hessian matrix is computationally challenging by singular value decomposition technique. It is well known that ten to twenty

normal modes in the low frequency spectrum is sufficient to describe the dominant motions of the system (Lee *et al.*, 2006). Therefore, the three-dimensional displacement vectors of the nodes are calculated for the slowest ten modes using the computationally efficient software package BLZPACK (Marques and Sanejouand, 1995) using a block Lanczos algorithm (Grimes *et al.*, 1991). This algorithm facilitates calculating the generalized and standard eigenvalue problem and allows obtaining corresponding eigenvectors and eigenvalues of a very large matrix within a considerably short time, especially using a large memory of 28 GB with a cluster of a 32-processor SGI Origin 2800.

Flow charts summarizing the mixed coarse-graining model with procedure I and II are displayed in Figure 3.4. It is clear that the second procedure is more advantageous as compared to the first procedure, since the former requires only one step of diagonalization of the Hessian matrix. Furthermore, in the first procedure, all  $3N-6$  normal modes should be calculated to estimate the force constant of the low- or high-resolution regions, whereas in the second mode, only first ten to twenty normal modes are sufficient to describe functional globular motions.

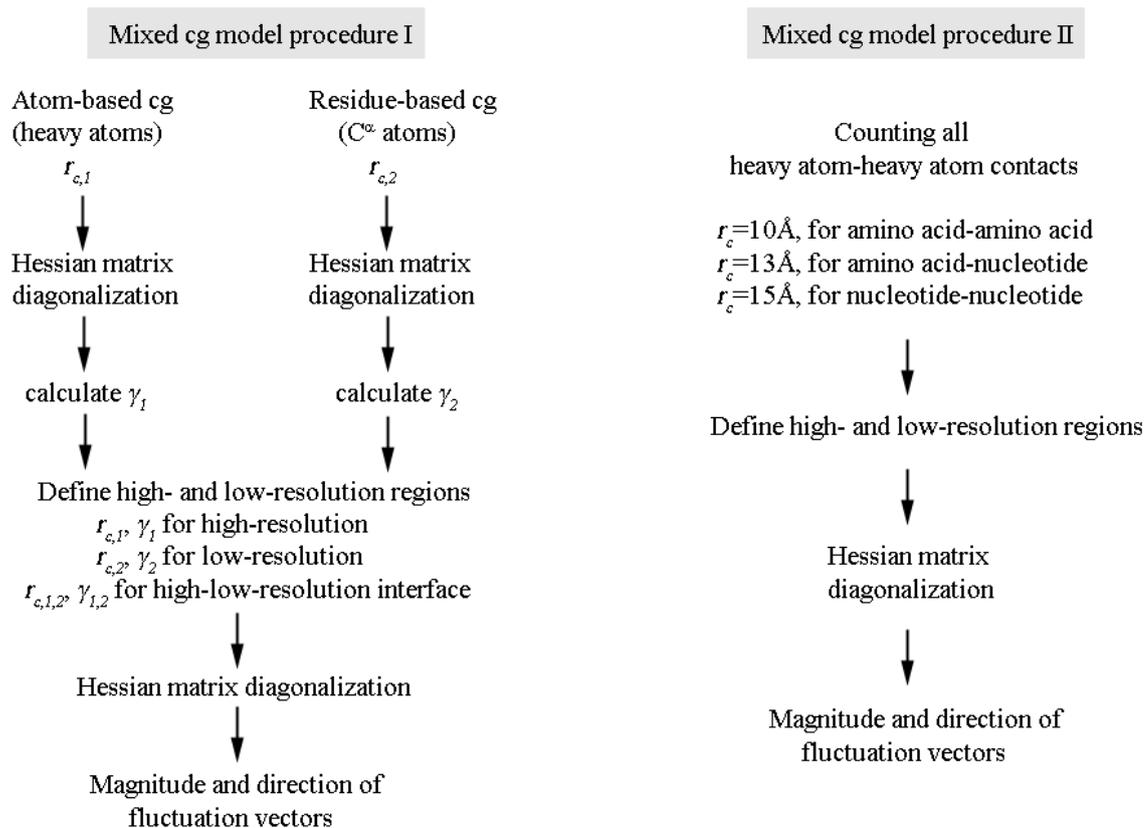


Figure 3.4. Flow charts of the mixed coarse-graining procedures I and II

## 4. DYNAMICS OF TRIOSEPHOSPHATE ISOMERASE

In elastic network models, the collective dynamics of large macromolecules have been effectively investigated by the residue-based uniform coarse-graining (cg), where the  $C^\alpha$  atom of each amino acid represents a single node in the elastic network (Keskin *et al.*, 2002; Wang *et al.*, 2004; Yildirim and Doruker, 2004; Chennubhotla and Bahar, 2006). However, recently developed mixed cg method (Kurkcuoglu, 2003; Kurkcuoglu *et al.*, 2004) enables representation of the system with high- and low-resolution regions. In this approach, the functionally important parts such as the active site are described at atomistic detail (high-resolution). As a result, a detailed analysis on interesting parts can be performed at atomistic level with high computational efficiency for supramolecular assemblages.

Two separate procedures are developed for mixed cg. In the first approach, appropriate cutoff distances are estimated for the coarse-grained (low-resolution) and atomic regions (high-resolution) of the protein. Then, force constants are determined separately with atom-based and residue-based ANM for high- and low-resolution regions, respectively (Kurkcuoglu *et al.* 2005; 2006). On the other hand, in the second procedure, a single cutoff value for atomic interactions is fixed, and the force constants linking atoms or residues are estimated based on atom-atom interactions throughout the elastic network.

In this chapter of the thesis, an important enzyme in glycolysis pathway, triosephosphate isomerase (TIM) in inhibitor-bound and unbound forms is studied. Functionally important parts i.e. active site, loop 6 (residues 166-176) and inhibitor phosphoglycolohydroxamate (PGH) on one monomer of TIM's free and complex forms are analyzed with atomic detail. To provide a complete account, both residue-based (i.e. only  $C^\alpha$  atoms are considered) and atom-based (i.e. all heavy atoms are considered) elastic network models of TIM and classical normal mode analysis of the free enzyme with full-atom force field CHARMM (Brooks *et al.*, 1983) are explored.

#### 4.1. Determination of Parameters for Mixed Coarse-Graining Procedure I

In the current mixed resolution representation of structure, heavy atoms constitute the high-resolution region and the low-resolution is coarse-grained (cg) at the residue level. As a node in the mixed elastic network either stands for an atom or a residue, the range and force of interactions, i.e. the cutoff distances and spring constants, should be separately defined for the high- and low-resolution regions and the interface. In previous studies (Atilgan *et al.*, 2001; Kurkcuoglu *et al.*, 2004), cutoff was defined as 13 Å to describe the low-resolution network in the residue-based ANM. It is therefore necessary to define an appropriate cutoff distance for the atomistic network.

For this purpose, motions of several proteins of relatively small sizes (total number of residues ranging between 78 and 238), namely glutaredoxin, HIV-1 protease, Che Y protein, dehydrofolate reductase (DHFR) (free and bound forms) and LAO binding protein, are studied with atom-based ANM at two different cutoff distances 6 Å and 9 Å. Their normalized frequency distributions are compared with the results obtained from residue-based ANM with a cutoff of 13 Å. The normalized density of vibrational normal modes  $g(\omega)$  (number of modes within a frequency range divided by the total number of nodes) (Ben-Avraham, 1992) is shown in Figure 4.1 (a) for DHFR. Over the whole frequency range, the results for  $r_c=6$  Å (atomistic model) are in accordance with that of residue-based ANM with 13 Å cutoff for DHFR and all proteins listed above. Figure 4.1 (b) displays  $g(\omega)$  for TIM, where the atom-based results with  $r_c=9$  Å agree better with residue-based results. This finding points that 9 Å may be more appropriate for a larger system. However, it should be noted that at low-frequency range, where collective motions of proteins are observed, both cutoffs 6 Å and 9 Å are in agreement with residue-based ANM results for 13 Å.

As a result, for the cutoff distance  $r_{c,1}$  of high-resolution region there are two choices 6 Å or 9 Å, and for the low-resolution system  $r_{c,2}$  is accepted as 13 Å. And, the cutoff distance between high- and low-resolution regions  $r_{c1,2}$  is determined by Equation 3.11, considering that the masses of spherical nodes are proportional to their volume assuming constant density.

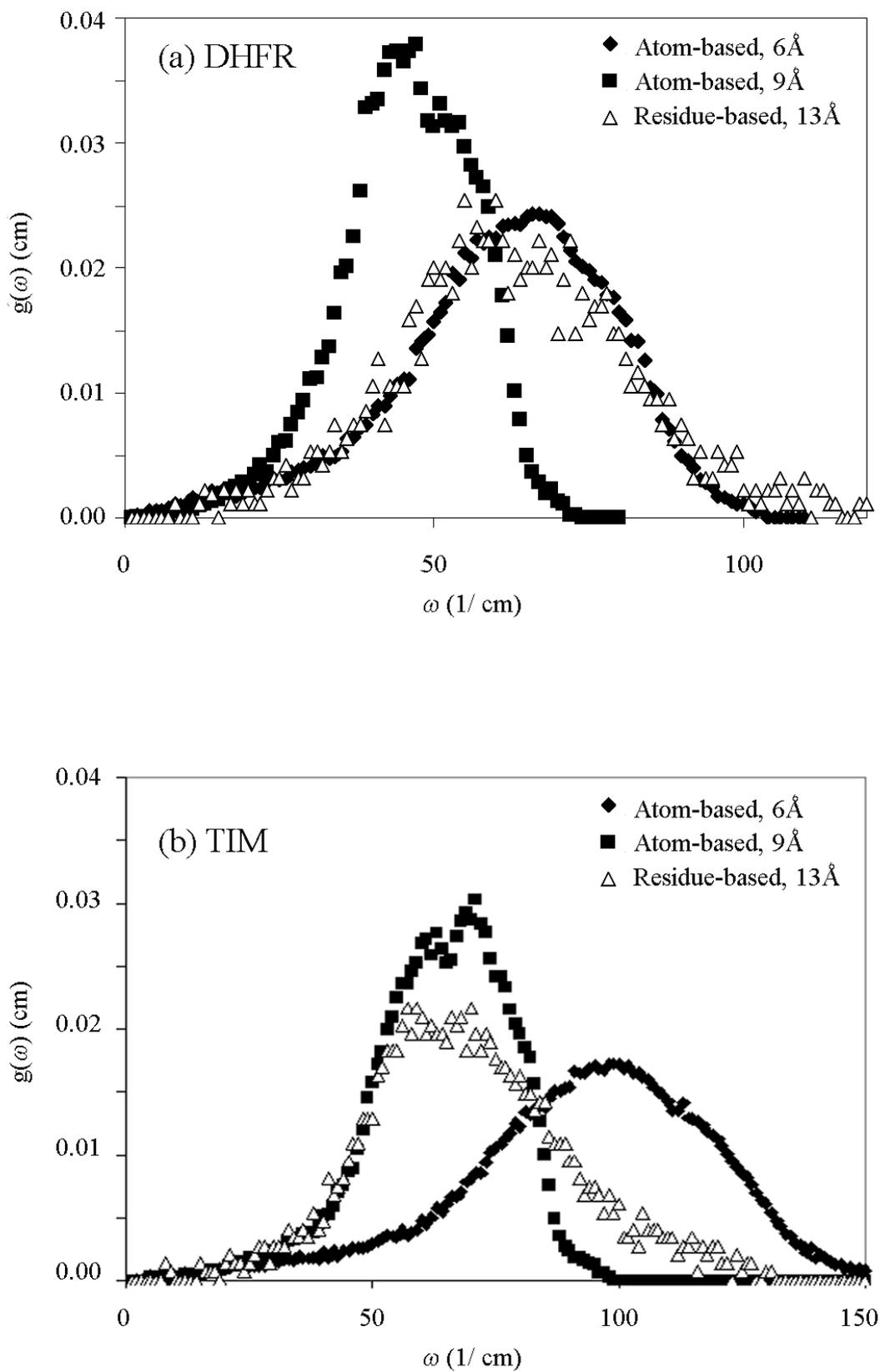


Figure 4.1. Density of vibrational modes,  $g(\omega)$  (cm), of (a) DHFR, and (b) TIM

After the cutoff distance is set for the uniform cg model of TIM, the force constant  $\gamma$  is calculated as a scale factor by comparing computed and experimental B-factors. As a result, cutoffs of 6 Å or 9 Å in the high-resolution correspond to a force constant  $\gamma_1$  of 0.64 or 1.45 kcal/mol Å<sup>2</sup> respectively. For the low-resolution region,  $\gamma_2$  is calculated as 3.1 kcal/mol Å<sup>2</sup>. The force constant linking different nodes, is found by linear interpolation between  $(r_{c1}, \gamma_1)$  and  $(r_{c2}, \gamma_2)$  pairs. Table 4.1 summarizes all interaction parameters used in residue-based, atom-based and mixed cg models.

Table 4.1. ANM interaction parameters for coarse-grained systems

Cg system	High-resolution		Low-resolution		High/ low-resolution interface	
	$r_{c1}$ (Å)	$\gamma_1$ (kcal/molÅ <sup>2</sup> )	$r_{c2}$ (Å)	$\gamma_2$ (kcal/molÅ <sup>2</sup> )	$r_{c1,2}$ (Å)	$\gamma_{1,2}$ (kcal/molÅ <sup>2</sup> )
Residue-based	-	-	13	3.1	-	-
Atom-based I	6	0.64	-	-	-	-
Atom-based II	9	1.45	-	-	-	-
Mixed case I	6	0.64	13	3.1	10.64	2.27
Mixed case II	9	1.45	13	3.1	11.35	1.88
Mixed case III	9	0.64	13	3.1	11.35	2.08

## 4.2. Residue- and Atom-Based Models of TIM

Inhibitor-bound and unbound structures of the enzyme is studied with residue-based ANM at 13 Å. Mean-square fluctuations calculated from residue-based ANM and experimental B-factors are in good agreement with a correlation coefficient of 0.65 (Figure 4.2). Slowest mode shapes are obtained for both structures and quite similar motions are observed in low frequencies (Figure 4.3 and 4.4). Probably due to local packing of the structures, opening/ closing motion of loop 6 seems to be more hindered in the bound structure, regardless of only one node representing the inhibitor PGH. Similar findings were mentioned in a recent computational study on different loop conformations (Guallar *et al.*, 2004).

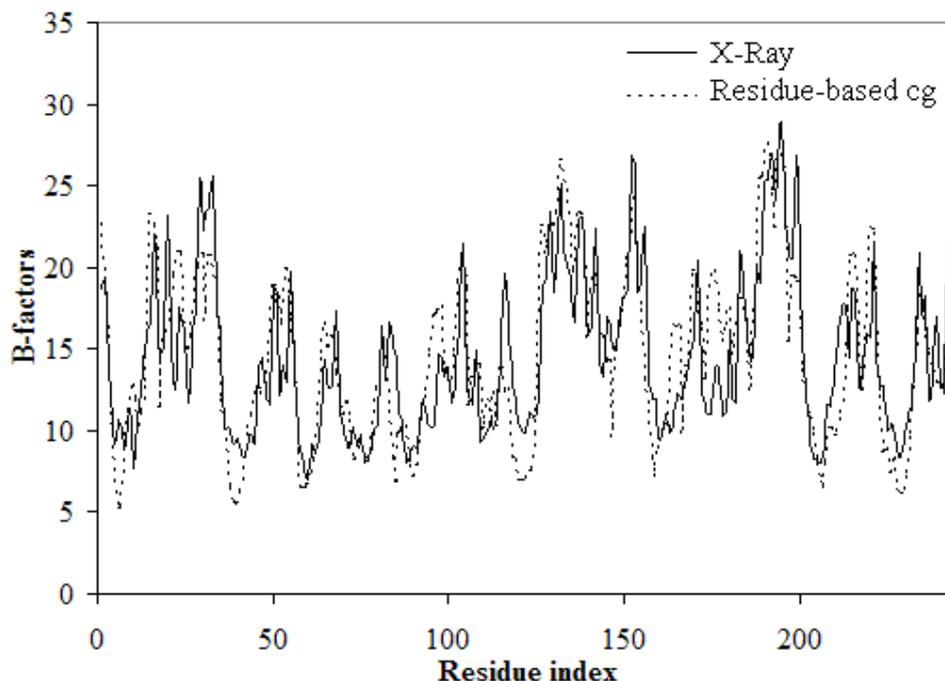


Figure 4.2. B-factors of one monomer of bound-TIM with residue-based cg from x-ray experiment and ANM calculations

Moreover, free and complex structures of TIM are analyzed with atom-based ANM at 6 Å and 9 Å considering all heavy atoms including two inhibitors PGH. First and second mode shapes of TIM-PGH complex shown in Figure 4.3 and Figure 4.4 are same for 6 Å and 9 Å results. Similar globular deformations of the dimer structure have been observed from essential motions of molecular dynamics simulations recently performed for 60 ns duration (Cansu and Doruker, *in press*).

Eigenvectors calculated from normal modes correspond to displacement vectors of residues at different frequencies. To estimate the similarity between modes of motion of two systems, the overlap value is calculated. This is defined as the inner dot product between two superimposed eigenvector sets  $(v_1, \dots, v_k)$  and  $(w_1, \dots, w_k)$  normalized over  $k$  modes given as (Amadei *et al.*, 1999)

$$overlap = \left( \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k (v_i w_j)^2 \right)^{1/2} \quad (4.1)$$

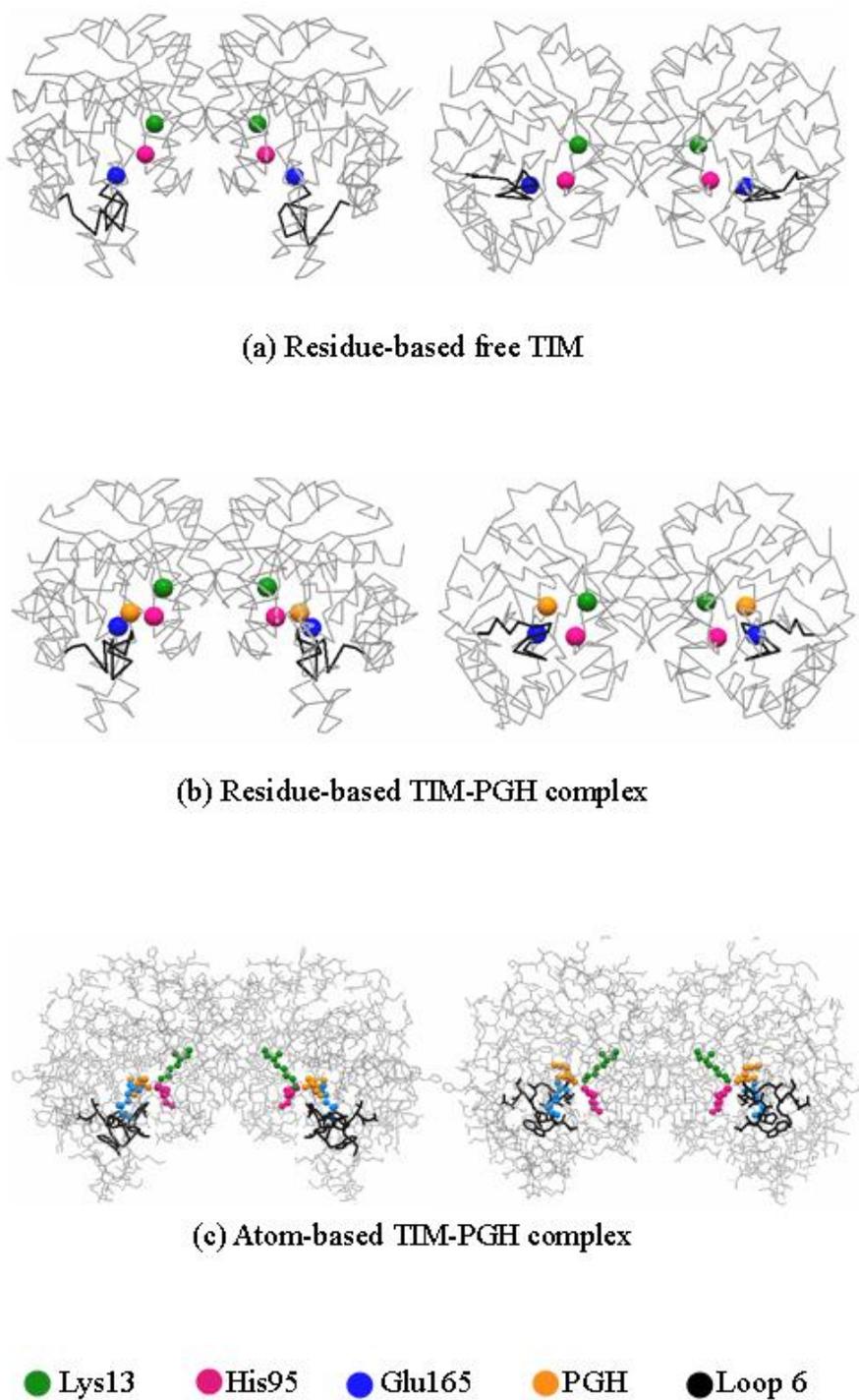


Figure 4.3. Alternative conformations at the first mode of (a) residue-based free TIM, (b) residue-based TIM complex and (c) atom-based TIM complex

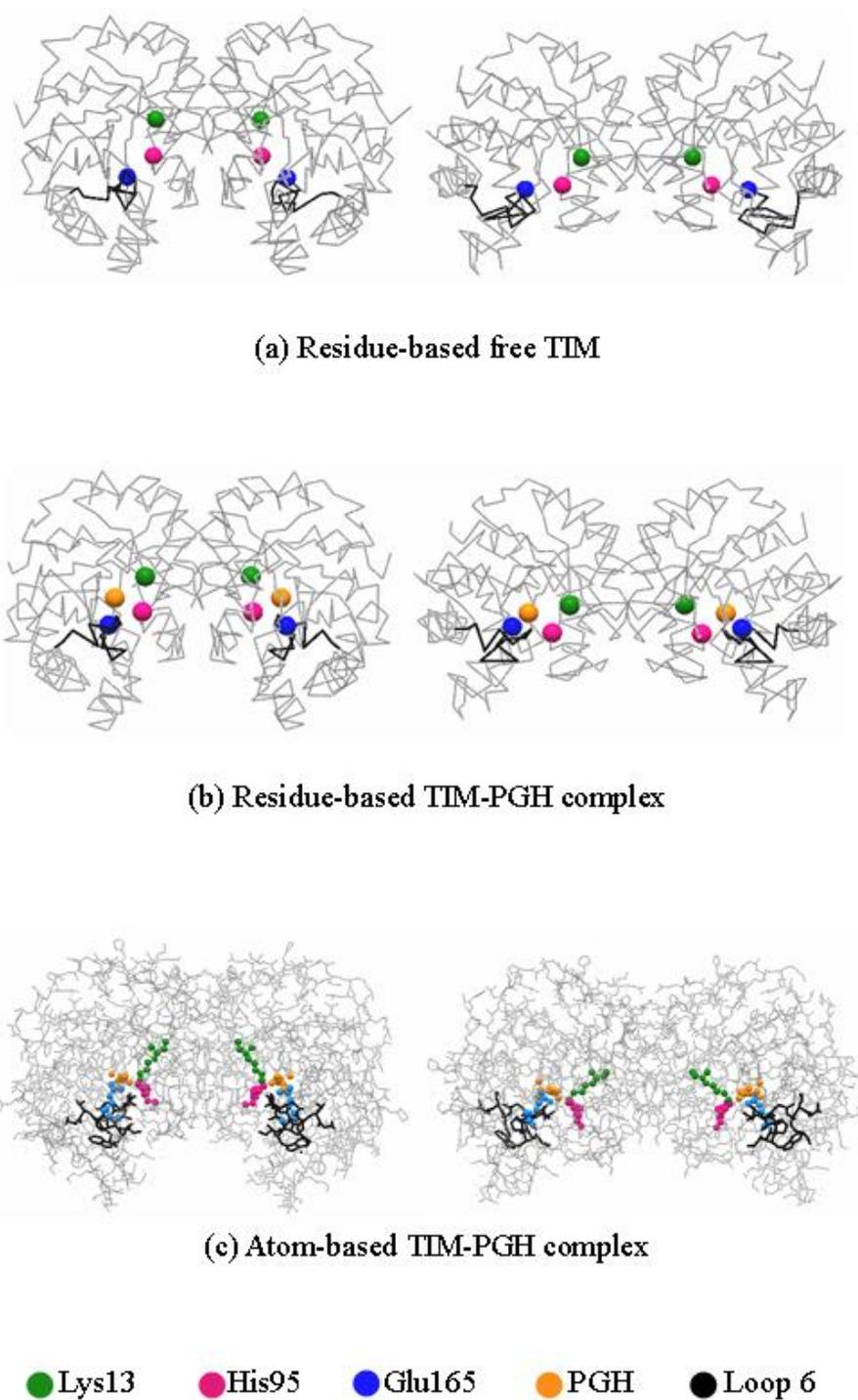


Figure 4.4. Alternative conformations at the second mode of (a) residue-based free TIM, (b) residue-based TIM complex and (c) atom-based TIM complex

The extent of control of slow motions is directly reflected in a very small number (~10-20) of the normal modes as dominant motions shown by experimental evidence (Lee *et al.*, 2006). For this reason, overlap values averaged over first ten modes (i.e.  $k=10$ ) is given for the studied models in Table 4.2. The overlaps are based only on C $^{\alpha}$  structures. They are calculated between different structures of TIM-PGH complex and residue-based bound structure is taken as reference. An overlap of 0.93 points to a high correspondence between slow modes of free and bound TIM analyzed with residue-based ANM. Similarly, slowest modes of atom-based approaches are successfully obtained with 0.87 at 6 Å and 0.97 at 9 Å.

Table 4.2. Overlap values for residue- and atom-based model averaged over first ten modes

Cg system	$r_c$ (Å), $\gamma$ (kcal/mol Å <sup>2</sup> )	Overlap
Residue-based	13, 3.1	0.93
Atom-based I	6, 0.64	0.87
Atom-based II	9, 1.45	0.97

Individual elements of the overlap (absolute value of  $(v_i \cdot w_j)$ ) constitute an overlap matrix of all  $(i,j)$  pairs. The overlap matrices of inhibitor-bound and unbound enzyme for residue- and atom-based models are given in Figure 4.5 (a)-(c) for first twenty five modes. At the low frequency modes, same direction of motions (i.e. overlap value close to unity) are observed for residue- and atom-based models of TIM. This finding is reasonable because collective motions are very robust and they can be reproduced as long as the shape of the structure is conserved.

### 4.3. Mixed Coarse-Grained Models of TIM

#### 4.3.1. Mixed Coarse-Graining Procedure I

Two alternative mixed cg models are applied to TIM, namely asymmetric and symmetric mixed cg models. In asymmetric mixed cg model, active site, loop 6 and one inhibitor PGH are taken in atomic detail on one monomer. In symmetric mixed cg model,

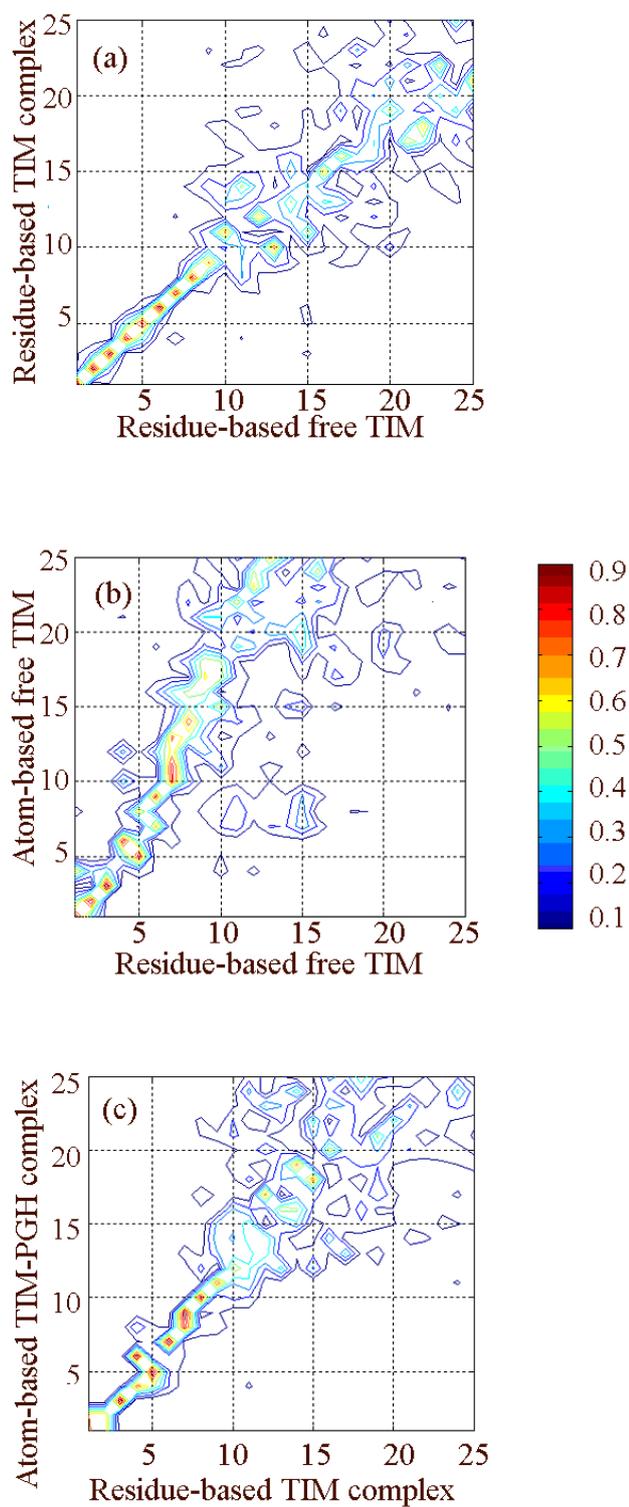


Figure 4.5. Overlap matrices shown for studied residue-based models

high-resolution region includes the active site and two inhibitors on both monomers. In both models, the remaining of the structure is modeled at one-node-per-residue level.

In Table 4.1, the cutoff distances and force constants used in mixed cg models are shown. The overlap values are calculated between asymmetric mixed cg and residue-based structures of TIM-PGH complex taken as the reference (Table 4.3). Here, overlap values are based only on  $C^\alpha$  structures. In the overlap matrix for the asymmetric mixed cg model case I ( $r_{cl}=6\text{\AA}$ ,  $\gamma_1=0.64\text{ kcal/mol \AA}^2$ ) and residue-based cg of bound enzyme, the first modes shift and spread out but the overlap value is 0.85 (Figure 4.6); and, functionally important motions are still retained. However, the overlap between asymmetric mixed cg case II ( $r_{cl}=9\text{\AA}$ ,  $\gamma_1=1.45\text{ kcal/mol \AA}^2$ ) and residue-based cg is slightly lower than that of case I, which is 0.80. On the other hand, when the force constant of high-resolution is lowered to  $0.64\text{ kcal/mol \AA}^2$  for a cutoff of  $9\text{ \AA}$  (case III), modes of motion are successfully reproduced with one-to-one correspondence and a high-overlap (0.89) as shown in Figure 4.6. A cutoff of  $9\text{ \AA}$  results in more interaction of residues and relatively smaller force constant of high-resolution allows more flexibility throughout the structure. In all mixed cg models, a smaller force constant ( $0.32\text{ kcal/mol \AA}^2$ ) is assigned for the ligand-enzyme interface flexibility.

Table 4.3. Overlap values for mixed cg models averaged over first ten modes

Cg system	$r_c$ ( $\text{\AA}$ ), $\gamma$ ( $\text{kcal/mol \AA}^2$ )		Overlap
	High-resolution	Low-resolution	
Mixed cg model, case I	6, 0.64	13, 3.1	0.85
Mixed cg model, case II	9, 1.45	13, 3.1	0.80
Mixed cg model, case III	9, 0.64	13, 3.1	0.89

The first and second mode shapes are displayed in Figure 4.7 for asymmetric and symmetric mixed cg models of TIM. The opening/ closing motions of loop 6 (in black) are clearly observed in the slowest modes of mixed-resolution description. Moreover, the local dynamics of active site residues Lys13 (in green), His95 (magenta), Glu165 (blue) and inhibitor PGH (orange) are efficiently calculated at atomic level.

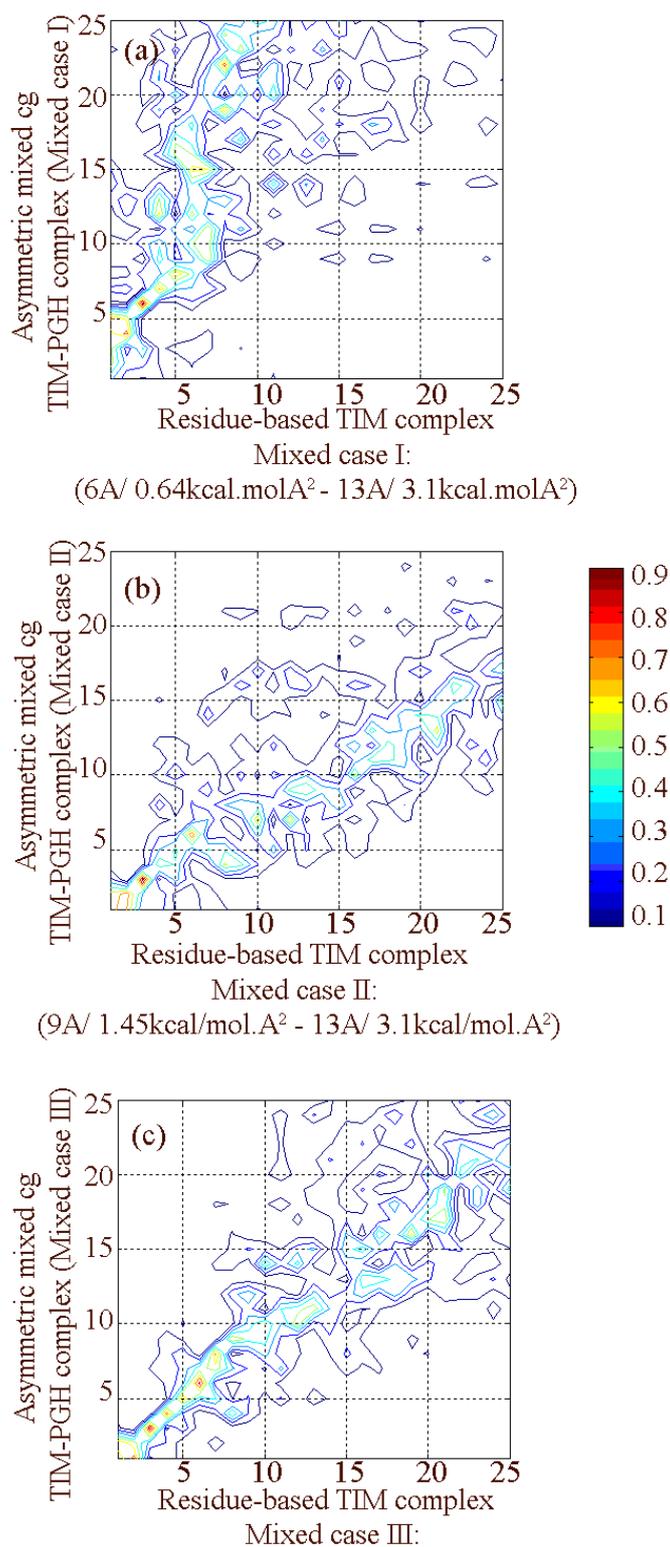
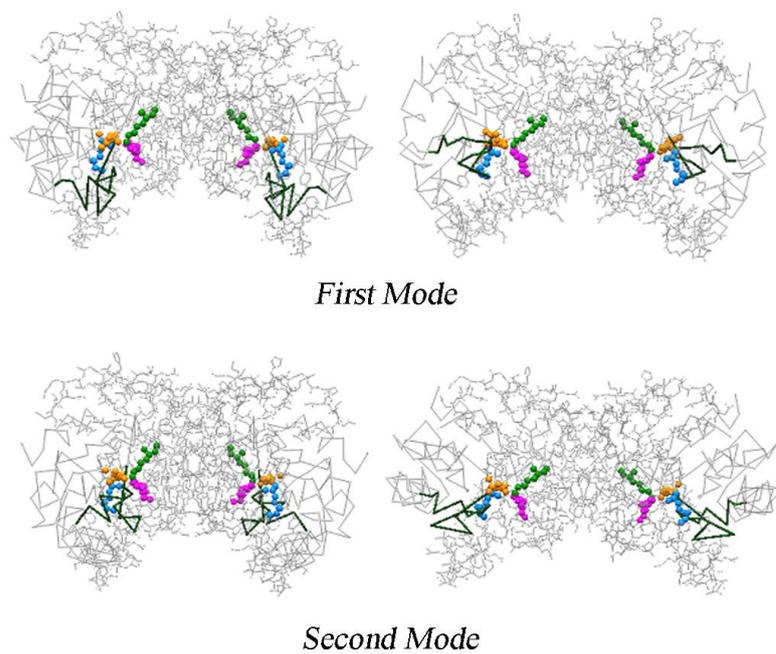
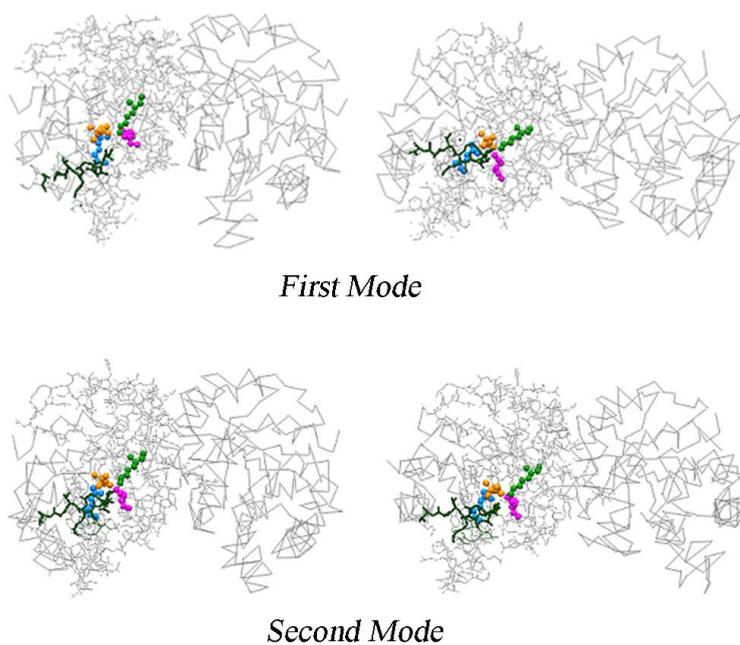


Figure 4.6. Overlap matrices shown for studied mixed cg models



(a) Symmetric mixed cg TIM-PGH complex



(b) Asymmetric mixed cg TIM-PGH complex

Figure 4.7. First and second mode alternative conformations for (a) symmetric and (b) asymmetric mixed cg of TIM-PGH complex

### 4.3.2. Mixed Coarse-Graining Procedure II

Mixed cg approach incorporating atom details to elastic networks is successful to predict the collective motions of the enzyme TIM with atomic details (Kurkcuoglu *et al.*, 2005; 2006). However, one drawback of this procedure is the prediction of force constant in the high-resolution region, calculated with atom-based ANM (i.e. heavy atoms as the nodes of the elastic network). This step consumes a significant amount of computational time depending on the size of the system. As a result, another approach is developed to implicitly assign the interaction forces throughout the elastic network of a supramolecular assemblage, as explained in the third chapter.

Overlap values are calculated for residue-based and mixed cg models of TIM with two procedures, only for the C<sup>α</sup> atoms. A high overlap of 0.94 is obtained between residue-based models from first and second procedures, showing the reliability of the new approach (Table 4.4). Moreover, the overlap value between residue-based and mixed cg model with the procedure II (p. II) is found as 0.96. This finding implies a progress in predicting the collective motions with mixed resolution when compared to residue-based modeling, since a similar calculation yielded a value ~ 0.85 (Table 4.2). Additionally, if the cg level of the low-resolution region is increased to one-node-per-five residues (i.e.  $n/5$ ,  $n$  being the number residues in low-resolution region) in mixed cg model, globular motions are still dominant with an overlap of 0.89. Furthermore, the alternative mode shapes at low frequencies can be obtained successfully at different levels of mixed cg (Figure 4.8).

Table 4.4. Overlap values calculated for various models of TIM

<b>Models</b>	<b>Overlap</b>
Residue-based (p. I)-Residue-based (p. II)	0.94
Residue-based (p. I)-Mixed cg (p. I)	0.79
Residue-based (p. II)-Mixed cg (p. II)	0.96
Mixed cg (p. I)-Mixed cg (p. II)	0.96
Mixed cg (p. II) -Mixed cg (p. II) ( $n/5$ )	0.89

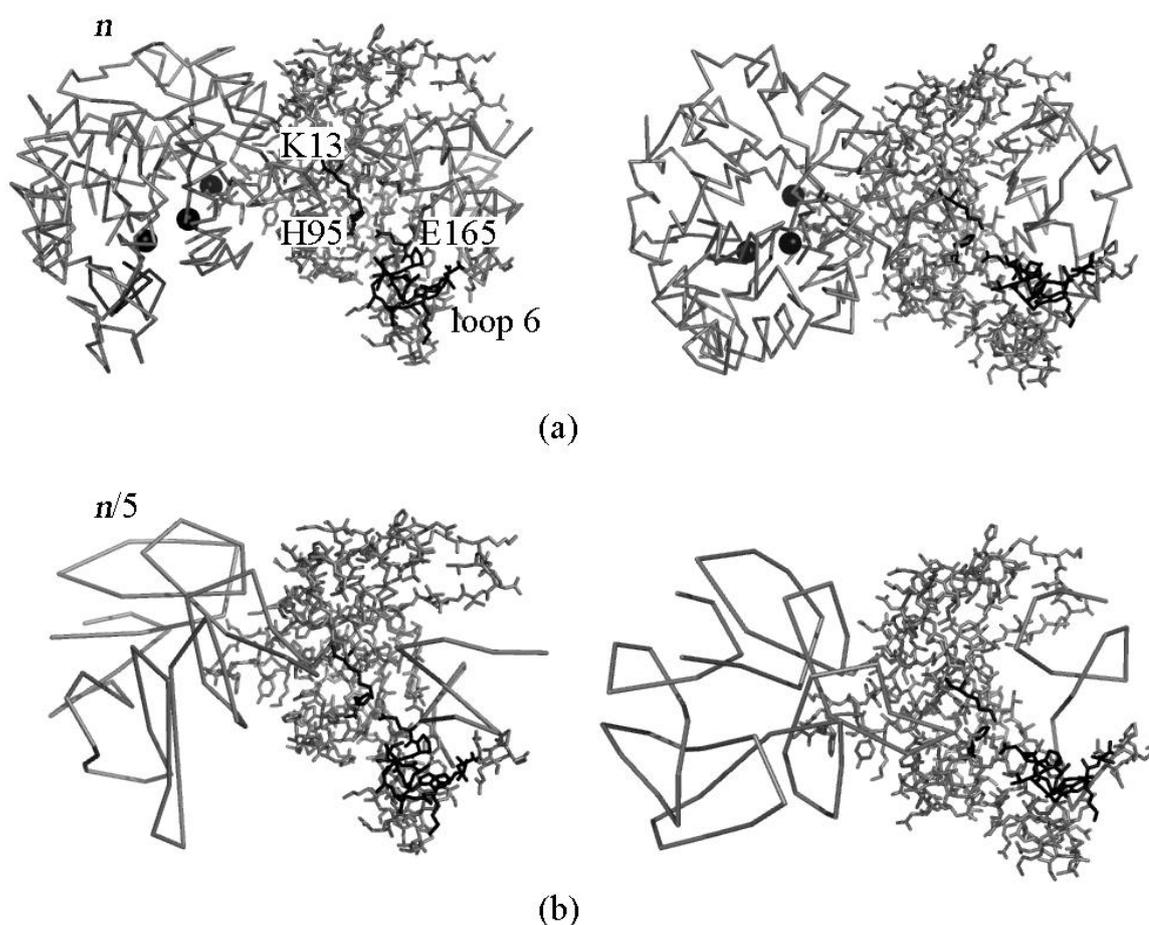


Figure 4.8. Slowest mode alternative mode shapes of TIM from mixed cg procedure II with low-resolution modeled at (a)  $n$  and (b)  $n/5$  level

#### 4.4. Collective Dynamics with Atomistic Detail

##### 4.4.1. Dominant Motions as a Dimer

TIM barrel proteins, due to their stability, are ideally suitable for protein engineering studies, especially in the design of new loops for variety of properties (Thanki *et al.*, 1997). During the design process TIM, which is fully active as a dimer (Waley, 1973), is usually monomerized into a stable fold with residual but significant activity. In this respect, collective dynamics of triosephosphate isomerase is investigated in detail to understand the role of the dimeric form for controlling the active site atomic motions conforming to the requirements of the enzymatic reaction mechanism.

For the four slowest modes, symmetrical counter rotations of two monomers around two axes are shown schematically in Figure 4.9. Similar deformations for dimer TIM are observed from crystal structures from PDB, namely 1M6J (Rodriguez-Romero *et al.*, 2002), 1B9B (Maes *et al.*, 1999) and 1TRE (Noble *et al.*, 1993) superimposed on their backbone (Figure 4.10) and from molecular dynamics simulations of 60 ns (Cansu and Doruker, *in press*). However, these globular motions are not observed for the monomeric TIM with MD simulations. As a result, the cooperativity of the motions in dimeric form may have an impact in constructing the geometry in the active site and forcing loop 6 open and close over the ligand for effective catalysis. This subject will be further discussed in the following sections.

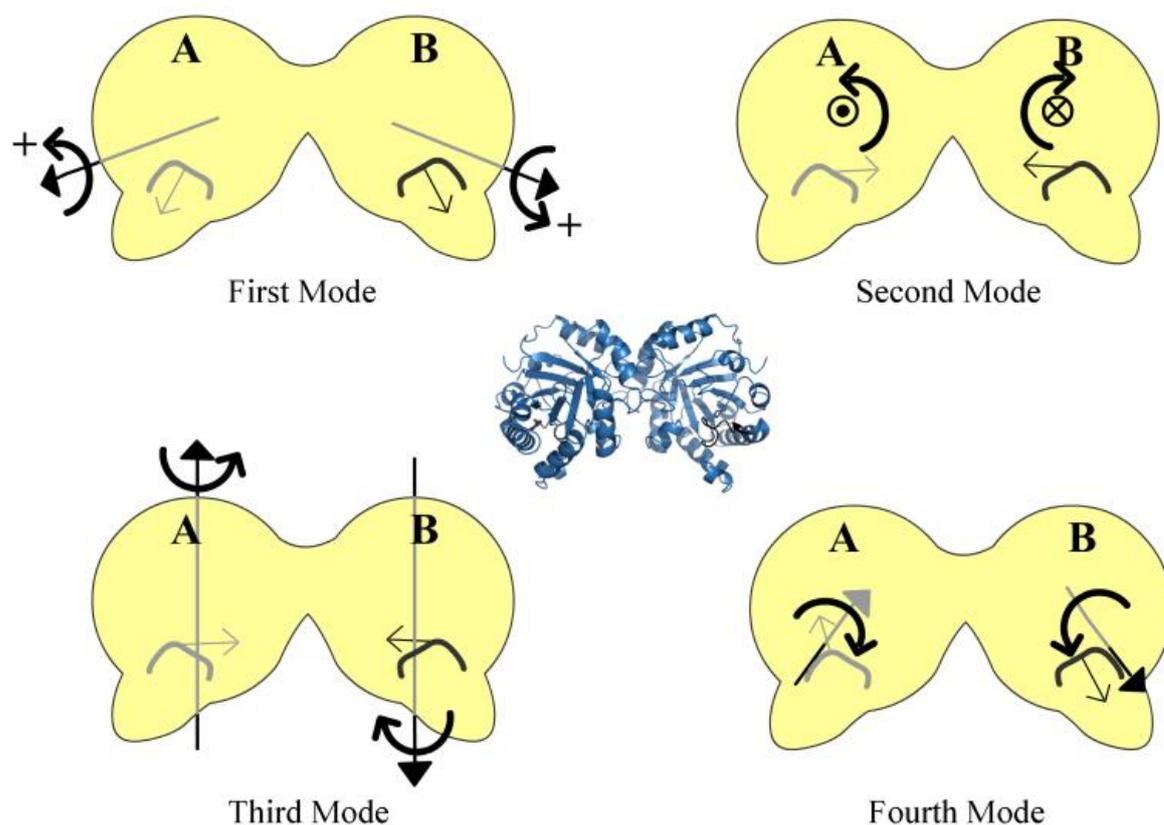


Figure 4.9. Twisting of monomers A and B, and loop 6 around rotation axes at first fourth modes. Arrows show displacement direction around axes ( $\odot$ ,  $\otimes$ ) following right hand rule

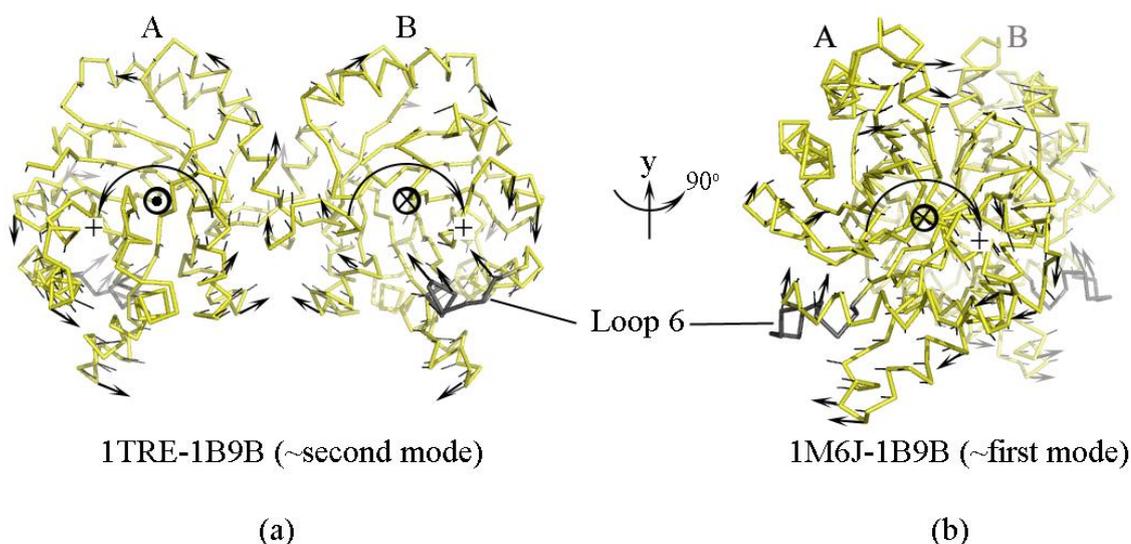


Figure 4.10. Twisting motions observed from superimposed experimental x-ray structures (a) 1TRE and 1B9B (similar to second mode), (b) 1M6J and 1B9B (similar to first mode). Arrows indicate direction of deformation around axes ( $\odot$ ,  $\otimes$ ) following right hand rule

#### 4.4.2. Classical Normal Mode Analysis

A vibrational analysis of the complex structure with NMA is performed to further confirm the reliability of the slow modes obtained by ANM. A fully atomistic empirical energy function with the software CHARMM (version 30b2) is used to define atomic interactions (Brooks *et al.*, 1983). The enzyme complex structure without the inhibitor is modeled with the standard PARAM19 force field, considering all atoms including hydrogens. Simulation temperature is taken as 300 K, and the cutoff distance for non-bonded interactions is set at 13 Å, with respective spline and buffer widths of 4 and 1 Å. To eliminate unrealistic atomic interactions originating from crystal structure, an energy minimization is performed. The energy minimizers used for the calculations are the steepest descent (1100 steps) and the adapted basis-Newton Raphson ( $10^9$  steps) methods. Steepest descent is a linear method to quickly lower the energy, but it does not reach the local minimum. Newton-Raphson is a second order method to converge efficiently to the nearest local minimum. No constraint is applied to the system.

After the total energy of the system is minimized, the VIBRAN module of CHARMM program is used to obtain normal-mode vectors, which are later compared with

the results of residue-based ANM. The average overlap including the first 10 eigenvectors is 0.79 (with dot products of the first and second mode eigenvectors being 0.75 and 0.67), where only the C<sup>α</sup> atoms are considered. Similar domain motions and local fluctuations of the loop 6 and the active-site residues are obtained in the first mode. The range for the ten slowest vibrational frequencies is calculated as 2.0-5.3 cm<sup>-1</sup> with CHARMM, which is somewhat below that of the atom-based model (7.9-21.8 cm<sup>-1</sup>). This difference is based on the fact that the parameters of ANM, i.e., the force constant and the cutoff distance, have not been adjusted to fit the corresponding frequency ranges; however, this could readily be modified if desired. As a result, the collective motions observed with ANM agree with the results obtained from classical NMA.

#### 4.4.3. Active Site and Loop 6 Motions

Various x-ray structures and experimental findings exhibit common features concerning TIM's catalytic mechanisms: (i) When loop 6 (amino acid 166-176) is closed, the amide group of Gly171, located at the loop, makes a hydrogen bond with the phosphate group of the ligand. (ii) The tip of the loop (labeled by the carbon of Thr172) displaces more than 7 Å between open and closed conformations. (iii) During the opening/closing motion of loop 6, the catalytic base Glu165 moves about 2 Å forcing the ligand into optimal orientation for catalysis, which is observed for various substrate enzyme-complexes including PGH (Zhang *et al.*, 1994). (iv) The Trp168 indole ring rotates about 50° upon binding, which has also been monitored by fluorescence (Sampson and Knowles, 1992) and T-jump relaxation spectroscopy (Desamero *et al.*, 2003).

The opening/ closing direction of loop 6 observed from slow modes is compared with x-ray open-closed structures. Table 4.5 compares the direction of loop motion between free (unbound, open) and complex (bound, closed) forms of TIM and the displacement vectors from slow modes of studied models for the specific region on loop 6, residues 169-173. Especially, first and fourth modes strongly overlap with experimental residue displacements, implying that these modes may have functional importance in catalysis. Combining these four modes' eigenvectors ( $S_i$ ) into single displacement vector  $\langle \Delta q_i \rangle$  as described by Kundu and Jernigan (2004),

$$\langle \Delta q_i \rangle = \sum_{k=1}^4 \frac{S_{ik}}{\lambda_k} \quad (4.2)$$

results to an overlap value between 0.75-0.79. An important conclusion is that the functional loop of TIM may be controlled to move in different directions than the domains, but under the control of the collective motions.

Table 4.5. Overlap values observed between displacement vectors of residues (169-173) from different models and x-ray open-closed structures

<b>Coarse-grained system</b>	<b>Mode indices</b>	<b>Overlap value</b>
Residue-based cg (complex form)	1	0.65
	2	0.39
	3	0.14
	4	0.84
Residue-based cg (free form)	1	0.58
	2	0.24
	3	0.16
	4	0.65
Asymmetric mixed cg case III parameters (complex form)	1	0.55
	2	0.68
	3	0.29
	4	0.81
Symmetric mixed cg case I parameters (complex form)	1	0.66
	2	0.12
	3	0.16
	4	0.82

The collective motions of the active-site residues, the flexible loop, and the inhibitor are analyzed and compared with the experimental findings, to verify the validity of the mixed cg model. The active-site details are displayed in Figure 4.11 for four slowest modes only for the asymmetric model for the complex enzyme since all models give same collective motions. The displacement vectors of residues in all slowest modes are multiplied with the same scaling factor, to enhance and better visualize the internal

motions. The alternative conformations are superimposed on the basis of the structure backbone, and only one of the PGH molecules is shown for clarity.

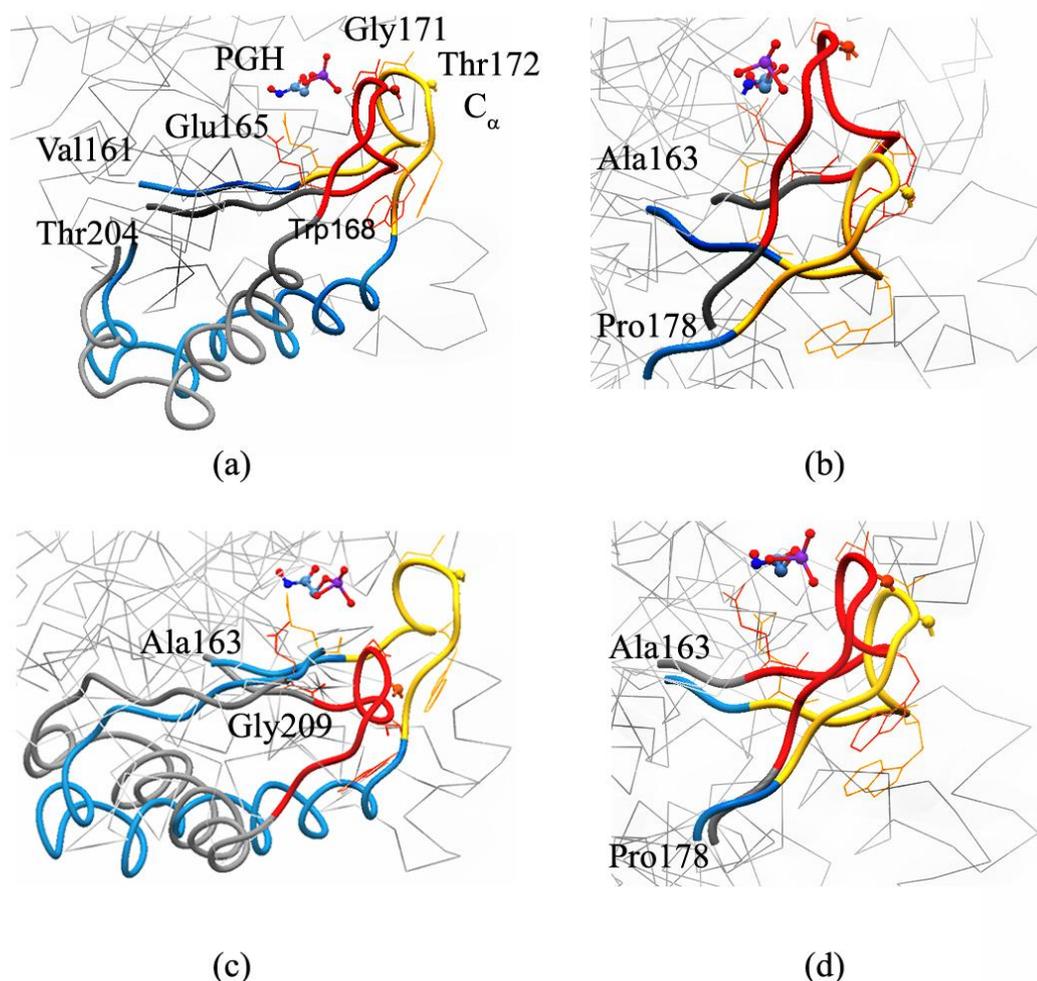


Figure 4.11. Opening (yellow)/ closing (red) motion of loop 6 with active site residues, and PGH at (a) first, (b) second, (c) third and (d) fourth modes

In Figure 4.11 (a), alternative conformations of the active site and loop 6 at the slowest mode are displayed for the asymmetric mixed cg model (case III). Driven by the domain motions, flexible loop (loop 6) opens and closes over the active site. When the tip of the loop (i.e.,  $C^\alpha$  of Thr172) moves about 7 Å, the catalytic base Glu165 shifts about 4-4.5 Å, and the indole ring of Trp168 rotates about 50°. Similarly, the distance between the amide group of Gly171 and the phosphate group of PGH which make a hydrogen bond in the bound-structure, varies between 4.40 Å (yellow loop) and 2.09 Å (red loop) in the two alternative conformations in the first mode (Figure 4.11 (a)). These deformation values are

scaled changes between two alternative conformations in the first mode, however they agree significantly well with the experimental findings listed above (Zhang *et al.*, 1994; Desamero *et al.*, 2003). This indicates the remarkable control of the protein in affecting this motion.

Rescaled with the same factor used in the first mode, the displacements of active-site region in the second and third modes of the asymmetric mixed cg model, are displayed in Figure 4.11 (b) and (c), respectively. In these modes, two monomers rotate around symmetrical axes perpendicular to the longest axis of TIM (see Figure 4.9). As a result, the loop movement is more along the horizontal direction in contrast to the vertical displacement of the loop observed among the X-ray structures. Respectively in second and third modes, the tip of the loop moves by about 9.84 and 11 Å, relatively larger compared to the first mode. The indole ring of Trp168 rotates by about 50° and 25°, and the distance between the amide group of Gly171 and the phosphate group of PGH changes from 6.77 Å (yellow loop) to 2.42 Å (red loop) and from 5.09 Å (yellow loop) to 2.84 Å (red loop) in the second and third modes, respectively.

The active-site region for the fourth mode is shown in Figure 4.11 (d). Similar to the first mode, loop 6 opens and closes in vertical direction as observed in crystal structures. In this mode, there are two symmetric hinge axes passing from Thr177 on each monomer forcing the tip of the loop to move about 4 Å, Glu165 to shift 1.5 Å and the indole ring of Trp168 to rotate slightly. The distance between the amide group of Gly171 and the phosphate group of PGH varies between 4.18 Å (yellow loop) and 2.18 Å (red loop).

In summary, due to domain motions involving around different rotation axes as shown in Figure 4.9, the first and fourth modes clearly relate more closely to the opening/closing motion of loop 6 observed from crystal structures 8TIM (open) and 1TPH (closed). Therefore, these modes seem to be important in catalytic activity of the enzyme, whereas the second and third modes also involve loop motion but in other directions as compared to that observed in the x-ray conformations. However, a common feature to all of these four slowest modes is that the flexible loop moves nearly as a rigid body: the root-mean-square deviation (rmsd) between the alternative conformations of loop 6 in a specific mode (as shown in Figure 4.11) is below 0.4 Å.

#### 4.4.4. Twisting Motion and Hinges in Slow Modes

The N and C termini of loop 6 (amino acids 166 and 176) were reported to act as hinges during its rigid-body displacement between the open and closed x-ray conformations (Joseph *et al.*, 1990). Moreover, the mutation of these end residues to glycine resulted in reduced catalytic activity (Xiang *et al.*, 2004). However, when various crystal structures of ligand-bound and unbound TIM (such as structures with PDB names 1M6J, 1TRE, 1TIM, 1WYI, and 1NF0) are superimposed on their backbones, C-terminus of the loop (amino acid 176) seems to be relatively more mobile [with an rmsd of about 2 Å among different structures] compared to the N-terminus (amino acid 166). This observation agrees with the findings in the slowest four modes: one hinge lies close to N-terminus (around residues 160-163 in the first, third, and fourth modes) but not close to the C-terminus (only observed in the fourth mode around residue 178).

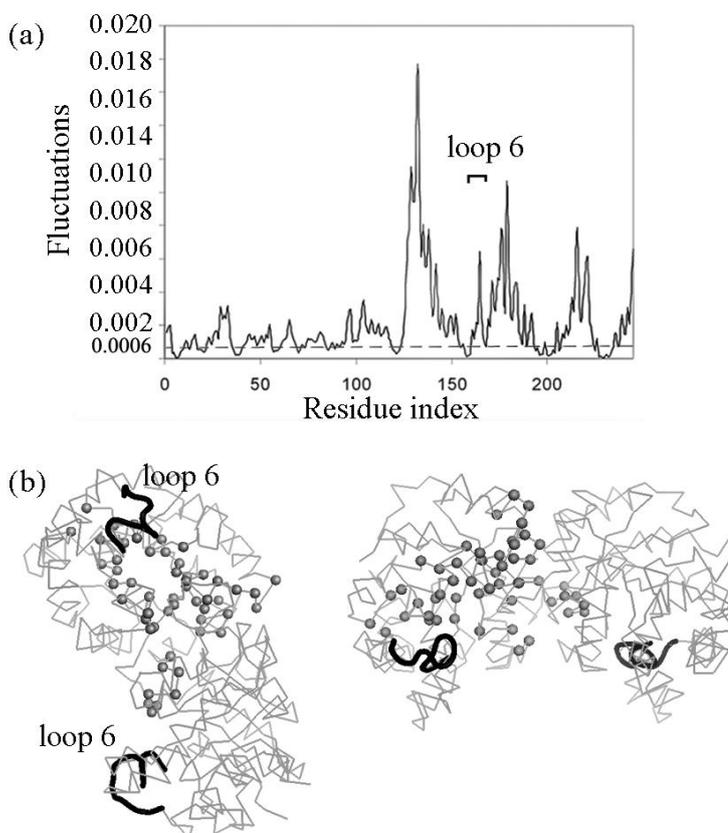


Figure 4.12. (a) Mean-square fluctuations at the slowest mode. Dashed lines show the threshold for hinge residues shown by spheres in (b) different perspectives

Domain motions (shown in Figure 4.9) indicate other inner residues that may act as hinges rather than relatively mobile ends of the loop 6. These sites are recognized with their small displacement, some of which are labeled in Figure 4.11. The mean-square fluctuations of residues in the slowest mode are shown in Figure 4.12 (a) for one of the identical monomers (the area under the curve is normalized to 1). The residues with small fluctuations (below 0.0006) are shown in Figure 4.12 (b) as spheres on one monomer from different perspectives. In the most favorable conformation of the structure, relatively immobile hinge-like residues are located in the middle of the TIM barrel and at the subunit interface linking both monomers' active sites. Large deformations occur at the outer sites of the TIM barrel, leading to loop opening and closure.

#### 4.4.5. Importance of Retaining Low-Resolution Regions

In previous mixed cg calculations (Kurkcuoglu *et al.*, 2004), the fluctuation dynamics of the high resolution regions have been captured successfully even when the remaining elements of the structure are kept at lower resolutions. In this study, a “high-resolution-only” model is considered, where only the atomistic region of the asymmetric mixed cg model is retained and the low-resolution region is left out. The first mode deformations of the mixed model with the corresponding second mode of the high-resolution-only model are compared for all available heavy atoms. Although the flexible loop makes a similar opening/ closing movement in both models as it is on the surface of the structure, the overall collective motions of the systems are significantly different (see Table 4.6). This points out the importance of maintaining the shape of the entire structure, even if cg at high levels.

Table 4.6. Correlation coefficients with the atom-based elastic network model taken as the reference

Correlation coefficients	Asymmetric mixed cg versus atom-based model	High-resolution region only versus atom-based model
First four cumulative modes	0.95	0.25
B-factors	0.99	0.40

When the shape of the enzyme is altered, atom-based ANM calculations result in low correlation coefficients for atomic fluctuations in the slowest modes. However, correlation coefficients for B-factors are relatively higher; the reason behind this result is that slow motions concern the shape of the structure, whereas B-factors are related with the local packing. In contrast, the correlation values for the asymmetric mixed cg structure are extremely satisfactory, being at least 0.95, as given in Table 4.6.

#### 4.5. Effect of the Crystal Packing on Collective Dynamics of TIM

When various crystal structures of TIM (bound and unbound forms) are superimposed, the only apparent deformation is observed at loop 6. For this reason, computational studies focus only on the active site dynamics while constraining the rest of the structure (Wade *et al.*, 1993; Guallar *et al.*, 2004). However, all cg models discussed here point out the importance of domain motions on local fluctuations in active site of TIM. Specifically first and fourth modes, driven by hinge motions, strongly overlap with the displacement vectors obtained from open-closed x-ray structures of TIM used in this study.

The reason behind the restricted motion of the enzyme may be the effect of crystal contacts. To address this issue, an artificial crystal structure containing 17 TIM molecules using the crystal structure 8TIM.pdb is constructed, as shown in Figure 4.13 (a) and (b) from different perspectives. Here, the enzyme at the center (shown in black in (a) and by dashed box in (b)) has the maximum number of neighbors. The structure is modeled at single-node-per-residue level, and displacement vectors are obtained at low frequency spectrum. Motions of individual enzymes in the ‘crystal’ are compared to dynamics of single TIM coarse-grained with residue-based ANM. Overlap values change within a range of 0.27 and 0.39 with the center enzyme having the lowest value, indicating dense packing around a structure like in a crystal may drastically inhibit mechanistic globular motions.

On the other hand, loop 6’s opening/ closing motion in a dense packing is investigated. Surprisingly, even the magnitude of motions is small, displacement directions of loop residues strongly overlap with the findings of single TIM modeled with residue-based ANM (~0.90), and with displacements between x-ray open-closed structures (0.60-

0.73). In summary, limited globular motions and flexibility of loop 6 observed from crystal structures may be due to effect of dense packing in a closely ordered medium.

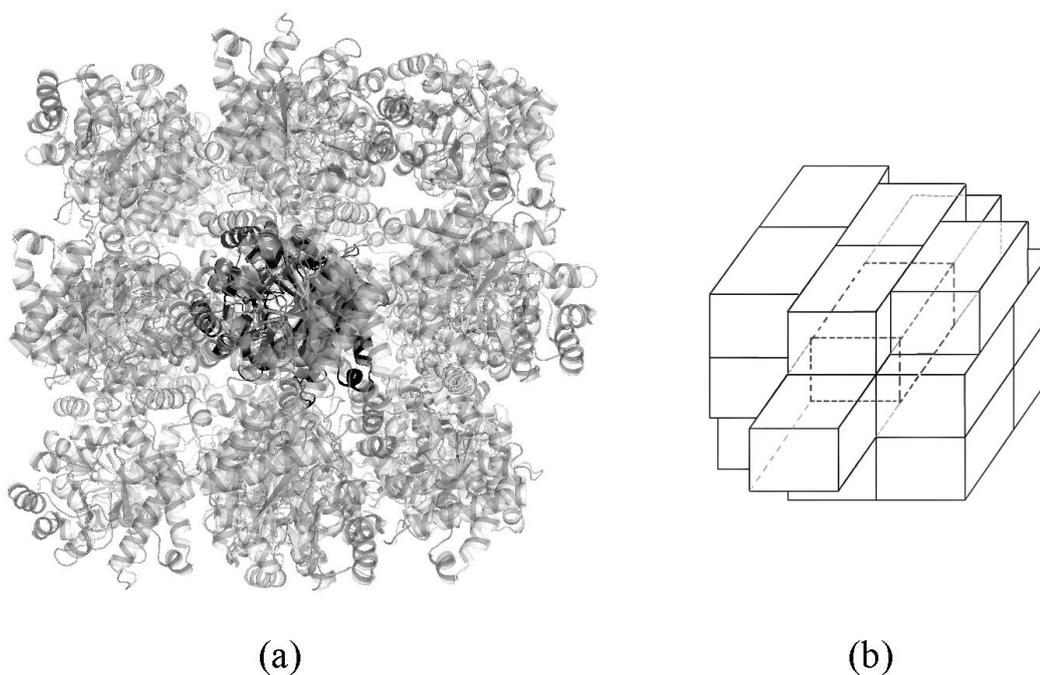


Figure 4.13. TIM shown in (a) black and (b) dashed in the generated crystal packing

In this chapter, two different procedures are discussed to determine the adjustable parameters of the mixed cg model for TIM, which are the cutoffs and force constants for the high- and low-resolution regions. TIM structure is also studied with atom-based ANM and classical NMA to validate the results obtained from mixed cg model. The collective dynamics of the enzyme observed from atom-based models, and experimental observations are successfully reproduced with the mixed cg model by both procedures. In addition, mixed resolution model reduces the computational time significantly compared to the full-atom approach. Specifically, the asymmetric mixed cg model for TIM is six times faster than the atom-based model.

One drawback in mixed cg procedure I is the calculation of the force constant for the high-resolution region. Here, this parameter is estimated by comparing experimental and theoretical B-factors calculated over  $3N-6$  normal modes, which requires the

diagonalization of the Hessian matrix by singular value decomposition. However, cutoff distances and predicted force constants are somewhat arbitrary, and cannot be used for another protein or DNA/RNA complex. Additionally, the diagonalization step is time consuming for a system like ribosome with ~10000 residues. On the other hand, in the procedure II, the use of a single cutoff for all interactions enables calculating necessary forces to link the nodes and modeling supramolecular assemblages with mixed cg method. Once this approach is validated by successfully exploring the dynamics of TIM, it will be applied to mixed cg the ribosome structure in the following chapter.

Finally, the most significant conformational change in the crystal structures of the enzyme is the opening/ closing motion of the flexible loop (loop 6) over the catalytic site. However, experimental evidences from few TIM structures, and results from the mixed cg model and recent MD simulations of 60 ns (Cansu and Doruker, *in press*) indicate that collective motions of the dimeric enzyme may have a role on the catalytic activity. Therefore, TIM is modeled in a generated crystal packing. The results indicate that the enzyme remains rigid due to crystal contacts, but still maintains the opening/closing motion of loop 6. This finding implies that proteins are flexible structures and their motions may be hindered due to crystal packing, so they should be investigated with molecular modeling techniques in detail to understand their function and mechanism.

## 5. DYNAMICS OF BACTERIAL RIBOSOME

### 5.1. Structure and Function of Bacterial Ribosome

Ribosome is a very large protein-RNA complex of size 2.5 MDa. It is a ‘molecular machine’ that synthesizes proteins based on the genetic information encoded in the mRNA in the cell. Ribosome is the drug target in bacterial infections (Hermann, 2005; Sutcliffe, 2005), thus it has a particular importance in antibiotic design and development.

The complex 70S composed of two subunits, namely 30S (blue) and 50S (gray) is displayed in Figure 5.1 where functionally important regions subject to the thesis are explicitly shown. Each subunit has a unique role in the protein synthesis. The mRNA carrying the genetic code from nucleus wraps around the neck of the 30S subunit. This region forms part of the three tRNA binding sites, called A- (aminoacyl), P- (peptidyl) and E- (exit). During translation process, helical structure of the mRNA strand is disrupted by ribosomal proteins S3 and S4 (Takyar *et al.*, 2005), while protein S5 forms a solid basis for the mRNA for correct frame reading (Kirthi *et al.*, 2006). Translation of genetic code occurs by the selection of the tRNAs carrying amino acids in decoding center A1492 and A1493, located on top region of helix 44 (h44) in 30S subunit (Figure 5.1). Cognate tRNA-amino acid assisted by the elongation factor (EF) Tu, enters the ribosome to bind at the A-site. During elongation step, the peptidyl transferase center residing in the 50S subunit catalyzes the peptide bond formation between the two amino acids attached to the A- and P-tRNAs (Figure 5.1). This center is formed of rRNA responsible for catalysis, so the ribosome is also referred as a ribozyme (Cech, 2000). After peptide-bond formation, the nascent polypeptide grows by one amino acid pushing into an exit tunnel formed mainly of rRNAs (Frank and Spahn, 2006).

After the peptide bond formation is completed, the deacylated tRNA remains bound at the P-site with the polypeptide remaining attached to the A-tRNA. Translocation of tRNAs is required to vacate the A-site for an incoming tRNA-amino acid. In this process, A- and P-tRNAs with mRNA attached to them are moved forward by one codon, placing tRNAs to P- and E-sites respectively. Translocation is catalyzed by EF-G. GTP complex

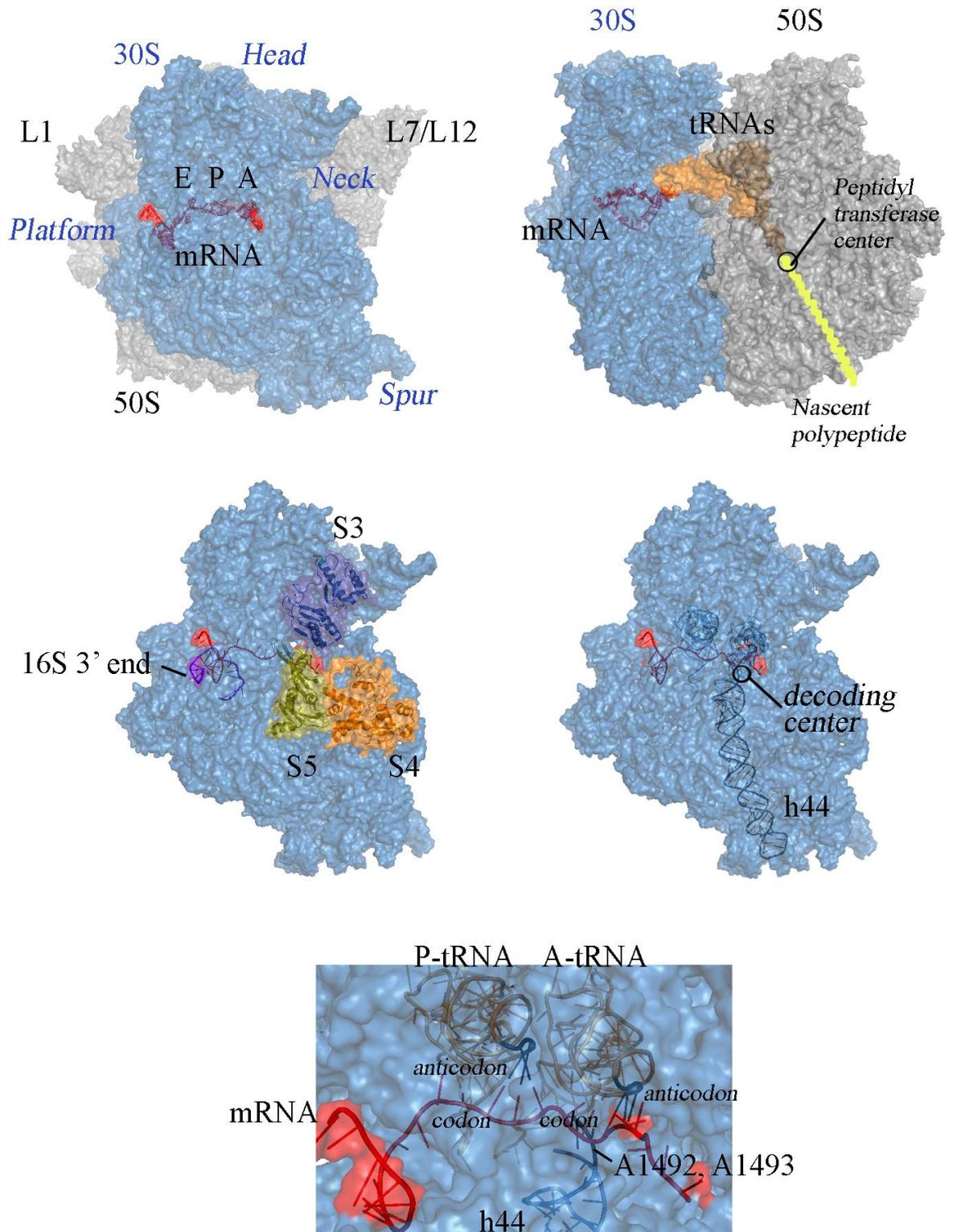


Figure 5.1. Bacterial ribosome 70S and functionally important regions

binding to the L7/L12 stalk base on 50S subunit (Joseph, 2003), but the exact mechanism is not known. In recent *in vitro* studies, it has been stated that EF-G in cell is likely to be in GDP form and binds ribosome in this structure. And ribosome acts as the missing guanine nucleotide exchange factor for the EF (Zavialov *et al.*, 2005). According to the hybrid-states model suggested almost two decades ago (Moazed and Noller, 1989), first the regions of the A- and P-tRNAs interacting with the large subunit move to the P- and E-sites while their anticodons stay bound to the mRNA in the 30S subunit. This transition state is called A/P and P/E for two tRNAs. Then, the anticodon stems in the 30S subunit move to the P- and E-sites. At the end of translocation the deacylated tRNA is located at the E-site and the polypeptide carrying the tRNA at the P-site (Joseph, 2003). These intermediate positions have also structural evidence in literature (Agrawal *et al.*, 1999a; Valle *et al.*, 2003).

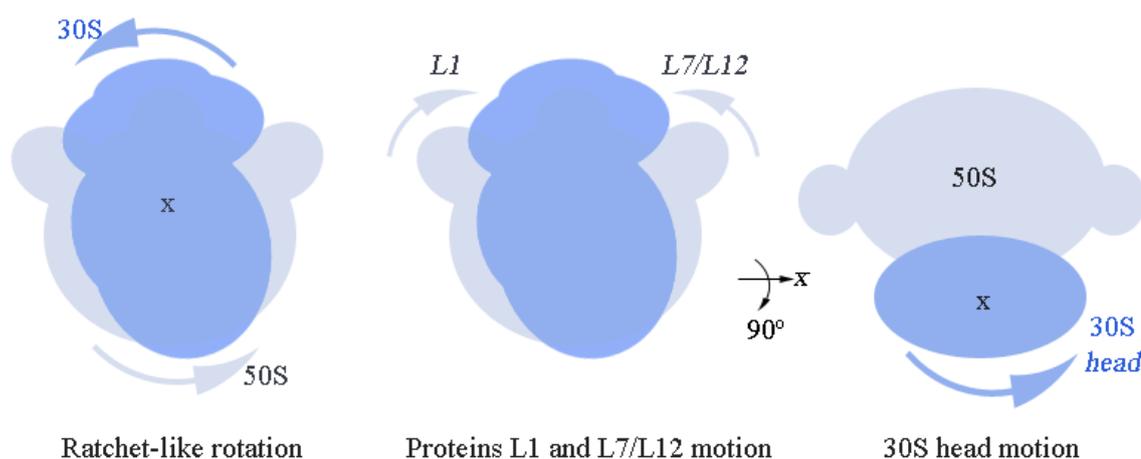


Figure 5.2. Schematic representation of experimentally observed large motions of subunits 30S (blue) and 50S (gray)

During protein synthesis, some regions of the ribosome undergo large conformational changes, such as 30S head (Agrawal *et al.*, 1999b), h44 on subunit 30S (VanLoock *et al.*, 2000), and proteins L1 and L7/L12 on 50S (Valle *et al.*, 2003; Agrawal *et al.*, 2001) having important roles for the translocation of the tRNAs and mRNA (Figure 5.1). In Figure 5.2, schematic representations of experimentally observed large motions such as ratchet-like motion of subunits, 30S head rotation, and mobility of proteins L1 and L7/L12 are displayed. However, the conformational dynamics of the ribosome complex have not

been extensively investigated by computational approaches. The collective motions of the ribosome and its complexes have been explored with normal modes using a cg force field (Tama *et al.*, 2003), and an elastic network model, namely the anisotropic network model (ANM) (Wang *et al.*, 2004), a coarse-grained molecular dynamics (MD) up to 500 ns (Trylska *et al.*, 2005), and full-atom targeted MD simulations in explicit water (Sanbonmatsu and Tung, 2006).

In this thesis, the globular rearrangements of the bacterial ribosome are explored with the anisotropic network model (ANM) (Atilgan *et al.*, 2001) and mixed coarse-graining (cg) method incorporating atomic detail to ‘interesting’ regions (Kurkcuoglu *et al.*, 2005; 2006). First, various model systems based on 70S complex crystal structures are analyzed with residue-based model in order to understand the details of conformational changes. Coarse-grained (cg) models include 70S and mRNA, but tRNAs, EF-Tu, and ribosomal proteins are selectively excluded to consider their effects on the ribosomal dynamics. The various 70S complex cg models used for the analysis are discussed in the next section.

Collective dynamics of bacterial ribosome are analyzed thoroughly to demonstrate the validity of the residue-based and mixed cg model by comparing results with experimental observations. Another aim is to explore the mechanical motions of ribosome at atomic detail, difficult to attain with full-atom approaches, and to report local and collective motions which may have role in protein synthesis. Recent computational studies have concentrated on the general dynamics of the mobile L1 and L7/L12 stalks, tRNAs and the ratchet-like rotations of the subunits (Tama *et al.*, 2003; Wang *et al.*, 2004). Here the focus is on (1) the dynamic domains and displacement directions of the large subunit stalks, (2) the conformational changes at the entry and exit channels of the mRNA on the ribosome in the 30S subunit, (3) the dynamics of tRNAs and (4) conformational rearrangements in ribosomal exit tunnel, with residue-based approach. After a comprehensive analysis of collective dynamics of the cg complex, high-resolution regions are selected from functionally important sites and investigated in atomistic detail with the mixed cg model.

## 5.2. Residue-Based Models of Ribosome

Ribosome structure 70S of *Thermus Thermophilus* in its complex with the A-, P-, E-tRNAs and mRNA, using PDB structures 1JGO and 1GIY (Yusupova *et al.*, 2001; Yusupov *et al.*, 2001) is studied with normal mode analysis based on elastic networks as an extension of earlier work (Wang *et al.*, 2004). In addition, the crystal structure of EF-Tu in its complex with tRNA (1MJ1 (Stark *et al.*, 2002)), originally fit to the 1GIY structure, is used to explore the ribosome dynamics. The low-resolution crystal structure of the bacterial ribosome at 5.5 Å contains only P and C<sup>α</sup> atoms of 30S and 50S subunits, together with, all heavy atoms of the tRNAs and mRNA.

One major advantage of cg elastic network models over full-atom approaches is the possibility of using low-resolution protein structures. Therefore, bacterial ribosome structure is modeled as one-node per interaction site, i.e. P for nucleotides and C<sup>α</sup> for amino acids from available crystal structures at 5.5 Å resolution. Similar to a previous study (Wang *et al.*, 2004), the cutoff distance to determine interacting C<sup>α</sup>-C<sup>α</sup> residue pairs is taken as 15Å, whereas it is 24Å for C<sup>α</sup>-P and P-P pairs in order to have a sufficient number of interactions to constrain the structure (Doruker *et al.*, 2002a). There is no specificity to the interactions, so the harmonic force, i.e. spring constant linking all these nodes is uniform throughout the elastic network.

To observe the effect of some functional members of the complex using the elastic network model, various structures of the bacterial ribosome have been constructed for the purpose of simulations. The model systems are 70S-mRNA complexes (i) with A-, P-, E-tRNAs, called 70S; (ii) with P-, E-tRNAs, called 70S\_no\_A; (iii) with A-, P-tRNAs, called 70S\_no\_E; (iv) without tRNAs, called 70S\_no\_A\_P\_E; (v) with A-, P-, E-tRNAs and EF-Tu, called 70S+EF-Tu and (vi) without ribosomal proteins, called 70S\_no\_proteins. The details and structures of models are given in Figure 5.3 and Table 5.1, respectively. In Figure 5.3, rRNAs in the structure are shown by blue (30S) and gray (50S), ribosomal proteins in dark colors, tRNAs in orange, mRNA in red and EF-Tu in magenta. These model structures are all generated from the same PDB files 1JGO-1GIY and 1MJ1 containing three tRNAs, a 27 nucleotide long mRNA and the EF-Tu.tRNA, respectively.

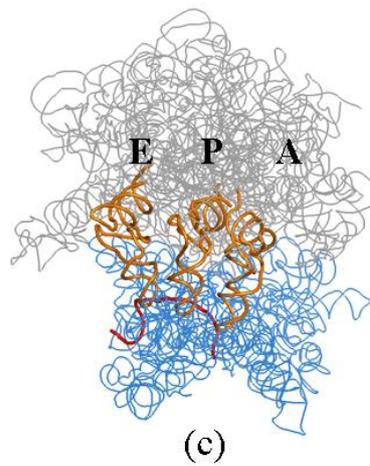
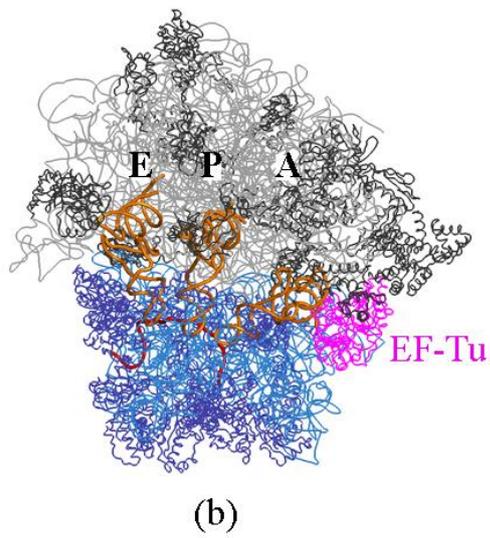
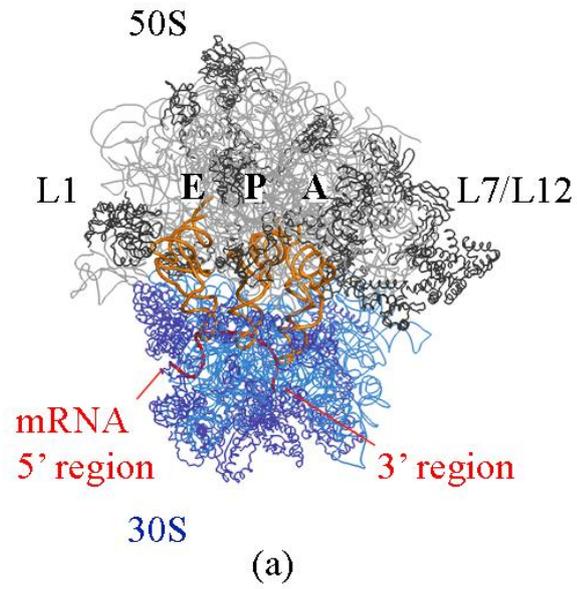


Figure 5.3. Ribosome model (a) 70S, (b) 70S+*EF-Tu*, and (c) 70S\_*no\_proteins*

There may be some distortions originating from deleted parts, but they will not be accounted for in these calculations due to the robustness of the cg model.

Table 5.1. Various ribosome models used in the coarse-grained model

<b>Model name</b>	<b>Model Details</b>	<b>PDB</b>	<b>Number of Nodes</b>	<b>Figure</b>
70S	70S with A-, P-, E-tRNAs and mRNA	1JGO, 1GIY	9808	5.3(a)
70S <sub>no_A</sub>	70S with P-, E-tRNAs and mRNA	1JGO, 1GIY	9732	5.3(a)
70S <sub>no_E</sub>	70S with A-, P-tRNAs and mRNA	1JGO, 1GIY	9734	5.3(a)
70S <sub>no_A_P_E</sub>	70S without A-, P-, E-tRNAs, with mRNA	1JGO, 1GIY	9582	5.3(a)
70S+ <i>EF-Tu</i>	70S with A-, P-, E-tRNAs, mRNA and EF-Tu	1JGO, 1GIY, 1MJ1	10213	5.3(b)
70S <sub>no_proteins</sub>	70S with A-, P-, E-tRNAs, mRNA, without ribosomal proteins	1JGO, 1GIY	4763	5.3(c)

### 5.3. Conformational Dynamics of Bacterial Ribosome

From various cg model systems displayed in Figure 5.3, it is clear that the presence of tRNAs, EF-Tu (or EF-G) and the ribosomal proteins do not seriously affect the overall shape of ribosome. It is well known that shape is the dominant effect in determining the functionally important slowest motions of macromolecules (Doruker and Jernigan, 2003; Lu and Ma 2005). Therefore, it can be predicted that all these ribosome structures are likely to have similar slowest motions. Figure 5.4 shows the displacement vectors of the nodes for three of the model systems (70S, 70S+*EF-Tu* and 70S<sub>no\_proteins</sub>) for some of their comparable slowest frequency modes. Two main global motions are observed from the collective dynamics of different models of the ribosome: the ratchet-like rotation of the

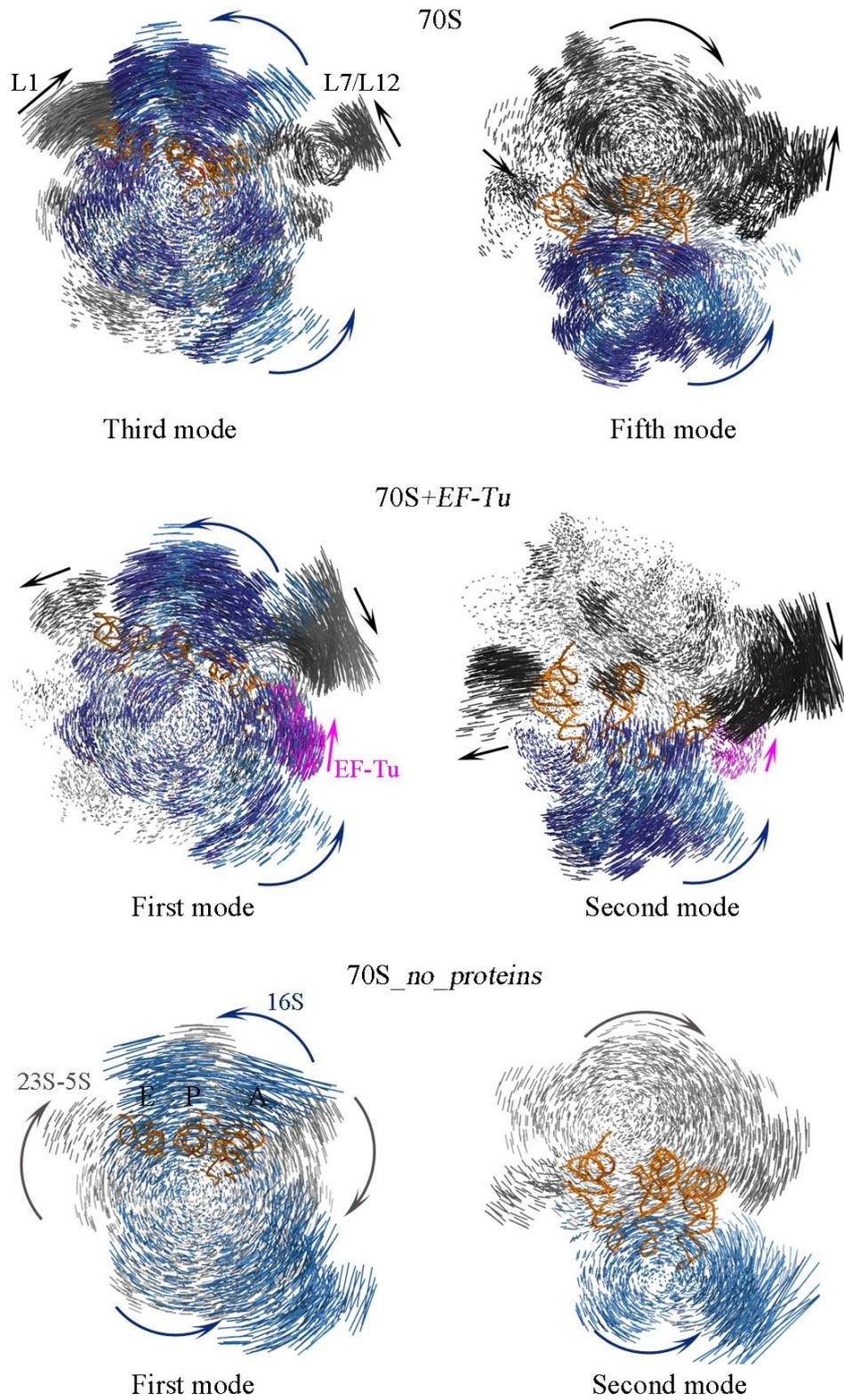


Figure 5.4. Comparable slow motions of three ribosome models, 70S, 70S+*EF-Tu* and 70S\_no\_proteins

two subunits -30S and 50S- as shown in the first column of Figure 5.4 and the mobility of the L1 and L7/L12 stalks displayed in first and second columns of Figure 5.4, which will be discussed next.

### 5.3.1. Ratchet-Like Rotation of Subunits

During protein synthesis, the largest conformational rearrangement observed is the ratchet-like rotation of two subunits upon binding of EF-G, revealed with cryo-electron microscopy (cryo-EM) (Frank and Agrawal, 2000). And, other experimental studies showed that small subunit head rotates during tRNA selection (Ogle *et al.*, 2002), translocation (Agrawal *et al.*, 1999b), and subunit association (Lata *et al.*, 1996), even in the absence of ligands (Schurwith *et al.*, 2005). The ratchet-like motion of the 30S and 50S subunits were previously reported for the 70S-tRNAs-mRNA complex from various computational studies (Wang and Jernigan, 2005; Tama *et al.*, 2003; Wang *et al.*, 2004; Trylska *et al.*, 2005). In this elastic network analysis, these motions are clear in the slowest modes for 70S, but also for all the different model systems used here, i.e. the 70S with/without tRNAs or ribosomal proteins, and even together with the EF-Tu. In Table 5.2, the overlap values between different models with 70S used as a reference are given. All values are averaged over the first ten vibrational normal modes excluding the six modes with zero eigenvalues accounting for translation and rotation of the whole molecule.

Table 5.2. Overlap values averaged over ten slowest modes in comparison with the reference structure 70S.

<b>Model</b>	<b>Overlap</b>
70S_no_A	0.99
70S_no_E	0.97
70S_no_A_P_E	0.92
70S+EF-Tu	0.96
70S_no_proteins	0.87

The ratchet motion of the subunits and the small subunit head motion are dominant even in the absence of ribosomal proteins constituting about half of the actual ribosome complex (Figure 5.4). The ribosome has a high level of molecular complexity, so it is

reasonable to argue that its evolutionary process may involve many multiple steps (Polacek and Mankin, 2005). Ribosome catalytic functions are mostly carried out by rRNA (i.e. peptidyl transferase center). Moreover, *in vitro* experiments show the viability and activity of rRNA by itself showing that the peptidyl transferase activity is preserved in 50S-like particles even in the absence of ribosomal proteins (Noller *et al.*, 1992; Khaitovich *et al.*, 1999). These may imply that rRNA was present first and the ribosomal proteins added later in evolution (Polacek and Mankin, 2005; Korostelev and Noller, 2007). The ANM simulations on the system *70S\_no\_protein* have shown that as the ribosome preserves its three-dimensional shape; it is capable of fulfilling its functional motions such as ratchet-like rotation of subunits, small subunit head rotation, and the L1-L7/L12 like stalks' fluctuations, overall with an overlap value of 0.87.

### 5.3.2. Motions of L1 and L7/L12 Stalks

Cryo-EM experiments have observed the mobility of the L1 and L7/L12 stalks (Agrawal *et al.*, 1999b, Valle *et al.* 2003) and reported the importance of these collective motions during translation by the ribosome. L1 stalk plays an important role during the translocation process helping E-tRNA to leave the ribosome complex (Valle *et al.*, 2003). L7/L12 stalk, at the base of which EF-Tu or EF-G binds to deliver tRNA-amino acid complex or to catalyze the translocation, respectively, undergoes the largest conformational change in the subunit (Agrawal *et al.*, 2001). *In vitro* experiments have shown that the lack of proteins L1 and L7/L12 from the ribosome complex decreases the rate of protein synthesis (Pettersson and Kurland, 1980; Subramanian and Dabbs, 1980). Additionally, previous computational studies on the ribosome complex with cg MD (Trylska *et al.*, 2005), cg NMA (Tama *et al.*, 2003), and cg ANM (Wang *et al.*, 2004; Wang and Jernigan, 2005) have also pointed out the large conformational rearrangements in the L1 and L7/L12 stalks, as well as the biologically functional importance of their anticorrelated movements. In this study, the dynamics of these two stalks and their neighbors are further analyzed, and different domains in particular regions that may have biological importance in the protein synthesis are investigated.

The displacement vectors of the L1 stalk, L7/L12 stalk, mRNA and E-tRNA are obtained with ANM, and their dot products are calculated to investigate their auto- and cross-correlations ( $C_{ij}$ ) in different slowest modes  $k=1, \dots, 3N-6$  or cumulatively by,

$$C_{ij} = \left( \sum_k \frac{\langle \Delta R_{ik} \cdot \Delta R_{jk} \rangle}{\lambda_k \sqrt{\Delta R_{ik}^2 \Delta R_{jk}^2}} \right) \frac{1}{\sum_k 1/\lambda_k} \quad (5.1)$$

Cross-correlations over cumulative ten slow modes (Figure 5.5 (a)) and alternative mode shapes in different modes (not shown) point out three domains on L7/L12: domain 1 (amino acids Met1-Thr57 (L7)) in green, domain 2 (Glu58 (L7)-Pro43 (L12)) in purple and domain 3 (Val44-Lys128 (L12)) shown in blue, in Figure 5.5 (e), (f). Domain 1 is distinguished with low mean-square fluctuations at slow modes compared to other domains of L7/L12 (not shown). As a result, this area may act as a hinge region between the dynamic domains 2 and 3, which are anti-correlated.

On the other hand, cross-correlations over cumulative ten slow modes indicate that L1 stalk moves as a single domain (Figure 5.5 (b)). And, displacement vector dot products for L1 and L7/L12 plotted for the cumulative ten modes indicate various dynamical behaviors of L1 stalk. Firstly, L1 and L7/L12 proteins are entirely anticorrelated in the second mode (not shown). But in the other modes, the anticorrelation is partial: only the regions comprising amino acids Thr40-Lys200 of the ribosomal protein L1 and domain 1 and 2 are anticorrelated (Figure 5.5 (c)-(e)). In Figure 5.5 (d), the correlation maps for the cumulative ten modes for mRNA and L7/L12 are displayed. Domains 1 and 3 together, despite a long distance, are anticorrelated (shown in blue) with A- and P-sites on mRNA especially at two slowest modes. In addition, the mRNA P-site motions may be coupled with all domains of L7/L12. These findings might point to a biologically important functional mechanism for these dynamical domains.

Another functional dynamical feature that is observed is a correlation between the L1 stalk and the E-tRNA, which are known to be closely interacting (Valle *et al.*, 2003) and presumably related to removal of the E-site tRNA from the ribosome. In the slowest mode, they are positively correlated (not shown) but an analysis of the cumulative first ten modes

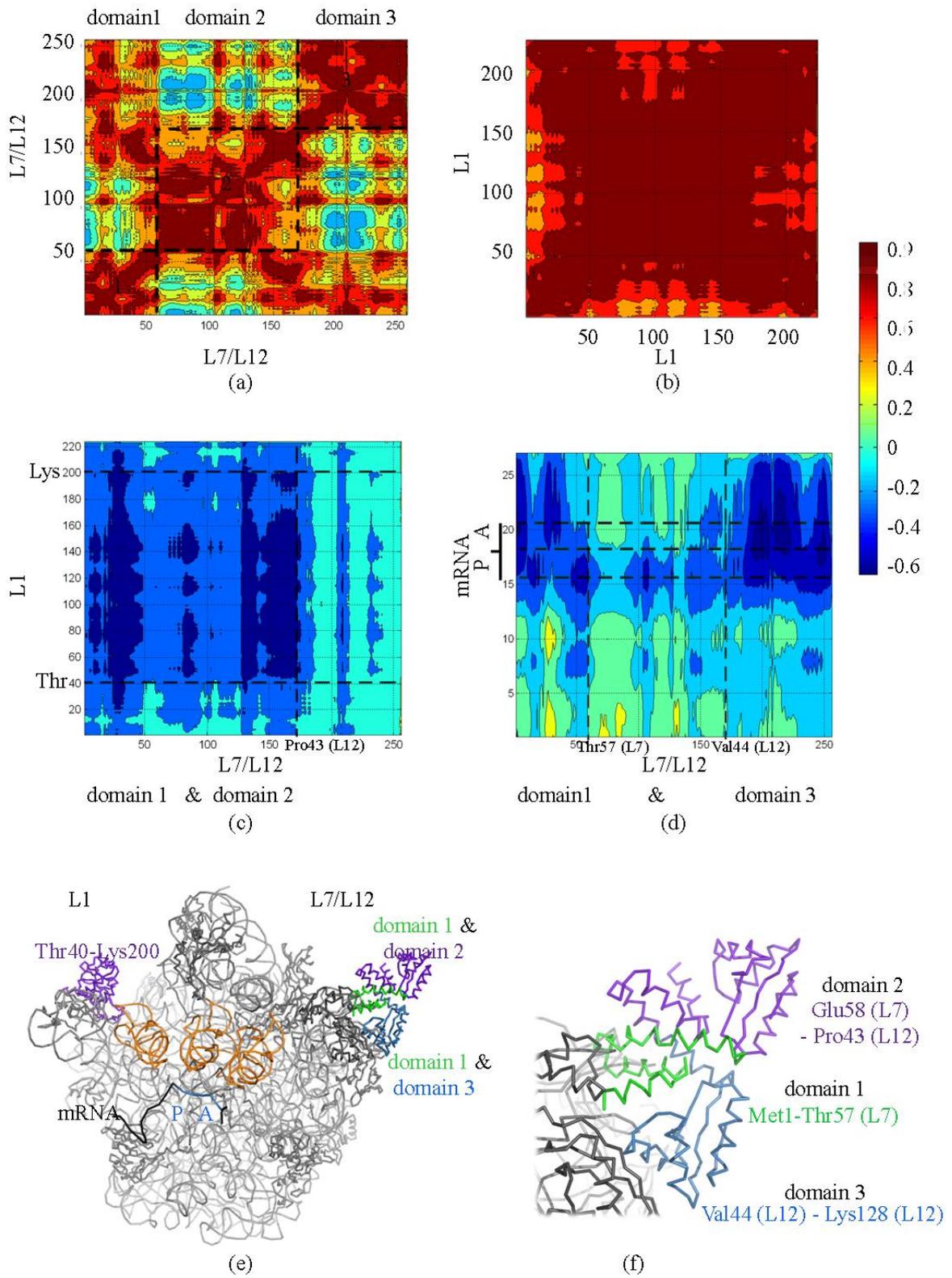


Figure 5.5. (a)-(d) Correlation maps for L1, L7/L12 stalks, and mRNA. (e)-(f) Domains on L1 and L7/L12 stalks

reveals an anticorrelated domain on the E-tRNA comprising nucleotides A10-G15 and U20-C40 (Figure 5.6). This domain is buried in the 30S subunit, and the other parts of the E-tRNA closely interact with the opening/closing L1 protein, which may help the tRNA exit the complex as proposed in various references (Valle *et al.*, 2003; Wang *et al.*, 2004; Trylska *et al.*, 2005). Also, in the slowest modes the L1 stalk moves usually towards or occasionally perpendicular to the inter-subunit cavity, but consistently in a direction opposite to the 30S head motion driven by the ratchet-like rotation (See for example Figure 5.4, fifth mode of 70S). These findings provide computational evidence for the hypothesis that L1 conformational changes may be coupled to the tRNA translocation as previously proposed (Valle *et al.*, 2003).

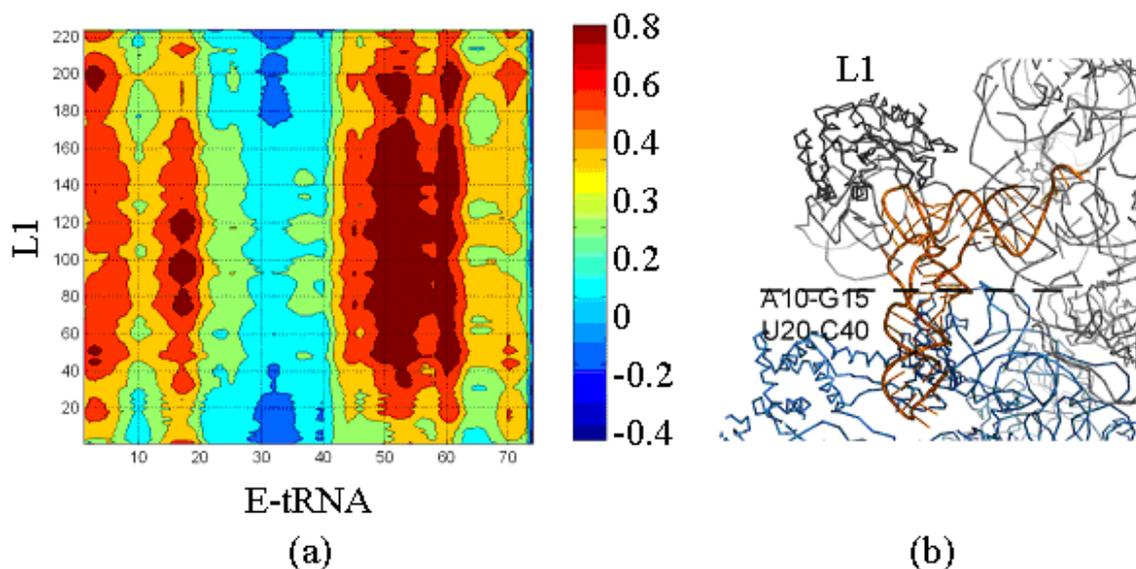


Figure 5.6. (a) Correlation map, and (b) domains for L1 stalk and E-tRNA

The root-mean-square fluctuations are plotted for both the 70S and the 70S<sub>no\_A\_P\_E</sub> structures (also for 70S<sub>no\_A</sub> and 70S<sub>no\_E</sub>). Fluctuation magnitudes, especially for the 30S head and spur, as well as for the L1-L7/L12 stalks, increase in the absence of tRNAs. This finding implies that binding of tRNAs stabilizes several regions in both 30S and 50S: the 30S head, especially ribosomal proteins S3, S4, S5, and their interactions with the mRNA, neck, spur, and the L1-L7/L12 stalks of 50S. In addition, binding of an elongation factor stabilizes 30S even further.

### 5.3.3. Motions of Helix 44

Helix 44 (h44) (C1399-G1503 in the *Thermus Thermophilus* 30S subunit), which carries the universally conserved decoding center A1492 and A1493 (see Figure 5.1), and Helix 69 (H69) (G1906-C1924 in the *Thermus Thermophilus* 50S subunit) are reported to be highly dynamic during the translocation and ribosome recycling (VanLoock *et al.*, 2000).

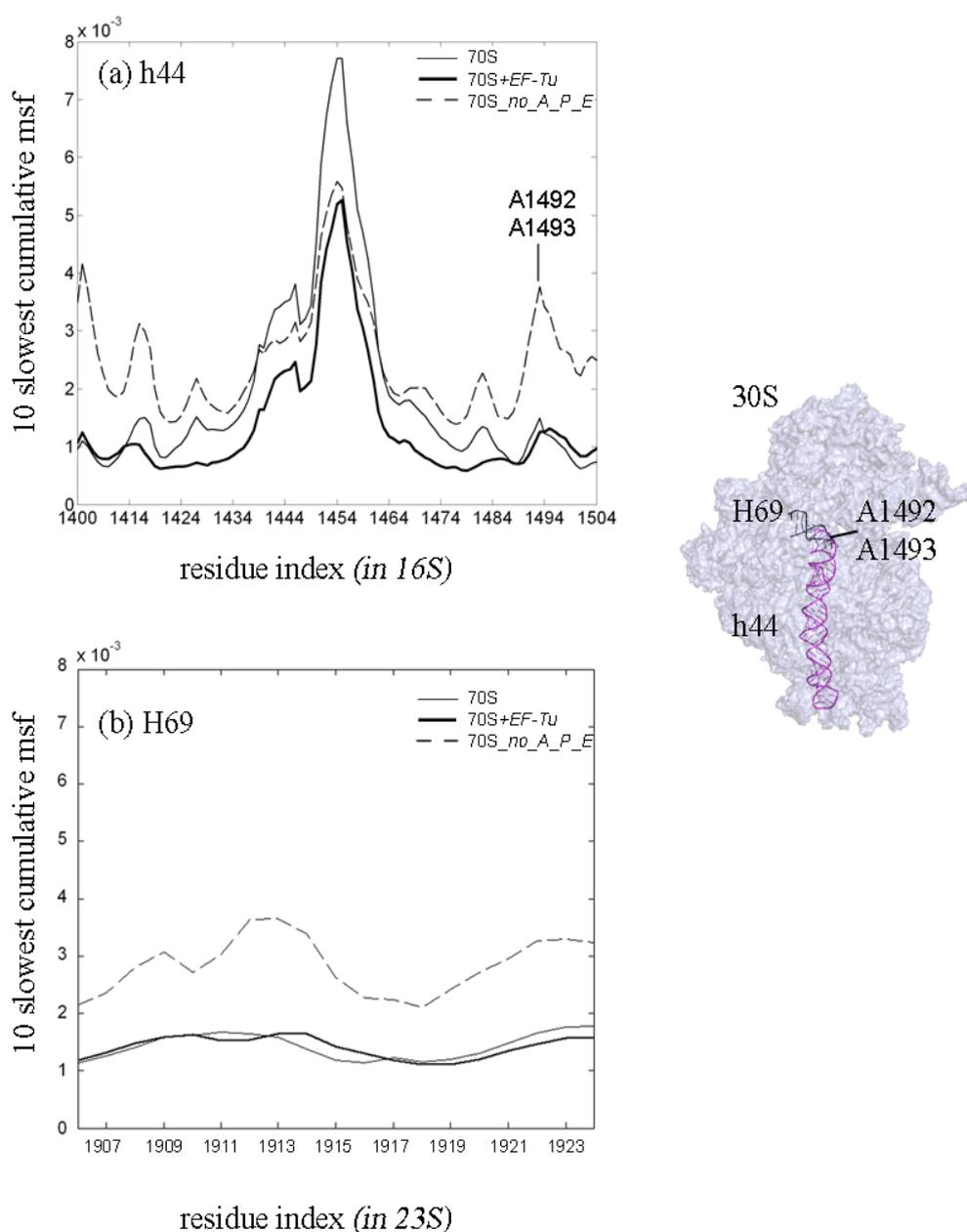


Figure 5.7. Mean-square fluctuations averaged over ten slowest modes for (a) h44, (b) H69

ANM simulations on the model *70S+EF-Tu* suggest that including the EF-Tu (or EF-G as the two molecules are quite similar) to the ribosome structure reduces somewhat the fluctuations of h44. On the other hand, absence of the tRNAs in binding sites may significantly increase the motions of h44 and H69 (Figure 5.7). In addition, the lower end of h44 (nucleotides 1440-1460) is highly mobile in the low frequency motions, contrary to the experimental observations by VanLoock *et al.* (2000). This difference in observations can be explained by the size of the cavity in which h44 is confined. Elastic network model allows a larger dynamics space around h44 due to ribosomal globular motions, in contrast to the observations by VanLoock *et al.* The lower part of h44 may still function as an ‘anchoring point’ as suggested by VanLoock *et al.*, yet not as a static anchor but rather a mobile anchor moving as part of its surroundings.

#### 5.3.4. mRNA and its Interactions with Ribosomal Proteins

A recent experimental study has suggested the helicase activity of ribosome to unwind helical structure of mRNA for correct frame reading. In this activity, ribosomal proteins S3 and S4 in the 30S, mediate residues to fulfill this task by acting like a ‘clamp’ around the mRNA as displayed in Figure 5.8 (Takyar *et al.*, 2005). In addition, the mRNA

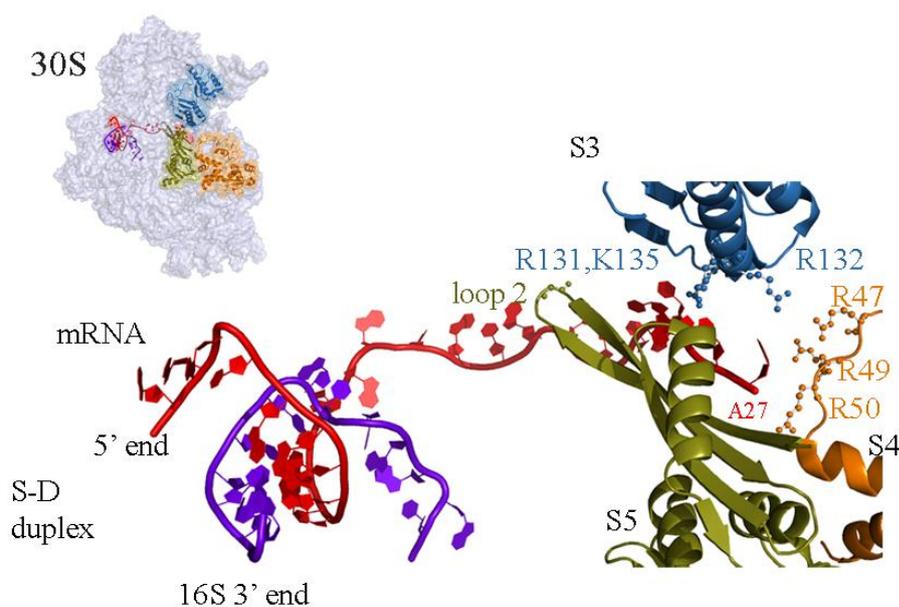


Figure 5.8. mRNA interacting with 16S 3' end together forming the Shine-Dalgarno (S-D) duplex and ribosomal proteins S3, S4, S5 on 30S subunit

helicase can function even in the absence of GTP and elongation factors (Takyar *et al.*, 2005) implying that the collective motions of the complex are critical for this activity.

Mean square fluctuations of the distance between  $i^{\text{th}}$  and  $j^{\text{th}}$  residues  $\langle \Delta R_{ij}^2 \rangle = \langle \Delta R_i^2 \rangle + \langle \Delta R_j^2 \rangle - 2\langle \Delta R_i \rangle \langle \Delta R_j \rangle$  averaged over the first ten modes are calculated for proteins S3, S4 and S5. The fluctuation of distances between parts of the ribosomal proteins S3 ( $i=1, \dots, 207$ ), S4 ( $i=1, \dots, 208$ ), S5 ( $i=1, \dots, 150$ ) and the closest nucleotide ( $j=A27$  in 1JGO) on the mRNA are shown in Figure 5.9. These fluctuations have not been rescaled to compare with experimental B-factors. The nodes at Arg131, Arg132, and Lys135 on S3, and also Arg47 and Arg50 on S4 (Figure 5.8) are known to have helicase activity (Takyar *et al.*, 2005). These residues are close to the mRNA and move in coordination with the mRNA for the ten slowest modes summed cumulatively (Figure 5.9). During this global motion, S3 also comes into close proximity with the mRNA strand in such a way that its side chains Glu161, Gln162 and Arg164, as well as that of Arg49 of S4, are near enough to mRNA to interact. These residues might be playing an important role in the helicase activity of the ribosome.

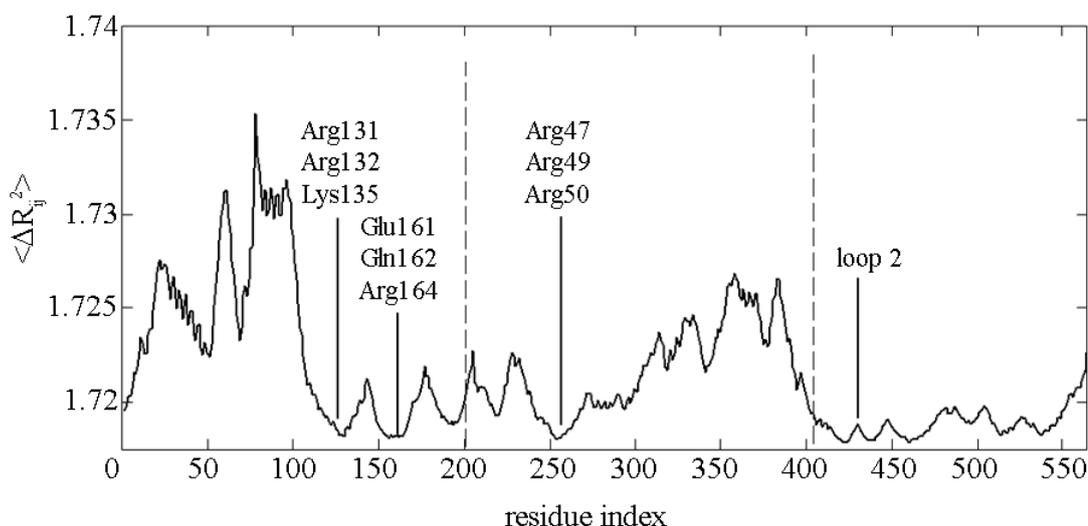


Figure 5.9. Mean-square distance fluctuations averaged over ten slowest modes for the nucleotide A27 on mRNA and residues of S3, S4, S5 proteins

The ribosomal protein S5 in the small subunit (Figure 5.8) is close to the helicase active site and forms a ring around the entrance channel. However, this protein was proposed not to play any role in unwinding the mRNA strand (Takyar *et al.*, 2005). But, recent ambiguous ribosome mutations on S5 have revealed this protein's functional importance in translation fidelity, and specifically, loop 2's role in RNA binding (Kirthi *et al.*, 2006). Mean-square distance fluctuations between S5 residues and the nucleotide A27 (Figure 5.9) reveal that the entire S5 ribosomal protein remains closer to the mRNA than S3 and S4 do. This suggests that the tight positional alignment of S5 with mRNA is functionally important to orient the strand for correct frame reading at the A-site, confirming the previous work.

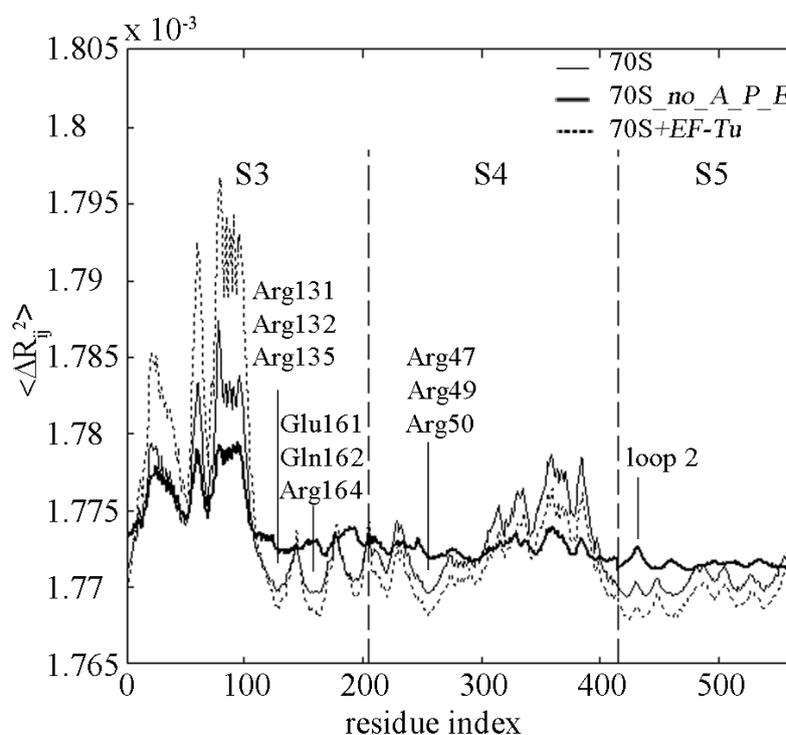


Figure 5.10. Mean-square distance fluctuations averaged over ten slowest modes for mRNA nucleotide A27 and residues of S3, S4 and S5 from three different model systems

The presence of bound tRNAs and EF-Tu affects the fluctuations of the small subunit head significantly. Figure 5.10 displays the mean-square distance fluctuations between ribosomal proteins S3, S4, S5 and the nucleotide A27, based on cumulative effect of the sum of the slowest ten modes for model systems 70S, 70S\_no\_A\_P\_E and 70S+EF-Tu.

The  $\langle \Delta R_{ij}^2 \rangle$  values averaged over residues of the specific region are  $6.05 \text{ \AA}^2$ ,  $1.72 \text{ \AA}^2$  and  $1.07 \text{ \AA}^2$  for *70S\_no\_A\_P\_E*, *70S* and *70S+EF-Tu* respectively. However, for comparison purposes, the area under the curve is normalized to unity, specifically for proteins S3, S4 and S5. In the absence of tRNAs, the positional stabilities of helicase active site residues on S3-S4 and residues on S5 are disturbed. On the other hand, the stability of this region is enhanced in the presence of EF-Tu (with the tRNAs present). This observation indicates that the binding of the elongation factor may help the helicase active site to seize the mRNA stronger and S5 to further stabilize the structure to assure correct codon reading at the A-site.

Motions of the mRNA and the 16S rRNA 3' end, which form the Shine-Dalgarno duplex with the 5' mRNA in the initiation phase, are analyzed with residue-based model. The mean-square fluctuations  $\langle \Delta R_i^2 \rangle$  over the ten slowest modes are plotted for this region in Figure 5.11 (a) and (b). There is no experimental B-factor information for 1JGO, so B-factor values from 2HGR.pdb, 30S *Thermus Thermophilus* crystal structure at  $4.5 \text{ \AA}$  (Yusupova *et al.*, 2006) are given in inserts of Figure 5.11. Both cumulative ten modes and B-factor plots reveal that the A- and P-site codons are less mobile sites of mRNA strand in the low frequency motions. In contrast, the 5' and 3' regions of the mRNA are more flexible, as previously observed by difference Fourier calculations (Yusupova *et al.*, 2006). The 3' end of the 16S rRNA has functional importance in the initiation complex to help the start codon be accommodated in the correct position (Yusupova *et al.*, 2006). And, this region shows higher mobility, possibly to facilitate the movement of the ribosome along mRNA after initiation (Uemura *et al.*, 2007, Korostelev *et al.*, 2007).

To further validate the results of normal mode analysis with residue-based model, the available crystal structures of the mRNAs (1JGO (blue, (Yusupova *et al.*, 2001)), 2B64 (orange, (Petry *et al.*, 2005)), 2J00 (green, (Selmer *et al.*, 2006)), 1YL4 (purple, (Jenner *et al.*, 2005)) and 2HGR (black, (Yusupova *et al.*, 2006)) are superimposed in Figure 5.12. Single stranded mRNA shows flexibility at both the 5' and 3' regions of mRNA by assuming a range of conformations except where it is bound to the P-site, and to a lesser extent at the A site; and recent conformational analyses also point to similar results (Jenner *et al.*, 2007). Moreover, the superimposition reveals that side chains of the helicase active

site residues are flexible and adjust their conformations in coordination with those of mRNA.

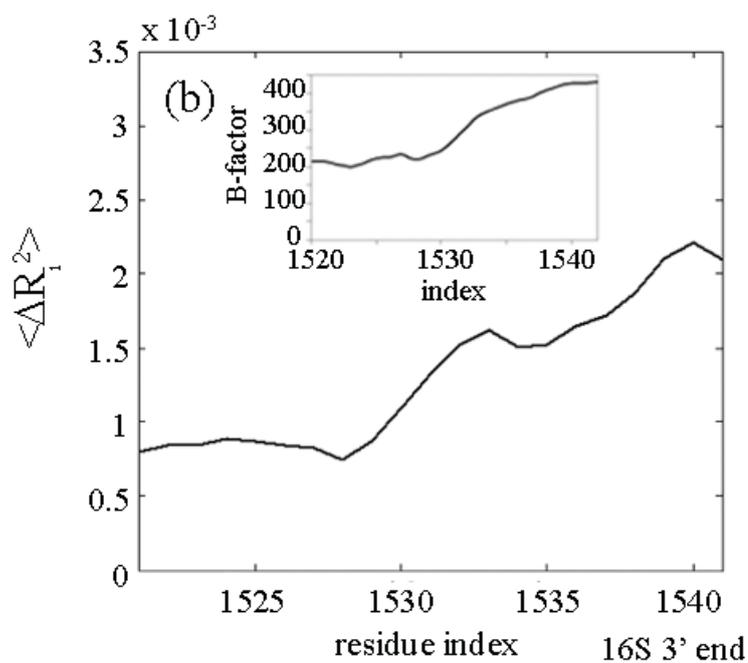
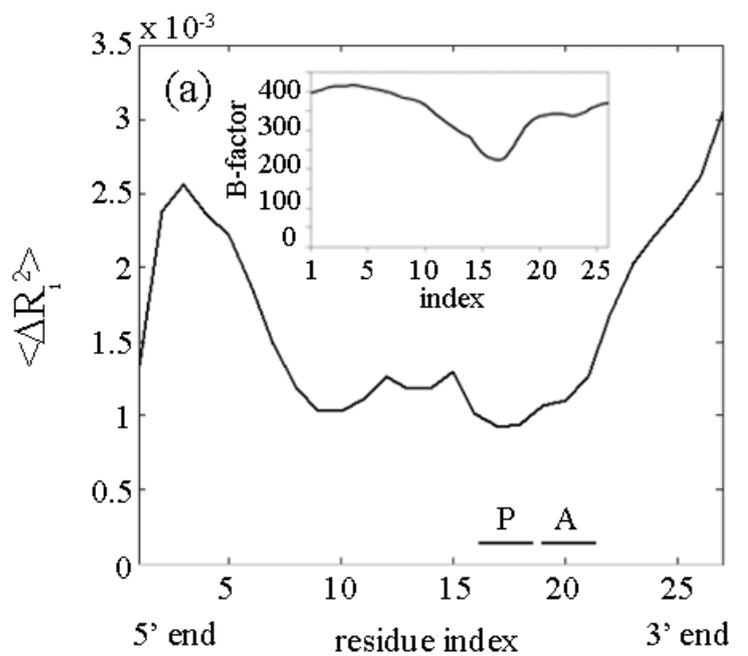


Figure 5.11. Mean-square fluctuation averaged over ten slowest modes for (a) mRNA, and (b) 16S 3' end. Inserts display experimental B-factors for these regions

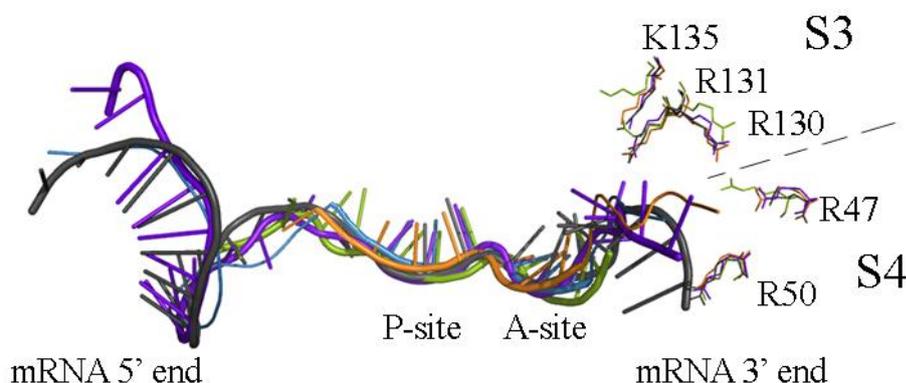


Figure 5.12. Superimposed crystal structures of mRNA with helicase active site residues of ribosomal proteins S3 and S4

### 5.3.5. Collective Motions of tRNAs

In Figure 5.13 (a)-(c), the displacement vectors for the first three modes are shown for tRNAs in ribosome model 70S. The black arrows indicate the alternative displacement directions of domains. In the first mode (Figure 5.13 (a)), all three tRNAs move coordinately together, but in the other modes, A- and P-tRNAs have various combinations of correlated motions. In the second mode (Figure 5.13 (b)), while the anticodon stem loops of A- and P-tRNAs seem to detach from the mRNA, their elbows tend to translocate to the hybrid-states A/P and P/E. However, as discussed above, E-tRNA translocates in coordination with the L1 stalk, seemingly independently from the motions of the other tRNAs (Figure 5.13 (c)).

Elbows interacting with regions of the 50S undergo large conformational changes; whereas the anticodon stem loops move in concert with the 30S head and neck. On the other hand, because the amino acid carrying ends of the A- and P-tRNAs are buried in the peptidyl transferase center at the core of the 70S complex; they are the least fluctuating parts of tRNAs (also observed in MD simulations (Trylska *et al.*, 2005)). Based on these observations, under large conformational changes like the 30S head movement or ratchet-like rotation of subunits, the elbows might be the regions translocating first, rather than the anticodon stem loops and aminoacyl ends of A- and P-tRNAs.

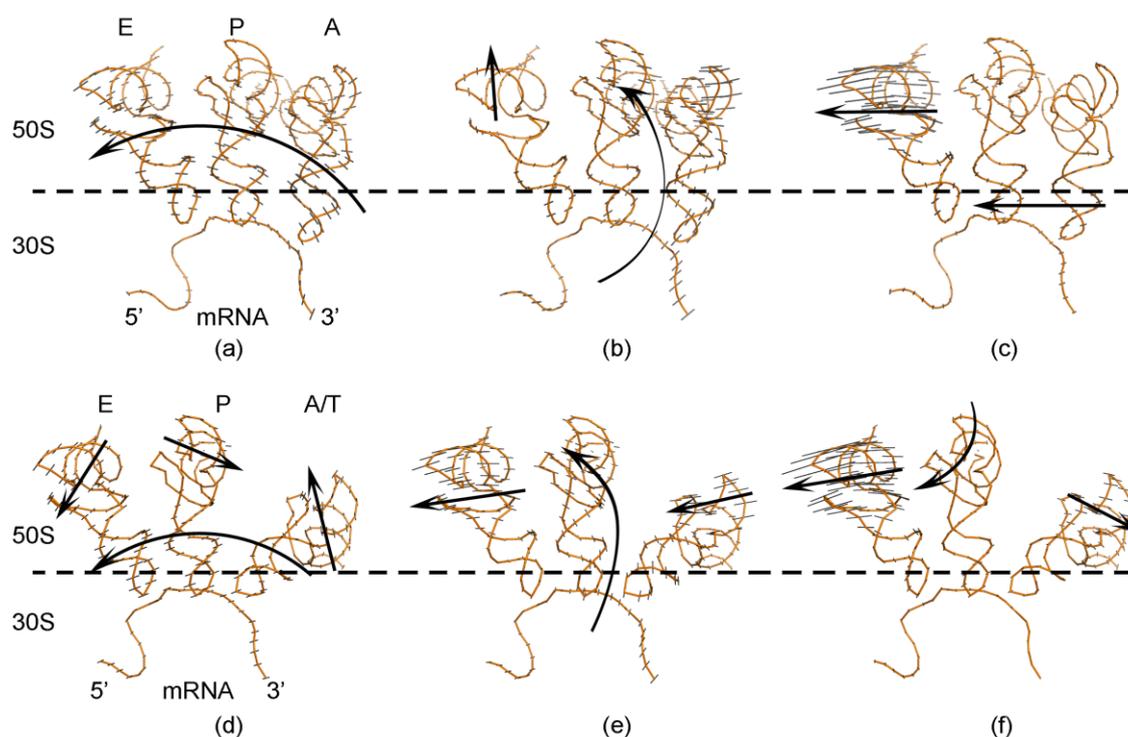


Figure 5.13. Displacement vectors of tRNAs at first three modes displayed for model (a)-(c) 70S and (d)-(f) 70S+*EF-Tu*

In Figure 5.13 (d)-(f), the displacement vectors for the first three modes are shown for the model 70S+*EF-Tu*. This model is important as it displays the changes in motions arising from the presence of bound *EF-Tu* with tRNA (and implicitly those of *EF-G* because *EF-Tu* and *Ef-G* structures are so similar). In the first mode (Figure 5.13 (d)), while the mRNA and anticodon stem loops of A/T- and P-tRNAs move towards the 5' end of the mRNA, their elbows approach one another. In the second and third modes (Figure 5.13 (e) and (f) respectively), considering the A/T-tRNA as the tip of *EF-G* (i.e. domain IV), P- and E-tRNAs seem to be co-translocating. These observations also point out the three major domains forming the *EF-Tu*.tRNA complex being the anticodon stem loop, the elbow, and the aminoacyl end with *EF-Tu*. It should be noted that previously the internal tRNA motions at all three sites are shown to be substantially similar, although different from the intrinsic motions of an independent tRNA (Wang and Jernigan, 2005).

These results show that the collective motions of the ribosome trigger the translocation of tRNAs and mRNA, supporting the experimental observation that the

translocation occurs at a slow rate even in the absence of the elongation factor G and GTP hydrolysis (Joseph, 2003). Finally, A- and P-tRNAs elbows move in correlation with the 50S and the anticodons oppositely with the 30S motions in agreement with the hybrid-states model for tRNA translocation.

### 5.3.6. A Mechanistic Model Proposed for Translocation

After combining the experimental data and the normal mode analysis for the elastic network models, mRNA translocation may comprise a possible sequence of events as follows: (i) During the elongation phase, the ribosome globular motions disrupt the mRNA helical structure due to ‘clamping’ motion by the small subunit proteins S3 and S4. Meanwhile stable S5 orients mRNA for correct frame reading at the A-site. (ii) Due to dense packing and strong interactions critical for peptide bond synthesis, the mRNA strand is particularly tightly held at the P-site. But after the head and ratchet-like rotation of 30S, the mRNA with bound anticodons of tRNAs is forced to translocate one codon. (iii) The S3 and S4 clamp seizes again the incoming mRNA strand to unwind it. (iv) After leaving P-site, the mRNA strand finds enough space to adjust its conformation so that E-tRNA can detach itself with the help of the L1 stalk. (v) Meanwhile, due to the 30S head motion, interactions between the 3’ end of the 16S rRNA and the 5’ region of the mRNA are already weakened following initiation. And the intrinsically mobile 3’ end of 16S acts like a hook to reel the folded mRNA into the inter-subunit region.

### 5.3.7. Dynamics of Ribosomal Exit Tunnel

Proteins are synthesized at the core of ribosome large subunit 50S at the peptidyl transferase center (PTC) mainly formed of rRNA. Despite several other gates within the structure (Gabashvili *et al.*, 2001), the nascent polypeptide chain seems to emerge from PTC through the ribosomal tunnel of  $\sim 15\text{\AA}$  diameter and  $\sim 100\text{\AA}$  long out of the complex (Figure 5.14) (Voss *et al.*, 2006). After many experimental evidences of its existence (Milligan and Unwin 1986, Yonath *et al.*, 1987; Beckmann *et al.*, 1997), the dynamics of the exit tunnel and the mechanism of nascent polypeptide secretion have gained importance in understanding protein synthesis (Gabashvili *et al.*, 2001; Agmon *et al.*, 2003; Berisio *et al.*, 2003; Woolhead *et al.*, 2004; Lu and Deutsch, 2005).

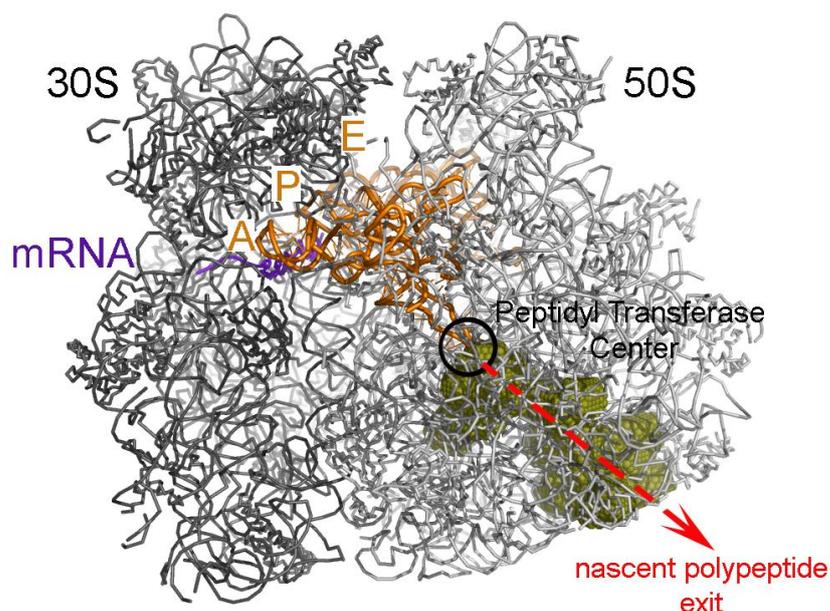


Figure 5.14. Ribosome complex and nascent polypeptide exit tunnel (in green)

Ribosomal tunnel wall is mainly formed of 23S rRNA and few extensions of several ribosomal proteins. This particular structure has led to the assumption of ‘non-sticky’ property which may allow the passage of any polypeptide (Nissen *et al.*, 2000). However, at the narrowest part of tunnel not far from PTC, extensions of ribosomal proteins L4 and L22 are located. These highly conserved regions function as a discriminating gate for specific nascent chains which can cause ribosomal stalling (Nakatogawa and Ito, 2002; Tenson and Ehrenberg, 2002). In addition, this gate is a dynamic region undergoing open (diameter  $\sim 16\text{-}20\text{\AA}$ ) and closed (diameter  $\sim 10\text{\AA}$ ) conformations exposed with cryo-EM of ligand bound/unbound, and L4-L22 mutated structures (Gabashvili *et al.*, 2001). Furthermore, other mutation studies on L4 and L22 conserved tips (Chittum and Champney, 1994), and the ‘swing motion’ of the  $\beta$ -hairpin of L22 to block the nascent polypeptide passage, imply the role of L4 and L22 in signaling and channeling the nascent polypeptide through the exit tunnel (Agmon *et al.*, 2003; Berisio *et al.*, 2003; Brodersen and Nissen, 2005).

Biological evidences (Woolhead *et al.*, 2004; Kosolapov *et al.*, 2004) and computational studies (Ziv *et al.*, 2005) support the existence of  $\alpha$ -helical structures in the ribosomal tunnel. A recent experimental study has reported possible secondary structure folding zones along the tunnel (Lu and Deutsch, 2005). However, a tertiary structure seems

not to fit in due to geometrical restrictions (Voss *et al.*, 2006). On the other hand, proteins at the tunnel exit, L22, L23, L24 and L29 are strategically located to interact with membrane proteins (Beckmann *et al.*, 2001; Mitra *et al.*, 2005), chaperone trigger factor (Kramer *et al.*, 2002; Ferbitz *et al.*, 2004) and signal recognition particle (Gu *et al.*, 2003).

Here, residue-based ANM results of the model structure 70S is used to investigate the harmonic dynamics of the ribosomal tunnel wall residues. The ribosomal tunnel volume is extracted from the low-resolution structure of large subunit using program  $^3V$  (Voss *et al.*, 2006) based on spherical probes searching the solvent accessible areas. The exit tunnel conformations are displayed via the cg residues with 10 Å distance to the tunnel wall (379 residues in total). Principal component analysis is used to place the y-axis onto the tunnel nascent polypeptide exit route, and to sort the residue index accordingly (Figure 5.15 (a)). The extracted tunnel, mainly formed of 23S rRNA with several ribosomal proteins' extensions, is in funnel shape with small secondary openings to solvent observed with cryo-EM (Gabashvili *et al.*, 2001), which are not discussed in this study.

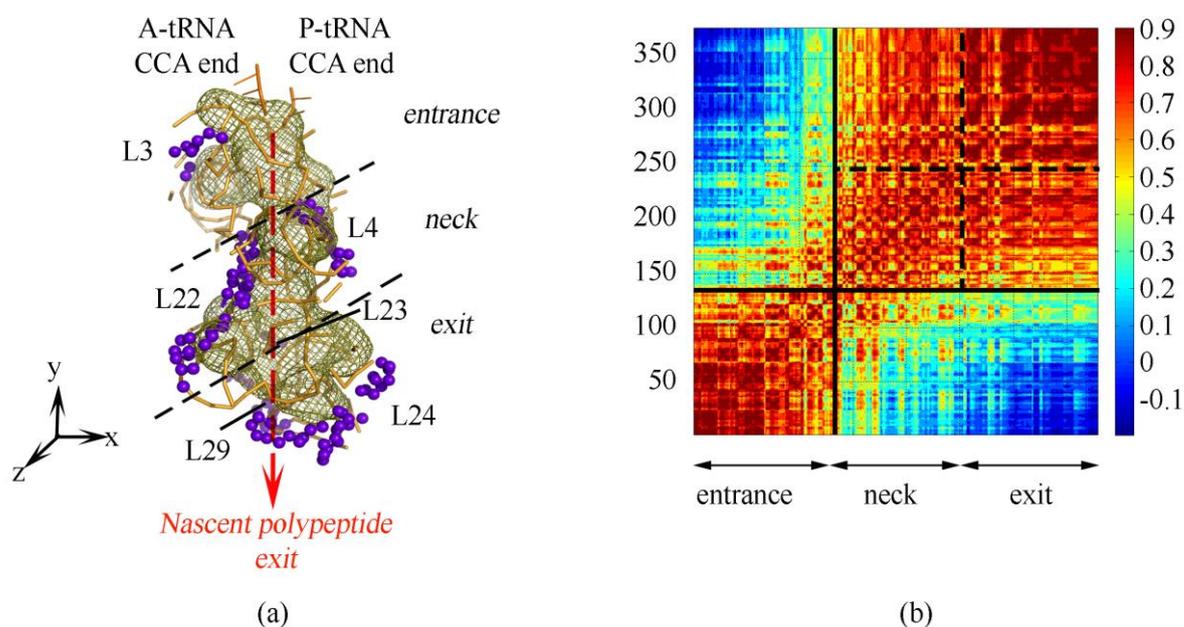


Figure 5.15. (a) Ribosomal tunnel wall with 23S rRNA and few ribosomal proteins. (b) Domains of tunnel shown on the cross-correlation map averaged over first five modes

5.3.7.1. Motions From Normal Modes. Embedded in the large subunit 50S, the ribosomal tunnel undergoes translational motions in slow normal modes. To observe any difference in directional vectors of tunnel residues, cross-correlation maps of tunnel residues are calculated over first twenty modes. There are two major zones distinguished from Figure 5.15 (b): (i) entrance including the PTC; (ii) neck comprising the narrowest region with extensions of proteins L4 and L22, and the broad region near the exit. The latter region can be also divided into two sub-domains as the neck and exit region, based on high values of correlation coefficients within residues 250-379 (i.e. exit zone).

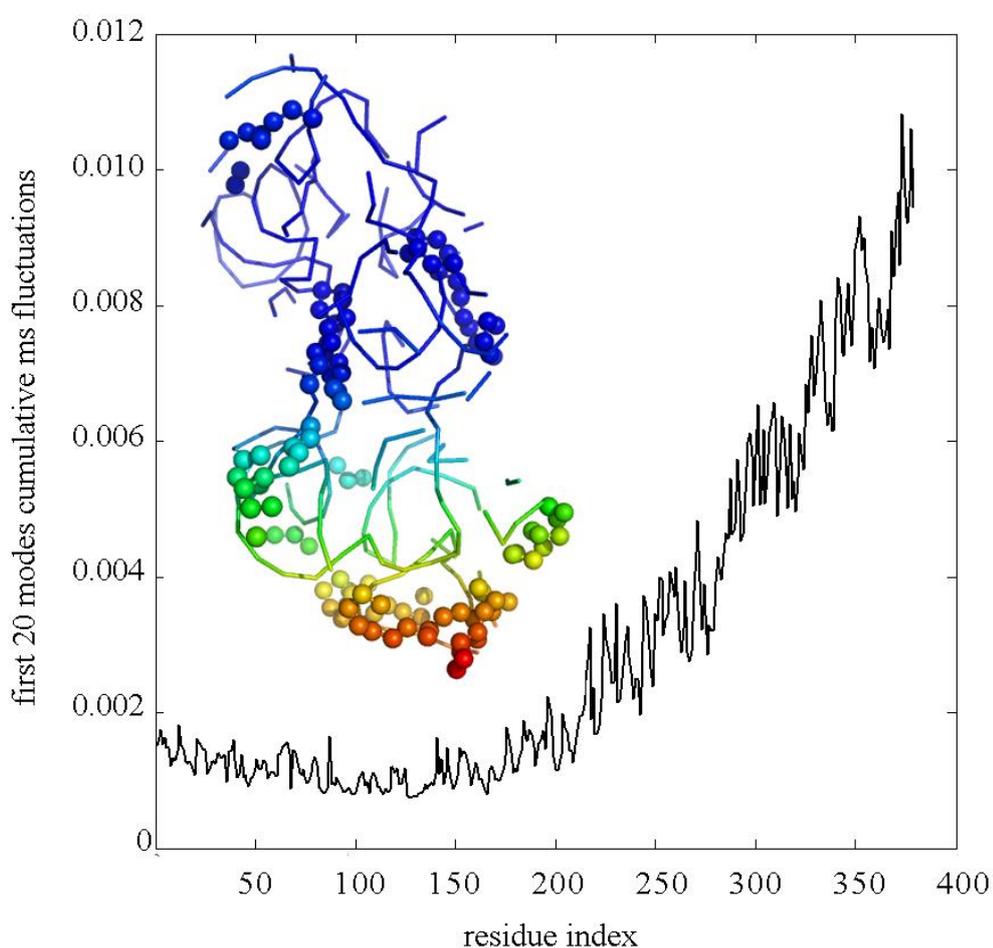


Figure 5.16. Mean-square fluctuations averaged over first twenty modes of exit tunnel, color coded with blue to red indicating increase in mobility

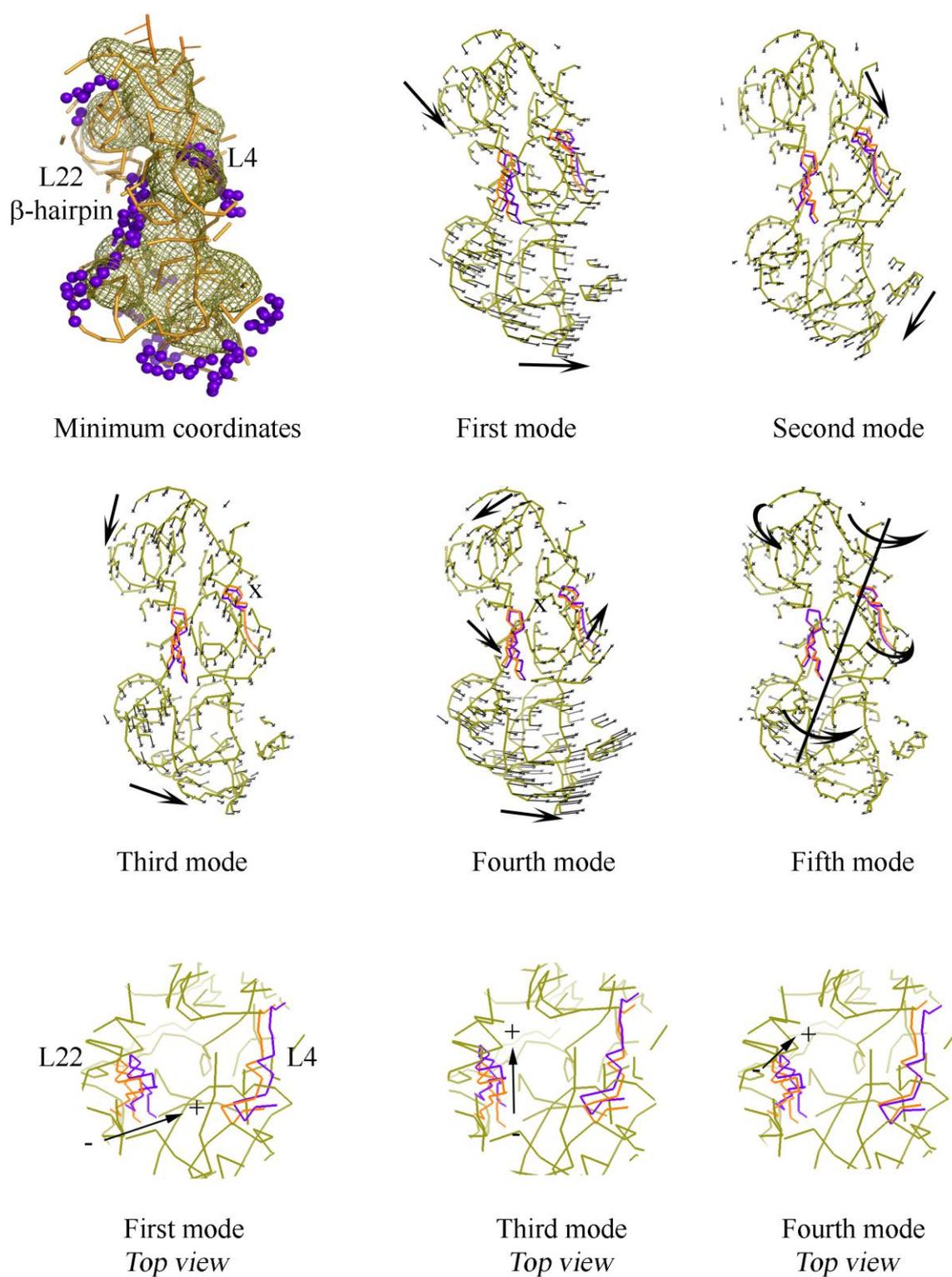


Figure 5.17. Displacement vectors of tunnel wall residues for first five modes with front and top views. Alternative conformations for L22 and L4 are shown in orange and purple

The exit domain exhibits high amplitude motions compared to the rest of the tunnel, observed from mean-square fluctuations averaged over slowest twenty modes (Figure 5.16). Collective motions at first five normal modes point that in general the entrance region moves downwards (assuming y-axis is along the tunnel), whereas the residues at the exit move perpendicular to the polypeptide propagation route. Motions of the neck region are relatively more mixed including twisting, tangential and radial motions (Figure 5.17). Another interesting and probably important finding is the location of a rotational axis usually around the narrowest part of the tunnel close to ribosomal proteins L4 and L22 (Figure 5.17, third, fourth and fifth modes, axis shown by an x or a bar).

The existence of three defined domains (i.e. entrance, neck and exit) is investigated over first twenty modes. For this reason, orientational correlations are calculated for both the entrance and exit domains. In order to have a clear picture of motions, correlation coefficients are classified as positive ( $>0.7$ ), negative ( $-0.7 <$ ) and not-correlated ( $-0.4 < \text{coefficient} < 0.4$ ). Figure 5.18 (a) and (b) show the distribution of the positively correlated residue pair fractions for entrance and exit zones, respectively. For example, in Figure 5.18 (a), in eight normal modes, 90 per cent of residues in the entrance domain fluctuate in the same direction. Both entrance and exit region moves as one domain, and low fraction (0.4) of positive correlation usually corresponds to a rotation around an axis (for example fourth and fifth modes in Figure 5.17). Surprisingly, three domains defined as entrance, neck and exit are similar to three proposed folding zones of the ribosomal tunnel which may have role on nascent polypeptide folding (Lu and Deutsch, 2005).

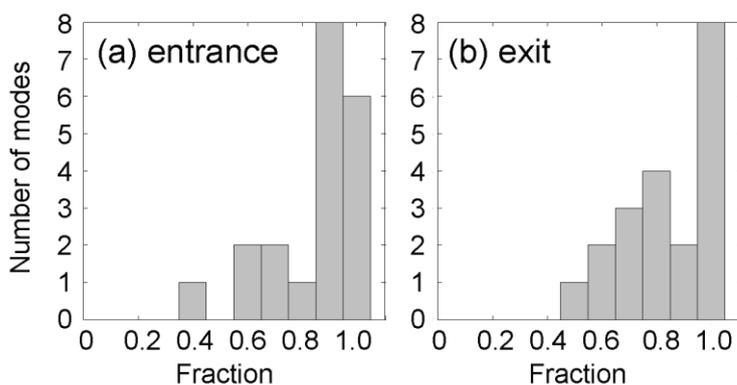


Figure 5.18. Distribution of positively correlated residue pairs in (a) entrance and (b) exit domains of ribosomal tunnel from normal modes

The downwards motions of the entrance-neck area in various modes may direct the nascent polypeptide through the tunnel. And, ribosomal proteins L4 and L22 are located at the narrowest region of tunnel and are known to function as a gate, as explained above. At first three normal modes, the tip of L4 (residues Phe61-Ala68) and L22 (residues Leu86-Ala93) move together (with a correlation of  $\sim 0.90$ ) whereas at the fourth and fifth modes their motion is not correlated (values in the range of  $-0.55$ - $0.10$  according to residue location) as shown in Figure 5.17. Over first twenty normal modes, tips of L4 and L22 show various conformational rearrangements. They are positively correlated (32 per cent of cases), negatively correlated (24 per cent of cases) and show no correlation (44 per cent of cases). Most of the non-correlated motions (13 per cent of cases) consist of perpendicular motion ( $-0.1 < \text{correlation} < 0.1$ ). Furthermore, the ‘swing motion’ of L22  $\beta$ -hairpin could be observed in the first, third and fourth modes (Figure 5.17, bottom panel), where one alternative direction is indicated with an arrow. As they are at the periphery of 50S, the residues at the tunnel exit undergo relatively large harmonic motions at low frequencies. This may have an important role in polypeptide secretion through the tunnel to the cytoplasm or in protein-protein interactions with membranes and trigger factor.

5.3.7.2. Motions from Combined Normal Modes. In reality, biological system dynamics should be described as a superposition of harmonic motions with different phase angles  $\varepsilon$  given by  $Q_i = A_i \cos(\omega_k t + \varepsilon_k)$  for a residue  $i$ , with vibrational frequency  $\omega$ , and time  $t$  at a specific mode  $k$ . To have a robust idea about the dynamics of the ribosomal tunnel area, first five non-zero normal modes are linearly combined with different coefficients ( $a_i = -1, 0, 1; i=1, \dots, 5$ ) multiplied with the eigenvector sets  $\mathbf{w}_i$  of five modes. This combination mimics in a crude way the superposition of displacement vectors of normal modes at their maximum ( $-1, 1$ ) or equilibrium ( $0$ ) values,

$$\mathbf{v}_k = \sum_{i=1}^5 a_i \mathbf{w}_i \quad (5.1)$$

This procedure leads to  $k=3^5$  eigenvector sets but in fact to  $(3^5-1)/2=121$  distinct conformations, since  $(a_{1-5}=0, 0, 0, 0, 0)$  gives the native state and combinations of  $(a_{1-5}=1, 1, 1, 1, 1)$  and  $(a_{1-5}=-1, -1, -1, -1, -1)$  represent identical conformations.

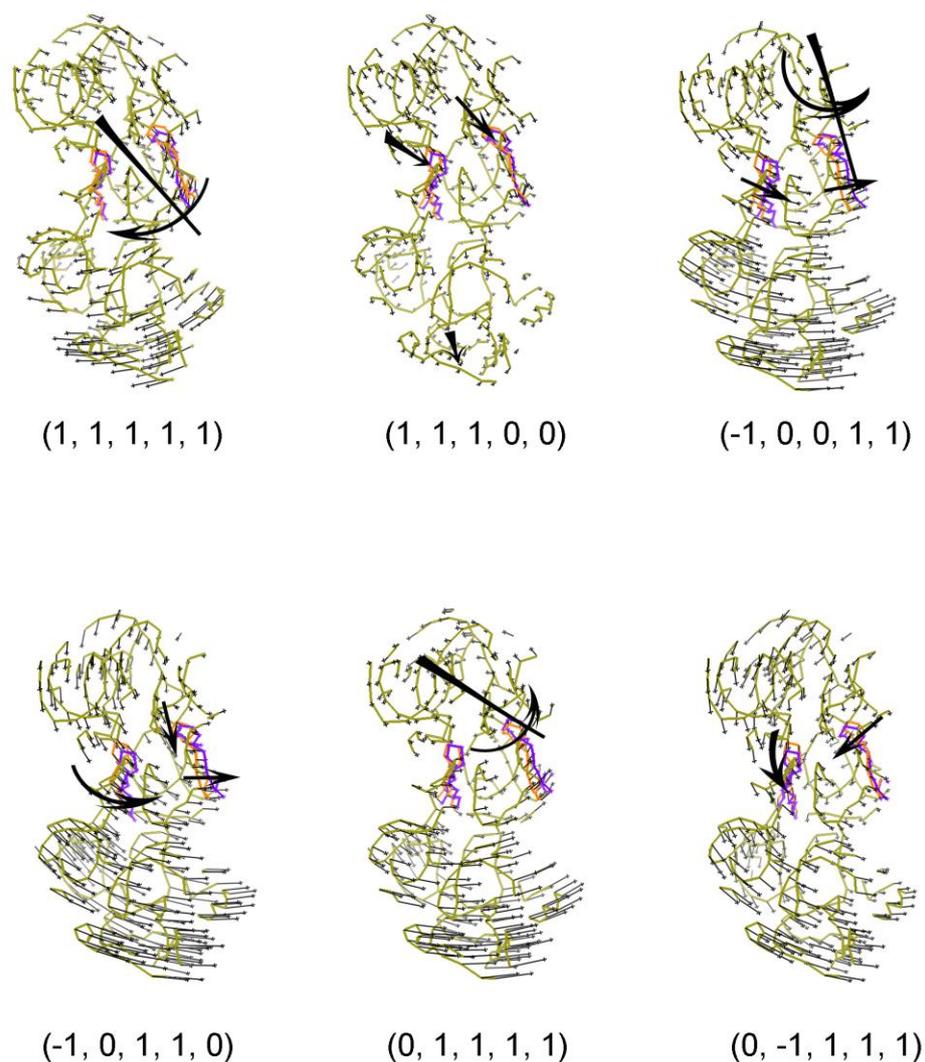


Figure 5.19. Displacement vectors of tunnel wall residues calculated from different combinations of first five normal modes

In Figure 5.19, randomly selected conformations are shown and coefficients of the eigenvector sets are given in brackets from first to fifth mode. For example, in a combined-mode  $(a_1=1, a_2=1, a_3=1, a_4=1, a_5=1)$  where all five normal modes contribute positively, a rotation axis passes through the narrowest region of the ribosomal tunnel around which all residues rotate. In the combination  $(a_{1-5}=1, 1, 1, 0, 0)$ , where only first three modes are used,  $\beta$ -hairpin of protein L22 oscillates in tangential motion while the tip of L4 move in y direction, such as directing a nascent polypeptide through the tunnel. Another interesting motion is observed in combined-mode  $(a_{1-5}= -1, 0, 1, 1, 0)$  where L22  $\beta$ -hairpin displaces

in tangential direction with respect to tunnel wall and L4 oscillates perpendicular to L22 as to open the gate.

Over 121 conformations, correlations for tips of L4 and L22 proteins are calculated. They are positively and negatively correlated in 48 per cent and 15 per cent of cases respectively, and show no correlation in 37 per cent, where 9 per cent consists of perpendicular motion as displayed by combined-modes (1, 1, 1, 0, 0) and (-1, 0, 1, 1, 0) in Figure 5.19. This perpendicular motion seems to have importance in the opening/closing motion of the tunnel entrance observed in various cryo-EM data (Gabashvili *et al.*, 2001), that could channel the nascent polypeptide.

The entrance and exit domains are further investigated in all generated 121 conformations. In Figure 5.20, the distribution of the positively correlated residue pair fractions over combined modes for entrance (Figure 5.20 (a)) and exit zones (Figure 5.20 (b)) are shown. As in normal modes, these zones move as separate domains. And small fractions of positive correlation in one domain correspond to rotational movement around an axis such as in combined-modes (1, 1, 1, 1, 1) and (0, 1, 1, 1, 1) displayed in Figure 5.19.

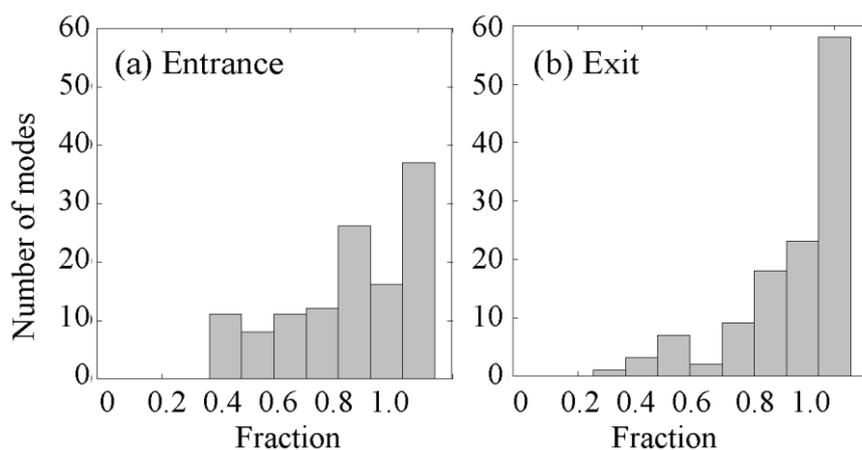


Figure 5.20. Distribution of positively correlated residue pairs in (a) entrance and (b) exit domains of ribosomal tunnel from combined-modes

#### 5.4. Mixed Coarse-Grained Model of Ribosome

A high-resolution structure of ribosome complex is chosen to study in this part of the work. The PDB file names of the complex are 2HGP (30S subunit, mRNA and three tRNA) and 2HGQ (50S subunit) (Yusupova *et al.*, 2006), respectively. Extended protein L9 of ribosome is excluded from the model and the missing protein L7/L12 (1DD4.pdb (Wahl *et al.*, 2000)), forming an important domain in collective dynamics (Tama *et al.*, 2003; Wang *et al.*, 2004; Trylska *et al.*, 2005), is added to the crystal structure.

Mixed cg model calculating interaction forces from atom-atom contacts (i.e. procedure II) is shown to be successful to obtain the collective motions of TIM and local fluctuations of its active site as discussed in previous chapter. Here, the universally conserved nucleotides A1492, A1493 on subunit 30S, codons of mRNA and anticodons of tRNAs at A- and P-sites constitute the high-resolution with atomistic details (see Figure 5.1), whereas the other parts are modeled at low-resolution with cg level  $n$  and  $n/5$ . The ribosome complex is modeled with residue-based model with procedure II of mixed cg technique for comparison. Table 5.3 lists the system size for residue-based and mixed cg ribosome structures.

Table 5.3. Sizes of mixed coarse-grained ribosome models

Model	Procedure	Number of nodes		
		Total	High-resolution	Low-resolution
Residue-based cg	II	11001	-	11001
Mixed cg	II	11285	298	10987
Mixed cg ( $n/5$ )	II	2496	298	2198

Table 5.4 lists the overlap values between different ribosome models, calculated only for  $C^\alpha$  of amino acids and P of nucleotides. Mixed cg of 70S with procedure I is not performed due to size limit. Slow modes from residue-based uniform cg of the two approaches yielded in a high overlap of 0.88 for the ribosome complex. On the other hand, the overlap value between the residue-based and mixed cg with procedure II is almost unity as expected. Moreover, low-resolution of the mixed resolution is modeled at one-

node-per-five residue level and still a satisfactory result is obtained (overlap=0.80) despite a poorer description of the assemblage (see Table 5.3).

Table 5.4. Overlap values calculated for various models of 70S

Models		Overlap
Residue-based model	Residue-based model (procedure II)	0.88
Residue-based model	Mixed cg (procedure I) ( $n$ )	-
Residue-based model (procedure II)	Mixed cg (procedure II) ( $n$ )	0.99
Mixed cg (procedure I)	Mixed cg (procedure II) ( $n$ )	-
Mixed cg (procedure II)	Mixed cg (procedure II) ( $n/5$ )	0.80

As a result, collective dynamics of a large protein complex can be successfully and efficiently described with the procedure II of cg.

#### 5.4.1. Dynamics of Decoding Center with Atomistic Detail

Cumulative first ten mode mean-square fluctuations are calculated for different levels of mixed cg. Figure 5.21 displays the mean-square fluctuations for small subunit 30S with tRNAs and mRNA plotted for mixed cg with cg level  $n$  and  $n/5$ . For the high-resolution region, i.e. A1492, A1493 on 16S rRNA, anti stem loop (ASL) of A-tRNA and P-tRNA, and corresponding codons of mRNA, the mean-square fluctuations are described in detail (Figure 5.21 (b)) when compared to residue-based model mean-square fluctuations with interested residues indicated (Figure 5.21 (c)). Although the fluctuations seem to be lower in magnitude in mixed cg with low-resolution (lr) at cg level  $n/5$  when compared to  $n/1$  (one-node-per-residue), the trend is the same, showing the efficiency of the model.

The displacement vectors of ribosome complex are displayed in Figure 5.22 for one alternative conformation of slowest third mode obtained with the mixed cg model at cg level  $n/5$ . Schematic representation of ribosome complex with experimentally observed motions indicated with arrows is also given for comparison at the right. Experimental observations during the protein synthesis such as the ratchet-like rotation of the subunits (Frank and Agrawal, 2000) indicated by arrows, the mobility and counter motion of L1 and L7/L12 stalks (Valle *et al.*, 2003; Agrawal *et al.*, 2001; Bocharov *et al.*, 2004) are observed

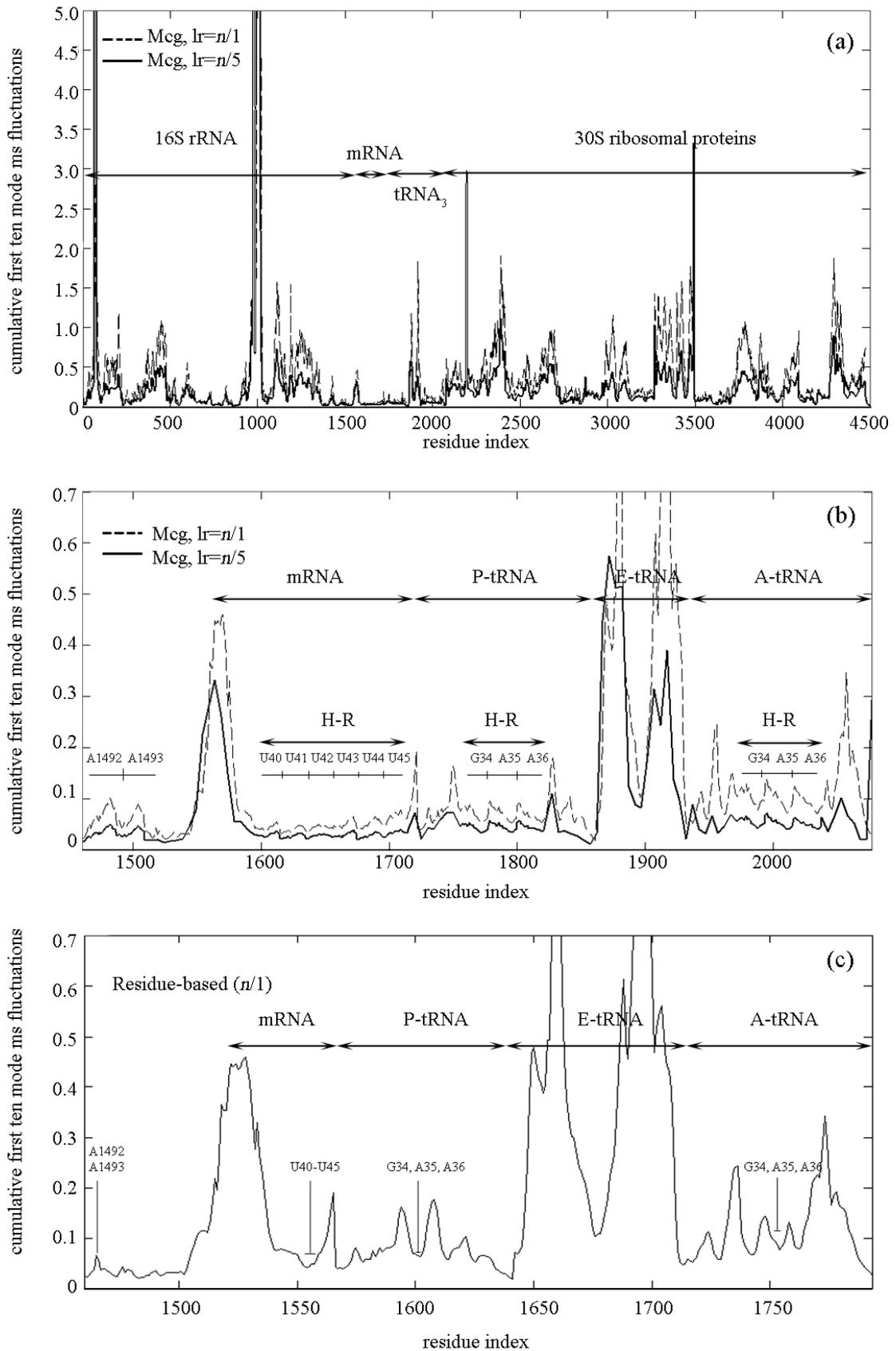


Figure 5.21. Cumulative first ten mode mean-square fluctuations are calculated for different levels of mixed cg for ribosome complex

clearly with atomistic detail at the decoding center and codons/ anticodon region, which would not be analyzed in biological importance aspect here.

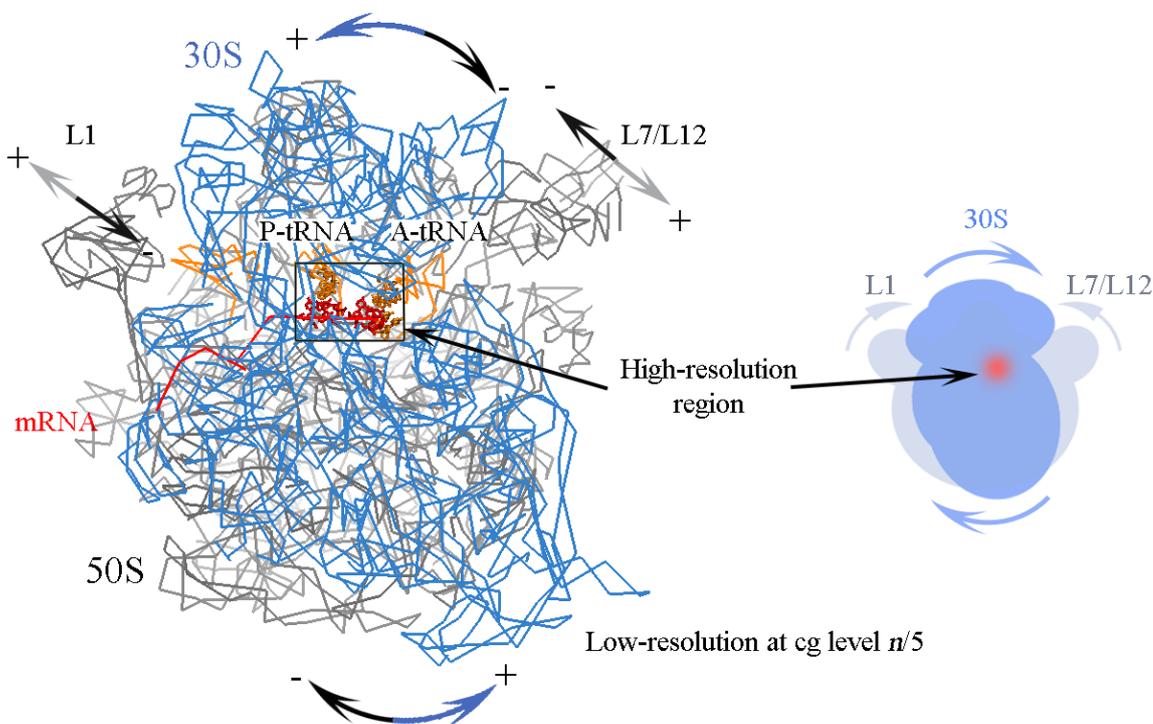


Figure 5.22. Displacement vectors of alternative directions (+,-) from the third mode shown in ribbon presentation of ribosome complex. A simplified scheme of motions is given on the right

Figure 5.23 (a) displays the mobility at the high-resolution region calculated from ten cumulative mode fluctuations, where red implies high mobility. The side chains of A1492, A1493 and ASL of A-tRNA seem to possess more mobility in slow modes when compared to P-tRNA and mRNA. In addition, A-site of mRNA is more mobile when compared to its P-site, all in accord with experimental B-factors. These detailed results on the side chains could not be observed with low-resolution normal mode analysis (Tama *et al.*, 2003), elastic network models (Wang *et al.*, 2004) and MD simulations (Trylska *et al.*, 2005).

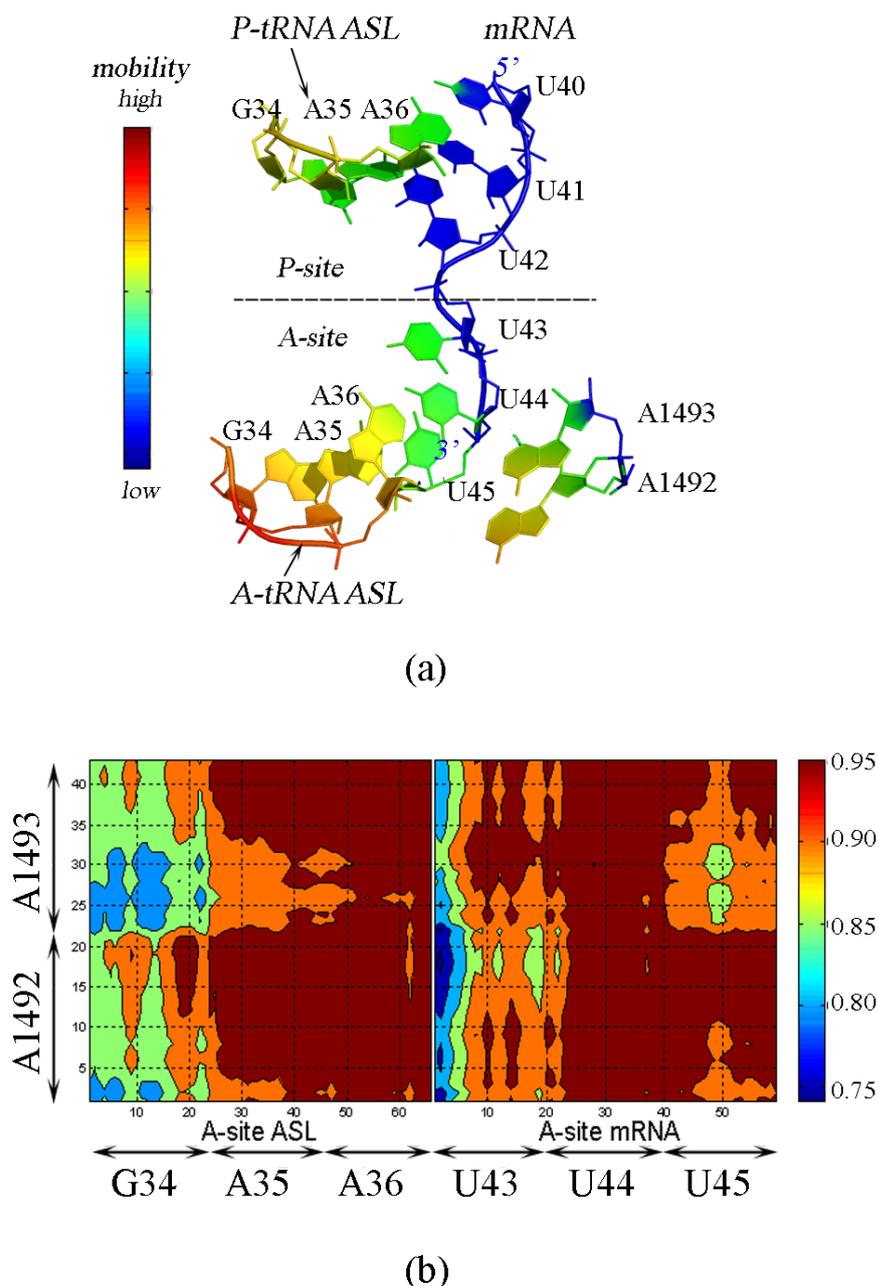


Figure 5.23. (a) Mobility of the high-resolution region with red color indicating the highest value. (b) Cross-correlation map of the high-resolution region over first ten modes

The inner dot products of displacement vectors averaged over first ten modes are calculated for the high-resolution decoding center and A -site codons and anticodons (Figure 5.23 (b)). Harmonic vibrational analysis points that G34 of A-tRNA and U43 of mRNA show relatively lower correlation with the decoding center, also related to the orientation of the side chains in native structure. However, this detailed picture can not be

observed with the residue-based cg model where the cross correlations only between the phosphorous nodes of A1492 (A1493) and G34 on A-tRNA, U43 on mRNA are calculated as 0.90 (0.84) and 0.96 (0.93) respectively. Moreover, in Figure 5.24, the cross-correlations calculated over first ten slowest modes of decoding center and anticodons of A- and P-tRNAs are displayed. Mixed cg results indicate not-correlated regions between the decoding center and P-site anticodons, which cannot be depicted from residue-based calculations.

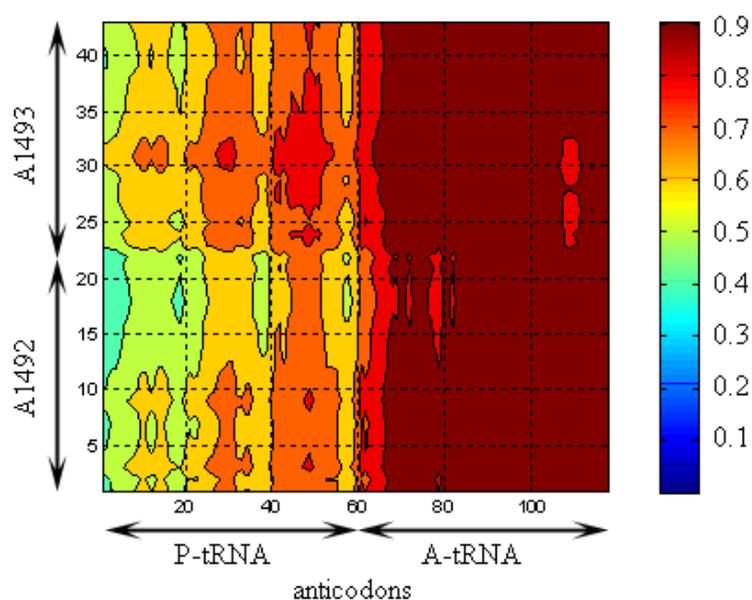


Figure 5.24. Cross-correlation map calculated from cumulative ten modes for the decoding center A1492, A1493 and anticodons of A- and P-tRNAs

Based on the experimental observations and computational results from elastic network models, there are other ‘interesting’ regions that may be described at atomistic detail with the mixed cg model. The aminoacyl ends of A- and P-tRNA and the peptidyl transferase center is the catalytic center of the ribosome, and the local atomic fluctuations at low frequencies can give insights about the protein synthesis mechanism. Another important region is the narrow gate of the ribosomal exit tunnel. Here, universally conserved extensions of ribosomal proteins L22 and L4 are functionally important to open or close the gate to nascent polypeptides. According to residue-based model, this region undergoes important harmonic motions at low frequencies, which is discussed in detail in

this chapter. Moreover, these extensions of proteins are located at the antibiotic binding site, therefore they can be studied in detail with the mixed cg model to generate realistic conformations for drug binding sites. Finally, regions that are thought to have a role in the translation-translocation mechanism, such as helicase active sites of S3 and S4 can be modeled with this computationally efficient method. Therefore, the functionally important motions at atomistic level can be obtained, where a classical NMA and atom-based ANM would be inadequate.

## 6. CONCLUSIONS AND RECOMMENDATIONS

### 6.1. Conclusions

In this thesis, the mixed coarse-graining technique for the elastic network model has been developed to explore the collective dynamics of extremely large proteins with atomic detail. In this approach, the choice of high-resolution region's residues depends on the purpose of the study; for example the functionally important regions of the native structure, such as an active site, can be described at atomic detail with the remaining parts at lower resolution. As a result, each node in the network represents a heavy atom (high-resolution), a residue or a collection of residues (low-resolution); and sufficiently close nodes within a cutoff are connected by springs. Thus, mixed-resolution representation of a structure provides computational efficiency and atomistic detail at the same time for determining function related collective motions that is difficult to attain with classical molecular modeling methods. The mixed coarse-grained calculations have been performed on ligand-bound and -unbound triosephosphate isomerase (TIM), a homodimeric enzyme composed of 494 residues, and bacterial ribosome with mRNA, tRNAs and elongation factor Tu, constituting a supramolecular assemblage of ~11000 amino acids and nucleotides in total.

In the elastic network model, there are two important interaction parameters to be adjusted: the cutoff distance and the force constant. Two different approaches are developed to assign suitable interaction parameters to high- and low-resolution regions. In the first approach of mixed coarse-graining (i.e. procedure I), the cutoff distances and force constants for pair wise interactions are estimated based on residue-based elastic network model. However, the cutoffs and force constants may depend on system size and this method may not be applied to large macromolecules due to limits in the computational power. Second procedure of mixed coarse-graining is developed to compensate the discrepancies derived from residue-based approach. With a single cutoff distance fixed for atomic interactions of amino acids or nucleotides, the force constants between any pair of nodes (high-high, low-low, and high-low resolution pairs) are determined according to the total number of interatomic interactions surviving between the node pairs. As a result, the

second approach is a better method because any prior adjustment of cutoffs and force constants, which may depend on the system size, is not required, and this technique can be applied to DNA/ RNA complexes.

TIM is chosen as the model system to validate the mixed coarse-grained model with atomistic detail due to its manageable size and wealth of experimental data available in the literature. The opening/ closing motion of functionally important loop 6 over the catalytic site during catalysis is obtained successfully. In addition, the first and fourth normal modes are found to be related with loop 6 motion observed in various crystal structures. The mixed resolution model has elucidated the dynamic motions of the dimeric enzyme having an important role on loop 6 motion and active site residues, which is confirmed by few crystal enzyme structures, classical normal mode analysis and recent molecular dynamics simulations (Cansu and Doruker, *in press*). Investigating the catalytic site at atomistic detail without constraints on the structure has been informative about enzyme mechanics concerning active site residue side chains and ligand geometries in harmonic motions (Kurkcuoglu *et al.*, 2005; 2006).

TIM is modeled within a generated crystal packing to explain the unexpected rigidity of the available TIM structures in the protein data bank (Berman *et al.*, 2000). This rigidity is revealed by relatively small ( $\sim 0.85\text{-}2 \text{ \AA}$ ) root mean-square deviation between various crystal structures. However, in these native structures, the flexible loop 6 adopts significantly different conformations. In the thesis, the residue-based model elucidated that functionally important opening/closing motion of loop 6 is retained while the whole structure remained rigid due to dense packing. This finding explains that the collective motions of TIM in a crystal may be hindered due to the crystal contacts. In addition, only the high-resolution region of TIM is simulated without the remaining parts of the structure, to study the collective motions. The results have demonstrated that it is not so important to know all the details of a structure to calculate its motions as long as the whole protein structure is kept intact. This finding indicates the effectiveness of mixed coarse-graining in providing a large protein's globular deformations with atomic detail, where an incomplete full-atom classical approach would remain insufficient.

The ultimate aim of mixed coarse-grained model was to apply this methodology to a supramolecular assemblage difficult to work with a full-atom technique. Ribosome in its complex with mRNA, three tRNAs and elongation factor Tu is studied with both residue-based, and mixed coarse-grained model using the second procedure, which is based on atom-atom contact. Experimental observations such as ratchet-like rotation of subunits, high flexibility and anti-correlated motions of proteins L1 and L7/L12, and head rotation of subunit 30S are obtained successfully at low frequency spectrum. Moreover, a mechanism for translocation of mRNA and bound tRNAs is proposed by combining experimental and computational data. Different from previous computational studies on the ribosome complex, the harmonic dynamics of nascent polypeptide exit tunnel is observed with normal modes and combined-modes. The results point to three dynamic domains on the tunnel that may force the nascent polypeptide to exit or the ribosome to stall. Finally, mixed coarse-graining technique is used to explore the dynamics of the decoding center, and codons/ anticodons at A- and P-sites with atomistic detail; as a result the utility and the computational efficiency of the mixed resolution approach are validated.

In summary, elastic network models using normal mode analysis provide a simple way to sample protein conformations, especially the low frequency range motions involving large domains. The control of these motions is so strong that it is clear that globular conformations are related to biological mechanisms. On the other hand, elastic network models are insensitive to the details of the protein structure in extracting large-scale motions. Therefore, the mixed coarse-grained model is a powerful technique to explore the harmonic dynamics of a supramolecular assemblage with its functionally important regions at atomistic detail and the remaining parts at extreme levels of coarse-graining (Jernigan *et al.*, 2008, *in press*).

## 6.2. Recommendations

The present mixed coarse-grained model gives insights about the conformational dynamics of a large macromolecule and a supramolecular assemblage, which cannot be regenerated by conventional fully-atomistic molecular dynamics simulations and normal mode analysis. This method can be applied to study the vibrational dynamics of other parts of ribosome complex at atomistic level, such as the peptidyl transferase center and

ribosomal tunnel entrance which constitutes one of the antibiotic binding regions in bacteria. Collective motions of other supramolecular assemblages, such as a viral capsule, can be also studied with atomistic details while maintaining computational efficiency.

The mixed coarse-grained model provides a particularly efficient way to generate new protein conformations with atomistic detail. The generated conformations after energy minimization can be used as initial structures with minimum energy in molecular dynamics simulations to better sample the conformational space and to reach the true minimum on the energy surface.

Finally, this approach may be an important tool to investigate atomic rearrangements at protein-ligand, protein-DNA or protein-RNA interfaces by incorporating flexibility. The mixed coarse-graining method may also be used for flexible protein-ligand docking if realistic binding region conformations can be generated along the slow modes, even for supramolecular assemblages that are unfeasible to work with full-atom approaches.

## REFERENCES

- Agmon, I., T. Auerbach, D. Baram, H. Bartels, A. Bashan, R. Berisio, P. Fucini, H. A. S. Hansen, J. Harms, M. Kessler, M. Peretz, F. Schlutzen, A. Yonath and R. Zarivach, 2003, "On Peptide Bond Formation, Translocation, Nascent Protein Progression and the Regulatory Properties of Ribosomes" Delivered on 20 October 2002 at the 28<sup>th</sup> FEBS Meeting in Istanbul, *European Journal of Biochemistry*, Vol. 270, pp. 2543–2556.
- Agrawal, R. K., A. B. Heagle, P. Penczek, R. A. Grassucci and J. Frank, 1999b, "EF-G-Dependent GTP Hydrolysis Induces Translocation Accompanied by Large Conformational Changes in the 70S Ribosome", *Nature Structural Biology*, Vol. 6, pp. 643–647.
- Agrawal, R. K., J. Linde, J. Sengupta, K. H. Nierhaus and J. Frank, 2001, "Localization of L11 Protein on the Ribosome and Elucidation of its Involvement in EF-G Dependent Translocation", *Journal of Molecular Biology*, Vol. 311, pp. 777-787.
- Agrawal, R. K., P. Penczek, R. A. Grassucci, N. Burkhardt, K. H. Nierhaus and J. Frank, 1999a, "Effect of Buffer Conditions on the Position of tRNA on the 70 S Ribosome as Visualized by Cryoelectron Microscopy", *Journal of Biological Chemistry*, Vol. 274, pp. 8723–8729.
- Alakent, B., M. C. Camurdan and P. Doruker, 2007 "Mimicking Protein Dynamics by the Integration of Elastic Network Model with Time Series Analysis", *The International Journal of High Performance Computing Applications*, Vol. 21, pp. 59–65.
- Amadei, A., M. A. Ceruso and A. D. Nola, 1999, "On the Convergence of the Conformational Coordinates Basis Set Obtained by the Essential Dynamics Analysis of Proteins' Molecular Dynamics Simulations", *Proteins*, Vol. 36, pp. 419-424.

- 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, pp. 505-515.
- Bahar, I. and R. L. Jernigan, 1998, "Vibrational Dynamics of Transfer RNAs: Comparison of the Free and Synthase Bound Forms", *Journal of Molecular Biology*, Vol. 281, pp. 871-885.
- Bahar, I., A. R. Atilgan, B. Erman, 1997, "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single Parameter Harmonic Potential", *Folding and Design*, Vol. 2, pp.173-181.
- Bahar, I., A. R. Atilgan, M. C. Demirel and B. Erman, 1998b, "Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability", *Physical Review Letters*; Vol. 80, No. 12.
- Bahar, I., A. Wallqvist, D. Covell and R. L. Jernigan, 1998a, "Correlation between Native State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model", *Biochemistry*, Vol. 37, pp. 1067.
- Ban, N, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, 2000, "The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4Å Resolution", *Science*, Vol. 289, pp. 905-920.
- Banner, D. W., A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle and S. G. Waley, 1975, "Structure of Chicken Muscle Triosephosphate Isomerase Determined Crystallographically at 2.5 Å Resolution Using Amino Sequence Data", *Nature*, Vol. 255, pp. 609-614.
- Beckmann R, D. Bubeck, R. Grassucci, P. Penczek, A. Verschoor, G. Blobel and J. Frank, 1997, "Alignment of Conduits for the Nascent Polypeptide Chain in the Ribosome-Sec61 Complex", *Science*, Vol. 278, pp. 2123–2126.

- Ben-Avraham, D., 1993, "Vibrational Normal-Mode Spectrum of Globular Proteins", *Physical Review*, Vol. 47, No. 21.
- Berisio, R., F. Schluenzen, J. Harms, A. Bashan, T. Auerbach, D. Baram and A. Yonath, 2003, "Structural Insight into the Role of the Ribosomal Tunnel in Cellular Regulation", *Nature Structural Biology*, Vol. 10, pp. 366-370.
- 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, pp. 235-242.
- Bocharov, E. V., A. G. Sobol, K. V. Pavlov, D. M. Korzhnev, V. A. Jaravine, A. T. Gudkov and A. S. Arseniev, 2004, "From Structure and Dynamics of Protein L7/L12 to Molecular Switching in Ribosome", *Journal of Biological Chemistry*, Vol. 279, pp. 17697-17706.
- Branden, C. and J. Tooze, 1999, *Introduction to Protein Structure*, 2nd ed., Garland Publishing, Inc., New York.
- Brodersen, D. E. and P. Nissen, 2005, "The Social Life of Ribosomal Proteins", *FEBS Journal*, Vol. 272, pp. 2098-2108.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, 1983, "CHARMM", *Journal of Computational Chemistry*, Vol. 4, pp. 187-217.
- Cansu, S. and P. Doruker, "Dimerization Affects Collective Dynamics of Triosephosphate Isomerase", *Biochemistry*, *in press*.
- Cech, T. R., 2000, "The Ribosome is a Ribozyme", *Science*, Vol. 289, pp. 878-879.

- Chacon, P., T. Florence and W. Wriggers, 2003, "Mega-Dalton Biomolecular Motion Captured from Electron Microscopy Reconstructions", *Journal of Molecular Biology*, Vol. 326, 485-492.
- Chennubhotla, C. and I. Bahar, 2006, "Markov Propagation of Allosteric Effects in Biomolecular Systems: Application to GroEL-GroES", *Molecular Systems Biology*, Article number: 36.
- Chennubhotla, C. and I. Bahar, 2007a, "Markov Methods for Hierarchical Coarse-Graining of Large Protein Dynamics", *Journal of Computational Biology*, Vol.14, pp. 765-76.
- Chennubhotla, C. and I. Bahar, 2007b, "Signal Propagation in Proteins and Relation to Equilibrium Fluctuations", *PLOS Computational Biology*, Vol. 3: e172.
- Chittum, H. S. and W. S. Champney, 1994, "Ribosomal Protein Gene Sequence Changes in Erythromycin-Resistant Mutants of *Escherichia coli*", *Journal of Bacteriology*, Vol. 176, pp. 6192-6198.
- Cheng, R. H. and L. Hammar (eds.), 2004, *Conformational Proteomics of Macromolecular Architecture : Approaching the Structure of Large Molecular Assemblies and their Mechanisms of Action*, World Scientific Publishing Company, Inc., New Jersey, London.
- Darst, S. A., N. Opalka, P. Chacon, A. Polyakov, C. Richter, G. Zhang and W. Wriggers, 2002, "Conformational Flexibility of Bacterial RNA Polymerase", *Proceedings of National Academy of Sciences USA*, Vol. 99, pp. 4296-4301.
- DeLano, W. L., 2002, *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA.
- Derreumaux, P. and T. Schlick, 1998, "The Loop Opening/Closing Motion of the Enzyme Triosephosphate Isomerase", *Biophysical Journal*, Vol. 74, pp. 72-81.

- Desamero, R., S. Rozovsky, N. Zhadin, A. McDermott and R. Callender, 2003, "Active Site Loop Motion in Triosephosphate Isomerase: T-Jump Relaxation Spectroscopy of Thermal Activation", *Biochemistry*, Vol. 42, pp. 2941-2951.
- Doruker, P., A. R. Atilgan and I. Bahar, 2000, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to  $\alpha$ -Amylase Inhibitor", *Proteins*, Vol. 40, pp. 512-524.
- Doruker, P. and R. L. Jernigan, 2003, "Functional Motions can be Extracted from On-Lattice Construction of Protein Structures", *Proteins*, Vol. 53, pp. 174-181.
- Doruker, P., L. Nilsson and O. Kurkcuoglu, 2006, "Collective Dynamics of *EcoRI*-DNA Complex by Elastic Network Model and Molecular Dynamics Simulations", *Journal of Biomolecular Structure & Dynamics*, Vol. 24, pp. 1-15.
- Doruker, P., R. L. Jernigan and I. Bahar, 2002a, "Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures", *Journal of Computational Chemistry*, Vol. 23, pp. 119-127.
- Doruker, P., R. L. Jernigan, I. Navizet and R. Hernandez, 2002b, "Important Dynamics of Large Protein Structures Are Preserved upon Coarse-Grained Renormalization", *International Journal of Quantum Chemistry*, Vol. 90, pp. 822-837.
- Elcock, A. H., 2002, "Modeling Supramolecular Assemblages", *Current opinion in Structural Biology*, Vol. 12, pp. 154-160.
- 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, pp. 590-596.
- Flores, S., N. Echols, D. Milburn, B. Hespeneide, K. Keating, J. Lu, S. Wells, E. Z. Yu, M. Thorpe and M. Gerstein, 2006, "The Database of Macromolecular Motions: New Features Added at the Decade Mark", *Nucleic Acids Research* Vol. 34 (Database issue), pp. D296-D301.

- Frank, J. and R. K. Agrawal, 2000, "A Ratchet-like Inter-Subunit Reorganization of the Ribosome During Translocation", *Nature*, Vol. 406, pp. 318–22.
- Frank, J. and C. M. T. Spahn, 2006, "The Ribosome and the Mechanism of Protein Synthesis", *Reports on Progress in Physics*, Vol. 69, pp. 1383–1417.
- Frauenfelder, H., A. R. Bishop, A. Garcia, A. Perelson, P. Schuster, D. Sherrington, P. J. Swart. (eds.), 1997, *Landscape paradigms in physics and biology: Concepts, structures, and dynamics*, North-Holland, Amsterdam.
- Frauenfelder, H., S. G. Sligar, P. G. Wolynes, 1991, "The Energy Landscape of Motions of Proteins", *Science*, Vol. 254, pp. 1598-1603.
- Fry, E. E., J. W. Newman, S. Curry, S. Najjam, T. Jackson, W. Blakemore, S.M. Lea, L. Miller, A. Burman, A.M. King and D.I. Stuart, 2005, "Structure of Foot-and-Mouth Disease Virus Serotype A1061 Alone and Complexed with Oligosaccharide Receptor: Receptor Conservation in the Face of Antigenic Variation", *Journal of General Virology*, Vol. 86, pp. 1909-1920.
- Gabashvili I. S., S. T. Gregory, M. Valle, R. Grassucci, M. Worbs, M. C. Wahl, A. E. Dahlberg and J. Frank, 2001, "The Polypeptide Tunnel System in the Ribosome and Its Gating in Erythromycin Resistance Mutants of L4 and L22", *Molecular Cell*, Vol. 8, pp. 181–188.
- Gabashvili, I. S., R. K. Agrawal, C. M. T. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank and P. Penczek, 2000, "Solution Structure of the *E.Coli* 70S Ribosome at 11.5 Å Resolution", *Cell*, Vol. 100, pp. 537-549.
- Gerstein, M., A. M. Lesk and C. Chothia, 1994, "Structural Mechanisms for Domain Movements in Proteins", *Biochemistry*, Vol. 33, pp. 6739-6749.

- Grimes, R. G., J. G. Lewis and H. D. Simon, 1991, *A Shifted Block Lanczos Algorithm for Solving Sparse Symmetric Eigenvalue Problems*, Seattle, WA: Boeing Computer Services.
- Gu, S. Q., F. Peske, H.-J. Wieden, M. V. Rodnina and W. Wintermeyer, 2003, "The Signal Recognition Particle Binds to Protein L23 at the Peptide Exit of the *Escherichia coli* Ribosome", *RNA*, Vol. 9, pp. 566–573.
- Guallar, V., M. Jacobson, A. McDermott and R. A. Friesner, 2004, "Computational Modeling of the Catalytic Reaction in Triosephosphate Isomerase", *Journal of Molecular Biology*, Vol. 337, pp. 227-239.
- Haliloglu, T. and I. Bahar, 1999, "Structure-Based Analysis of Protein Dynamics. Comparison of Theoretical Results for Hen Lysozyme with X-ray Diffraction and NMR Relaxation Data", *Proteins*, Vol. 37, pp. 654-667.
- Haliloglu, T., I. Bahar and B. Erman, 1997, "Gaussian Dynamics of Folded Proteins", *Physical Review Letters*, Vol. 79, pp. 3090-3093.
- Hayward, S., 2001, "Normal Mode Analysis of Biological Molecules" in Becker, O. M., A. D. Jr. MacKerell, B. Roux, and M. Watanabe (eds.), *Computational Biochemistry and Biophysics*, pp. 153-168, Marcel Dekker, Inc. New York.
- Hermann, T., 2005, "Drugs Targeting the Ribosome", *Current Opinion in Structural Biology*, Vol. 15, pp. 355-366.
- Hinsen K., 1998, "Analysis of Domain Motions by Approximate Normal Mode Calculations", *Proteins*, Vol. 33, pp. 417–29.
- Hinsen, K., A. Thomas and M. J. Field, 1999, "Analysis of Domain Motions in Large Proteins", *Proteins*, Vol. 34, pp. 369-382.

- Jenner, L., P. Romby, B. Rees, C. Schulze-Briese, M. Springer, C. Ehresmann, B. Ehresmann, D. Moras, G. Yusupova and M. Yusupov, 2005, “Translational Operator of mRNA on the Ribosome: How Repressor Proteins Exclude Ribosome Binding”, *Science*, Vol. 308, pp. 120-123.
- Jernigan, R. L., L. Yang, O. Kurkcuoglu, G. Song and P. Doruker, 2008, “Elastic Network Models of Coarse-Grained Proteins Are Effective for Studying the Structural Control Exerted over Their Dynamics”, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Taylor & Francis Group, LLC, *in press*.
- Jernigan, R. L., M. C. Demirel and I. Bahar, 1999, “Relating Structure to Function through the Dominant Modes of Motion of DNA Topoisomerase II”, *International Journal of Quantum Chemistry*, Vol. 75, pp. 301-312.
- Joseph, D., G. A. Petsko and M. Karplus, 1990, “Anatomy of a Conformational Change: Hinged Lid Motion of the Triosephosphate Isomerase Loop”, *Science*, Vol. 249, pp. 1425-1428.
- Joseph, S., 2003, “After the Ribosome Structure: How Does Translocation Work?”, *RNA*, Vol. 9, pp. 160-164.
- Karplus, M. and J. A. McCammon, 2002, “Molecular Dynamics Simulations of Biomolecules”, *Natural Structural Biology*, Vol. 9, pp. 646–652.
- Keskin, O., R.L. Jernigan and I. Bahar, 2000, “Protein with Similar Architecture Exhibit Large-Scale Dynamics Behavior”, *Biophysical Journal*, Vol. 78, pp. 2093-2106.
- Keskin, O., I. Bahar, D. Flatow, D. G. Covell and R. L. Jernigan, 2002a, “Molecular Mechanisms of Chaperonin GroEL-GroES Function”, *Biochemistry*, Vol. 41, pp. 491-501.

- Keskin, O., S. R. Durell, I. Bahar, R. L. Jernigan and D. G. Covell, 2002b, “Relating Molecular Flexibility to Function: A Case Study of Tubulin”, *Biophysical Journal*, Vol. 83, pp. 663-680.
- Khaitovich, P., A. S. Mankin, R. Green, L. Lancaster and H. F. Noller, 1999, “Characterization of Functionally Active Subribosomal Particles from *Thermus aquaticus*”, *Proceedings of National Academy of Sciences U. S. A.*, Vol. 96, pp. 85-90.
- Kim, M. K., W. Li, B. A. Shapiro and G. S. Chirikjian, 2003, “A Comparison Between Elastic Network Interpolation and MD Simulation of 16S Ribosomal RNA”, *Journal of Biomolecular Structure & Dynamics*, Vol. 21, pp. 395-405.
- Kirthi, N., B. Roy-Chaudhuri, T. Kelley and G. M. Culver, 2006, “A Novel Single Amino Acid Change in Small Subunit Ribosomal Protein S5 Has Profound Effects on Translational Fidelity”, *RNA*, Vol. 12, pp. 1–12.
- Kitao, A. and N. Go, 1999, “Investigating Protein Dynamics in Collective Coordinate Space”, *Current Opinion in Structural Biology*, Vol. 9, pp. 164-169.
- Knowles, J. R., 1991, “Enzyme Catalysis: Not Different, Just Better”, *Nature*, Vol. 350, pp. 121-124.
- Korostelev, A. and H. F. Noller, 2007, “The Ribosome in Focus: New Structures Bring New Insights”, *TRENDS in Biochemical Sciences*, Vol. 32(9), pp. 434-441.
- Korostelev, A., S. Trakhanov, H. Asahara, M. Laurberg, L. Lancaster and H. F. Noller, 2007, “Interactions and Dynamics of the Shine–Dalgarno Helix in the 70S Ribosome”, *Proceedings of National Academy of Sciences U. S. A.*, Vol. 104, pp. 16840-16843.
- Kosolapov, A. L., L. Tu, J. Wang and C. Deutsch, 2004, “Structure Acquisition of the T1 Domain of Kv1.3 During Biogenesis”, *Neuron*, Vol. 44, pp. 295–307.

- Kramer, G., T. Rauch, W. Rist, S. Vorderwulbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling and B. Bukau, 2002, "L23 Protein Functions as a Chaperone Docking Site on the Ribosome", *Nature*, Vol. 419, pp. 171-174.
- Kundu, S. and R. L. Jernigan, 2004, "Molecular Mechanism of Domain Swapping in Proteins: An Analysis of Slower Motions", *Biophysical Journal*, Vol. 86, pp. 3846-3854.
- Kuriyan, J. and M. O'Donnell, 1993, "Sliding Clamps of DNA Polymerases", *Journal of Molecular Biology*, Vol. 234, pp. 915-925.
- Kuriyan, J., G. A. Petsko, R. M. Levy and M. Karplus, 1986, "Effect of Anisotropy and Anharmonicity on Protein Crystallographic Refinement. An Evaluation by Molecular Dynamics", *Journal of Molecular Biology*, Vol. 190, pp. 227-254.
- Kurkcuoglu, O., 2003, *Mixed Coarse-Graining of Large Proteins Using Elastic Network Model*, M.S. Thesis, Boğaziçi University.
- Kurkcuoglu, O., P. Doruker and R. L. Jernigan, 2004, "Mixed Levels of Coarse-Graining of Large Proteins Using Elastic Network Model Succeeds in Extracting the Slowest Motions", *Polymer*, Vol. 45, pp. 649-657.
- Kurkcuoglu, O., P. Doruker and R. L. Jernigan, 2005, "Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model", *QSAR and Combinatorial Science*, Vol. 24, pp. 443-448.
- Kurkcuoglu, O., P. Doruker and R. L. Jernigan, 2006, "Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks", *Biochemistry*, Vol. 45, pp. 1173-1182.
- Lata, K. R., R. K. Agrawal, P. Penczek, R. Grassucci, J. Zhu and J. Frank, 1996, "Three-Dimensional Reconstruction of the Escherichia Coli 30 S Ribosomal Subunit in Ice", *Journal of Molecular Biology*, Vol. 262, pp. 43-52.

- Lee, B. M., J. Xu, B. K. Clarkson, M. A. Martinez-Yamout, H. J. Dyson, D. A. Case, J. M. Gottesfeld and P. E. Wright, 2006. "Induced Fit and 'Lock and Key' Recognition of 5S RNA by Zinc Fingers of Transcription Factor IIIA", *Journal of Molecular Biology*, Vol. 357, pp. 275-291.
- Locher, K. P., B. Rees, R. Koebnik, A. Mitschler, L. Moulinier, J. P. Rosenbusch and D. Moras, 1998, "Transmembrane Signaling Across the Ligand-Gated FhuA Receptor: Crystal Structures of Free and Ferrichrome-Bound States Reveal Allosteric Changes", *Cell*, Vol. 95, pp. 771-778.
- Lu, J. and C. Deutsch, 2005, "Folding Zones Inside the Ribosomal Exit Tunnel", *Nature Structural & Molecular Biology*, Vol. 12(12), pp. 1123-1129.
- Lu, M. and J. Ma, 2005, "The Role of Shape in Determining Molecular Motions", *Biophysical Journal*, Vol. 89, pp. 2395-2401.
- Maes, D., J. P. Zeelen, N. Thanki, N. Beaucamp, M. Alvarez, M. H. Thi, J. Backmann, J. A. Martial, L. Wyns, R. Jaenicke and R. K. Wierenga, 1999, "The Crystal Structure of Triosephosphate Isomerase (TIM) from *Thermotoga Maritima*: A Comparative Thermostability Structural Analysis of Ten Different TIM Structures", *Proteins*, Vol. 37, pp. 441-453.
- Marques, O. and Y. H. Sanejouand, 1995, "Hinge-Bending Motion in Citrate Synthase Arising from Normal Mode Calculations", *Proteins*, Vol. 23, pp. 557-560.
- McCammon, A. and S. C. Harvey, 1987, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
- Milligan, R. A. and P. N. Unwin, 1986, "Location of Exit Channel for Nascent Protein in 80S Ribosome", *Nature*, Vol. 319, pp. 693-695.
- Minakhin, L., S. Bhagat, A. Brunning, E. A. Campbell, S. A. Darst, R. H. Ebright and K. Severinov, 2001, "Bacterial RNA Polymerase Subunit Omega and Eukaryotic RNA

- Polymerase Subunit RPB6 are Sequence, Structural, and Functional Homologs and Promote RNA Polymerase Assembly, *Proceedings of National Academy of Sciences U.S.A.*, Vol. 98, pp. 892-897.
- Ming, D., Y. Kong, M. A. Lambert, Z. Huang and J. Ma, 2002, "How to Describe Protein Motion Without Amino-Acid Sequence and Atomic Coordinates", *Proceedings of National Academy of Sciences U.S.A.*, Vol. 99, pp. 8620-8625.
- Mitra, K., C. Schaffitzel, T. Shaikh, F. Tama, S. Jenni, C. L. Brooks III, N. Ban and J. Frank, 2005, "Structure of the E. coli Protein-Conducting Channel Bound to a Translating Ribosome", *Nature*, Vol. 438, pp. 318-324.
- Moazed, D. and H. F. Noller, 1989, "Intermediate States in the Movement of Transfer RNA in the Ribosome", *Nature*, Vol. 334, pp. 362-364.
- Nakatogawa, H. and K. Ito, 2002, "The Ribosomal Exit Tunnel Functions as a Discriminating Gate", *Cell*, Vol. 108, pp. 629-636.
- Nissen, P., J. Hansen, N. Ban, P. B. Moore and T. A. Steitz, 2000, "The Structural Basis of Ribosome Activity in Peptide Bond Synthesis", *Science*, Vol. 289, pp. 920-930.
- Noble, M. E., J. P. Zeelen, R. K. Wierenga, V. Mainfroid, K. Goraj, A. C. Gohimont and J. A. Martial, 1993, "Structure of Triosephosphate Isomerase from Escherichia Coli Determined at 2.6 Å Resolution", *Acta Crystallographica, Section D*, Vol. 49, pp. 403-417.
- Noller, H. F., V. Hoffarth and L. Zimniak, 1992, "Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures", *Science*, Vol. 256, pp. 1416-1419.
- Ogle, J. M., F. V. Murphy, M. J. Tarry and V. Ramakrishnan, 2002, "Selection of tRNA by the Ribosome Requires a Transition From an Open to a Closed form", *Cell*, Vol. 111, pp. 721-732.

- Petry, S., D. E. Brodersen, F. V. Murphy, C. M. Dunham, M. Selmer, M. J. Tarry, A. C. Kelley and V. Ramakrishnan, 2005, "Crystal Structures of the Ribosome in Complex with Release Factors RF1 and RF2 Bound to a Cognate Stop Codon", *Cell*, Vol. 123, pp. 1255-1266.
- Petsko, A. G. and D. Ringe, 2004, *Protein Structure and Function*, New Science Press Ltd., London.
- Pettersson, I. and C. G. Kurland, 1980, "Ribosomal Protein L7/L12 is Required for Optimal Translation", *Proceedings of National Academy of Sciences U.S.A.*, Vol. 77, pp. 4007-4010.
- Polacek, N. and A. S. Mankin, 2005, "The Ribosomal Peptidyl Transferase Center: Structure, Function, Evolution, Inhibition", *Critical Reviews in Biochemistry and Molecular Biology*, Vol. 40, pp. 285-311.
- Rodriguez-Romero, A., A. Hernandez-Santoyo, L. Del Pozo-Yauner, A. Kornhauser and D. A. Fernandez-Velasco, 2002, "Structure and Inactivation of Triosephosphate Isomerase from *Entamoeba Histolytica*", *Journal of Molecular Biology*, Vol. 322, pp. 669-675.
- Saibil, H. R., 2000, "Conformational Changes Studied by Cryo-Electron Microscopy", *Natural Structural Biology*, Vol. 7, pp. 711-14.
- Sampson, N. S. and J. R. Knowles, 1992, "Segmental Movement: Definition of the Structural Requirements for Loop Closure in Catalysis by Triosephosphate Isomerase", *Biochemistry*, Vol. 31, pp. 8482-8487.
- Sanbonmatsu, K. Y. and C.-S. Tung, 2006, "Large-Scale Simulations of the Ribosome: A New Landmark in Computational Biology", *Journal of Physics: Conference Series*, Vol. 46, pp. 334-342.

- Schurwith, B. S., M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton and J. H. D. Cate, 2005, "Structures of the Bacterial Ribosome at 3.5 Å Resolution", *Science*, Vol. 310, pp. 827-834.
- Selmer, M., C. M. Dunham, V. M IV Frank, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir and V. Ramakrishnan, 2006, "Structure of the 70S Ribosome Complexed with mRNA and tRNA", *Science*, Vol. 313, pp. 1935-1942.
- Song, G. and R. L. Jernigan, 2007, "vGNM: A Better Model for Understanding the Dynamics of Proteins in Crystals", *Journal of Molecular Biology*, Vol. 369, pp. 880-893.
- Stark, H., M. V. Rodnina, H. J. Wieden, F. Zemlin, W. Wintermeyer and M. van Heel, 2002, "Ribosome Interactions of Aminoacyl-tRNA and Elongation Factor TU in the Codon Recognition Complex", *Natural Structural Biology*, Vol. 9, pp. 849-854.
- Subramanian, A. R. and E. R. Dabbs, 1980, "Functional Studies on Ribosomes Lacking Protein L1 from Mutant *Escherichia coli*", *European Journal of Biochemistry*, Vol. 112, pp. 425-430.
- Suezaki, Y. and N. Go, 1975, "Breathing Mode of Conformational Fluctuations in Globular Proteins", *International Journal of Peptide and Protein Research*, Vol. 7, pp. 333-334.
- Suezaki, Y. and N. Go, 1976, "Fluctuations and Mechanical Strength of Alpha-Helices of Polyglycine and Poly(L-alanine)", *Biopolymers*, Vol. 15, pp. 2137-2153.
- Sutcliffe, J. A., 2005, "Improving on Nature: Antibiotics that Target the Ribosome", *Current Opinion in Microbiology*, Vol. 8, pp. 534-542.
- Takyar, S., R. P. Hickerson and H. F. Noller, 2005, "mRNA Helicase Activity of the Ribosome", *Cell*, Vol. 120, pp. 49-58.

- Tama, F. and C. L. Brooks III, 2006, "Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines", *Annual Review of Biophysics and Biomolecular Structure*, Vol. 35, pp. 115-133.
- Tama, F. and Y. H. Sanejouand, 2001, "Conformational Change of Proteins Arising from Normal Mode Calculations", *Protein Engineering*, Vol 14, pp. 1-6.
- Tama, F., O. Miyashita and C. L. III Brooks, 2004, "Flexible Multi-Scale Fitting of Atomic Structures into Low-Resolution Electron Density Maps with Elastic Network Normal Mode Analysis", *Journal of Molecular Biology*, Vol. 337, pp. 985-99.
- Tama, F., M. Valle, J. Frank and C. L. III Brooks, 2003, "Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy", *Proceedings of National Academy of Sciences U.S.A.*, Vol. 100(6), pp. 9319-9323.
- Tenson, T. and M. Ehrenberg, 2002, "Regulatory Nascent Peptides in the Ribosomal Tunnel", *Cell*, Vol. 108, pp. 591-594.
- Thanki, N., Z. P. Zeelen, M. Mathieu, R. Jaenicke, R. A. Abagyan, R. K. Wierenga and W. Schliebs, 1997, "Protein Engineering with Monomeric Triosephosphate Isomerase (Monotim): The Modeling and Structure Verification of a Seven-Residue Loop", *Protein Engineering*, Vol. 10 (2), pp. 159-167.
- Tirion, M. M, 1996, "Large Amplitude Elastic Motions in Proteins From a Single-Parameter, Atomic Analysis", *Physical Review Letters*, Vol. 77, pp. 1905-1908.
- Trylska J., V. Tozzini and J. A. McCammon, 2004, "A Coarse-Grained Model for the Ribosome: Molecular Dynamics Simulations", *Protein Science*, Vol. 13, pp. 121.
- Trylska, J., V. Tozzini and J. A. McCammon, 2005 "Exploring Global Motions and Correlations in the Ribosome", *Biophysical Journal*, Vol. 89, pp. 1455-1463.

- Uemura, S., M. Dorywalska, T. H. Lee, H. D. Kim, J. D. Puglisi and S. Chu, 2007, "Peptide Bond Formation Destabilizes Shine–Dalgarno Interaction on the Ribosome", *Nature*, Vol. 446, pp. 454-457.
- Valle, M., A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg and J. Frank, 2003, "Locking and Unlocking of Ribosomal Motions", *Cell*, Vol. 114, pp. 123-134.
- VanLoock, M. S., R. K. Agrawal, I. S. Gabashvili, L. Qi, J. Frank and S. C. Harvey, 2000, "Movement of the Decoding Region of the 16 S Ribosomal RNA Accompanies tRNA Translocation", *Journal of Molecular Biology*, Vol. 304, pp. 507-515.
- Voss, N. R., M. Gerstein, T. A. Steitz and P. B. Moore, 2006, "The Geometry of the Ribosomal Polypeptide Exit Tunnel", *Journal of Molecular Biology*, Vol. 360, pp. 893-906.
- Wade, R. C., M. E. Davis, B. A. Luty, J. D. Madura and J. A. McCammon, 1993, "Gating the Active Site of Triosephosphate Isomerase: Brownian Dynamics Simulations of Flexible Peptide Loops in the Enzyme", *Biophysical Journal*, Vol. 64, pp. 9-15.
- Wahl, M. C., G. P. Bourenkov, H. D. Bartunik and R. Huber, 2000, "Flexibility, Conformational Diversity and Two Dimerization Modes in Complexes of Ribosomal Protein L12", *EMBO Journal*, Vol. 19, pp. 174-186.
- Waley SG, 1973, "Refolding of Triosephosphate Isomerase", *Biochemical Journal*, Vol. 135, pp. 165-172.
- Wang, J. and D. C. Boisvert, 2003, "Structural Basis for GroEL-Assisted Protein Folding from the Crystal Structure of (GroEL-KMgATP)<sub>14</sub> at 2.0 Å Resolution", *Journal of Molecular Biology*, Vol. 327, pp. 843-855.
- Wang, Y. and R. L. Jernigan, 2005, "Comparison of tRNA Motions in the Free and Ribosomal Bound Structures", *Biophysical Journal*, Vol. 89, pp. 3399-3409.

- Wang, Y., A. J. Rader, I. Bahar and R.L. Jernigan, 2004, "Global Ribosome Motions Revealed with Elastic Network Model", *Journal of Structural Biology*, Vol. 147, pp. 302-314.
- Williams, J. C. and A. E. McDermott, 1995, "Dynamics of the Flexible Loop of Triosephosphate Isomerase: The Loop Motion is not Ligand Gated", *Biochemistry*, Vol. 34, pp. 8309-8319.
- Woolhead, C. A., P. J. McCormick and A. E. Johnson, 2004, "Nascent Membrane and Secretory Proteins Differ in FRET-Detected Folding Far inside the Ribosome and in Their Exposure to Ribosomal Proteins", *Cell*, 116, 725–736.
- Wriggers, W. and P. Chacon, 2001, "Modeling Tricks and Fitting Techniques for Multiresolution Structures", *Structure*, Vol. 9, pp. 779-788.
- Wriggers, W. and S. Birmanns, 2001, "Using Situs for Flexible and Rigid-Body Fitting of Multi-Resolution Single Molecule Data", *Journal of Structural Biology*, Vol. 133, pp. 193-202.
- Xiang, J., J. Jung and N. S. Sampson, 2004, "Entropy Effects on Protein Hinges: The Reaction Catalyzed by Triosephosphate Isomerase", *Biochemistry*, Vol. 43, pp. 11436-11445.
- Yildirim, Y. and P. Doruker, 2004, "Collective Motions of RNA Polymerases. Analysis of Core Enzyme, Elongation Complex and Holoenzyme", *Journal of Biomolecular Structure & Dynamics*, Vol. 22(3), pp. 267-280.
- Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman and P. R. Selvin, 2003, "Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-nm Localization", *Science*, Vol. 300, pp. 2061-2065.
- Yildiz, A., M. Tomishige, R. D. Vale and P. R. Selvin, 2004, "Kinesin Walks Hand-Over-Hand", *Science*, Vol. 303, pp. 676-678.

- Yonath A, K. R. Leonard and H. G. Wittmann 1987, “A Tunnel in the Large Ribosomal Subunit Revealed by Three-Dimensional Image Reconstruction”, *Science*, Vol. 236, pp. 813–816.
- Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate and H. F. Noller, 2001, “Crystal Structure of the Ribosome at 5.5 Å Resolution”, *Science*, Vol. 292, pp. 883-896.
- Yusupova, G. Z., M. M. Yusupov, J. H. Cate and H. F. Noller, 2001, “The Path of Messenger RNA through the Ribosome”, *Cell*, Vol. 106, pp. 233-241.
- Yusupova, G., L. Jenner, B. Rees, D. Moras and M. Yusupov, 2006, “Structural Basis for Messenger RNA Movement on the Ribosome”, *Nature*, Vol. 444, pp. 391-394.
- Zavialov, A. V., V. V. Hauryliuk and M. Ehrenberg, 2005, “Guanine-Nucleotide Exchange on Ribosome-Bound Elongation Factor G Initiates the Translocation of tRNAs”, *Journal of Biology*, Vol. 4, Article No: 9.
- Zhang, Z., S. Sugio, E. A. Komives, K. D. Liu, J. R. Knowles, G. A. Petsko and D. Ringe, 1994, “Crystal Structure of Recombinant Chicken Triosephosphate Isomerase-Phosphoglycolohydroxamate Complex at 1.8 Å Resolution”, *Biochemistry*, Vol. 33, pp. 2830-2837.
- Zhao, Q., L.-G. Öfverstedt, U. Skoglund and L. A. Isaksson, 2004, “Morphological Variation of Individual *Escherichia Coli* 50S Ribosomal Subunits in situ, as Revealed by Cryo-Electron Tomography”, *Experimental Cell Research*, Vol. 300, pp. 190-201.
- Ziv, G., G. Haran and D. Thirumalai, 2005, “Ribosome Exit Tunnel Can Entropically Stabilize {Alpha}-Helices”, *Proceedings of National Academy of Sciences U.S.A.*, Vol. 102, pp. 18956-18961.