

STRUCTURE AND DYNAMICS OF CATABOLITE GENE ACTIVATOR PROTEIN
BY MOLECULAR MODELING AND SIMULATIONS

by

AYŞE BURCU AYKAÇ FAS

B.S. in Chemical Engineering, Boğaziçi University, 1998

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

Graduate Program in Chemical Engineering
Boğaziçi University

2008

STRUCTURE AND DYNAMICS OF CATABOLITE GENE ACTIVATOR PROTEIN
BY MOLECULAR MODELING AND SIMULATIONS

APPROVED BY:

Prof. Türkan Haliloğlu
(Thesis Supervisor)

Prof. Pemra Doruker Turgut

Prof. Viktorya Aviyente

DATE OF APPROVAL: 18.01.2008

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis advisor, Prof. Türkan Haliloğlu for her guidance, encouraging and motivating attitude throughout my study. I'm also grateful to Asst. Prof. Yusuf Tutar for his helpful suggestions and discussions; Prof. Pemra Doruker Turgut and Prof. Viktorya Aviyente for their constructive criticisms and suggestions to my work.

I would like to thank Arzu Uyar, Ayşegül Özen, Bülent Balta, Ezgi Karaca, Nevra Özer, Nigar Kantarcı Çarşıbaşı, Özge Kürkçüoğlu Levitas and Sertan Cansu for their support and help along the whole study but especially during the preparation of this thesis. I would also like to thank Canan Dedeoğlu for her smiling face and positiveness I face everyday I step into PRC.

I would like to thank my friends Alkım Aytaman for her 24/7 hospitality and encouragement, Yelda Başkaya Kabaklarlı for her support and care over the phone and Dr. Bahar Bilgen Baç for her last minute suggestions over the Atlantic.

My special thanks are for my mother: Thank you for being such an incredible mom and a perfect role model for me.

Mr. Genco Fas, thank you for being there for me ∞ .

ABSTRACT

STRUCTURE AND DYNAMICS OF CATABOLITE GENE ACTIVATOR PROTEIN BY MOLECULAR MODELING AND SIMULATIONS

The activities of proteins are regulated by interactions with other molecules through various types of local to global motions. Such internal motions may occur at different time scales and involve complex correlations between the atomic motions. To this end, it is of importance to have the knowledge of different conformational states of a system and to search for the conformational transitions between these states. In the present thesis, the main focus has been to investigate the structural and dynamic features of the plausible apo forms of catabolite activator protein (CAP) in solution, whose crystal structure has not been resolved until now, by modeling and molecular simulations. Molecular dynamics (MD) simulations have been run and analyzed for the unbound monomer and the dimer structures of CAP. Besides the direct conventional analysis of the large amount of the conformations generated by MD simulations, these conformations have further been processed by clustering. The cluster best members have been analyzed with respect to their structural and dynamic peculiarities. The elastic network models are also used for the analysis of the cluster best members. The dynamic fluctuations, the conservation and variation of the fluctuations from one cluster to another have been analyzed. Comparison with the results from the direct analysis of the MD trajectories have been carried out. The combined analysis of the results by various means suggest that the monomeric structure could be a plausible structure for the apo state. The helicity content, the existence of the salt bridges, the shift of a hinge in the DNA binding domain agree with the experimental observations in the apo state, whereas the stabilization of the two domains, cAMP and DNA binding, and their association is enhanced by the DNA binding region upon dimerization. The predicted fluctuations for the cluster best members by elastic network models agree with the conformational space spanned by the cluster best members. The results in general provide evidence for the preexistence of the conformations of the monomer, in the dimer unbound state.

ÖZET

MOLEKÜLER MODELLEME VE SİMÜLASYON İLE KATABOLİT GEN AKTİVATÖR PROTEİNİN YAPISI VE DİNAMİĞİ

Proteinlerin aktivitesi diğer moleküllerle olan etkileşimlerinin sonucu oluşan çeşitli lokal ve global hareketlerle düzenlenir. Atomik hareketlerin karmaşık korelasyonunu (bağlılaşımını) gerektiren bu gibi dahili hareketler değişik zaman ölçeklerinde gerçekleşebilir. Bu amaçla proteinin farklı konformasyonel durumlarını belirlemek ve bu konformasyonlar arası geçiş mekanizmalarını araştırmak önem taşır. Bu çalışmada temel olarak, kristal yapısı henüz bulunmamış olan katabolit aktivatör proteininin (CAP) olası apo formlarının yapısal ve dinamik özellikleri modelleme ve moleküler simülasyon yöntemleriyle araştırılmıştır. CAP'ın ligantsız ve DNA'sız monomer ve dimer yapıları için moleküler dinamik (MD) simülasyonları ve analizleri gerçekleştirilmiştir. MD simülasyonlarının ürettiği yüksek sayıdaki konformasyonların standart analizinin yanısıra, bu konformasyonlar ek olarak kümeleme yöntemi ile analiz edilmiştir. Kümelerin en iyi elemanları yapılarına ve dinamik özelliklerine bakılarak elastik ağyapı modelleriyle incelenmiştir. Dinamik dalgalanmalar, bir kümeden başkasına geçerken dalgalanmaların korunumu ve varyasyonu analiz edilmiştir. Bu analizler MD gidimizi analiz sonuçlarıyla karşılaştırılmıştır. Tüm analizler sonucunda monomer yapının apo form için olası bir yapı olabileceğini işaret etmiştir. Alfa sarmal oluşumu, tuz köprülerinin varlığı ve DNA bağlanma bölgesindeki destek noktanın (hinge) kayması apo formu ile ilgili var olan deneysel verilerle uyumaktadır. cAMP ve DNA bağlanma bölgelerinin stabilizasyonunun ve iki bölge arasındaki ilişkinin, ikizleşme (dimerizasyon) sonucunda pekiştiği görülmüştür. En iyi küme elemanlarına uygulanan elastik ağyapı modelleri ile öngörülen dalgalanmalar, kümeleme sonucunda belirlenen en iyi küme elemanlarının konformasyonel uzayda taradığı alanlarla örtüşmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT.....	iv
ÖZET	v
LIST OF FIGURES	viii
LIST OF TABLES	xiv
LIST OF SYMBOLS/ABBREVIATIONS.....	xv
1. INTRODUCTION.....	1
2. CATABOLITE ACTIVATOR PROTEIN	3
2.1. Biological Context	3
2.2. Structure	4
3. METHODS	9
3.1. Molecular Dynamics	9
3.1.1. Force Field	9
3.1.2. Energy Minimization	11
3.1.3. Equation of Motion.....	12
3.1.4. Computational Details	14
3.2. Cluster Analysis.....	17
3.3. The Gaussian Network Model.....	18
3.4. The Anisotropic Network Model.....	23
4. RESULTS AND DISCUSSION	25
4.1. Molecular Dynamics Trajectory Analysis.....	25
4.1.1. Root -Mean-Square Deviation (RMSD)	25
4.1.2. Equilibrium Fluctuations	27
4.1.2.1. Mean-square fluctuations	27
4.1.2.2. Cross-correlation of fluctuations.....	30
4.1.2.3. Dihedral angle fluctuations.....	31
4.1.3. Salt Bridge Formations	34
4.2. Clustering MD Trajectory Snapshots.....	36
4.3. Analysis of the Best Members of the Clusters	43
4.3.1. Clusters of monomer MD snapshots with cluster radius of 3.5 Å ...	44

4.3.2.	Clusters of dimer MD snapshots with cluster radius of 3.5 Å	49
4.3.3.	RMSD of crystal structures versus cluster best members with cluster radius of 2.7 Å.....	53
4.3.4.	GNM.....	54
4.3.5.	ANM.....	60
4.3.6.	Secondary Structure Assignments.....	62
4.4.	General Discussion	63
5.	CONCLUSIONS AND RECOMMENDATIONS.....	65
5.1.	Conclusions	65
5.2.	Recommendations.....	65
APPENDIX A		66
A.1	Salt Bridges	66
A.2	Cluster Analysis.....	71
A.2.1	Monomer MD Simulation Snapshots	71
A.2.2	Dimer MD Simulation Snapshots.....	72
A.2.3	Monomer & Dimer MD Simulation Snapshots	74
REFERENCES		76

LIST OF FIGURES

Figure 2.1. The lac operon in E. Coli (cAMP is shown with red triangle) [8].	3
Figure 2.2. cAMP synthesis [8].....	4
Figure 2.3. Crystal structure of CAP-DNA complex at a resolution of 3.00 Å (PDB entry code is 1CGP [12, 13].....	5
Figure 2.4. CAP-cAMP complexes which have been found experimentally to exist.	6
Figure 2.5. Secondary structure assignments by DSSP and Stride [13].....	7
Figure 2.6. Sequence coloured by residue conservation [13].	7
Figure 2.7. Protein-ligand interactions [13].....	8
Figure 3.1. Schematic representation of the bonded and non-bonded interactions [24].	11
Figure 3.2. The algorithm to generate MD trajectories	14
Figure 3.3. TIP3P Water Model Parameter Set. The model uses a total of three sites for the electrostatic interactions [26]	15
Figure 3.4. Minimized starting structure of CAP _{dimer} in a truncated octahedron box (CAP _{dimer} ^{min}).	16
Figure 3.5. Virtual bond model	18
Figure 3.6. Description of the GNM. (a) Schematic and (b) EN representations of the equilibrium positions of the <i>i</i> th and <i>j</i> th nodes [37].	19
Figure 3.7. Elastic Network (EN) Model of CAP in the unbound form.	23

Figure 4.1.	Root-mean-square deviation (RMSD) of residue positions of (a) CAP _{monomer} (b) CAP _{dimer} , from the reference structures CAP _{monomer} ^{min} and CAP _{dimer} ^{min} , calculated for a time span of 60 ns (for each frame, every 0.4 ps).	26
Figure 4.2.	Comparison of (a) ms fluctuations and (b) B-factors for CAP _{monomer} and CAP _{dimer}	28
Figure 4.3.	Comparison of B-factors for CAP _{monomer} (subunit A) and CAP _{dimer} (subunit A and B), calculated from the MD trajectories with experimental B-factors of 1G6N, 2GZW and 2CGP [19, 40, 41].	29
Figure 4.4.	Cross correlation values calculated from the MD trajectories of CAP _{monomer}	30
Figure 4.5.	Cross correlation values calculated from the MD trajectories of CAP _{dimer} ...	31
Figure 4.6.	Ms fluctuations of phi angles calculated from the MD trajectories of CAP _{monomer} and CAP _{dimer}	32
Figure 4.7.	Correlation coefficients of phi angles calculated from the CAP _{monomer} MD trajectory.	33
Figure 4.8.	Correlation coefficients of phi angles calculated from monomer A of CAP _{dimer} MD trajectory.	33
Figure 4.9.	Correlation coefficients of phi angles calculated from monomer B of CAP _{dimer} MD trajectory.	34
Figure 4.10.	Cluster evolution in time for CAP _{monomer} on the RMSD plot with cluster radius of 3.0 Å.....	40
Figure 4.11.	Cluster evolution in time for CAP _{monomer} on the RMSD plot with cluster radius of 3.5 Å.....	40
Figure 4.12.	Cluster evolution in time for CAP _{dimer} on the RMSD plot with cluster radius of 3.0 Å.....	41
Figure 4.13.	Cluster evolution in time for CAP _{dimer} on the RMSD plot with cluster radius of 3.5 Å.....	41

Figure 4.14. Cluster evolution in time for CAP _{monomer} and CAP _{dimer} subunit A on the RMSD plot with cluster radius of 3.0 Å.....	42
Figure 4.15. Cluster evolution in time for CAP _{monomer} and CAP _{dimer} subunit A on the RMSD plot with cluster radius of 3.5 Å.....	43
Figure 4.16. (a) Superimposition of the cluster best members of CAP _{monomer} to the first cluster's best member and (b) to the X-ray structure 1CGP. Cluster radius is 3.5 Å.....	45
Figure 4.17. RMSDs of the cluster best members of CAP _{monomer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.	46
Figure 4.18. RMSDs of the cluster best members of CAP _{monomer} from the X-ray structure 1G6N. Superimposed structures are given below in the same order.	46
Figure 4.19. RMSDs of the cluster best members of CAP _{monomer} from the MD snapshot 1G6N (13 ns). Superimposed structures are given below in the same order.	47
Figure 4.20. RMSDs of the cluster best members of CAP _{monomer} from the MD snapshot 1G6N (40 ns). Superimposed structures are given below in the same order.	47
Figure 4.21. RMSDs of the cluster best members of CAP _{monomer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.	48
Figure 4.22. RMSDs of the cluster best members of CAP _{monomer} from the X-ray structure 2GZW. Superimposed structures are given below in the same order.	48
Figure 4.23. (a) Superimposition of the cluster best members of CAP _{dimer} to the first cluster's best member and (b) to the X-ray structure 1CGP. Cluster radius is 3.5 Å.....	49
Figure 4.24. RMSDs of the cluster best members of CAP _{dimer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.	50
Figure 4.25. RMSDs of the cluster best members of CAP _{dimer} from the X-ray structure 1G6N. Superimposed structures are given below in the same order.	50

Figure 4.26. RMSDs of the cluster best members of CAP _{dimer} from the MD snapshot 1G6N (13 ns). Superimposed structures are given below in the same order.	51
Figure 4.27. RMSDs of the cluster best members of CAP _{dimer} from the MD snapshot 1G6N (40 ns). Superimposed structures are given below in the same order.	51
Figure 4.28. RMSDs of the cluster best members of CAP _{dimer} from the X-ray structure 2CGP. Superimposed structures are given below in the same order.	52
Figure 4.29. RMSDs of the cluster best members of CAP _{dimer} from the X-ray structure 2GZW. Superimposed structures are given below in the same order.	52
Figure 4.30. Monomer MD cluster best members (cluster radius 2.7 Å) versus the crystal structures and other MD snapshots of CAP.	53
Figure 4.31. Dimer MD cluster best members (cluster radius 2.7 Å) versus the crystal structures and other MD snapshots of CAP.	54
Figure 4.32. GNM (a) slow mode and (b) cross-correlation map of the monomer MD cluster best member 1 for the cluster radius 3.5 Å.	55
Figure 4.33. GNM (a) slow mode plot and (b) cross-correlation map of the monomer MD cluster best member 2 for the cluster radius 3.5 Å.	56
Figure 4.34. GNM (a) slow mode plot and (b) cross-correlation map of the monomer MD cluster best member 3 for the cluster radius 3.5 Å.	57
Figure 4.35. GNM (a) slow mode plot and (b) cross-correlation map of the dimer MD cluster best member 1 for the cluster radius 3.5 Å.	58
Figure 4.36. GNM (a) slow mode plot and (b) cross-correlation map of the dimer MD cluster best member 2 for the cluster radius 3.5 Å.	59
Figure 4.37. GNM cross-correlation map of the crystal structure 1CGP.	60
Figure 4.38. ANM (a) mode 1 and (b) mode 2 of monomer MD Best Members.	61
Figure 4.39. ANM (a) mode 1 and (b) mode 2 of dimer MD Best Members.	61

Figure 4.40. Secondary structure assignments to the CAP _{monomer} and CAP _{dimer} (subunit A) trajectory cluster best members.....	62
Figure A.1. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in isolated monomer MD.....	66
Figure A.2. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in isolated monomer MD.....	67
Figure A.3. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in isolated monomer MD.....	67
Figure A.4. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in monomer A of dimer MD.....	68
Figure A.5. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in monomer A of dimer MD.....	68
Figure A.6. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in monomer A of dimer MD.....	69
Figure A.7. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in monomer B of dimer MD.....	69
Figure A.8. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in monomer B of dimer MD.....	70
Figure A.9. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in monomer B of dimer MD.....	70
Figure A.10. Cluster evolution in time for CAP _{monomer} on the RMSD plot with cluster of radius of 2.0 Å	71

Figure A.11. Cluster evolution in time for CAP _{monomer} on the RMSD plot with cluster radius of 2.5 Å	71
Figure A.12. Cluster evolution in time for CAP _{monomer} on the RMSD plot with cluster radius of 2.7 Å	72
Figure A.13. Cluster evolution in time for CAP _{dimer} on the RMSD plot with cluster radius of 2.0 Å	72
Figure A.14. Cluster evolution in time for CAP _{dimer} on the RMSD plot with cluster radius of 2.5 Å	73
Figure A.15. Cluster evolution in time for CAP _{dimer} on the RMSD plot with cluster radius of 2.7 Å.....	73
Figure A.16. Cluster evolution in time for CAP _{monomer} and CAP _{dimer} subunit A on the RMSD plot with cluster radius of 2.0 Å	74
Figure A.17. Cluster evolution in time for CAP _{monomer} and CAP _{dimer} subunit A on the RMSD plot with cluster radius of 2.5 Å	74
Figure A.18. Cluster evolution in time for CAP _{monomer} and CAP _{dimer} subunit A on the RMSD plot with cluster radius of 2.7 Å	75

LIST OF TABLES

Table 3.1. Details of the models constructed.....	15
Table 4.1. List of the salt bridges	34
Table 4.2. Overview of the clusters formed	37
Table 4.3. Cluster members and best members (BM) resulting from cluster radius of 2.7 Å	37
Table 4.4. Cluster members and best members (BM) resulting from cluster radius of 3.0 Å	38
Table 4.5. Cluster members and best members (BM) resulting from cluster radius of 3.5 Å	38
Table 4.6. Members of the clusters for the monomer and dimer (together) cluster analysis with a cluster radius of 3.0 Å	39
Table 4.7. Members of the clusters for the monomer and dimer (together) cluster analysis with a cluster radius of 3.5 Å	39

LIST OF SYMBOLS/ABBREVIATIONS

a_i	Acceleration of particle i
A_{ij}	van der Waals repulsion parameter
B_i	B-factor
B_{ij}	van der Waals attraction parameter
$C(i,j)$	Orientalional cross-correlations
C^α	Alpha carbon atom
d_i	Distance between the coordinates of atom i in the two structures
F_i	Force exerted on particle i
H	Hessian matrix
k	Mode
k_B	Boltzman constant
k_l	Force constant for the bond
k_θ	Force constant for the bond angle
l_{eq}	Equilibrium bond length
l_i	Bond length between atoms $i-1$ and i
m	Mass of particle i
N	Number of atoms
q_i	Partial atomic charge of atom i
r_c	Cutoff radius or interaction range
r_i	Position vector of residue i
r_{ij}	Interparticle distance
R_i	Instantaneous position
R_i^0	Equilibrium position vectors of node i
R_{ij}	Distance vector between residues i and j
t	Time
T	Temperature
u_k	Eigenvector
U	Force field, potential energy function
v_i	Velocity of atom i

V	Overall potential energy
V_n	Force constant
ΔR	Matrix of instantaneous fluctuations of ΔR_i
ΔR_i	Displacement vector for R_i
ΔR_{ij}	Fluctuation in the distance vector R_{ij}
$\text{\AA}$	Angstrom
α	Alpha helix
β	Beta strand
ϵ	Dielectric constant
γ	Uniform Hookean force constant
λ_k	Eigenvalue
Γ	Kirchhoff matrix
θ_i	Instantaneous bond angle of atom i
θ_{eq}	Equilibrium bond angle
ϕ_i, φ_i	Dihedral (torsional) angle of bond i
3D	Three dimensional
Amber	Assisted Model Building with Energy Refinement
ANM	Anisotropic Network Model
CAP	Catabolite (gene) activator protein
cAMP	Cyclic adenosine mono phosphate
CRP	cAMP receptor protein, cAMP regulatory protein, catabolite repressor protein, catabolite regulatory protein
CBD	cAMP binding domain
D	Asp, aspartic acid
DBD	DNA binding domain
E	Glu, glutamic acid
EN	Elastic Network
GNM	Gaussian Network Model
MD	Molecular Dynamics
ms	Mean-square
PDB	Protein Data Bank

R	Arg, arginine
RMSD	Root mean square deviation
RNAp	RNA polymerase
VMD	Visual Molecular Dynamics

1. INTRODUCTION

The importance of proteins and nucleic acids lies in the diverse functions they undertake in the course of life. They are largely responsible for the expression and transmission of biological information. The activities of a number of proteins are regulated by interactions with other molecules, through certain types of internal motions, to accomplish their specific functions. Such internal motions may occur at different time scales and involve complex correlations between the atomic motions. For example, a specific ligand will bind a particular protein to induce a conformational transition in order to activate binding and start a cascade of processes, ending in an expression of a certain gene [1, 2]. The study of the mechanism underlying these conformational transitions among multiple conformations of a protein is important for the understanding of its function. To this end, the relationship between sequence, structure and dynamics, which is intimately related to the function of proteins, play a crucial role.

Molecular dynamics (MD) simulations have given many insights into the internal motions of biomolecules, by providing information on the time evolution of their conformations. They also provide, kinetic and thermodynamic information about the systems studied. They can be used to address specific questions about the properties of a model system constructed in silico, often more easily than experiments on the actual system (especially when experimental methods are inadequate). Dynamic simulations are valuable tools in extending our understanding of model systems, since they can be utilized to quantify the properties of a system at a precision and on a time scale that is otherwise inaccessible [3]. An MD trajectory generally provides data only at the level of atomic positions, velocities, and single-point energies. To obtain the macroscopic properties in which one is usually interested requires the application of statistical mechanics, which connects microscopic simulations and macroscopic observables [2].

MD simulations, along with a range of complementary computational approaches, like Gaussian network model (GNM) [4, 5], anisotropic network model (ANM) [6] and cluster analysis [7] may offer insights into molecular fluctuations, conformational changes,

and allosteric mechanisms which are relevant for the function of the protein. In the present work, the main focus has been to investigate the structural and dynamic features of the plausible apo forms of catabolite activator protein (CAP) in solution, whose crystal structure has not been resolved until now.

Along, the dynamic behaviour of the structures with respect to binding in general has also been of interest. MD simulations have been run and analyzed for the unbound monomer and unbound dimer structures of CAP. The large amount of the conformations observed in time generated by MD simulations have further been processed by a cluster analysis. The most cooperative modes of motion have been obtained from MD trajectories by the elastic network models, GNM and ANM.

2. CATABOLITE ACTIVATOR PROTEIN

2.1. Biological Context

Organisms frequently compete with each other for scarce resources in order to survive. To take advantage of the nutrients available in the environment, they often switch between different sources of carbon and energy by altering gene expression, which in turn affects the type and amount of protein molecules produced in the cell.

When the intestinal bacterium *Escherichia coli* (*E. coli*) is exposed to both lactose and glucose as sources of carbon, the organisms first metabolize the glucose and then temporarily stop growing until the genes of the *lac* operon become induced to provide the ability to metabolize lactose as a usable energy source. Here glucose and lactose levels control the initiation of transcription of the *lac* operon through their effects on the lac repressor protein and catabolite activator protein (CAP) (Figure 2.1). It is now known that catabolite repression is in fact mediated by CAP in conjunction with cyclic adenosine mono phosphate (cAMP). CAP is also referred to as the cAMP regulatory protein (CRP)[8].

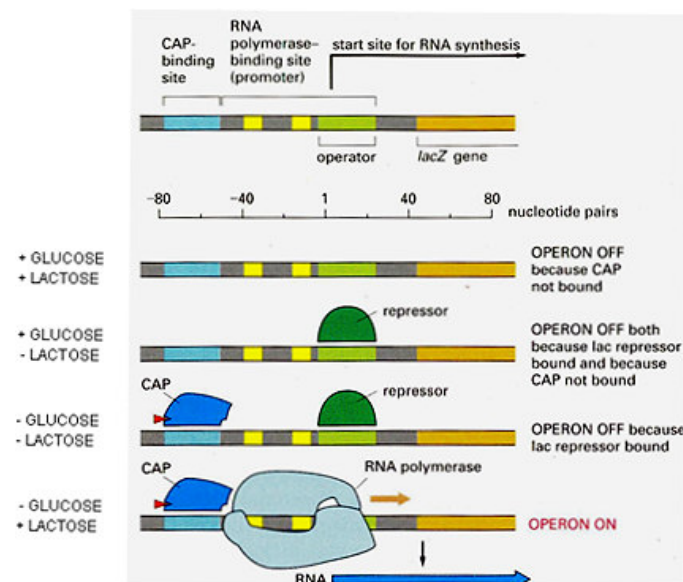


Figure 2.1. The lac operon in *E. Coli* (cAMP is shown with red triangle) [8].

In the presence of cAMP and its binding protein CAP, the RNA polymerase (RNAP) can transcribe the structural genes *lacZ*, *lacY*, and *lacA*, and the polycistronic (containing the coding sequences for the various proteins) mRNA molecule formed can be translated into the corresponding protein molecules β -galactosidase, permease, and transacetylase, allowing for the catabolism of lactose. By an independent mechanism, the bacterium accumulates cAMP only when it is starved for a source of carbon. In the presence of glucose (or of glycerol in concentrations sufficient for growth) the bacteria will lack sufficient cAMP to bind to CAP because the glucose inhibits adenyl cyclase, the enzyme that converts ATP to cAMP (Figure 2.2) [9, 10].

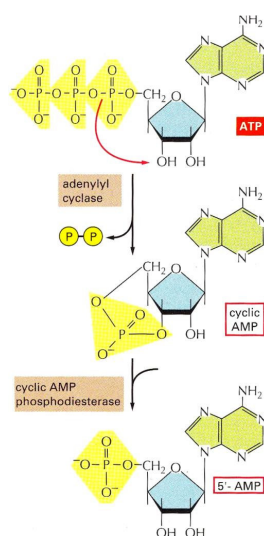


Figure 2.2. cAMP synthesis [8].

2.2. Structure

Catabolite activator protein (CAP) is a dimer made up of identical subunits (homo dimer), each consisting of 209 residues. The individual subunits fold into two domains. The aminoproximal domain, composed of amino acid residues 1-133, contains the cAMP binding site. The carboxyl-proximal domain, composed of amino acid residues 139-209, contains the helix-turn-helix DNA binding motif. Amino acid residues 134-138 form a flexible hinge that covalently couples these two domains.

In Figure 2.3, cAMP binding domain (CBD, residues 1-134), hinge region (residues 135-138) and DNA binding domain (DBD, residues 139-209) are represented in blue, green and dark pink in subunit A; and in light blue, yellow and light pink in subunit B, respectively. Figure is created by the VMD program [11].

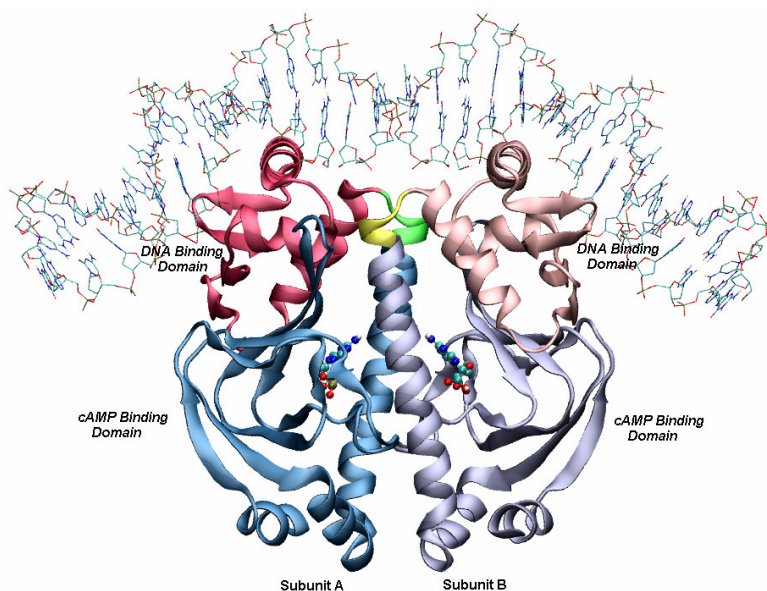


Figure 2.3. Crystal structure of CAP-DNA complex at a resolution of 3.00 Å (PDB entry code is 1CGP [12, 13].

Each CAP subunit contains a primary and a secondary cyclic nucleotide binding site, where anti and syn conformations of cAMP bind, respectively. The primary binding site, which has a μM affinity for cAMP, is located in the CBD, within the β -roll, and it is lined by the C-helices of both subunits: binding of cAMP (in its anti conformation) triggers a reorientation of the $\beta 4/\beta 5$ loops and of the C-helices, which leads to an allosteric change of the hinges that repositions the DBDs and results in DNA binding. Selectivity toward the target DNA sequence is achieved by cAMP binding to the primary sites. The secondary binding site, with mM affinity for cAMP that binds in its syn conformation, is formed by residues from both domains of the same subunit and from the hinge region of the opposite subunit. DNA binding to CAP depends on cAMP concentration.

In solutions containing cAMP, CAP undergoes sequential conformation changes from the inactive apo-form through the $\text{CAP}:(\text{cAMP})_1$ complex to the $\text{CAP}:(\text{cAMP})_2$

complex depending on the cAMP concentration. Apo-CAP binds DNA with low affinity and no apparent sequence specificity. The CAP:(cAMP)₂ complex exhibits high affinity, sequence specific DNA binding (consensus sequence is 5'-AAATGTGATCT-3') and interacts with RNA polymerase (RNAP), whether free in solution or complexed with DNA. The results of genetic, biochemical and biophysical studies have helped to uncover many of the details of cAMP-mediated allosteric control over CAP conformation and activity as a transcription factor. These studies indicate that cAMP binding produces only small, but significant, changes in CAP structure; changes that include subunit realignment and concerted motion of the secondary structure elements within the C-terminal DNA binding domain of each subunit. These adjustments promote CAP surface-patch interaction with RNA polymerase and protrusion of the F-helix to promote CAP site-specific interaction with DNA. Interactions between CAP and RNA polymerase at CAP-dependent promoters produce active ternary transcription complexes [14, 15, 16, 17, 18, 19].

In Figure 2.4 complexes of CAP-cAMP that are known to exist are given. Three dimensional structures of the apo CAP, CAP:(cAMP)₁ and CAP:(cAMP)₃ are not resolved until now.

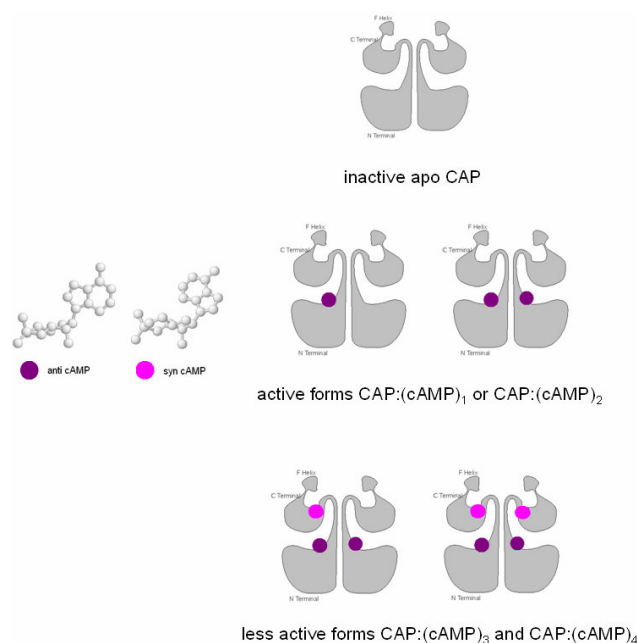


Figure 2.4. CAP-cAMP complexes which have been found experimentally to exist.

In Figure 2.5 and Figure 2.6 the primary sequence and the secondary structure of CAP is given, respectively. Details of the cAMP-CAP interaction is given in Figure 2.7 [13].

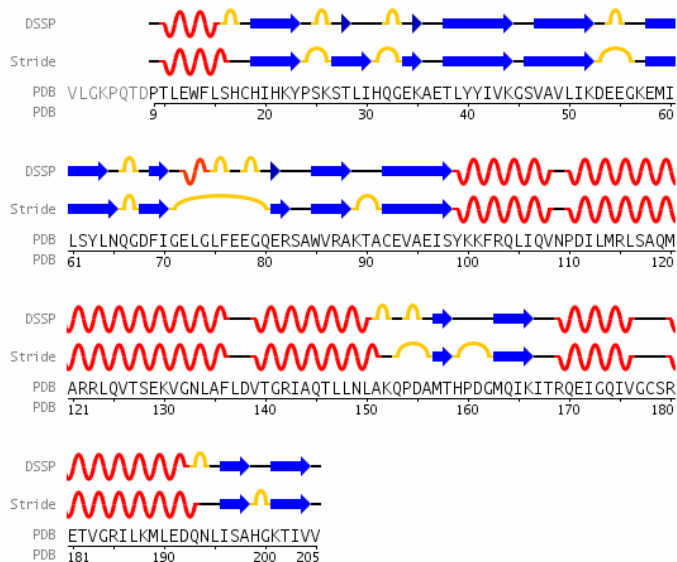


Figure 2.5. Secondary structure assignments by DSSP and Stride [13].

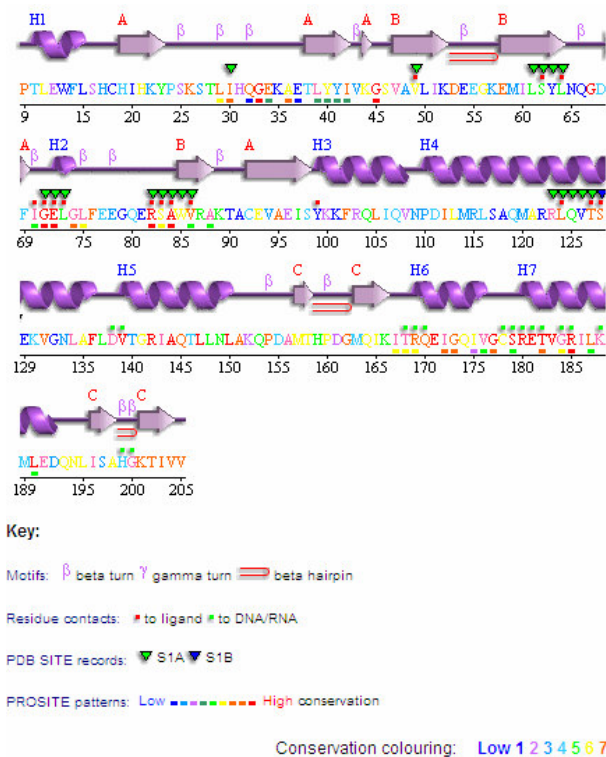


Figure 2.6. Sequence coloured by residue conservation [13].

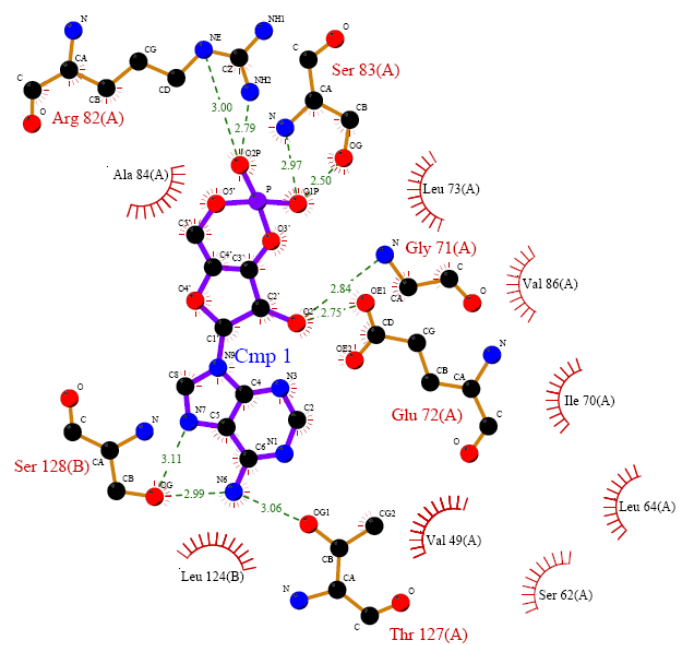


Figure 2.7. Protein-ligand interactions [13].

3. METHODS

The basic concepts of the methods used in this work are introduced in this chapter. These methods are molecular dynamics simulations [20, 21, 22], cluster analysis [7], Gaussian Network Model [4, 5] and Anisotropic Network Model [6].

3.1. Molecular Dynamics

Molecular dynamics simulations calculate the time dependent behaviour of a molecular system and start with a knowledge of the energy of the system as a function of the atomic coordinates. The forces acting on the atoms of the system, which are related to the first derivatives of the potential with respect to the atom positions, can be used to calculate the dynamic behaviour of the system by solving Newton's equations of motion for the atoms as a function of time. The classical newtonian equations of motion are adequate for the biomolecules in general, except for some cases where quantum corrections should be considered [21].

3.1.1. Force Field

The potential energy function or force field, U , is a function of the Cartesian coordinate set for N atoms, $(r_1, r_2, \dots, r_N)$, and is composed of bonding terms representing bond lengths, bond angles, and torsional angles, and non-bonding terms consisting of van der Waals and electrostatic interactions as seen in equation (3.1) (Figure 3.1).

$$\begin{aligned}
 U(r_1, r_2, \dots, r_N) = & \sum_{\substack{\text{bonds} \\ \text{bond stretching}}} \frac{1}{2} k_l (l_i - l_{eq})^2 \\
 & + \sum_{\substack{\text{bond angles} \\ \text{angle bending}}} \frac{1}{2} k_\theta (\theta_i - \theta_{eq})^2
 \end{aligned} \tag{3.1}$$

$$\begin{aligned}
& + \sum_{\substack{\text{torsions} \\ \text{bond rotation}}} \frac{1}{2} V_n [1 + \cos(n\phi_i - \gamma)] \\
& + \sum_{i=1}^N \sum_{j=i+1}^N \left\{ 4\varepsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] + \left(\frac{q_i q_j}{\varepsilon r_{ij}} \right) \right\}
\end{aligned}$$

Lennard-Jones
electrostatic

The first term in equation (3.1) represents instantaneous displacements of bond length l_i (length of the bond between atoms $i-1$ and i) from the equilibrium bond length, l_{eq} , by a Hooke's law (harmonic) potential. Harmonic potential is the first approximation to the energy of a bond as a function of its length. The bond force constant k_l determines the flexibility of the bond. The second term is the energy associated with alteration of bond angles, is also represented by a harmonic potential. k_θ is the force constant for the bond angles. θ_i and θ_{eq} are bond angle and equilibrium bond angle respectively. The third term is the torsion angle potential function for rotations about bonds. It is assumed to be periodic and modelled by a cosine or sum over cosine functions. ϕ_i is the dihedral angle. V_n is the corresponding force constant; the phase angle γ takes values of either 0° or 180° .

The nonbonded part of the potential is represented by van der Waals (A_{ij}) and London dispersion terms (B_{ij}) and interactions between partial atomic charges (q_i and q_j). ε is the dielectric constant that takes into account the effect of the medium that is not explicitly represented and usually equals to 1.0 in a typical solvated environment where solvent is represented explicitly. The nonbonded terms are calculated for all atom pairs that are either separated by more than three bonds or are not bonded [21, 23].

The internal coordinates used in the potential function; bond lengths, l_i (equation 3.2), bond angles, θ_i (equation 3.3), dihedral angles, ϕ_i (equation 3.4), and interparticle distances, r_{ij} (equation 3.5) are calculated from the Cartesian atomic coordinates.

$$l_i = |r_i - r_{i-1}| \quad 1 \leq i \leq n \quad (3.2)$$

$$\cos \theta_i = \frac{l_i \cdot l_{i+1}}{l_i l_{i+1}} \quad 1 \leq i \leq n-1 \quad (3.3)$$

$$\cos \varphi_i = \frac{\mathbf{l}_{i-1} \times \mathbf{l}_i}{|\mathbf{l}_{i-1} \times \mathbf{l}_i|} \cdot \frac{\mathbf{l}_i \times \mathbf{l}_{i+1}}{|\mathbf{l}_i \times \mathbf{l}_{i+1}|} \quad 2 \leq i \leq n-1 \quad (3.4)$$

$$r_{ij} = |\mathbf{r}_j - \mathbf{r}_i| \quad (3.5)$$

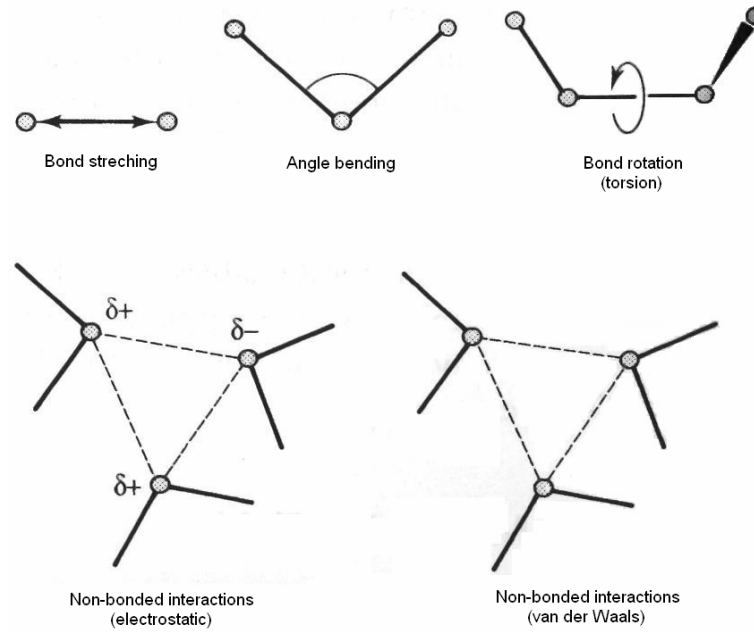


Figure 3.1. Schematic representation of the bonded and non-bonded interactions [24].

3.1.2. Energy Minimization

Before starting the simulation, the molecular structure should be refined by relieving local stresses due to overlaps of non-bonded atoms (steric conflicts) and relaxing bond lengths and bond angles to their canonical values. Energy minimization is performed given a set of N independent variables, $\mathbf{r}$, where $\mathbf{r} = (r_1, r_2, \dots, r_N)$, the task is to find $\mathbf{r}_{\min}$, for which the potential energy function, U , has its global minimum. This problem is a nonlinear optimization problem. Considering a system consisting of a protein of approximately 4000 atoms and a solvent (water) with 10000 atoms, the biomolecular system will be described by 42000 variables. It is thus impossible to locate the global

minimum. The most popular minimization methods include those that use derivatives of various orders, including the first order steepest descent and conjugate gradient methods and the second order Newton-Raphson method, which is a less useful one for nonquadratic systems with many local minima [2].

3.1.3. Equation of Motion

Newton's equations of motion gives the most exact and detailed information when solved for the atoms of the system and any surrounding solvent to generate successive conformations of the system [21]. Newton's equation of motion is given by

$$F_i = m_i a_i \quad (3.6)$$

where F_i is the force exerted on particle i , m_i is the mass of particle i and a_i is the acceleration of particle i . When a_i is expressed in terms of the second derivative of position of the particle r_i with respect to time, the equation becomes

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (3.7)$$

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

$$F_i = - \text{grad } U_i = - \nabla U_i \quad (3.8)$$

When the equations (3.7) and (3.8) are combined in equation (3.9), Newton's equation of motion can relate the derivative of the potential to the change in the position of the particle i as a function of time:

$$\frac{d^2 r_i}{dt^2} = - \frac{1}{m_i} \frac{\partial U(r_1, r_2, \dots, r_N)}{\partial r_i} = a_i \quad (3.9)$$

To calculate an MD trajectory the requirements are: the initial positions of the atoms, an initial distribution of velocities and the acceleration. The initial positions of the atoms

can be obtained from experimental structures, such as the X-ray crystal structure or NMR spectroscopy solution structure of the protein. The initial distribution of velocities are calculated using the Maxwell-Boltzmann distribution (equation 3.10), which gives the probability that an atom i has a velocity v_x in the x -direction at a temperature T , and corrected so that there is no overall momentum (equation 3.11). Initial distribution of the acceleration is determined by the gradient of the potential energy function (equation 3.12).

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad (3.10)$$

$$p = \sum_{i=1}^N m_i v_i = 0 \quad (3.11)$$

$$a_i = -\frac{1}{m_i} \frac{\partial U(r_1, r_2, \dots, r_N)}{\partial r_i} \quad (3.12)$$

The temperature is related to the kinetic energy of the system by the relation, where N is the number of atoms in the system, $\langle v_i^2 \rangle$ is the average velocity squared of the i th atom and k_B is the Boltzmann constant:

$$T = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{1}{2} m_i \langle v_i^2 \rangle \quad (3.13)$$

The last form of the equation of motion (equation 3.9) can be integrated by the Verlet algorithm [25], which is one of the most widely used integration algorithms. In Verlet algorithm, position of each atom is expressed by Taylor series expansion (equations 3.14 and 3.15) whose combination yields an expression for $r_i(t + \Delta t)$ (equation 3.16):

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2} a_i(t)\Delta t^2 + \dots \quad (3.14)$$

$$r_i(t - \Delta t) = r_i(t) - v_i(t)\Delta t + \frac{1}{2} a_i(t)\Delta t^2 - \dots \quad (3.15)$$

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)\Delta t^2 \quad (3.16)$$

Since the velocity does not appear in the last expression obtained, this algorithm does not require the usage of explicit velocities. For estimating the kinetic energy, velocities can be calculated from the following expression, which is obtained by the subtraction of equation 3.15 from equation 3.14:

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} \quad (3.17)$$

The summary of the algorithm to generate MD trajectories is given in Figure 3.. The equilibration period is considered finished when the temperature is stable for longer than about 10 ps. Continued integration of the equations of motion after equilibration generates the coordinates and velocities of the atoms as a function of time.

Algorithm 1 The algorithm to generate MD trajectories.

- 1: Given $r_i(t = 0)$ for all atom i
 - 2: Assign $v_i(t = 0)$ for all atom i , based on Boltzmann distribution at given T
 - 3: Calculate $r_i(\delta t) = r_i(t = 0) + \delta t v_i(t = 0)$
 - 4: Using new $r_i(\delta t)$ evaluate the total potential U_i on atom i
 - 5: Calculate $a_i(\delta t) = -\nabla U_i / m_i$
 - 6: Start Verlet algorithm using $r_i(t = 0)$, $r_i(\delta t)$, and $a_i(\delta t)$
 - 7: Repeat for all atoms, including solvent
 - 8: Go to **Step 4** and repeat the last three steps for $\sim 10^6$ successive times (MD steps)
-

Figure 3.2. The algorithm to generate MD trajectories

3.1.4. Computational Details

Molecular dynamics simulations were performed using the program Amber 8 [20, 22], and the force field parameter sets Amber ff03 [23] and TIP3P [26, Figure 3.3] for the proteins/ions, and water respectively.

$r(\text{OH}), \text{\AA}$	0.9572
HOH, deg	104.52
$q(\text{O})$	-0.834
$q(\text{H})$	0.417

Figure 3.3. TIP3P Water Model Parameter Set. The model uses a total of three sites for the electrostatic interactions [26]

MD trajectories of 60 nanoseconds (ns) have been generated for the two forms of unbound CAP in solution at 300 K temperature. The models used as the starting configurations are the unbound form of CAP in monomer (subunit A) form, **CAP_{monomer}**, and the unbound form of CAP in dimer form (subunits A and B), **CAP_{dimer}**. These structures are constructed by the removal of the ligands and the DNA from the X-ray structure of CAP-DNA complex at a resolution of 3.00 Å (PDB entry code 1CGP). Protonation states and pK of charged residues are predicted using H⁺⁺ server [27, 28, 29]. To neutralize the systems, **CAP_{monomer}** with overall charge of +6 and **CAP_{dimer}** with overall charge of +12 at the given pH, 6 and 12 Cl⁻ ions are added, respectively. The terminal residues are always left in their standard protonation states, even though their pKs may indicate otherwise. The details of the models are given in Table 3.1.

Table 3.1. Details of the models constructed

	based on					
model	PDB entry	no. of amino acids	box dimension (Å)	no. of water molecules	no. of Cl ⁻ ions	no. of total atoms
CAP_{monomer}	1CGP	197	77.93	28 839	6	32 009
CAP_{dimer}	1CGP	394	80.32	33 897	12	40 237

The truncated octahedron periodic boundary conditions were applied to the systems. The size of the simulation box is 77.93×77.93×77.93 Å³ for **CAP_{monomer}** and 80.32×80.32×80.32 Å³ for **CAP_{dimer}**. The SHAKE algorithm [30] was used for maintaining the bond constraints for the bonds involving hydrogens, and the interactions

involving hydrogens were omitted. The long-range interactions in electrostatic terms were corrected using the particle mesh Ewald method [31] with a cutoff distance of 9 Å.

Before starting the simulations, energy minimization of the systems are performed using 50 cycles of steepest descent algorithm to get closer to the minimum, followed by the conjugate gradient method to reach the minimum. The approximate times for the minimization of **CAP_{monomer}** and **CAP_{dimer}** systems to complete were 22 and 25 picoseconds (ps), respectively, with the resulting structures named as **CAP_{monomer}^{min}** and **CAP_{dimer}^{min}** (see Figure 3.4). In the steepest descent algorithm, the closer one gets to a minimum, the worse the performance becomes. Thus the algorithm may get trapped and oscillate back and forth endlessly. This deficiency is overcome by the conjugate gradient method, which takes into account the previous steps during the minimization.

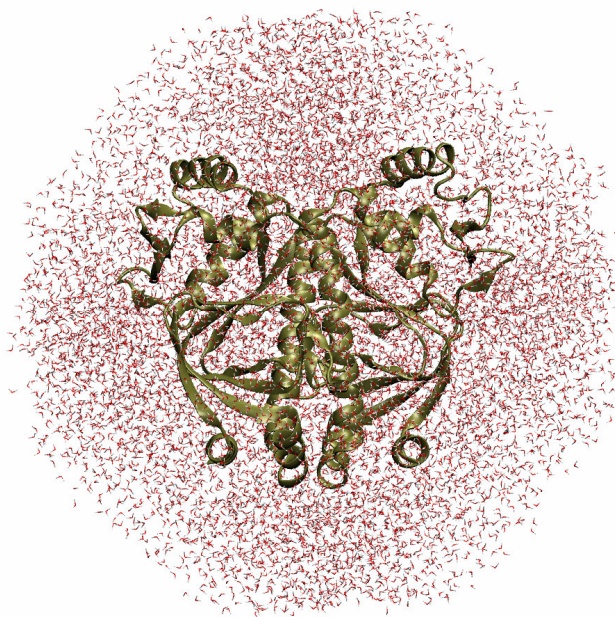


Figure 3.4. Minimized starting structure of **CAP_{dimer}** in a truncated octahedron box (**CAP_{dimer}^{min}**).

MD simulations are performed in the NPT (constant number of atoms, constant P, constant T) ensemble, using the Berendsen thermostat and barostat (weak coupling bath) [32] at $T = 300$ K and $p = 1$ bar. Initial velocity assignments are done according to the Maxwell-Boltzmann distribution at a temperature of 10 K for each system which was

gradually increased to 300 K. CAP_{monomer} and CAP_{dimer} are simulated for 60 ns with a time step of 2 femtoseconds (fs).

3.2. Cluster Analysis

A representative set of conformations can be selected among the large amount of conformations generated by MD simulations for subsequent analysis using the cluster analysis. Cluster analysis groups objects according to a similarity measure, which is root mean square distance (RMSD) in this case. In addition to RMSDs, a variety of parameters may be clustered, including dihedral angles and potential energies.

The cluster analysis procedure requires a similarity matrix in which each element represents the structural difference between a pair of structures. A similarity matrix is constructed by measuring the distance between frames using the root-mean-squared deviation (RMSD):

$$RMSD = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad (3.18)$$

where N is the number of atoms over which the RMSD is measured and d_i is the distance between the coordinates of atom i in the two structures, when they are superimposed [24].

In this work, MMTSB Toolset's [7] *kclust* utility is used to perform conformational clustering. It uses k-means clustering algorithm. k-means is a high-performance clustering algorithm, it starts by randomly choosing a collection of frames from the trajectory, each of which is assigned to its own cluster. It is then iterated over all other frames. Each frame is assigned to the cluster whose centroid is closest; the centroid for this cluster is then recomputed. The iterative procedure continues until all frames are assigned to their clusters. The number of clusters is a parameter which depends on the cutoff value of RMSD (cluster radius); as RMSD cutoff increases less number of clusters are found by the algorithm.

In this case, conformations generated from two 60 ns molecular dynamics trajectories of the isolated monomer and the dimer forms of CAP are clustered in order to group together “redundant” conformations and examine transitions common to many dynamics frames [33, 34, 35, 36].

3.3. The Gaussian Network Model

The Gaussian network model (GNM) [4, 5] is a coarse-grained one dimensional elastic network (EN) model, where the dynamics of proteins are entirely defined by the network topology of their three-dimensional (3D) structures. GNM assumes that residues undergo Gaussian fluctuations about their mean positions (or their native equilibrium states) and these fluctuations are isotropic, i.e., independent of direction. Interactions of the individual atoms are coarse grained using the virtual bond model where individual residues are represented by their backbone alpha carbon atoms (C^α) and the interactions between different types of amino acids are not differentiated (Figure 3.5).

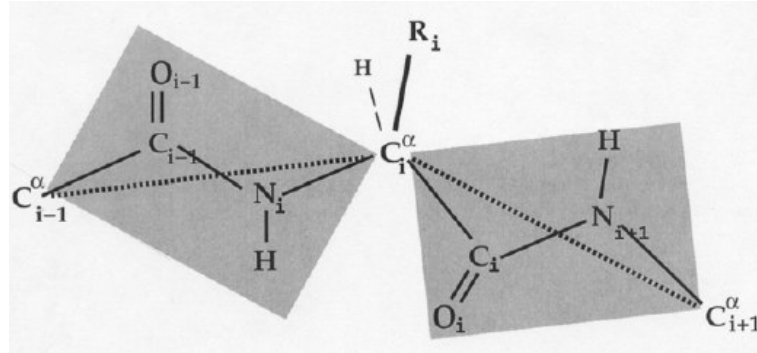


Figure 3.5. Virtual bond model

The C^α atom coordinates define the positions of the nodes and the springs connecting the nodes represent the covalent and noncovalent interactions between the pairs of residues located within an interaction range, or cutoff distance, r_c . If we define the equilibrium position vectors of node i by $\mathbf{R}_i^0$ and the instantaneous position by $\mathbf{R}_i$, the fluctuations from this mean position can be defined by the vector

$$\Delta \mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0 \quad (3.19)$$

The fluctuations in the distance vector $\mathbf{R}_{ij}$ between residues i and j can be expressed as

$$\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i \quad (3.20)$$

$\mathbf{R}_{ij}$ is the instantaneous separation vector and $\mathbf{R}_{ij}^0$ is the equilibrium distance between nodes i and j and [see Figure 3.6 (a)]. Figure 3.6 (b) shows the elastic network (EN) of GNM, where every residue is represented by a node and connected to spatial neighbors by uniform springs [37].

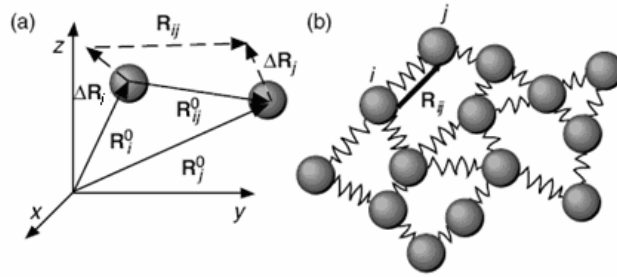


Figure 3.6. Description of the GNM. (a) Schematic and (b) EN representations of the equilibrium positions of the i th and j th nodes [37].

The total potential of the network of N nodes, V is given as

$$V = \frac{\gamma}{2} \left[\sum_i^N \sum_j^N \Gamma_{ij} (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \right] = \frac{\gamma}{2} \left[\sum_i^N \sum_j^N \Gamma_{ij} (\Delta\mathbf{R}_{ij})^2 \right] \quad (3.21)$$

$$V = \frac{\gamma}{2} \text{tr}[\Delta\mathbf{R}^T \mathbf{\Gamma} \Delta\mathbf{R}] \quad (3.22)$$

where γ is a force constant uniform for all springs, $\Delta\mathbf{R}$ is the N -dimensional vector of the fluctuations of $\Delta\mathbf{R}_i$ of the individual sites. $\Delta\mathbf{R}^T$ is its transpose and $\mathbf{\Gamma}$ is the Kirchhoff (connectivity) matrix, by the analogy with the statistic mechanical theory of elasticity originally developed by Flory and coworkers for polymer networks [38]. The dynamics of the interactions is controlled by Kirchhoff matrix, defined as

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (3.23)$$

where r_c is the cutoff distance defining the range of interaction of residues, each residue being represented by its α -carbon atom (C^α), and R_{ij} is the distance between i th and j th residues. A reasonable cutoff distance including all residue pairs within a first interaction shell is 7 Å.

In GNM the Kirchoff matrix is symmetric and sparse, therefore the determinant of it, $|\mathbf{\Gamma}|$, is zero, and the inverse $\mathbf{\Gamma}^{-1}$, which scales with the covariance, cannot be computed directly, but instead calculated by eigenvalue decomposition and reconstruction of the inverse using the $N - 1$ nonzero eigenvalues and associated eigenvectors. The inverse of $\mathbf{\Gamma}$ maybe written as

$$\mathbf{\Gamma}^{-1} = \mathbf{U}(\mathbf{\Lambda}^{-1})\mathbf{U}^T \quad (3.24)$$

where $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u}_i$, $1 \leq i \leq n$, are the eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues (λ_i) of $\mathbf{\Gamma}$, usually organized in ascending order $\lambda_1 = 0 < \lambda_2 < \dots < \lambda_n$.

Mean-square fluctuations of the C^α atoms and the cross correlations between the fluctuations of the C^α atoms are found using the diagonal and off-diagonal elements of $\mathbf{\Gamma}^{-1}$, respectively, in the following equation

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} [\mathbf{\Gamma}^{-1}]_{ij} = \frac{3k_B T}{\gamma} \sum_{k=1}^N [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \quad (3.25)$$

where λ_k is the k th eigenvalue of $\mathbf{\Gamma}$ and is representative of the frequency of the k th mode of motion, which is $\frac{\gamma \lambda_k}{2}$. $\mathbf{u}_k$ is k th eigenvector and gives the shape of the k th mode as a

function of residue index. k_B is the Boltzmann constant and T is the absolute temperature in degrees Kelvin. Equation 3.25 provides a simple means by which the dynamics can be decomposed into a series of N modes for N number of interacting residues. When $i = j$ (diagonal of Γ^{-1} matrix), equation 3.26 gives the mean-square fluctuations,

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ii} \quad (3.26)$$

It is possible to decompose Γ^{-1} as the sum of the contributions from individual modes as

$$\Gamma^{-1} = \sum_{k=2}^n \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T = \sum_{k=2}^n \mathbf{A}^{(k)} \quad (3.27)$$

Here $\mathbf{A}^{(k)}$ is the $n \times n$ matrix describing the contribution of the k th vibrational mode to atomic fluctuations. The first eigenvalue of Γ , which is equal to zero, is not included in the summation in equation 3.27.

$\langle \Delta \mathbf{R}_i^2 \rangle$ may be directly compared to the corresponding X-ray crystallography B-factor (Debye-Waller factor):

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{R}_i^2 \rangle \quad (3.28)$$

Normalized form of the cross-correlations with respect to the auto-correlations is evaluated as:

$$C(i, j) = \frac{\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle}{\left[\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle \langle \Delta \mathbf{R}_j \cdot \Delta \mathbf{R}_j \rangle \right]^{1/2}} \quad (3.29)$$

The lower-left and upper-right triangular portions of the correlation map gives the correlated ($C(i, j) > 0$) and the anticorrelated ($C(i, j) < 0$) pairs, respectively [4, 5].

The collective modes of motion accessible to structures in native state conditions are obtained by GNM. The eigenvalues obtained from diagonalization of the connectivity

matrix represent the frequencies of the individual modes, while the eigenvectors define the shapes of these modes. The first eigenvalue, λ_1 , is identically zero with the corresponding eigenvector comprised of elements all equal to a constant, which indicates that internal motions are absent in this *zero* mode and the molecule can be translated rigidly without any potential energy change.

The slowest motions are those having the highest *degree of collectivity* in addition to their largest contribution. The shapes of the slowest modes reveal the mechanisms of *cooperative* or *global* motions, and the most constrained residues in these modes play a critical role, such as a hinge-bending center, that govern the correlated movements of entire domains. Although these motions are slow, they involve substantial conformational changes distributed over several residues.

The fastest modes, on the other hand, involve the most tightly packed and hence most severely constrained residues in the molecule. Their high frequency does not imply a definitive conformational change, because they cannot effectively relax within their severely constrained environment. On the contrary, they enjoy extremely small conformational freedom, on a local scale, by undergoing fast, but small amplitude fluctuations. The slow modes involve almost the entire molecule, whereas the fastest modes are highly localized, with mode shapes that usually involve only a few peaks. These peaks refer to the residues that have a high concentration of local energy and are tightly constrained in motion. These residues are often conserved across species and may form the folding nuclei.

The C^α node representation of the protein studied in this work, CAP in the unbound form, is given in Figure 3.7 as an example protein network model. The position of the nodes of the EN are defined by the C^α -atom coordinates, which are drawn using a cutoff radius of 7 Å. Figures are created using the crystal structure 1CGP and PyMOL [39] as the visualization program. In the structure, red/pink, green/yellow and blue/cyan represent the DNA binding domain, the hinge region and the cAMP binding domain, respectively for the monomer A/B.

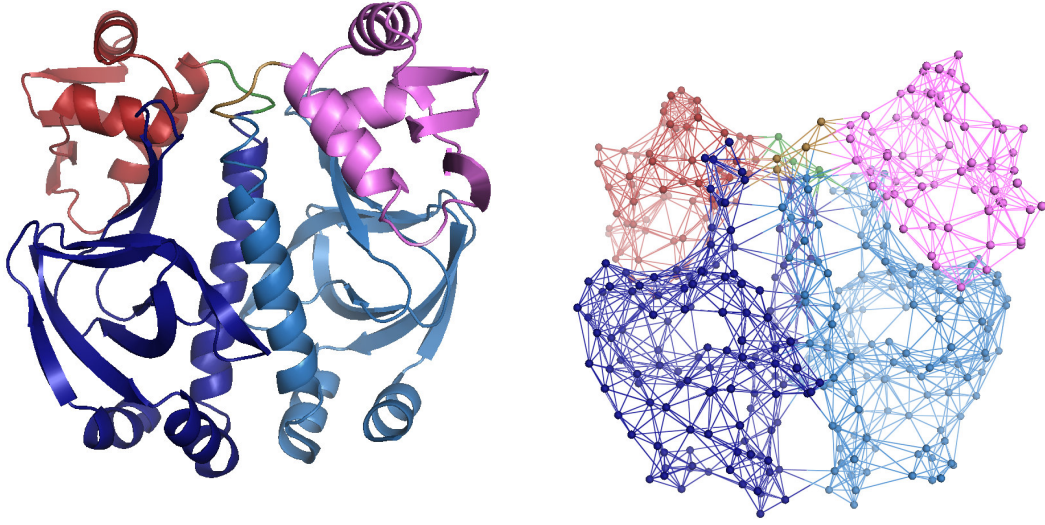


Figure 3.7. Elastic Network (EN) Model of CAP in the unbound form.

3.4. The Anisotropic Network Model

The anisotropic network model (ANM) [6] is an extension of GNM where the fluctuations are anisotropic (dependent on direction), which incorporates the X, Y, and Z components of the position vector, $\mathbf{R}_i$, independently. Therefore, the overall potential for the ANM calculations, includes the fluctuations for all components. The harmonic potential can be expressed as

$$V = \frac{\gamma}{2} (S_{ij} - S_{ij}^0)^2$$

$$V = \frac{\gamma}{2} \left(\left[(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2 \right]^{1/2} - S_{ij}^0 \right)^2 \quad (3.30)$$

where γ is the harmonic force constant and S_{ij}^0 is the instantaneous separation vector and S_{ij}^0 is the equilibrium separation vector of i th and j th residues, similar $\mathbf{R}_{ij}$ and $\mathbf{R}_{ij}^0$ in GNM. X_i , Y_i and Z_i , are the components of position vector $\mathbf{R}_i$.

In ANM, $\mathbf{\Gamma}$ of GNM is replaced by the Hessian matrix $\mathbf{H}$ of the second derivative of the intramolecular potential function in equation 3.30. $\mathbf{H}$ is a $3N \times 3N$ symmetric matrix

and composed of $N \times N$ super elements $\mathbf{H}_{ij}$ each of size 3×3 , given by the second derivatives of V with respect to $\mathbf{R}_i$ and $\mathbf{R}_j$ of C^α -atoms of respective i^{th} and j^{th} residues. The correlation between $\Delta \mathbf{R}_i$ and $\Delta \mathbf{R}_j$ decomposed into $3N - 6$ modes of motions is then given by

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} \text{tr}[\mathbf{H}^{-1}] = \frac{3k_B T}{\gamma} \sum_{k=1}^N \text{tr}[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \quad (3.31)$$

$\text{tr}[\mathbf{H}^{-1}]_{ij}$ is the trace of the ij^{th} submatrix $[\mathbf{H}^{-1}]_{ij}$ of $\mathbf{H}^{-1}$, which is the sum of the diagonal elements of this 3×3 matrix. It refers to three different components of $\Delta \mathbf{R}_i$ and $\Delta \mathbf{R}_j$; whereas, when $i = j$, the self correlations between the components $\Delta \mathbf{R}_i$ are obtained. Here the knowledge of fluctuation vectors permits us to construct and explicitly view pairs of alternative conformations sampled by the individual modes, simply by adding the fluctuation vectors $\pm \Delta \mathbf{R}_i$ to the equilibrium position vectors in the respective modes [6].

4. RESULTS AND DISCUSSION

MD trajectories have been generated for the unbound forms of CAP to investigate the structural features and dynamics of apo CAP in solution. The details of the models constructed are given in section 3.1.4, which are the unbound form of CAP in monomer form, **CAP_{monomer}**, and the unbound form of CAP in dimer form, **CAP_{dimer}**. To extract the key features of the MD trajectories, traditional methods of trajectory analysis were used like calculation of the root-mean-square deviations, mean-square fluctuations of the residue positions/dihedral angles and cross correlations of the mean-square fluctuations. Conformational clustering is performed to reduce the complexity of the systems while minimizing the loss of information.

4.1. Molecular Dynamics Trajectory Analysis

4.1.1. Root -Mean-Square Deviation (RMSD)

The root-mean-square deviation (RMSD) of residue positions of **CAP_{monomer}** and **CAP_{dimer}** from the structures **CAP_{monomer}^{min}** and **CAP_{dimer}^{min}**, respectively, are calculated for a time span of 60 ns after alignment based on C^α atoms. **CAP_{monomer}^{min}** and **CAP_{dimer}^{min}** are the energy minimized forms of the initial X-ray crystallographic structure 1CGP with both the ligands and the DNA removed. Figure 4.1 (a) and (b) display these RMSD plots.

RMSD values indicate that both structures seem to be equilibrated within the time span of 40 ns, yet the two display different trajectory profiles (see Figure 4.1) in terms of deviation from the initial structure as well as in their dynamic equilibrium fluctuations. In the monomer trajectory, RMSD values do not exceed 2.5 Å for the first 10 ns, after which a climb starts for 17 ns up to 4.5 Å. This jump may be an indication of a conformational transition. Since the starting model for the MD is formed by the removal of the ligands and the DNA from the crystal structure, the structure must adjust itself to the new conditions in a search for another equilibrium structure. After reaching the peak the RMSD decreases

back to about its starting value and increases again to around 3-4 Å and oscillates there for 10 ns till the end of the simulation.

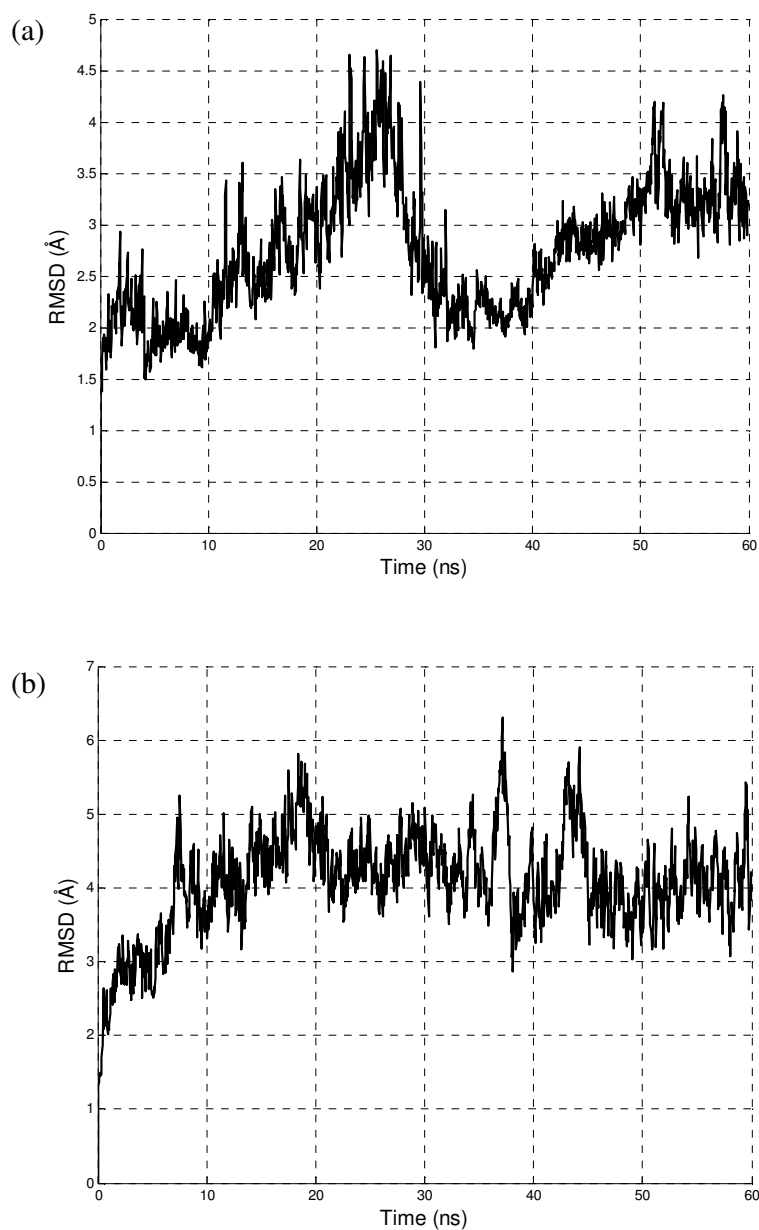


Figure 4.1. Root-mean-square deviation (RMSD) of residue positions of (a) CAP_{monomer} (b) CAP_{dimer}, from the reference structures **CAP_{monomer}^{min}** and **CAP_{dimer}^{min}**, calculated for a time span of 60 ns (for each frame, every 0.4 ps).

In the dimer trajectory, the RMSD value climbs gradually to 5.5 Å in the first 20 ns, and fluctuates between 3 Å and 6 Å till the end of 60 ns trajectory. Here, the amplitude of the conformational fluctuations are not as big as in the monomer trajectory. The details of the conformational samplings along the trajectory will be elaborated in the clustering section.

4.1.2. Equilibrium Fluctuations

4.1.2.1. Mean-square fluctuations

Figure 4.2 (a) compares the mean-square fluctuations (ms) ($\langle \Delta \mathbf{R}_i^2 \rangle$) in residue positions of the isolated CAP monomer and subunits of CAP dimer observed in MD simulations. In Figure 4.2 (b) the experimental B-factors from X-ray structure of 1CGP is displayed together with the B-factors of the isolated CAP monomer and the monomers of CAP dimer, calculated by following the relationship $B_i = 8\pi^2 / 3 \langle \Delta \mathbf{R}_i^2 \rangle$.

The ms fluctuations of the monomers of the dimer are very similar. The largest deviation between the fluctuations of the monomers of the dimer and the isolated monomer is at the dimerization interface (residues 110-134, C-helix, including the cAMP binding residues T127 and S128), as expected. Other significant differences in the ms fluctuation profiles are in the regions between the residues 54-67 and 170-181, which overlap with the secondary cAMP binding sites (residues 56-59, 170, 174, 177, 180) and the selective DNA binding residues (residues 180 and 181).

Most significant deviations from the experimental B-factors are also in the dimerization interface (residues 110-134). Throughout this region, the monomers of the dimer display lower amplitude fluctuations compared to the X-ray structure and the isolated monomer; whereas the isolated monomer displays higher mobility than the X-ray structure in this region. This implies that the hinges at the interface of cAMP and DNA binding domain are destabilized in the monomeric state. Additionally, close to this region, the region of residues 90-112, displays lower fluctuations in both the isolated monomer and the monomers of the dimer. This could be due to the absence of the cAMPs. The other regions that show deviations are observed for the residues 31-42 and 63-73, where the

fluctuations are restrained with respect to the crystal structure in both the isolated monomer and monomers of the dimer. The two regions contain an active site of residue 30 and cAMP primary binding site residues. Moreover, in the region between the residues 173-181 the monomer is more mobile compared to the X-ray structure and the dimer subunits, A and B, which again overlap the secondary cAMP and DNA binding regions.

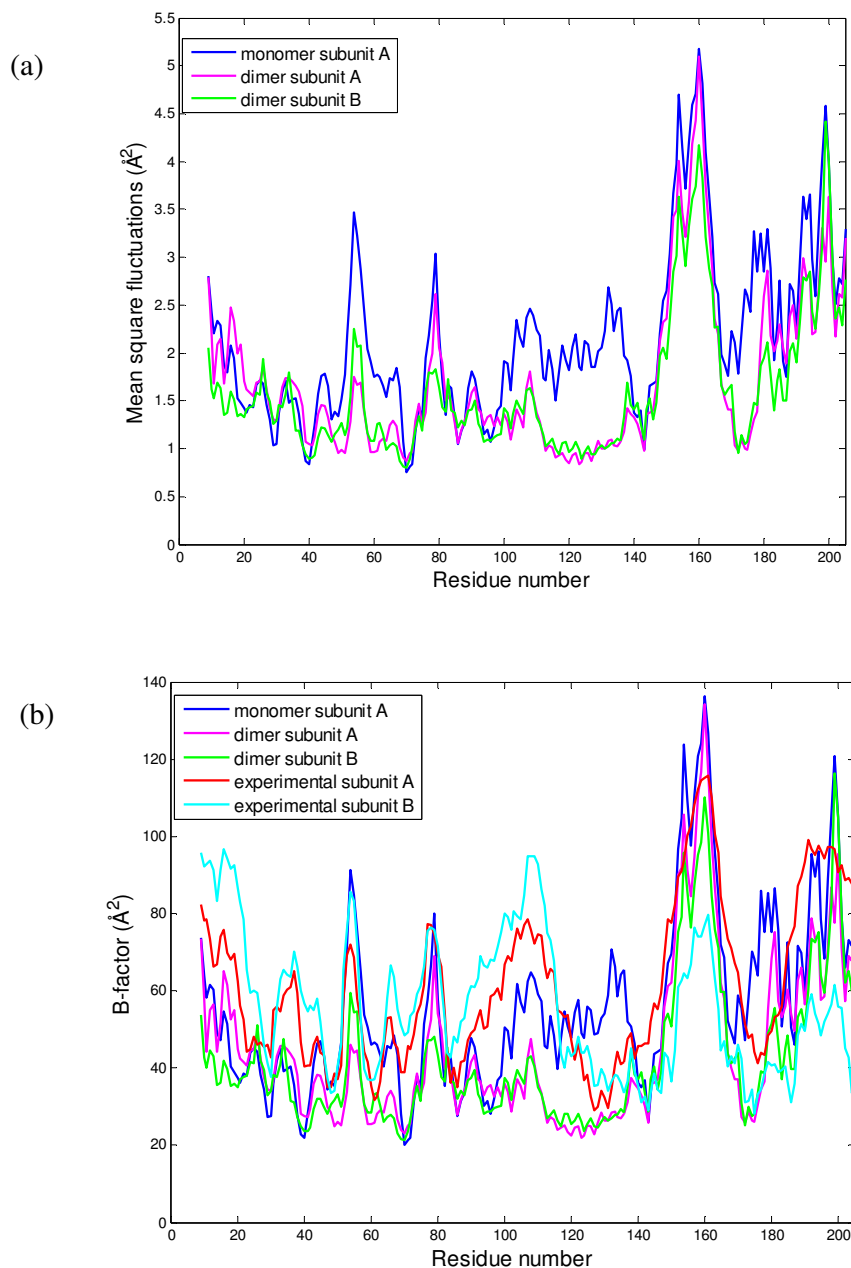


Figure 4.2. Comparison of (a) ms fluctuations and (b) B-factors for CAP_{monomer} and CAP_{dimer}

In Figure 4.3, the experimental B-factors of some of the known crystal structures of CAP are displayed together with the B-factors of the isolated CAP monomer and the monomers of CAP dimer. The PDB codes of the compared crystal structures are 1G6N; a CAP:(cAMP)₂ complex, 2GZW; a CAP:(cAMP)₄ complex, and 2CGP; a CAP:(cAMP)₄:DNA complex [19, 40, 41]. The regions of deviations of the B-factor profiles are the dimerization interface (residues 110-134), the RNAP interacting region-AR1 loop (residues 156-164), the secondary cAMP binding sites (residues 56-59, 170, 174, 177, 180) and selective DNA binding sites (residues 180-181). Structure 2CGP shows a different overall trend from the other proteins. 2CGP values are shifted downwards at all regions.

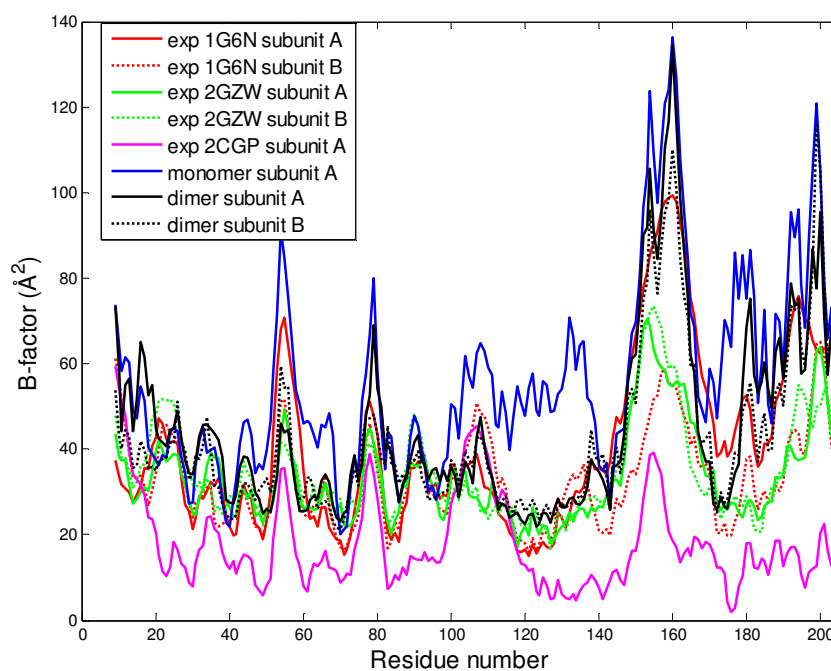


Figure 4.3. Comparison of B-factors for CAP_{monomer} (subunit A) and CAP_{dimer} (subunit A and B), calculated from the MD trajectories with experimental B-factors of 1G6N, 2GZW and 2CGP [19, 40, 41].

4.1.2.2. Cross-correlation of fluctuations

Figure 4.4 and Figure 4.5 display the correlations between the fluctuations of residues for the isolated monomer and the dimer. The isolated monomer does not have a significant pattern, whereas the dimer shows clear patterns. The most striking difference between the two is that the dimerizations stabilize the hinges at the dimerization interface (residues 110-134) and at the DNA binding site (residues 180-181). This leads to the appearance of the two domains, the cAMP binding domain (residues 1-134) and the DNA binding domain (residues 139-205), and also association of the two domains and the two monomers via the DNA binding site (residues 180-181).

While examining the cross-correlation maps, attention should be paid to the residue numbers, which should be shifted by 8 residues (+) when compared to the primary sequence.

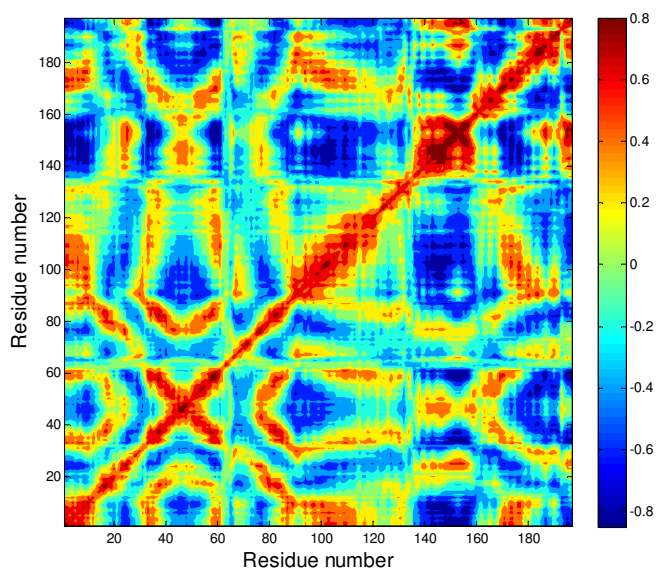


Figure 4.4. Cross correlation values calculated from the MD trajectories of CAP_{monomer}

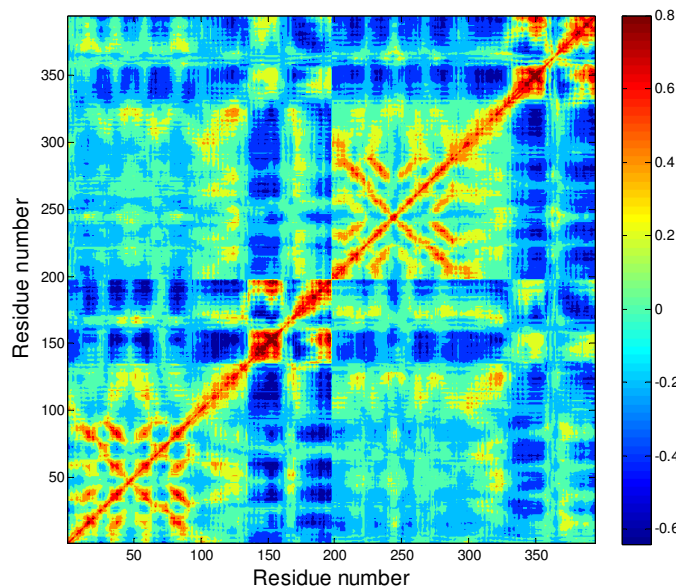


Figure 4.5. Cross correlation values calculated from the MD trajectories of CAP_{dimer}

4.1.2.3. Dihedral angle fluctuations

The mean-square fluctuations of the rotational angles ϕ and ψ are calculated from the MD trajectories of the isolated monomer and the dimer. The variations in ϕ and ψ are very similar, thus only the variations in ϕ are discussed here (Figure 4.6). The variations between the rotational fluctuations of the monomer and the monomers of the dimer are at the regions of residues 136 and 162. The former residue is in the hinge region near the end of C-helix and found to be responsible for secondary cAMP binding site formation by rotating and allowing cAMP to enter the site after the occupation of the two primary binding sites [42]. The latter is a residue from the region responsible for interacting with the RNAP during transcription activation [43]. On the other hand, the region of residues 25 and 35 are more stabilized in the monomeric state, which corresponds to the associating regions at the interface between the monomers.

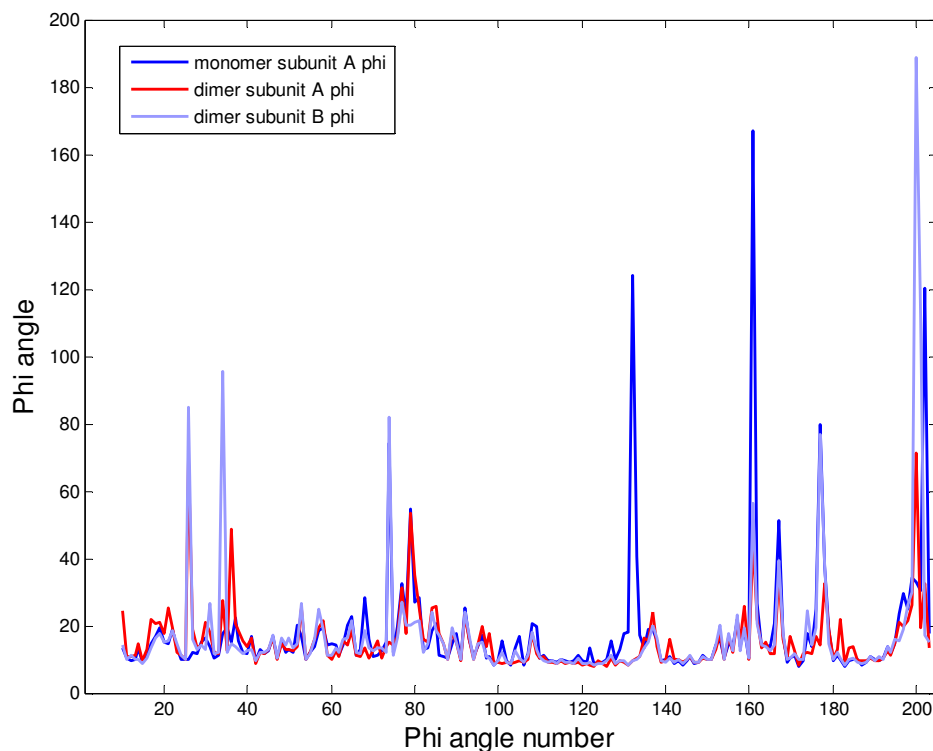


Figure 4.6. Ms fluctuations of phi angles calculated from the MD trajectories of $CAP_{monomer}$ and CAP_{dimer}

The correlation between the fluctuations of phi angles is presented in Figure 4.7, Figure 4.8 and Figure 4.9 for the isolated monomer and the monomers of the dimer. The psi angle correlations are very similar to the phi angle correlations, thus are not presented here.

While examining the correlation figures, attention should be paid to the residue numbers, which should be shifted by 8 residues (+) when compared to the primary sequence. For convenience, the residue numbers in the figures corresponding to the real residue numbers will be given in parenthesis.

The coupled correlations in the rotational motions of the monomer and the dimer are not similar. The most significant correlations are observed for the residues 160, 167, 175 and 198 (152, 159, 167 and 190) with the region 64-82 (58-74) for the isolated monomer.

Interestingly, the correlation of the phi angles behave differently in the different monomers of the dimer. In monomer B of the dimer, the emphasized correlations occur between the regions 170-175 (162-167), 23-36 (15-28) and 70-88 (62-80), whereas in monomer A of the dimer, only an insignificant correlation between the region 78-96 (70-88) and 16-35 (8-27) is observed.

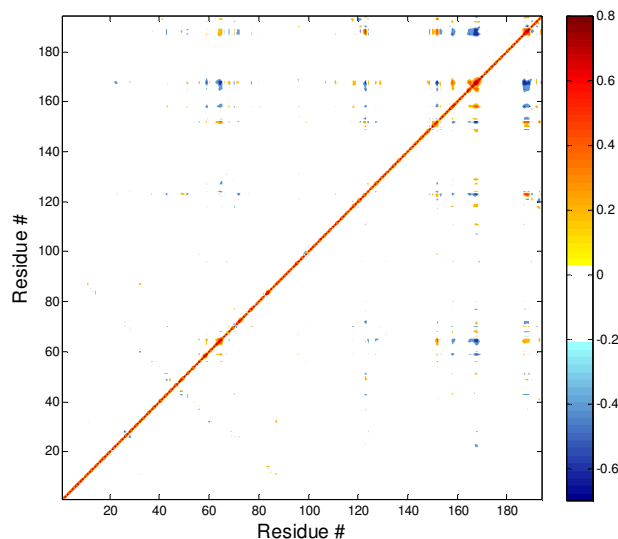


Figure 4.7. Correlation coefficients of phi angles calculated from the CAP_{monomer} MD trajectory.

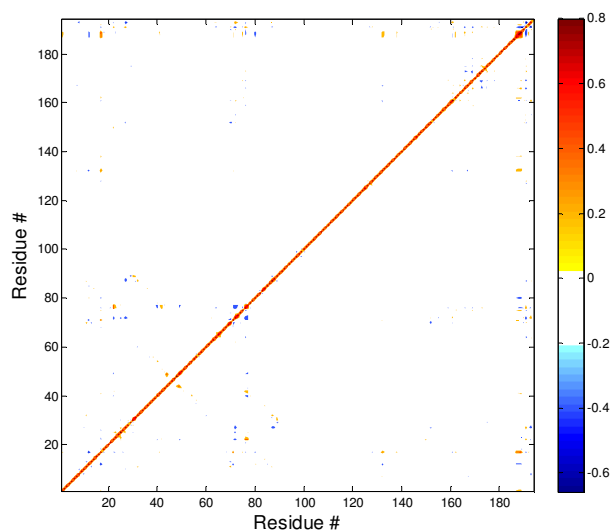


Figure 4.8. Correlation coefficients of phi angles calculated from monomer A of CAP_{dimer} MD trajectory.

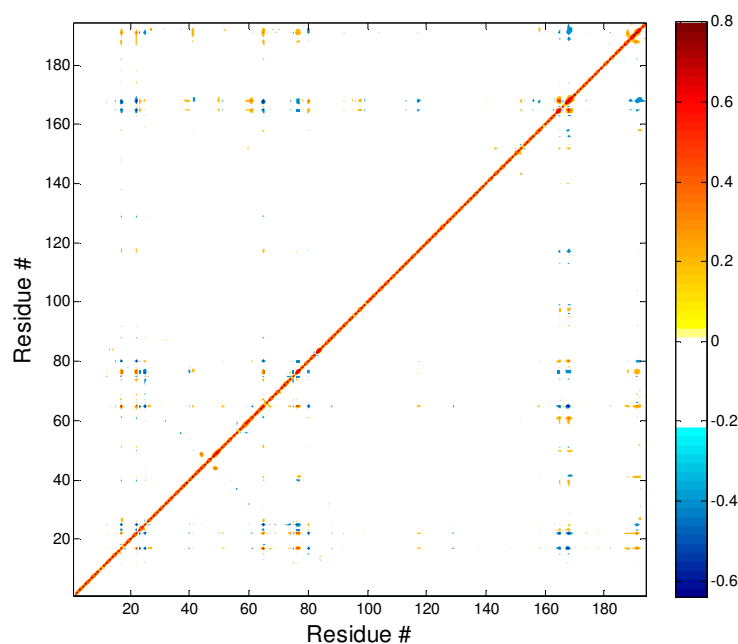


Figure 4.9. Correlation coefficients of phi angles calculated from monomer B of CAP_{dimer} MD trajectory.

4.1.3. Salt Bridge Formations

The salt bridges formed during the simulations are calculated by *Salt Bridges* module of VMD, with a oxygen-nitrogen distance cutoff of 3.2 Å. A new trajectory consisting of the 1/10th of the original trajectory was formed for convenience in the analysis. The list of the salt bridges formed along the monomer and dimer MD trajectories is given in Table 4.1.

Table 4.1. List of the salt bridges

	MONOMER	MONOMER A (DIMER)	MONOMER B (DIMER)	A/B (DIMER)	B/A (DIMER)
GLU12-ARG115		x			
GLU34-HIP31	x	x	x		
GLU34-LYS35			x		
GLU34-ARG82	x	x	x		
GLU37-LYS35	x	x	x		
GLU37-ARG82			x		
GLU37-LYS100	x	x	x		
GLU37-ARG103	x				
ASP53-LYS57	x	x	x		
ASP53-ARG142	x				

Table 4.1. (continued) List of the salt bridges.

	MONOMER	MONOMER A	MONOMER B	A/B	B/A
GLU54-LYS52	x	x	x		
GLU54-ARG142	x				
GLU55-LYS57	x	x	x		
GLU55-ARG142	x				
GLU58-LYS26	x				
GLU58-LYS52	x	x	x		
GLU58-LYS57	x				
GLU58-ARG87	x	x	x		
GLU58-ARG180		x	x		
ASP68-ARG115	x				
ASP68-ARG122	x	x	x		
ASP68-ARG123	x	x	x		
GLU72-ARG82	x				
GLU72-ARG123	x	x	x		
GLU77-ARG115				x	x
GLU77-ARG122				x	x
GLU78-LYS35	x		x		
GLU78-LYS100	x	x	x		
GLU78-ARG103	x	x	x		
GLU81-LYS35	x	x	x		
GLU93-LYS22	x	x	x		
GLU93-LYS44	x	x	x		
GLU96-HIP19	x	x	x		
GLU96-HIP21	x	x	x		
ASP111-ARG103				x	x
ASP111-ARG115	x	x	x		
GLU129-ARG122	x		x		
GLU129-LYS130	x	x	x		
GLU129-LYS152	x	x			
ASP138-ARG142			x	x	x
ASP155-LYS89	x	x	x		
ASP155-LYS166	x	x	x		
ASP161-HIP159	x	x	x		
ASP161-HIP199	x	x	x		
ASP161-LYS201	x	x	x		
GLU171-LYS26	x	x	x		
GLU171-ARG87	x	x	x		
GLU171-LYS89	x	x	x		
GLU171-LYS166	x				
GLU181-LYS57		x			
GLU181-ARG169	x	x	x		
GLU181-ARG180	x	x	x		
GLU181-ARG185	x	x	x		
GLU191-ARG169	x	x	x		
GLU191-LYS188	x	x	x		
GLU191-HIP199	x	x	x		
GLU191-LYS201	x				
ASP192-ARG185		x			
ASP192-LYS188	x	x	x		

The salt bridges with the important implications (showed with bold text in Table 4.1) are found to be between residues E72-R82, E72-R123 and D68-R123, which agree the literature [44, 45, 46].

These salt bridges have further been investigated by plotting their oxygen-nitrogen atom distances to observe the specific times of their existence in the duration of the simulations. (see Appendix A.1 for the plots). This will be clarified below together with the secondary structure content of the conformations in section 4.3.6, Secondary Structure Assignments.

4.2. Clustering MD Trajectory Snapshots

In this part, the results are presented from the clustering $CAP_{monomer}$, CAP_{dimer} and both $CAP_{monomer}$ and CAP_{dimer} (subunit A) snapshots taken from the two MD trajectories. The program used for cluster analysis is the *kclust* script of MMTSB toolset [7].

In the clustering of the conformations k-means clustering method was used. This method is nonhierarchical and uses an algorithm that goes through the trajectory and locates clusters of conformations that are the same by grouping them together using RMSD as the similarity measure. From this it creates centroids (cluster center) describing each cluster and then gives an RMSD for each structure in the trajectory with respect to each identified centroid. An average structure for each cluster centroid is produced, which is not quiet a physically meaningful one. The structure that is nearest to the centroid is a more physically relevant structure of the centroid, which is the cluster best member (BM).

Since the limit of the *kclust* tool is 50000 snapshots, the conformations were stored every 4 ps, resulting in 15000 snapshots for each of the two simulations. To cluster the 15000 snapshots from both trajectories (one by one and together), cluster radii of 2.0, 2.5, 2.7, 3.0 and 3.5 Å are used. An overview of the clusters formed is given in Table 4.2 and the details for the clusters formed with cluster radii of 2.7, 3.0 and 3.5 Å are given in Table 4.3, Table 4.4 and Table 4.5, respectively. Since the smaller cluster radius values produce

more clusters, often unmanageable and insignificantly different from each other, their results are not given in the text but can be found in the Appendix A2.

Table 4.2. Overview of the clusters formed

cluster radius (Å)	no. of clusters		
	monomer ^a	dimer ^b	both ^c
2.0	27	27	40
2.5	13	8	19
2.7	8	6	13
3.0	7	4	9
3.5	3	2	5
^a CAP _{monomer} trajectory, 15000 snapshots			
^b CAP _{dimer} trajectory, 15000 snapshots			
^c CAP _{monomer} and CAP _{dimer} trajectories, 30000 snapshots			

Table 4.3. Cluster members and best members (BM) resulting from cluster radius of 2.7Å

	Cluster#	RMSD from Cluster Centroid (Å)	BM Snapshot #	No. of Members
MONOMER	1	0.9118	19561	2827
	2	0.9644	46681	2159
	3	0.968	60091	1514
	4	0.9607	82221	1199
	5	0.9845	95351	2939
	6	1.0416	139381	3465
	7	0.977	4811	690
	8	0.9053	73211	207
DIMER	1	1.006	10241	1320
	2	1.1598	105101	1748
	3	1.1664	49701	5258
	4	1.1723	132801	5085
	5	1.1084	107241	1325
	6	0.9821	901	264
BOTH	1	1.0063	21601	1974
	2	0.9626	46681	2117
	3	0.967	60091	1496
	4	0.9898	82221	1219
	5	0.9983	103501	3194
	6	1.0395	139381	3257
	7	0.951	8081	1544
	8	1.1042	104291	7891
	9	0.9886	96151	2558
	10	1.2476	4771	692
	11	1.0025	36241	974
	12	0.8987	74271	162
	13	0.9357	55651	2922

Table 4.4. Cluster members and best members (BM) resulting from cluster radius of 3.0 Å

	Cluster#	RMSD from Cluster Centroid (Å)	BM Snapshot #	No. of Members
MONOMER	1	1.1137	21201	4561
	2	1.0121	60091	1832
	3	0.9919	103501	3016
	4	1.0405	139381	3242
	5	1.0034	5711	799
	6	0.9277	82221	1310
	7	0.9218	73211	240
DIMER	1	1.0381	10241	1589
	2	1.2403	77681	8774
	3	1.2443	138111	3170
	4	1.1471	108821	1467
BOTH	1	1.0781	21201	4483
	2	1.0699	60091	2076
	3	0.9809	102301	2931
	4	1.0401	140711	3109
	5	1.244	15151	2974
	6	1.0935	109981	11952
	7	1.8177	2521	797
	8	0.9537	82221	1441
	9	0.9426	73211	237

Table 4.5. Cluster members and best members (BM) resulting from cluster radius of 3.5 Å

	Cluster#	RMSD from Cluster Centroid (Å)	BM Snapshot #	No. of Members
MONOMER	1	1.1885	11491	7860
	2	1.1186	67981	2474
	3	1.0826	139391	4666
DIMER	1	1.0448	10241	1654
	2	1.2891	112391	13346
BOTH	1	0.9021	67861	709
	2	0.9887	60091	980
	3	2.0774	111651	7034
	4	1.0595	104271	13668
	5	1.2595	15041	7609

The highlighted cells in Table 4.3, Table 4.4 and Table 4.5 indicate that the best members are from the dimer trajectory. The details of the cluster analysis performed using snapshots from both of the trajectories can be seen in Table 4.6 and Table 4.7. Here, some clusters have members from both of the trajectories, which suggests the preexistent conformations of the monomers in dimer in unbound state. These clusters can be seen in Figure 4.14 and Figure 4.15, which are drawn with the same colors.

Table 4.6. Members of the clusters for the monomer and dimer (together) cluster analysis with a cluster radius of 3.0 Å

Cluster#	No. of Members	
	monomer	dimer
1	4478*	5*
2	2076	
3	2931	
4	3109	
5	32*	2942*
6		11952
7	697*	100*
8	1440*	1*
9	237	
* Bold text indicates common members.		

Table 4.7. Members of the clusters for the monomer and dimer (together) cluster analysis with a cluster radius of 3.5 Å

Cluster#	No. of Members	
	monomer	Dimer
1	709	
2	980	
3	5940*	1094*
4		13668
5	7371*	238*
* Bold text indicates common members.		

A more detailed picture of cluster evolution as a function of time is given in Figure 4.10 through Figure 4.15. The cluster conformations are given in the order of their appearance in the trajectory visualized on the RMSD plots [47].

Figure 4.10 and Figure 4.11 display the cluster evolution in time for the isolated monomer at cluster radii of 3.0 and 3.5 Å, respectively. The number of clusters are seven and three for the respective cases. The clustering approach extracts unique conformational states sampled throughout the time span of the simulations. As seen, a cluster could have members at different times of the trajectory, which may indicate the dynamic equilibrium between the conformational states. Here, with the larger cluster radius of 3.5 Å in Figure

4.11, the members of the yellow cluster could be seen both at early and later stages of the trajectory, between which another cluster (magenta) exists. The magenta cluster is indicative of one of the unique conformational states showing the largest deviation from the X-ray structure for both cases.

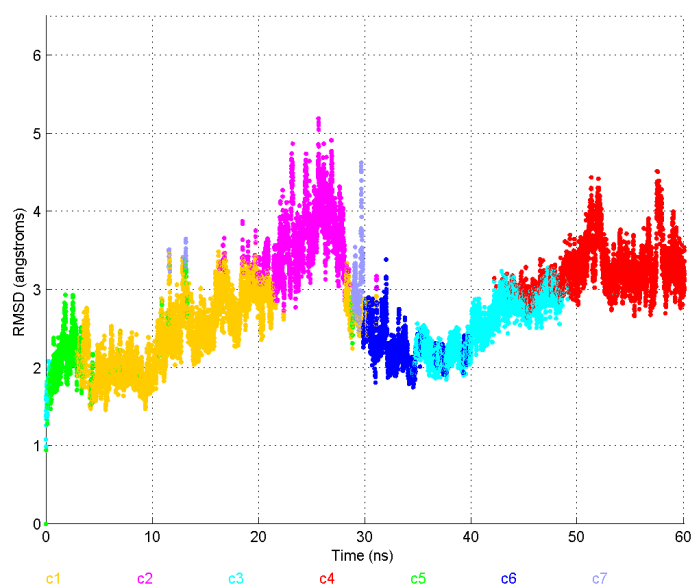


Figure 4.10. Cluster evolution in time for CAP_{monomer} on the RMSD plot with cluster radius of 3.0 Å

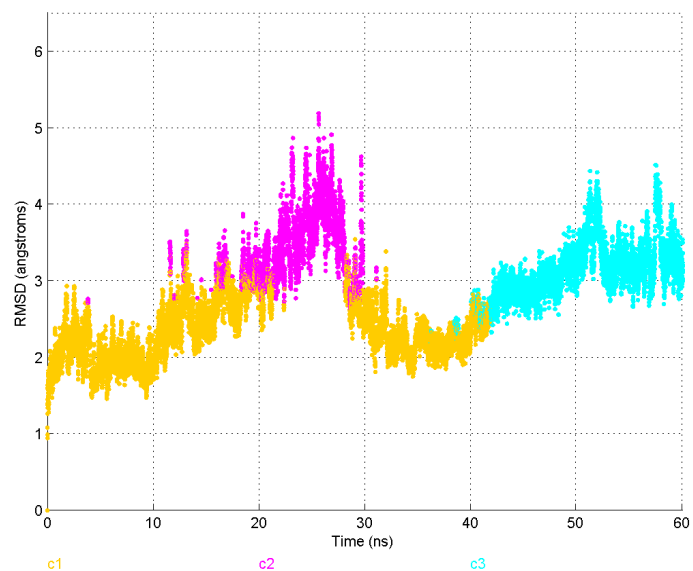


Figure 4.11. Cluster evolution in time for CAP_{monomer} on the RMSD plot with cluster radius of 3.5 Å

For the case of the dimer, Figure 4.12 and Figure 4.13 display the cluster evolution in time on the RMSD plot at cluster radii of 3.0 and 3.5 Å, respectively.

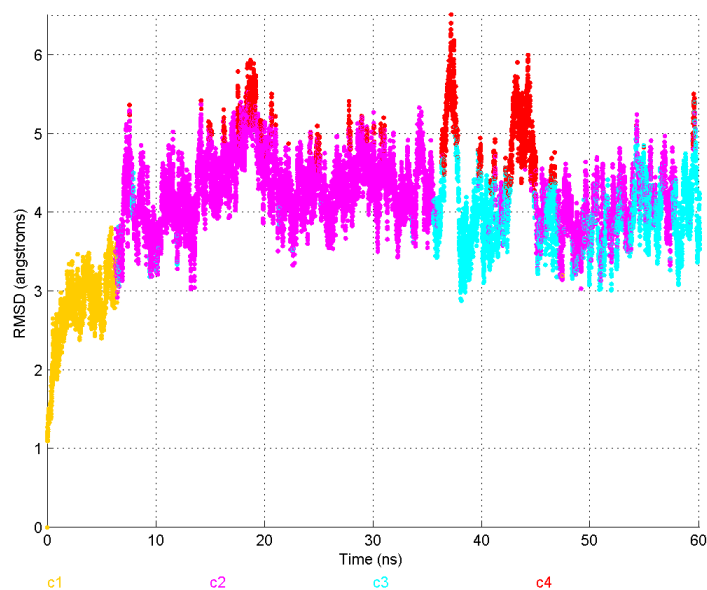


Figure 4.12. Cluster evolution in time for CAP_{dimer} on the RMSD plot with cluster radius of 3.0 Å

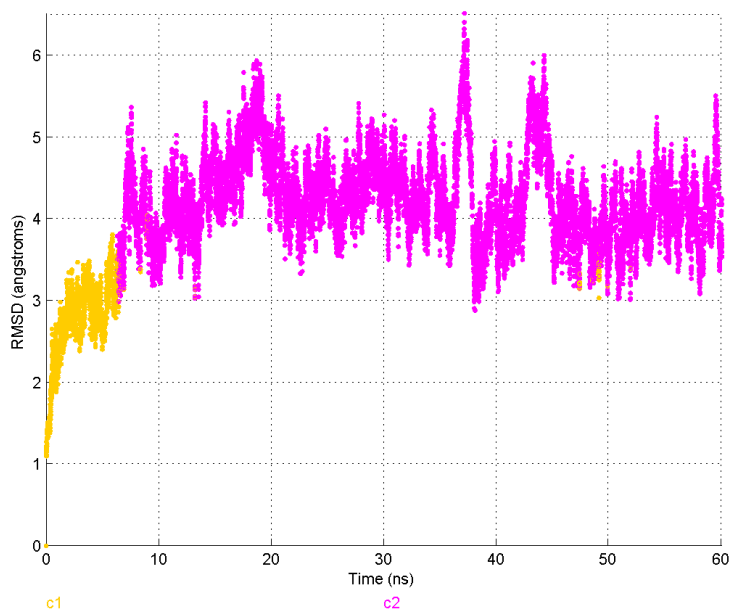


Figure 4.13. Cluster evolution in time for CAP_{dimer} on the RMSD plot with cluster radius of 3.5 Å

The comparison between the clusters of conformations from the monomer simulations and the dimer simulations shows that for all cluster radii studied, the number of clusters is lesser for the dimer case. This implies that dimerization restricts the conformational states accessible to the structure.

Figure 4.14 and Figure 4.15 display the clusters when the conformations of the isolated monomer and the monomer A of the dimer are clustered together with cluster radii 3.0 and 3.5 Å, respectively. There are 9 and 5 clusters appearing for the respective cases. The shared number of clusters decrease by the cluster radius, but the number of shared members increase with the cluster radius (The details of the shared members were given in Table 4.6 and Table 4.7 previously). Interestingly the conformations sampled at the later stage of the isolated monomer structure are similar to those sampled by the monomer of the dimer.

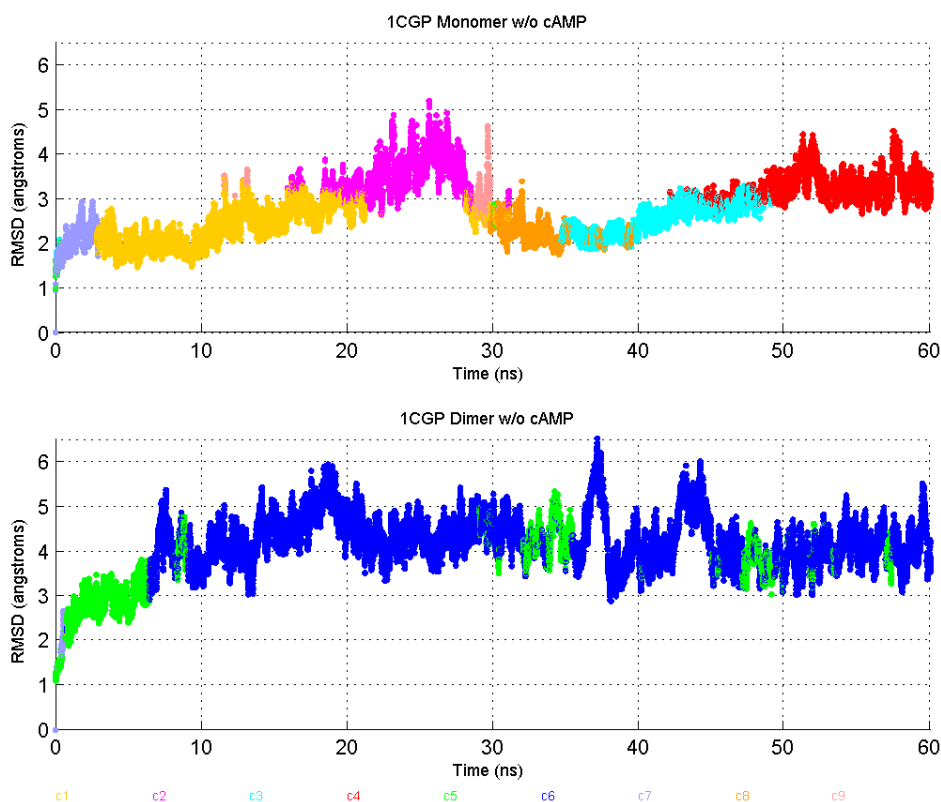


Figure 4.14. Cluster evolution in time for $CAP_{monomer}$ and CAP_{dimer} subunit A on the RMSD plot with cluster radius of 3.0 Å

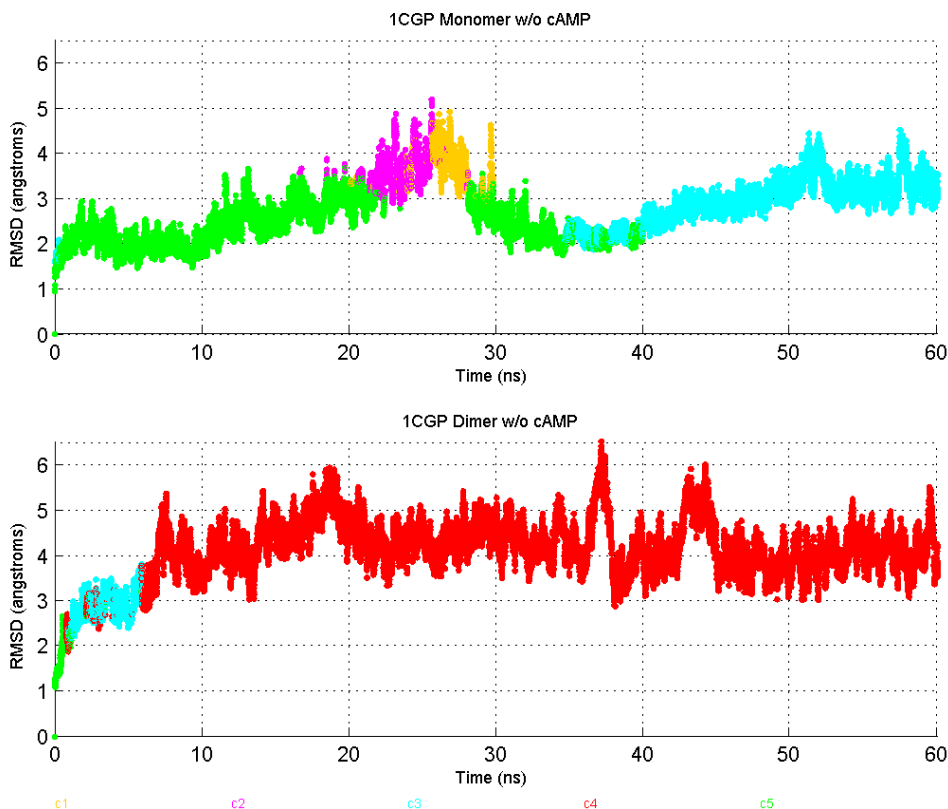


Figure 4.15. Cluster evolution in time for $CAP_{monomer}$ and CAP_{dimer} subunit A on the RMSD plot with cluster radius of $3.5 \text{ \AA}$

The cluster best members are elaborated in the next section, which further examines the differences between the conformations represented by the best members of the clusters.

4.3. Analysis of the Best Members of the Clusters

The cluster best members of the clusters formed with a cluster radii of 2.7 and $3.5 \text{ \AA}$ are chosen to be examined closely, representative of different levels of coarse-grained views of the structural changes/transitions occurring during the simulations.

Superimposition and RMSD calculation of the cluster best members to/from the known structures (crystal and MD snapshots) of CAP are performed. The results are first given for the cluster radius of $3.5 \text{ \AA}$, where very general features with respect to the

conformational differences are searched for. Then, for a more detailed comparison, the results for the lower cluster radius value are presented.

The other clusters formed will not be examined, since these smaller cluster radius values produced more clusters, which are not significantly different from each other. For example, for cluster radii of 2.0 and 2.5 Å, the monomer case has 27 and 13 clusters; the dimer case has 27 and 8 clusters; and the monomer and dimer case has 4 and 19 clusters.

4.3.1. Clusters of monomer MD snapshots with cluster radius of 3.5 Å

Figure 4.16 (a) displays the members of this cluster aligned to the first member and superimposed over the residues 1 to 196, from different perspectives. RMSDs are 2.892 and 2.892 Å, respectively for the second and the third members. This superimposition suggest that the most mobile regions correspond to the RNAP interacting AR1 loop (residues 156-162) and DNA recognition helix (residues 180-193). Interestingly, the second and third members of the cluster does not show helical tendency in the end region of C-helix, where C-helix also bends back and forth around the first member. Same colors are used with the RMSD cluster plots.

Figure 4.16 (b) displays the members of this cluster aligned to the X-ray structure of the starting conformation of MD, 1CGP, superimposed over the residues of monomer A. In the figure structures in yellow, magenta, cyan and gray structures represent first, second, third best members and the crystal structure, respectively.

Figure 4.17, Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21 and Figure 4.22 display the structures and RMSD values of the monomer cluster members aligned and superimposed to the structures 1CGP, 1G6N, 1G6N (13 ns), 1G6N (40 ns) [17], 2CGP and 2GZW, respectively. The largest deviations are observed in the comparison of 2CGP and 2GZW structures, which are the four liganded CAP complexes, the former including the DNA.

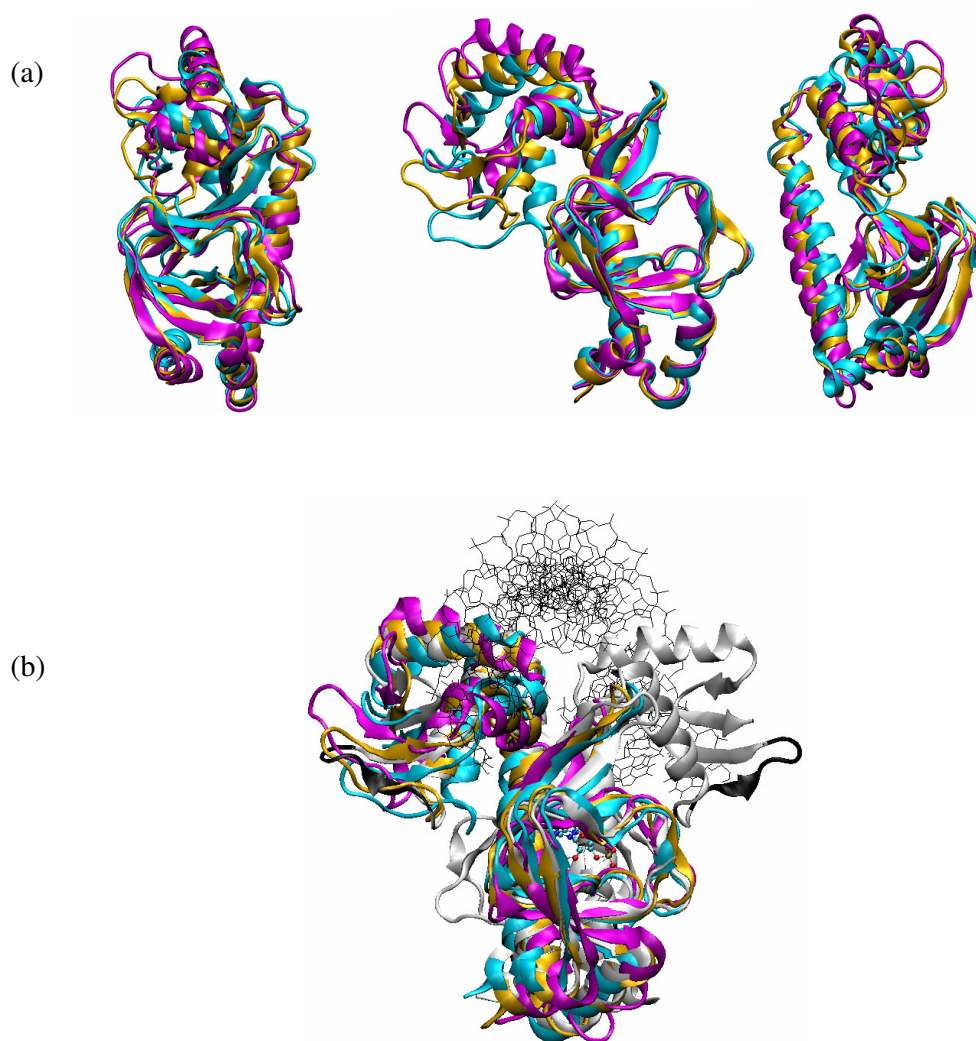


Figure 4.16. (a) Superimposition of the cluster best members of CAP_{monomer} to the first cluster's best member and (b) to the X-ray structure 1CGP. Cluster radius is 3.5 Å.

This set of figures suggest that, for the isolated monomer, the lack of dimerization creates instability, distortions and secondary structure changes in the C-helix (from α helix to turn).

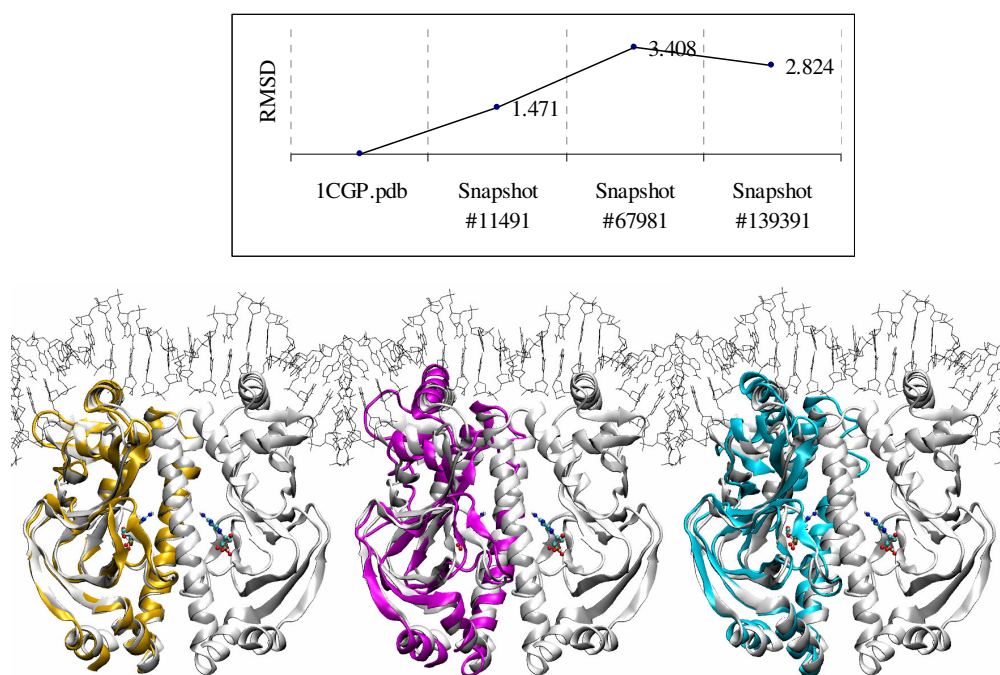


Figure 4.17. RMSDs of the cluster best members of CAP_{monomer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.

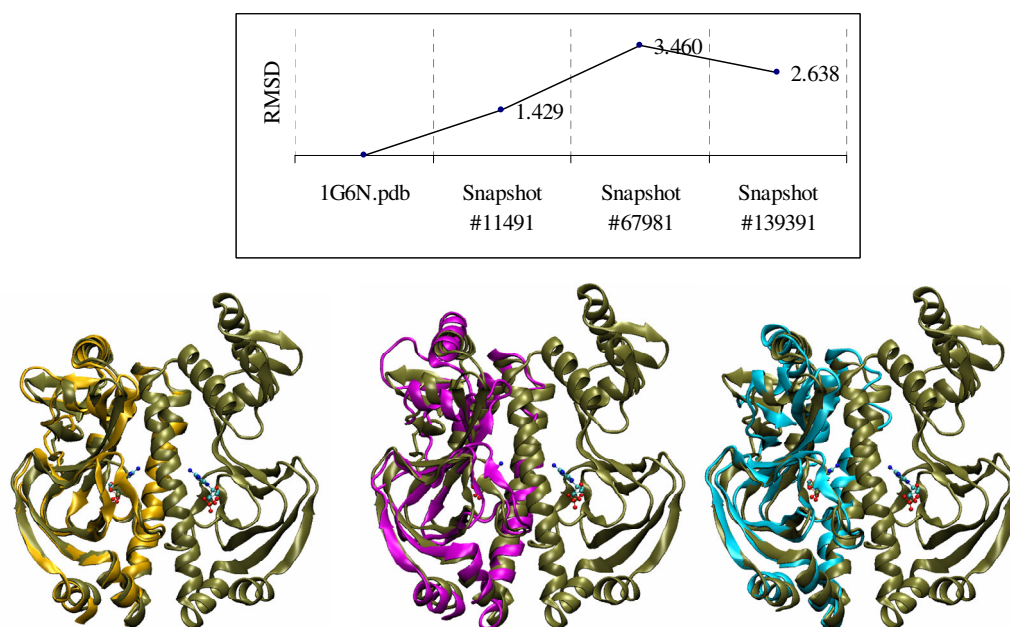


Figure 4.18. RMSDs of the cluster best members of CAP_{monomer} from the X-ray structure 1G6N. Superimposed structures are given below in the same order.

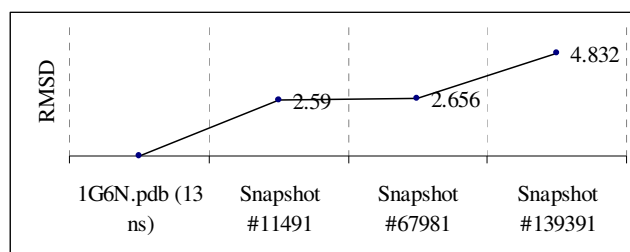


Figure 4.19. RMSDs of the cluster best members of CAP_{monomer} from the MD snapshot 1G6N (13 ns). Superimposed structures are given below in the same order.

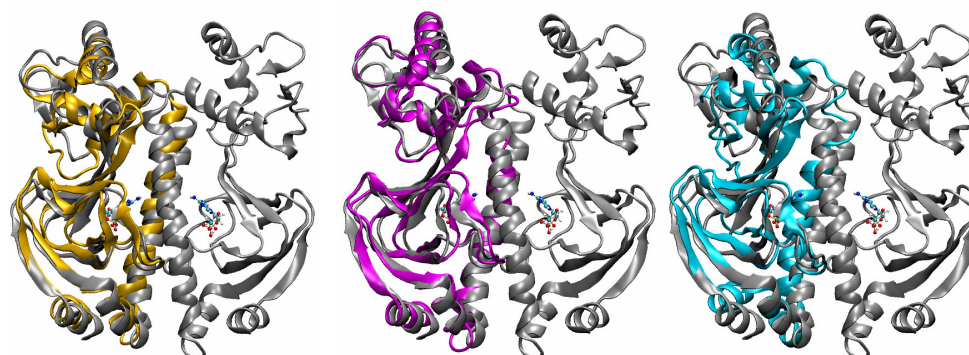
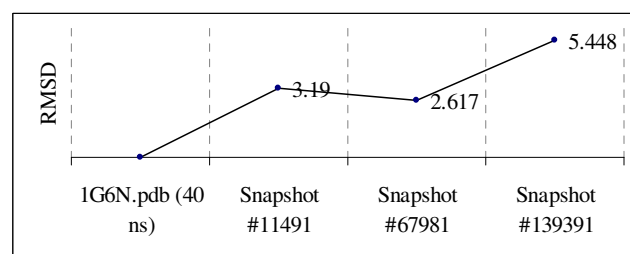


Figure 4.20. RMSDs of the cluster best members of CAP_{monomer} from the MD snapshot 1G6N (40 ns). Superimposed structures are given below in the same order.

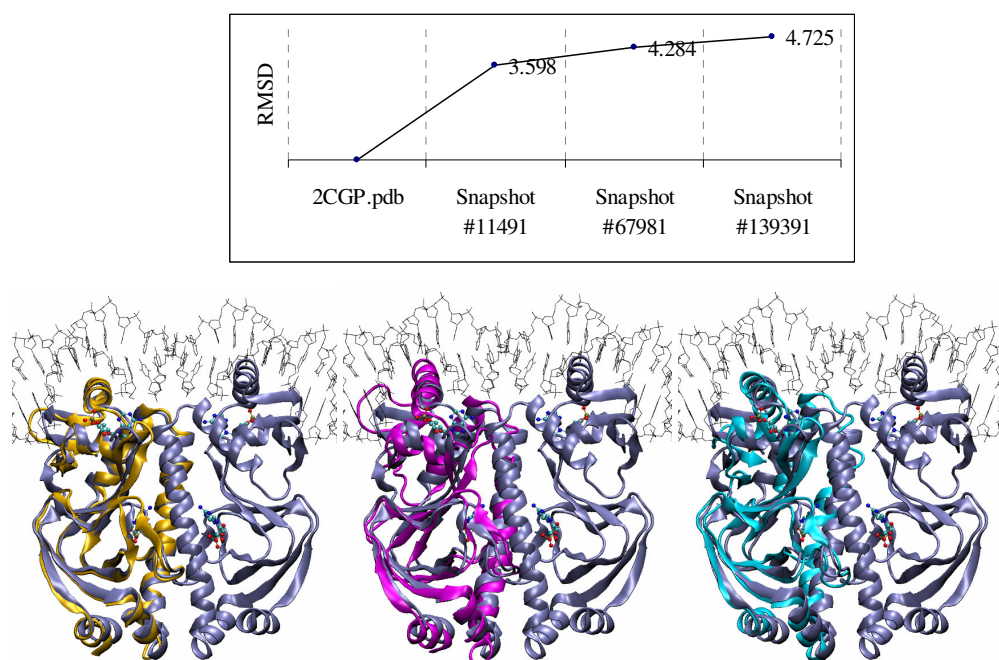


Figure 4.21. RMSDs of the cluster best members of CAP_{monomer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.

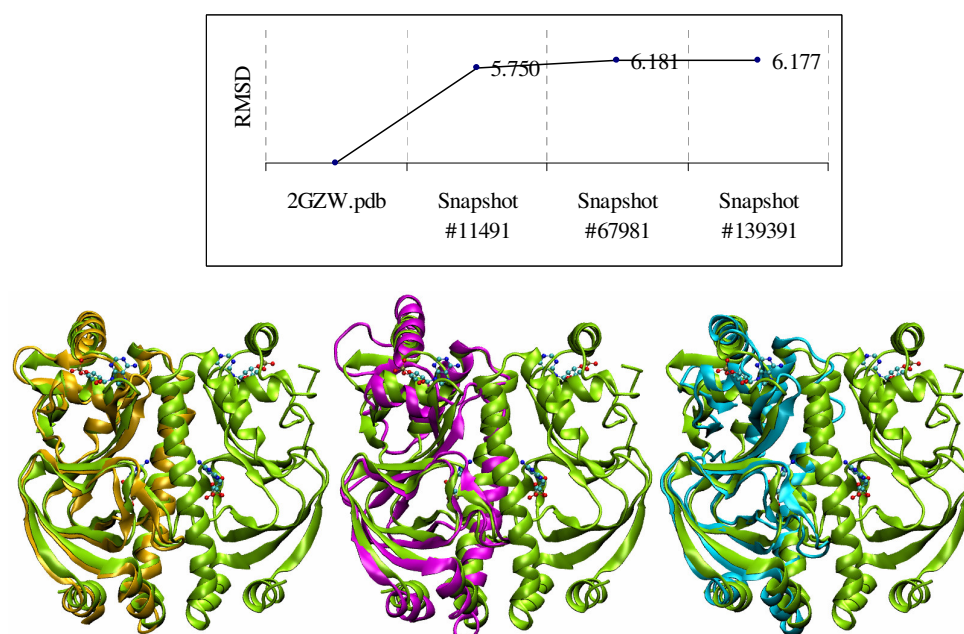


Figure 4.22. RMSDs of the cluster best members of CAP_{monomer} from the X-ray structure 2GZW. Superimposed structures are given below in the same order.

4.3.2. Clusters of dimer MD snapshots with cluster radius of 3.5 Å

Figure 4.23 (a) displays the members of this cluster aligned to the first member and superimposed over the residues 1 to 394 (RMSD = 3.205 Å), from different perspectives. Similar to the monomer case, the most mobile regions correspond to the AR1 loop (shown in red for the first member, purple for second member and black for the crystal structure). C-helix and DNA recognition regions also show some movement, but not as much as the monomer case.

Figure 4.23 (b) displays the members of this cluster aligned to the X-ray structure of the starting conformation of MD, 1CGP. In the figure structures in yellow, magenta, and gray structures represent first and second best members and the crystal structure, respectively. Same colors are used with the RMSD cluster plots.

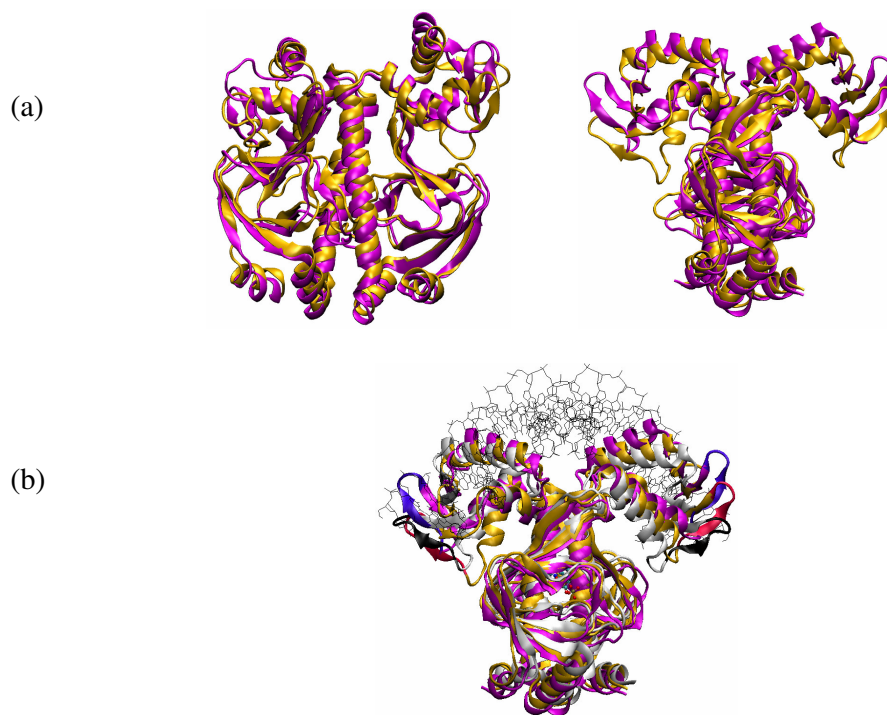


Figure 4.23. (a) Superimposition of the cluster best members of CAP_{dimer} to the first cluster's best member and (b) to the X-ray structure 1CGP. Cluster radius is 3.5 Å.

Figure 4.24, Figure 4.25, Figure 4.26, Figure 4.27, Figure 4.28 and Figure 4.29 display the structures and RMSD values of the dimer cluster members aligned and

superimposed to the structures 1CGP, 1G6N, 1G6N (13 ns), 1G6N (40 ns), 2CGP and 2GZW respectively.

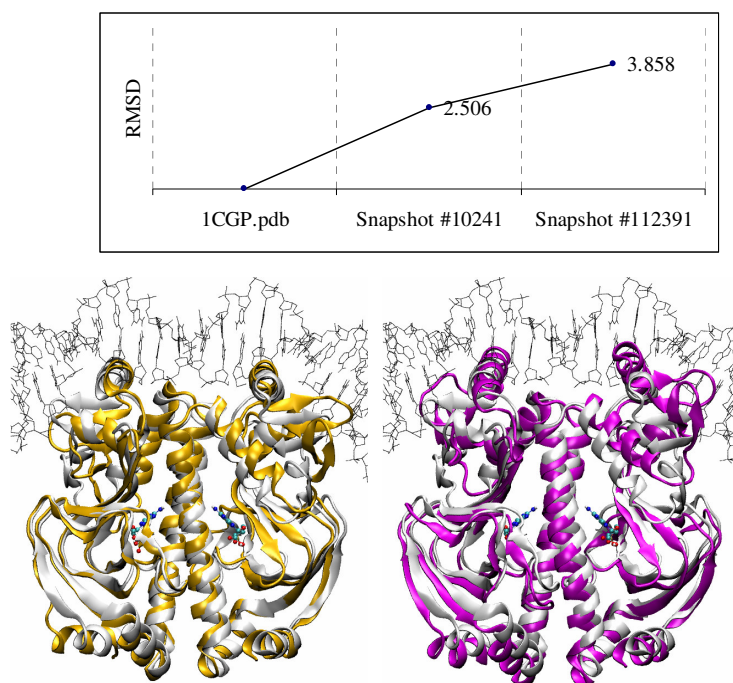


Figure 4.24. RMSDs of the cluster best members of CAP_{dimer} from the X-ray structure 1CGP. Superimposed structures are given below in the same order.

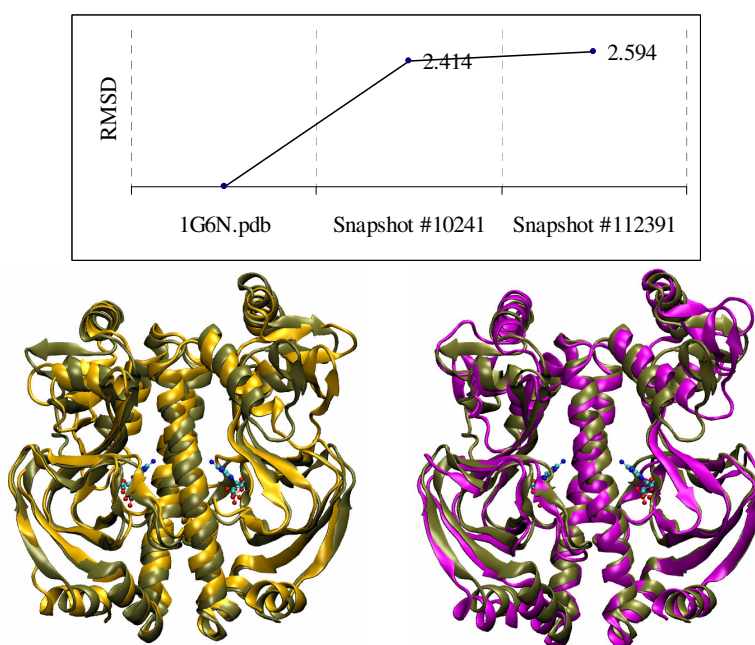


Figure 4.25. RMSDs of the cluster best members of CAP_{dimer} from the X-ray structure 1G6N. Superimposed structures are given below in the same order.

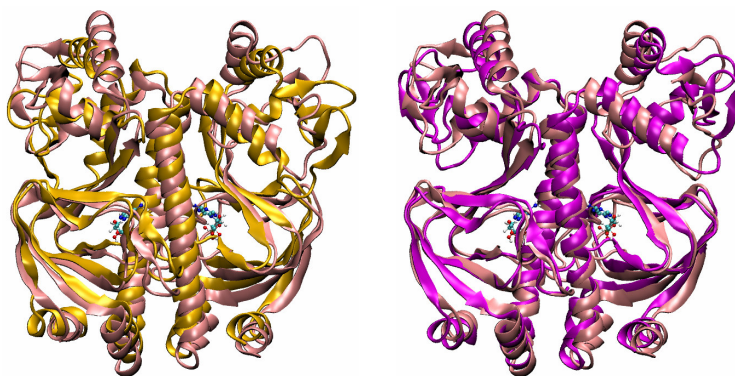
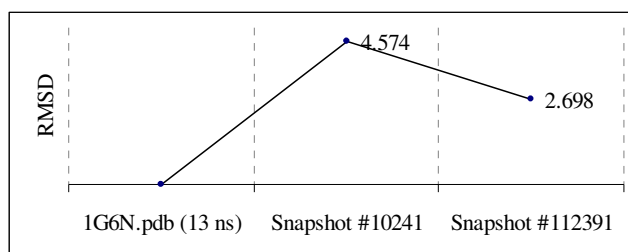


Figure 4.26. RMSDs of the cluster best members of CAP_{dimer} from the MD snapshot 1G6N (13 ns). Superimposed structures are given below in the same order.

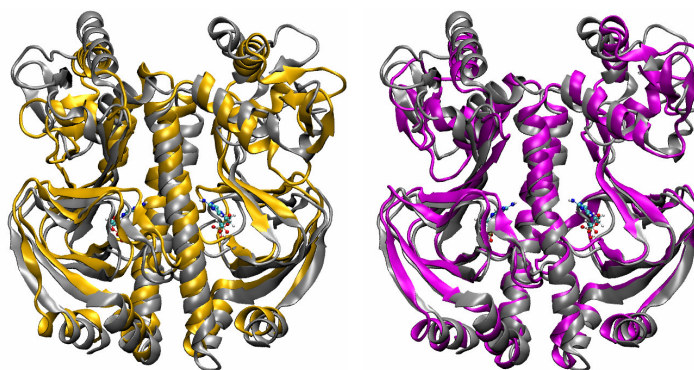
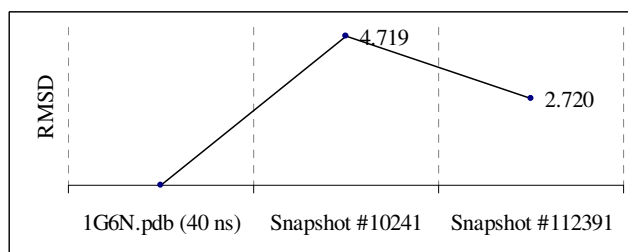


Figure 4.27. RMSDs of the cluster best members of CAP_{dimer} from the MD snapshot 1G6N (40 ns). Superimposed structures are given below in the same order.

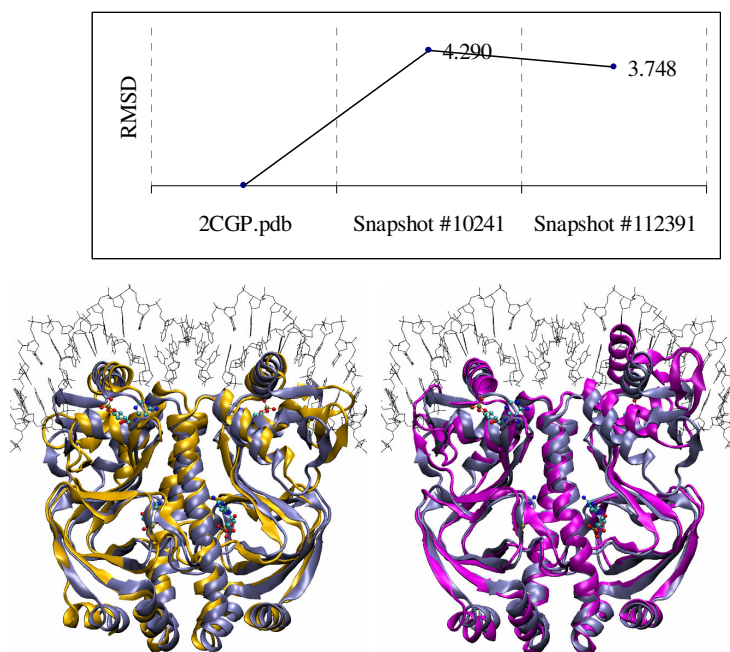


Figure 4.28. RMSDs of the cluster best members of CAP_{dimer} from the X-ray structure 2CGP. Superimposed structures are given below in the same order.

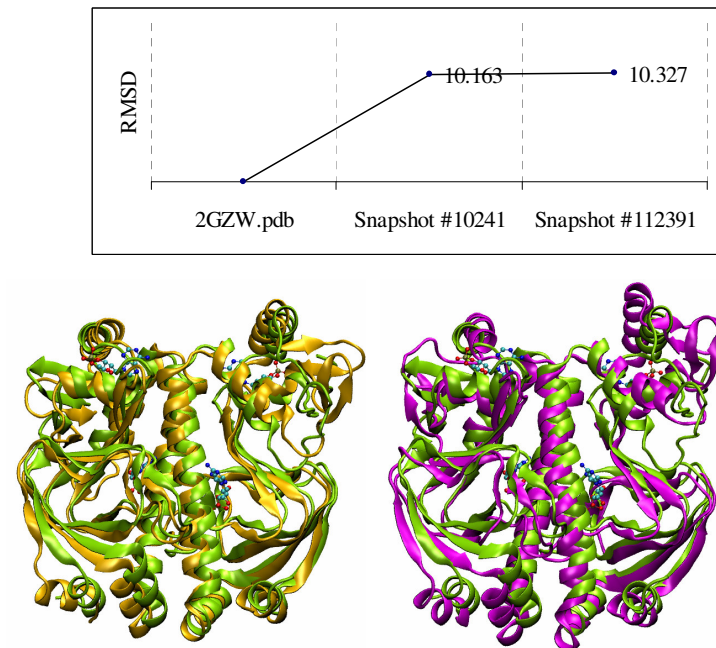


Figure 4.29. RMSDs of the cluster best members of CAP_{dimer} from the X-ray structure 2GZW. Superimposed structures are given below in the same order.

Similar to the monomer case, the largest RMS deviations are observed from the 2CGP and 2GZW structures. This set of figures suggest that, being dimer stabilizes the structure by stabilizing the C-helix.

4.3.3. RMSD of crystal structures versus cluster best members with cluster radius of 2.7 Å

Figure 4.30 and Figure 4.31 display the RMSD values of the cluster best members (with cluster radius 2.7 Å) from the structures 1CGP, 1G6N, 1G6N (13 ns), 1G6N (40 ns), 2CGP and 2GZW, for the monomer and the dimer cases, respectively. In comparison with the higher cluster radius value (3.5 Å), a transition between the different conformational states can clearly be seen.

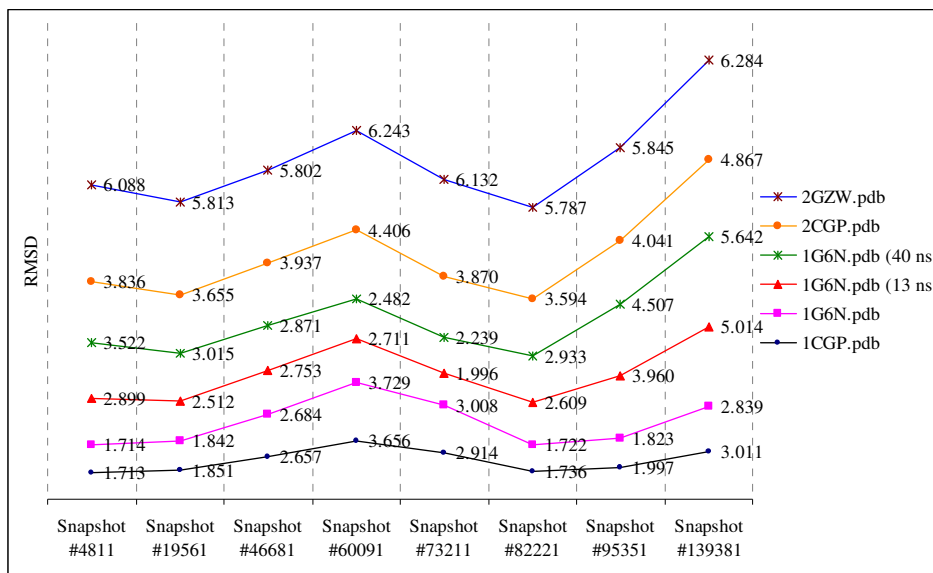


Figure 4.30. Monomer MD cluster best members (cluster radius 2.7 Å) versus the crystal structures and other MD snapshots of CAP.

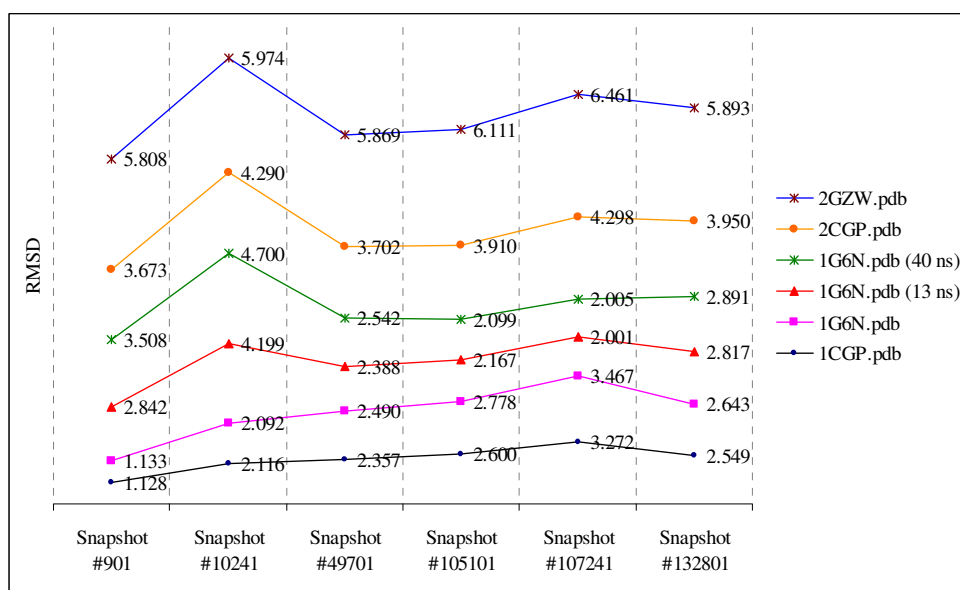


Figure 4.31. Dimer MD cluster best members (cluster radius 2.7 Å) versus the crystal structures and other MD snapshots of CAP.

4.3.4. GNM

Figure 4.32, Figure 4.33 and Figure 4.34 display the mean-square fluctuations (a) and the correlation between the fluctuations (b) in the slowest two modes for the three cluster best members (with cluster radius 3.5 Å) from the monomer simulations by GNM. In the first two modes, the slow mode shapes are conserved, whereas one of the hinge sites on the slowest mode of the third cluster best member is shifted to the left in the DNA binding domain. The correlations between the fluctuations display that the two domains of cAMP and DNA binding in all cluster members. Nevertheless, the hinge at the DNA binding site, which associate the two domains, are present in the first two cluster best members but shifted to the N terminus of the DNA binding domain in the third cluster best member. This is in accordance with the slowest mode shapes.

This shift in the hinge in the DNA binding domain is not observed when the trajectory is analyzed directly. In that analysis, all three clusters are analyzed together and this shift is not obvious due to the averaging. Additionally, in the first two clusters the

DNA binding site is strongly interacting with the secondary cAMP sites but this is not the case in the third cluster best member.

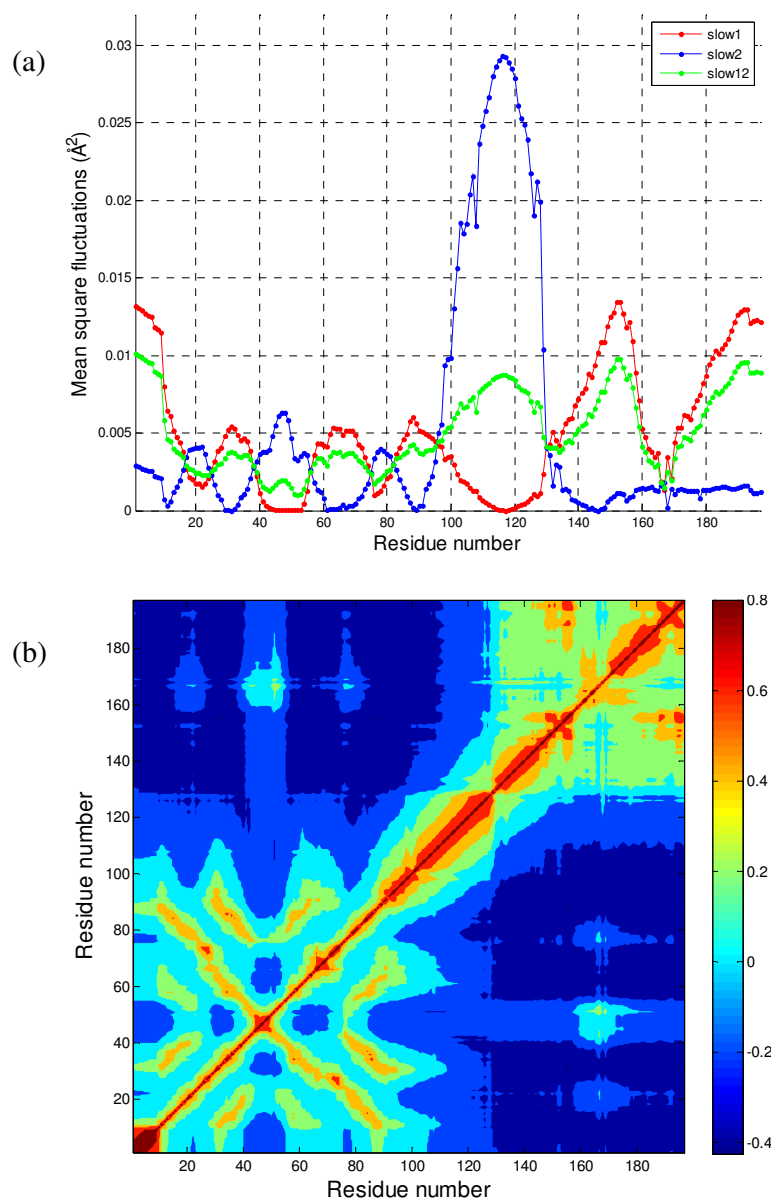


Figure 4.32. GNM (a) slow mode and (b) cross-correlation map of the monomer MD cluster best member 1 for the cluster radius 3.5 $\text{\AA}$.

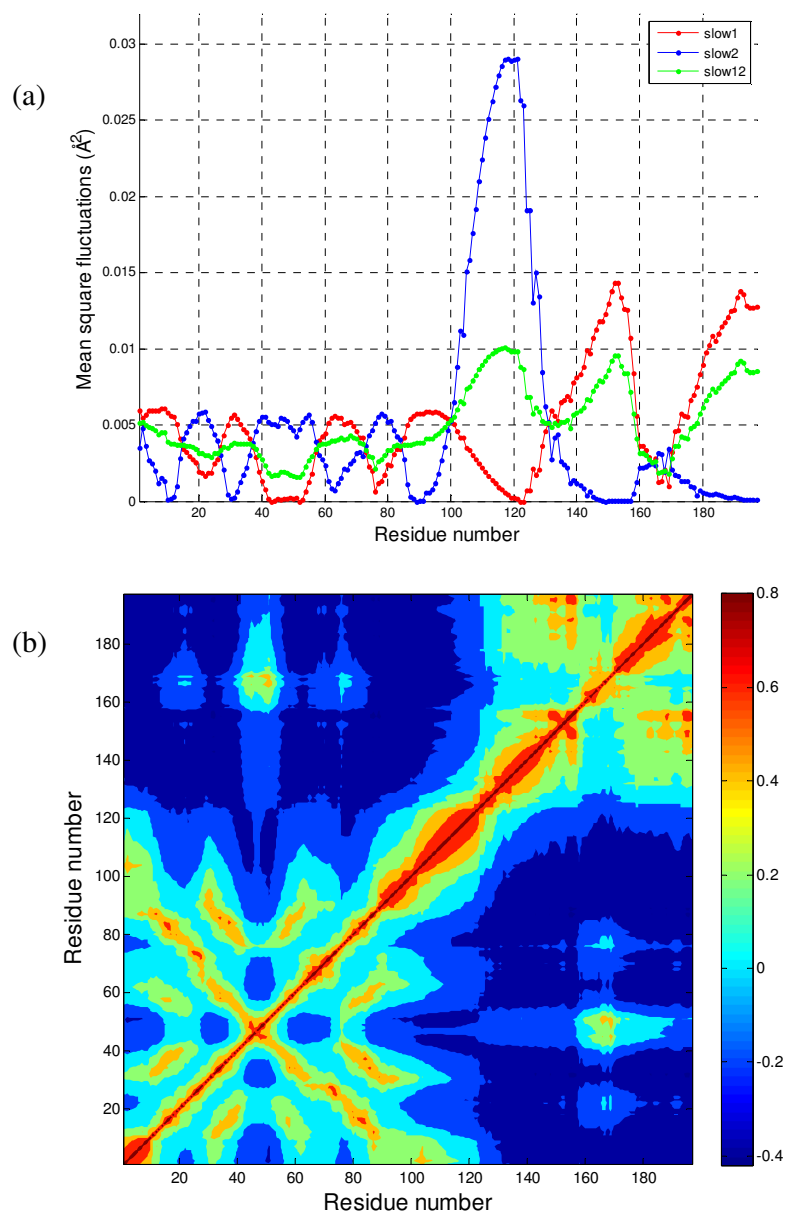


Figure 4.33. GNM (a) slow mode plot and (b) cross-correlation map of the monomer MD cluster best member 2 for the cluster radius 3.5 $\text{\AA}$.

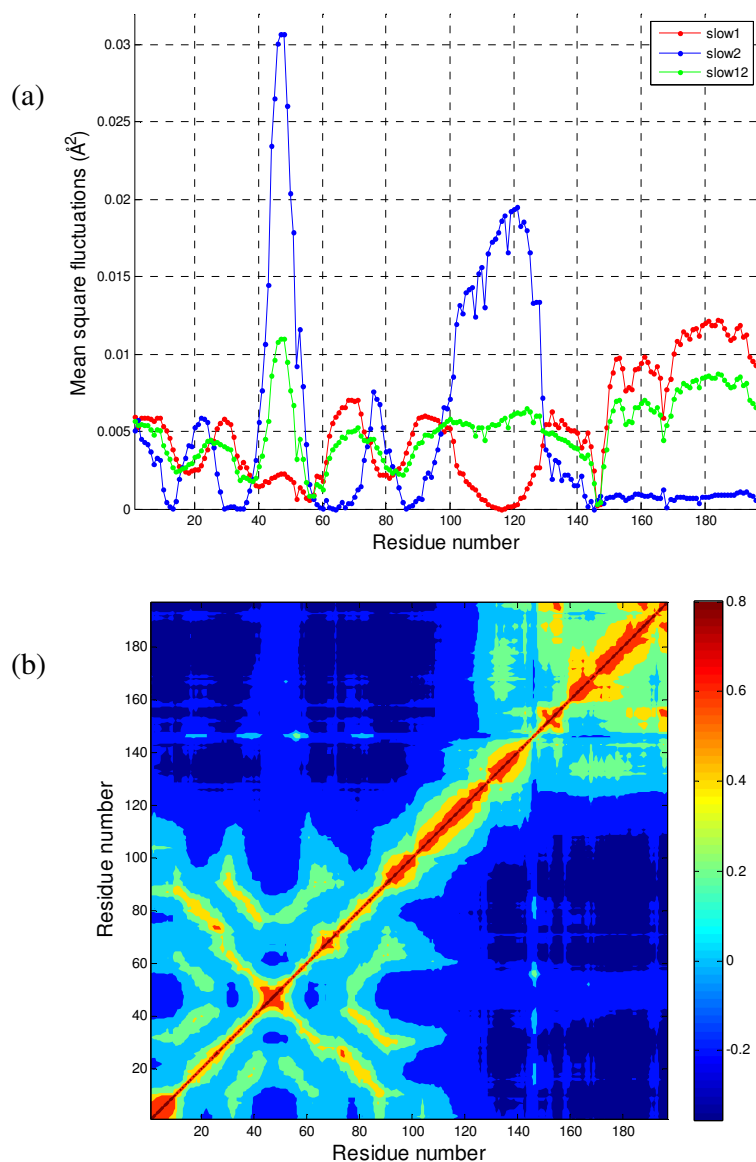


Figure 4.34. GNM (a) slow mode plot and (b) cross-correlation map of the monomer MD cluster best member 3 for the cluster radius 3.5 $\text{\AA}$.

Figure 4.35 and Figure 4.36 display the mean-square fluctuations (a) and the correlation between the fluctuations (b) in the slowest two modes for the two cluster best members (with cluster radius 3.5 $\text{\AA}$) from the dimer simulations by GNM. The slowest mode shapes and the correlation between the fluctuations are very similar for the two cluster best members. The hinge behaviour of the DNA binding region and its correlation with the primary and the secondary cAMP binding sites are observed in both clusters.

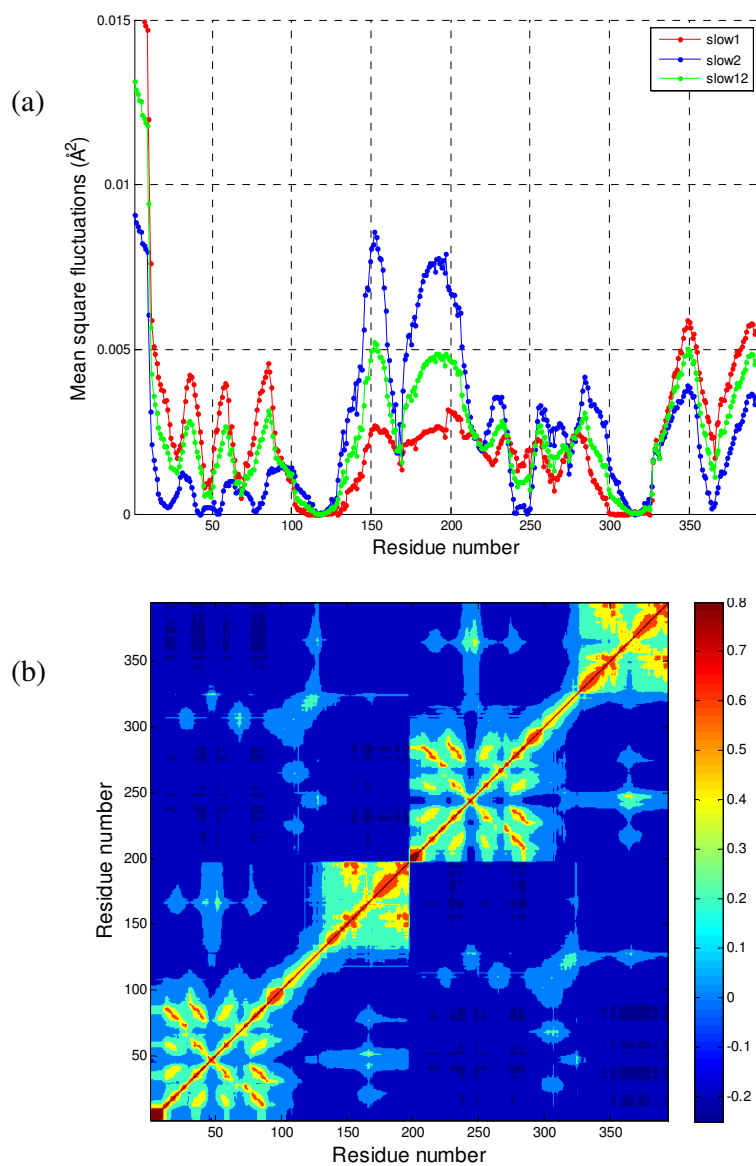


Figure 4.35. GNM (a) slow mode plot and (b) cross-correlation map of the dimer MD cluster best member 1 for the cluster radius 3.5 $\text{\AA}$.

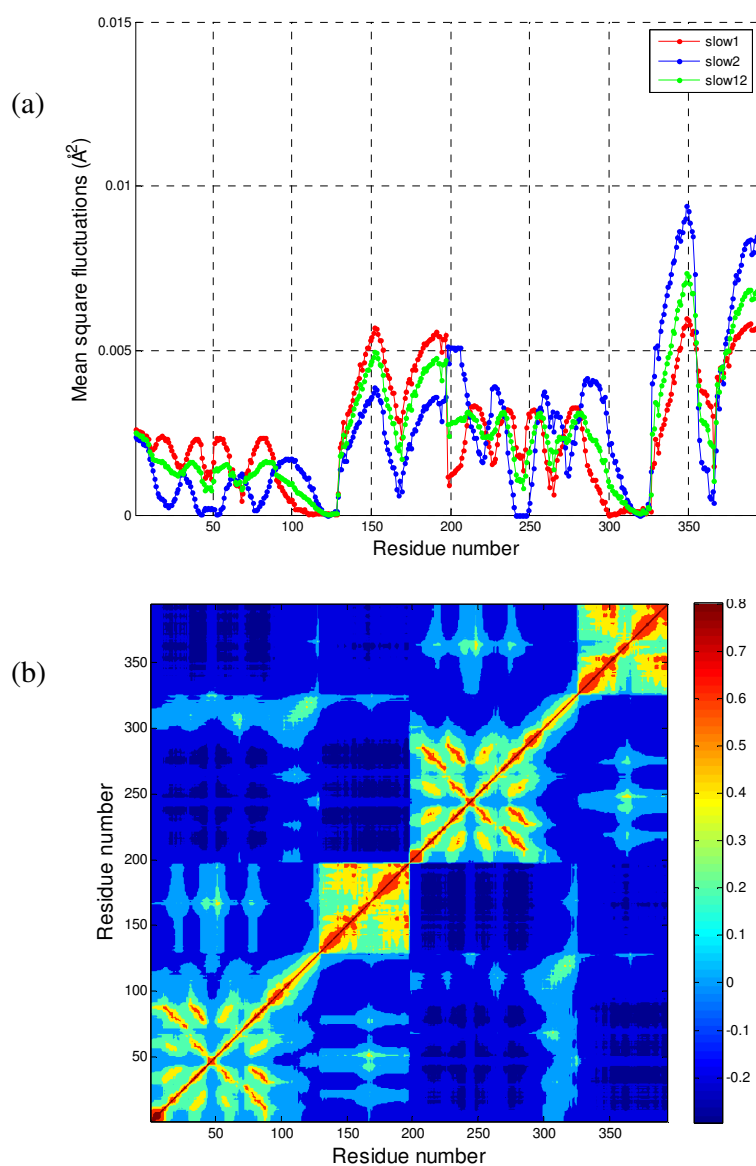


Figure 4.36. GNM (a) slow mode plot and (b) cross-correlation map of the dimer MD cluster best member 2 for the cluster radius 3.5 $\text{\AA}$.

Figure 4.37 displays the correlation between the fluctuations of residues for the X-ray crystal structure with the two cAMPS and the DNA bound. The DNA and cAMPS stabilizes the structure with the hinge regions around the secondary cAMP sites and the DNA binding site.

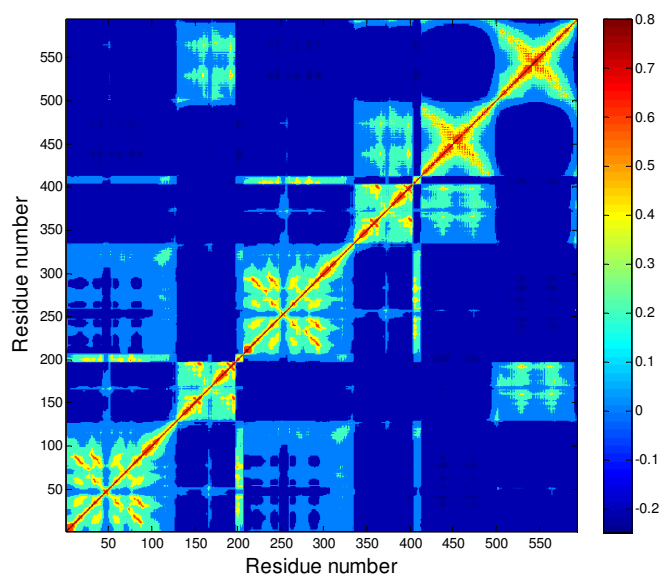


Figure 4.37. GNM cross-correlation map of the crystal structure 1CGP.

4.3.5. ANM

The fluctuations along the first and second ANM modes are examined for the best members of the clusters formed from the isolated monomer and dimer MD trajectories with the cluster radius of 3.5 Å. Figure 4.38 displays the predicted fluctuations of the three cluster best members of the monomer simulations for the first mode and the second mode. Significant fluctuations are obtained along the C-helix, AR1 loop and F-helix.

Figure 4.39 (a) and (b) displays the predicted fluctuations of the two cluster best members of the dimer simulations for the first mode and the second mode. The main fluctuation obtained from both modes is the big movement of the AR1 loop, especially for the second monomer (B) of the dimer.

Dimer fluctuations for both modes 1 and 2 differ from the monomer at all regions. The dimer does not fluctuate much at the dimerization interface (C-helix), whereas in the monomer C-helix bends from the middle while fluctuating. In the AR1 loop region, the monomer B of the dimer fluctuates at a smaller magnitude compared to the monomer, whereas show larger magnitude fluctuations compared to the monomer A of the dimer.

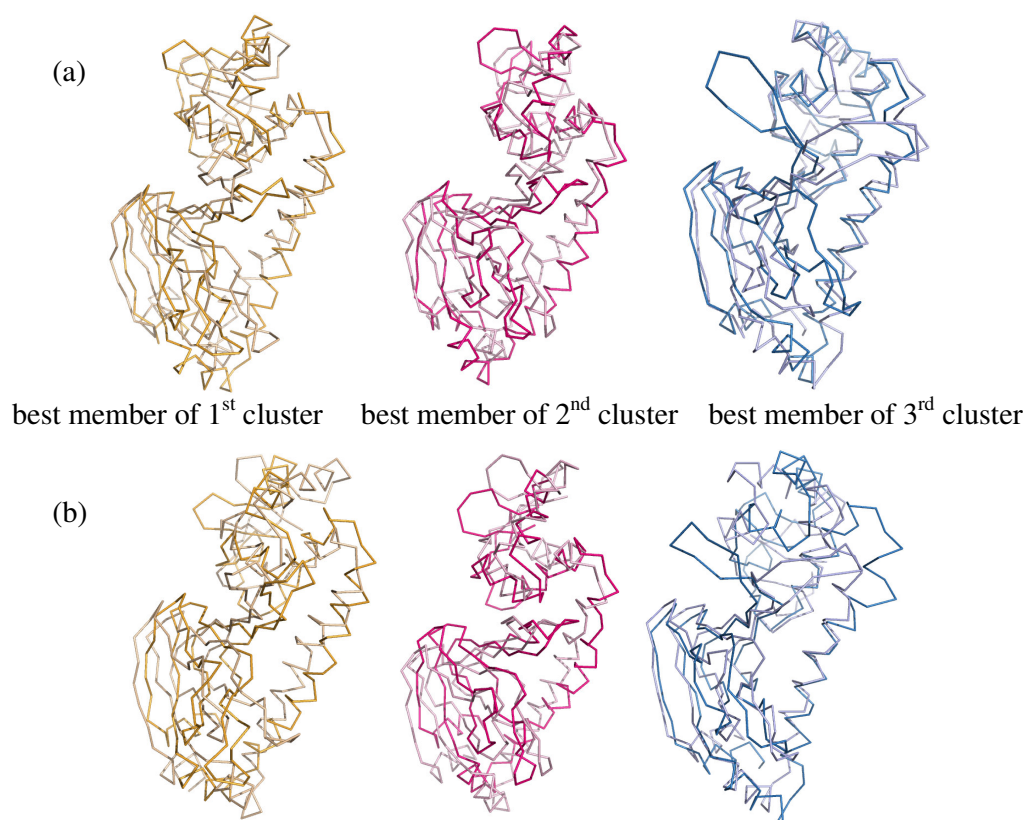


Figure 4.38. ANM (a) mode 1 and (b) mode 2 of monomer MD Best Members

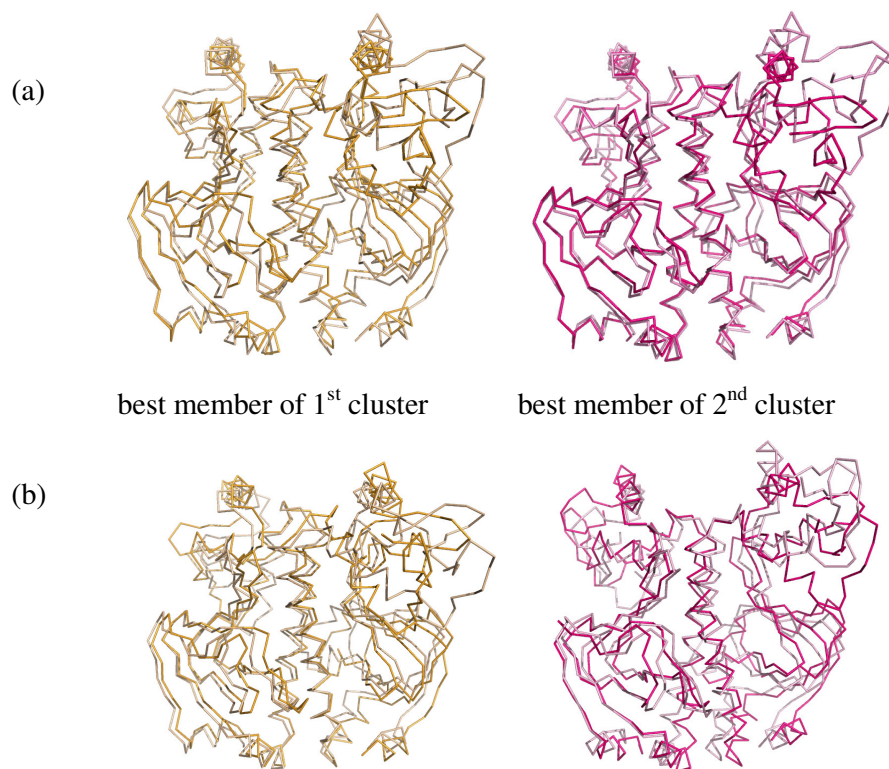


Figure 4.39. ANM (a) mode 1 and (b) mode 2 of dimer MD Best Members

4.3.6. Secondary Structure Assignments

The secondary structure assignments of the CAP_{monomer} and CAP_{dimer} (subunit A) is done for all the cluster best members formed by cluster radii 2.0, 2.5, 2.7, 3.0 and 3.5 Å. For this, STRIDE [48] utility of VMD is used.

In what follows, we use the clusters computed from our trajectory analysis as a framework for studying the secondary structure formation during the simulations for the monomer in both dimer and isolated states. As seen in Figure 4.40, the secondary structure of the monomer structure changes in some regions along the different snapshots (increasing snapshot number corresponds to increasing time). Between the residues 120-130 the monomer does not show helical tendencies after the 15th snapshot (corresponding to 20 ns in the trajectory) and between the residues 80-85 strand deformation occurs. These findings agree with the experimental findings addressing that the apo form of CAP shows reasonable differences from the holo form in secondary structure content in the regions D68-S83 and T127-N133 by Won et al. [49].

For the dimer case, except some small changes, the secondary structure stays intact.

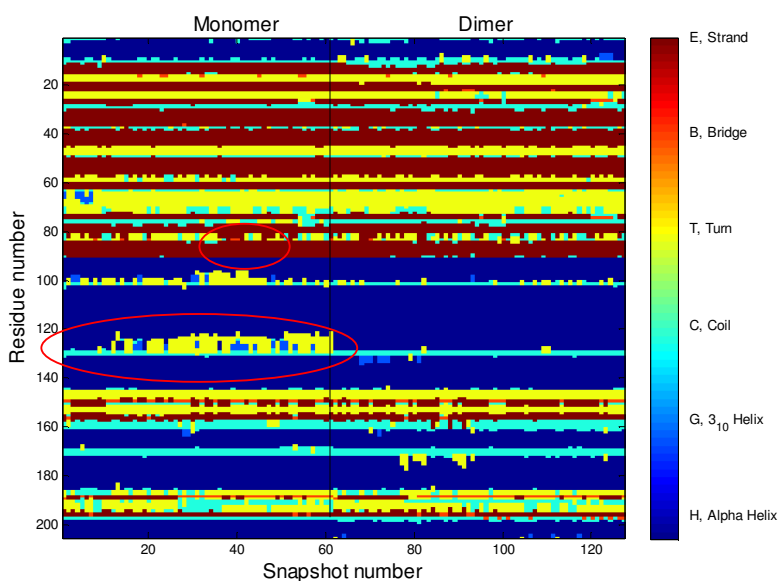


Figure 4.40. Secondary structure assignments to the CAP_{monomer} and CAP_{dimer} (subunit A) trajectory cluster best members.

In Figure 4.40, the set of snapshots analyzed are formed from the cluster best members of the clusters formed by cluster radii 2.0, 2.5, 2.7, 3.0 and 3.5 Å. Snapshot numbers 1-61 and 62-127 belong to the monomer and dimer md trajectories, respectively.

4.4. General Discussion

The RMSDs from the initial X-ray crystal structures follows a different behavior in time for the monomer and the dimer simulations. The RMSD stabilizes around 3 and 4 Å in the respective cases. Although the RMSD is lower for the unbound monomer, the structure samples a larger conformational space reflected by a large amplitude fluctuation profile.

The mean-square fluctuations between the unbound and bound monomer structures and with comparison to X-ray crystal structures display that the deviations are at the most in the binding interface, including cAMP binding sites. The other observed variations are at the secondary cAMP binding sites and the selective DNA binding residues. These regions are destabilized in the monomeric state. Interestingly, the selective binding DNA region is stabilized in the dimer state in the absence of the cAMP and the DNA.

The correlation between the fluctuations of residues in the unbound and bound states displays that the dimerization enhances the appearance of the domains, the cAMP and DNA binding domains, and the association of the domains through the DNA binding site. The association of the two monomers in the dimers are also maintained by the DNA binding site. This is in accordance with the mean-square fluctuations.

In the fluctuations of the backbone rotational angles, the significant deviations that are observed between the two states correspond to two sites: Secondary cAMP binding sites and the RNAP binding sites. The bonds of these two regions display enhanced rotational mobility in the unbound state. Although these regions are highly mobile with respect to the mean-square fluctuations in both states, the dimerization restricts the rotations of the bonds in this region.

Clustering of the conformations in both states individually and together allow the analysis of a large number of conformations in a comprehensive and convenient format. The number of clusters decreases with larger cluster radius and as going from the monomeric to the dimeric state. When they are clustered together, a cluster may have members from the unbound monomer simulations and the dimer simulations. That give evidence for the preexisting conformations of the monomer of the dimer in unbound states.

The analysis of the cluster best members suggests that the variations in the early stage is associated with DNA binding, which follows the conformational adjustments due to the absence of the cAMPs. This is concluded by the RMDS analysis of the best member with respect to several known X-ray crystal structures and snapshots from another MD simulations of CAP with cAMPs.

The analysis of the cluster best members by ANM suggests that the predicted fluctuations around the cluster best members largely overlap the conformational space spanned by the cluster best members.

The analysis of the cluster best member by GNM suggests that the dynamic fluctuations are more conserved in the dimeric state than in the monomeric state. The shift of a hinge in DNA binding domain is observed by the GNM slowest mode fluctuations for one of the best members in the monomeric state. This also agrees the experimental observation.

The use of the MD method before the analysis with EN models was crucial for this study. The models of the apo form of CAP were constructed by the removal of the cAMPs and the DNA using the crystal structure of CAP in complex with the cAMPs and the DNA. Therefore analyzing these structures directly with EN methods wouldn't give as much of an idea of the features of the apo form, since these methods mainly depend on the 3D topology of the structure. By performing MD, the structures adjust themselves to the new conditions (absence of ligands and DNA) in search for other equilibrium structures and EN methods applied afterwards give more insight into the dynamics of the newly formed models.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The combined analysis of the calculations suggest that the monomeric state is a plausible structure for the apo state in comparison with the experimental observations of salt bridge formation and the secondary structure content of the apo form. Nevertheless, the dimerization stabilizes the domains and the association of the two domains through the DNA binding site.

Clustering analysis of the MD trajectories provide a convenient means to study the structural and dynamics peculiarities of the structures.

Clustering the conformations of the isolated monomer and the monomer of the dimer from MD simulations together suggest the preexistence of the bound conformations in the unbound state.

5.2. Recommendations

Further MD simulations can be performed for the unknown structures of CAP like CAP:(cAMP)₁ and CAP:(cAMP)₃ which are determined to exist experimentally. Also MD simulations of apo CAP models constructed from other crystal structures can be carried out. This way it would be possible to get a complete picture of the allosteric transitions of CAP upon ligand, DNA and RNAP binding.

Clustering analysis of the monomer and the dimer MD conformations together was performed by selecting the monomer A of the dimer. It would complement the work performed in this thesis if the other monomer (B) of the dimer is included in the cluster analysis, since both in the trajectory and the ANM mode analysis the different behaviour of the two monomers of the dimer is observed.

APPENDIX A

A.1 Salt Bridges

The distances between the oxygen and hydrogen atoms of the residue pairs E72-R82, D68-R123 and E72-R123 have been plotted along the monomer and dimer MD trajectories. The four plots for each residue pair indicate the four combinations of possible atom interactions.

Figure A.1, Figure A.2 and Figure A.3 display the distance plots of the monomer case for the residue pairs E72-R82, D68-R123 and E72-R123, respectively. Figure A.4, Figure A.5 and Figure A.6 display the distance plots of the monomer A of the dimer and Figure A.7, Figure A.8 and Figure A.9 display the distance plots of the monomer B of the dimer, respectively for the residue pairs E72-R82, D68-R123 and E72-R123.

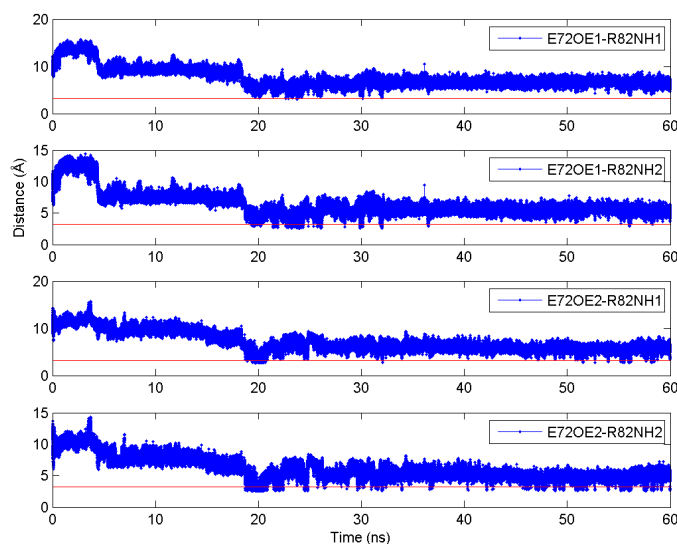


Figure A.1. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in isolated monomer MD

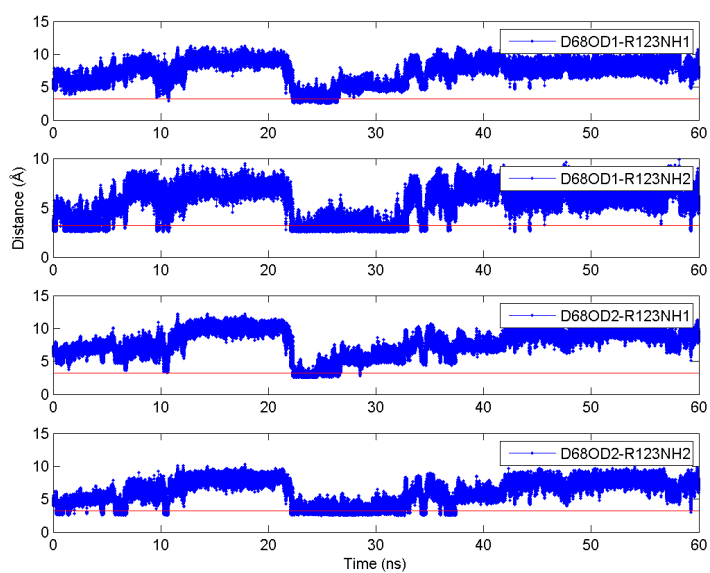


Figure A.2. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in isolated monomer MD

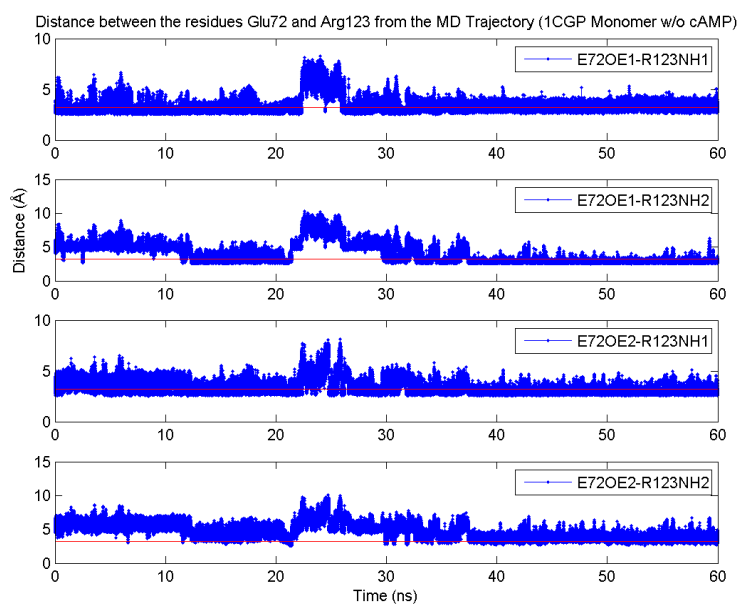


Figure A.3. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in isolated monomer MD

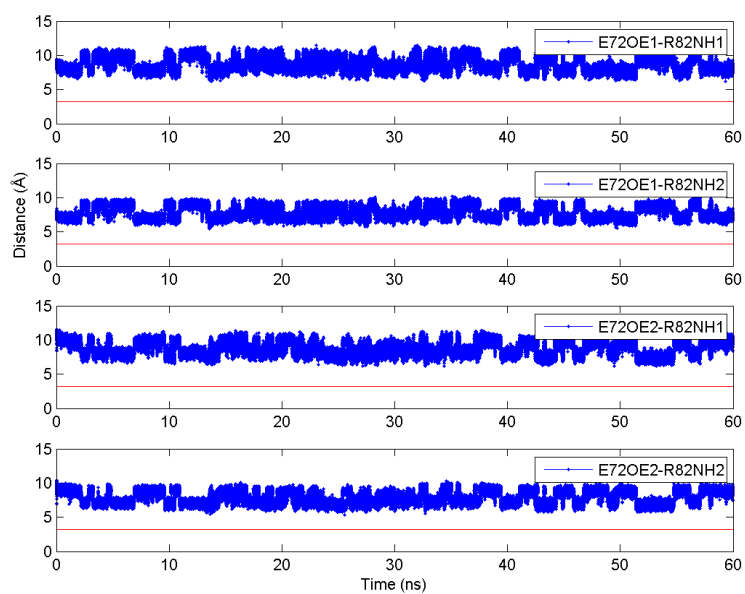


Figure A.4. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in monomer A of dimer MD

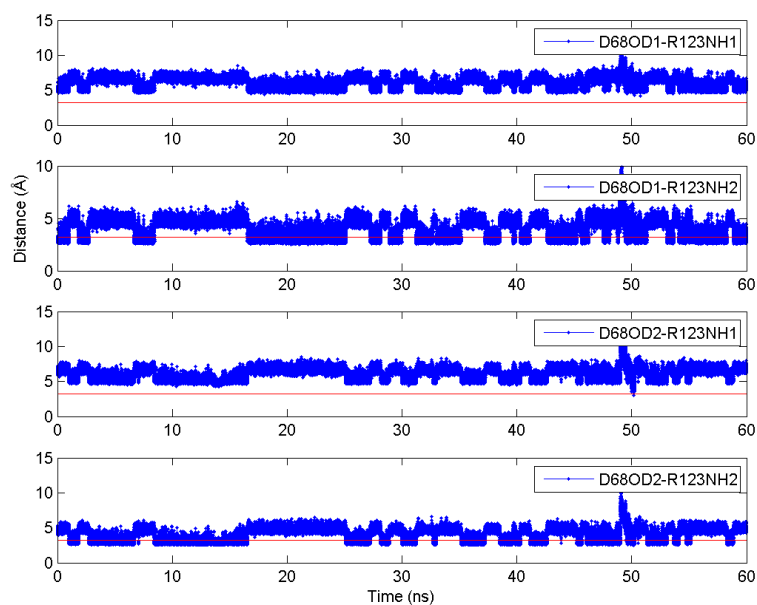


Figure A.5. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in monomer A of dimer MD

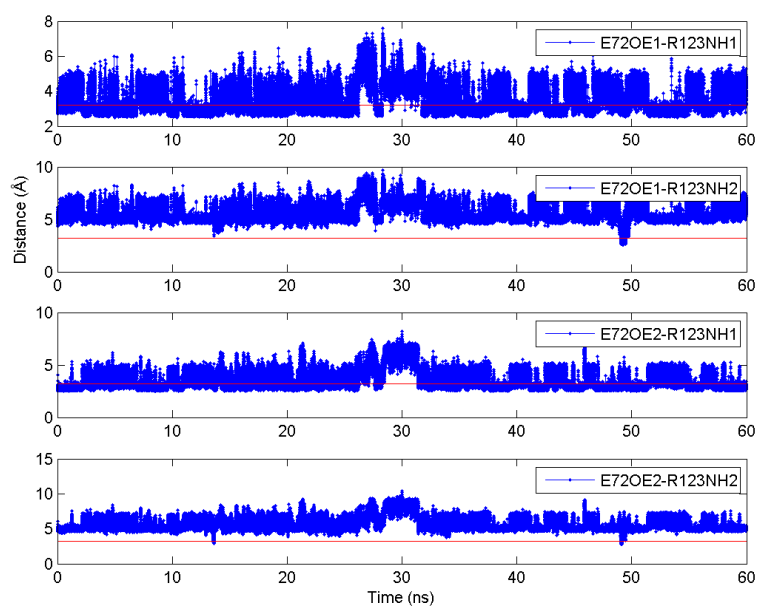


Figure A.6. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in monomer A of dimer MD

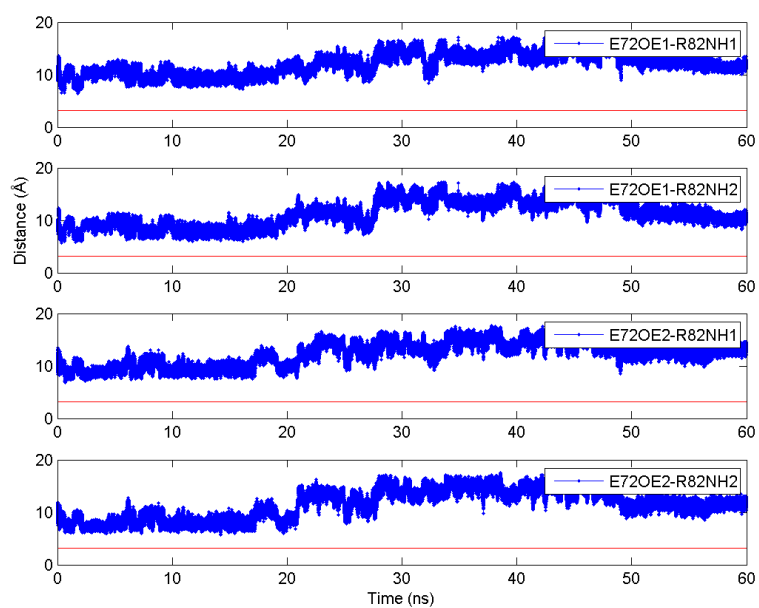


Figure A.7. Oxygen-Nitrogen atom distances between Glu72 and Arg82 in monomer B of dimer MD

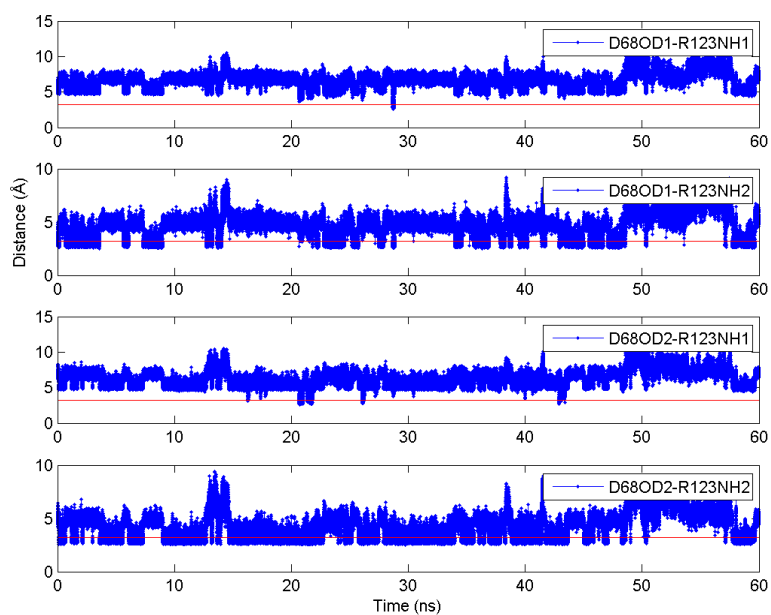


Figure A.8. Oxygen-Nitrogen atom distances between Asp68 and Arg123 in monomer B of dimer MD

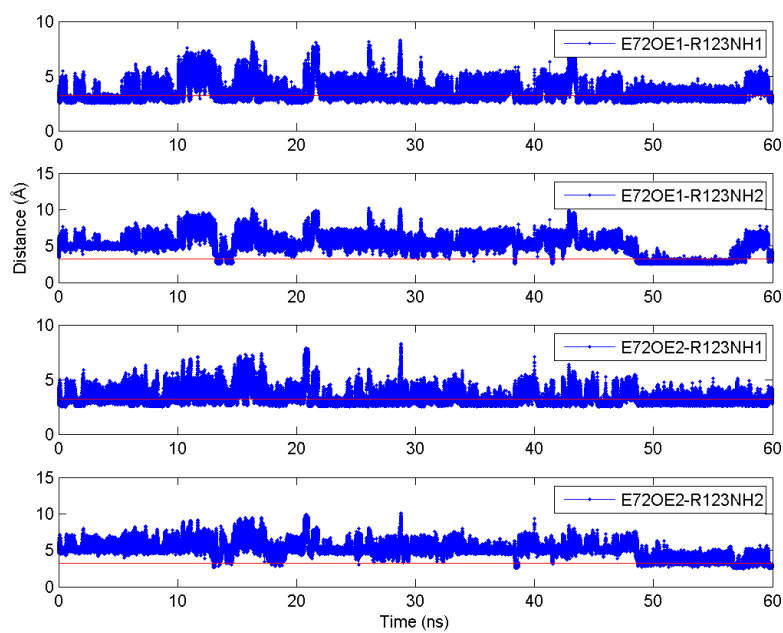


Figure A.9. Oxygen-Nitrogen atom distances between Glu72 and Arg123 in monomer B of dimer MD

A.2 Cluster Analysis

Cluster evolution in time for the monomer and the dimer MD trajectories with the cluster radii of 2.0, 2.5 and 2.7 Å are given in Figure A.10 through Figure A.18.

A.2.1 Monomer MD Simulation Snapshots

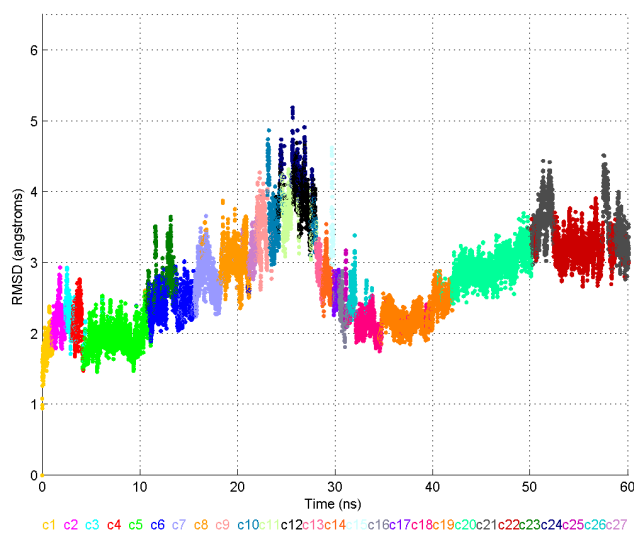


Figure A.10. Cluster evolution in time for CAP_{monomer} on the RMSD plot with cluster radius of 2.0 Å

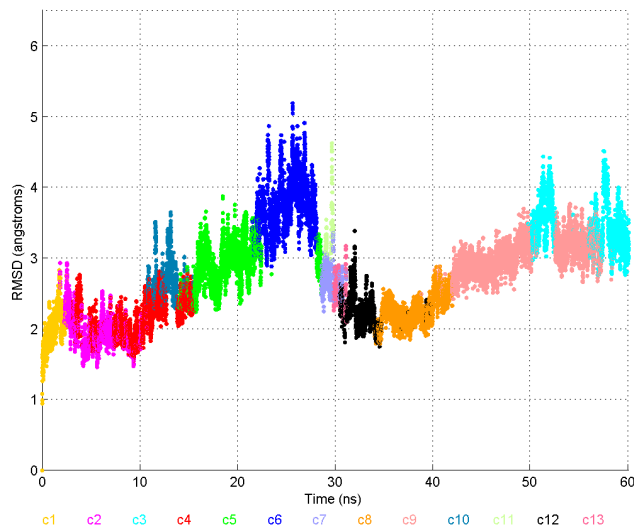


Figure A.11. Cluster evolution in time for CAP_{monomer} on the RMSD plot with cluster radius of 2.5 Å

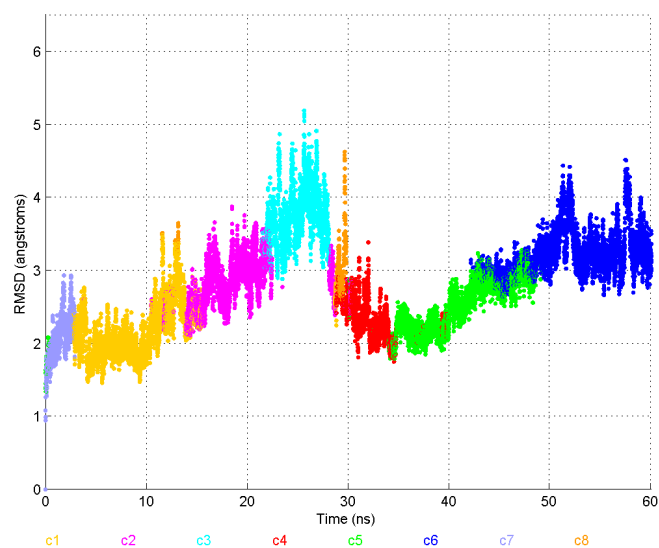


Figure A.12. Cluster evolution in time for CAP_{monomer} on the RMSD plot with cluster radius of 2.7 Å

A.2.2 Dimer MD Simulation Snapshots

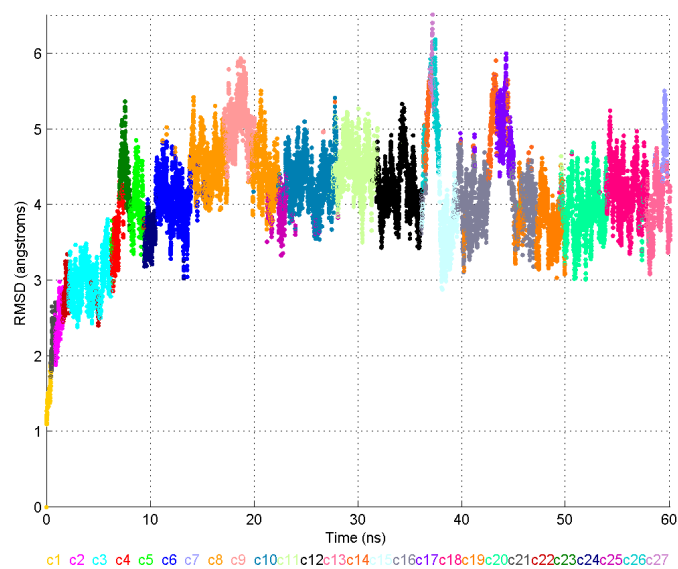


Figure A.13. Cluster evolution in time for CAP_{dimer} on the RMSD plot with cluster radius of 2.0 Å

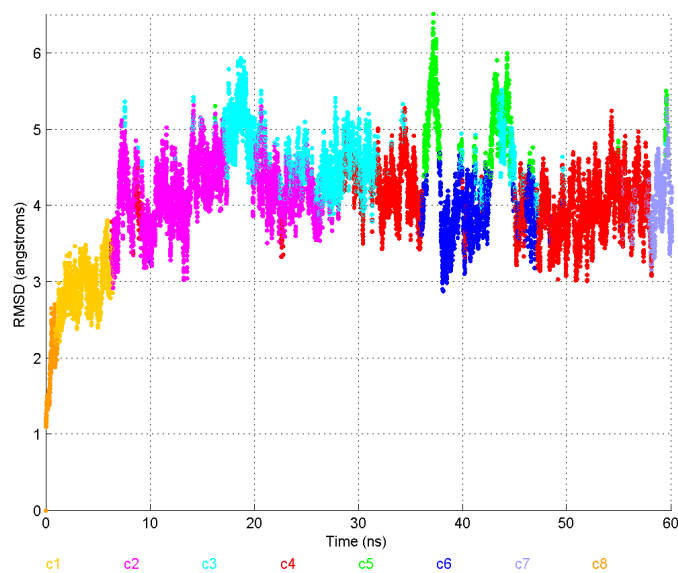


Figure A.14. Cluster evolution in time for CAP_{dimer} on the RMSD plot with cluster radius of 2.5 Å

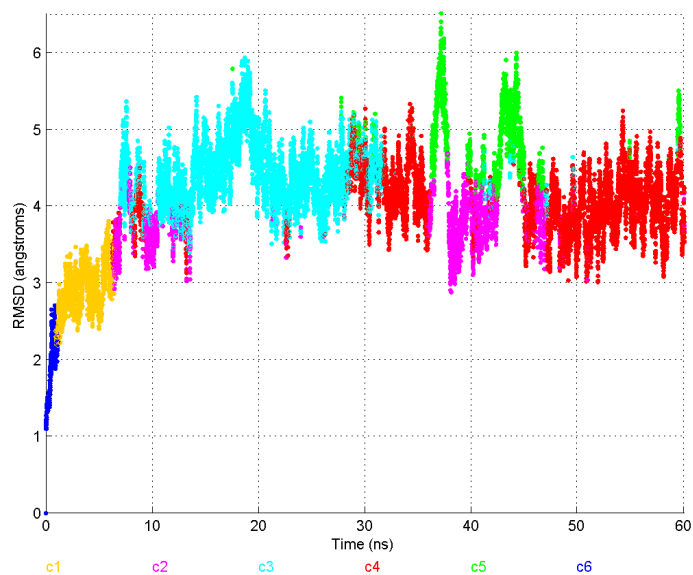


Figure A.15. Cluster evolution in time for CAP_{dimer} on the RMSD plot with cluster radius of 2.7 Å

A.2.3 Monomer & Dimer MD Simulation Snapshots

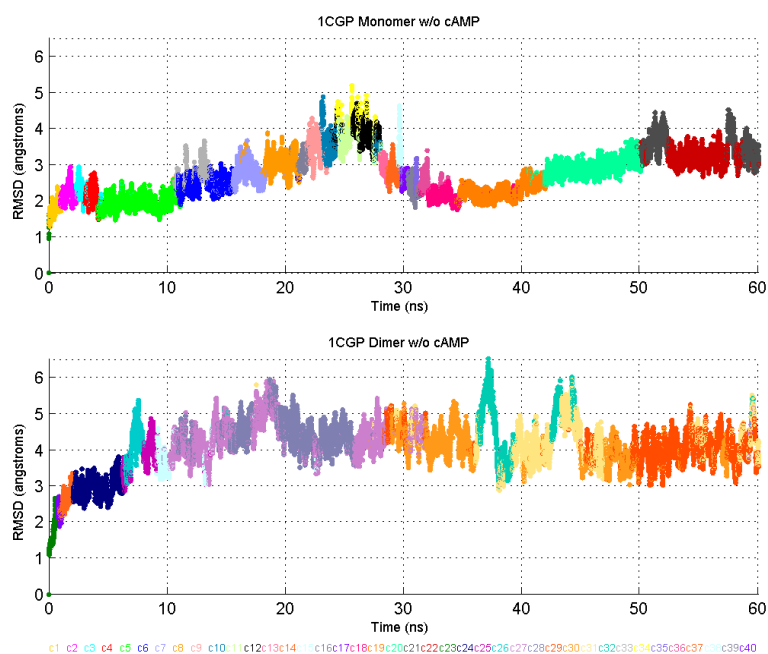


Figure A.16. Cluster evolution in time for CAP_{monomer} and CAP_{dimer} subunit A on the RMSD plot with cluster radius of 2.0 Å

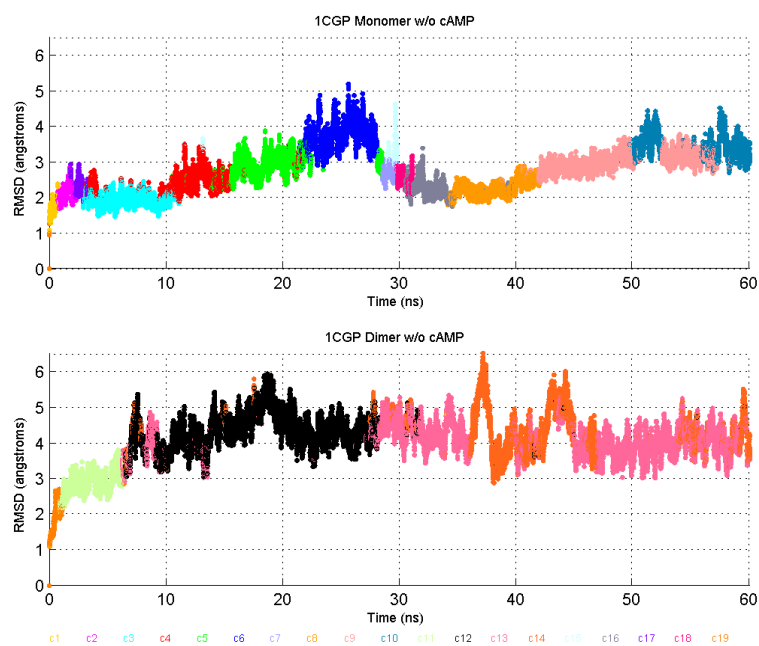


Figure A.17. Cluster evolution in time for CAP_{monomer} and CAP_{dimer} subunit A on the RMSD plot with cluster radius of 2.5 Å

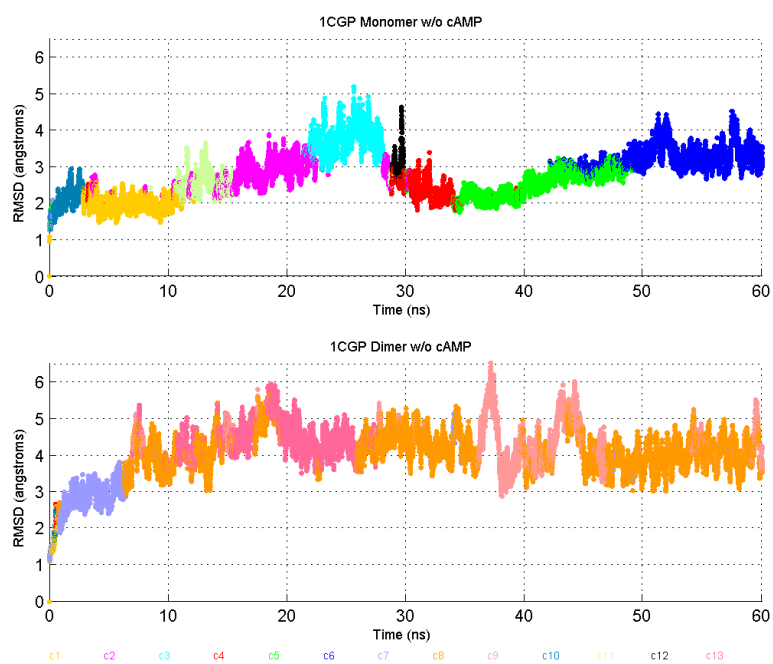


Figure A.18. Cluster evolution in time for CAP_{monomer} and CAP_{dimer} subunit A on the RMSD plot with cluster radius of 2.0 Å

REFERENCES

1. McCammon, A. and C.H. Harvey, *Dynamics of proteins and nucleic acids*, Cambridge University Press, 1987.
2. Adcock, S.A. and A. McCammon, “Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins”, *Chem. Rev.*, vol. 106, pp. 1589-1615, 2006.
3. Karplus, M. and A. McCammon, “Molecular dynamics simulations of biomolecules”, *Nature Structural Biology*, vol. 9-9, pp. 646-652, 2002.
4. Bahar, I., Atilgan, A.R., and Erman, B. “Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential”, *Fold. Des.*, vol. 2, pp. 173-181, 1997.
5. Haliloglu, T., I. Bahar and B. Erman, “Gaussian dynamics of folded proteins”, *Phys. Rev. Lett.*, vol. 79, pp. 3090-3093, 1997.
6. Atilgan, A.R., S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model”, *Biophys. J.*, vol. 80, pp. 505-515, 2001.
7. Feig, M., J. Karanicolas and C.L. Brooks, “MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology”, *J. Mol. Graph. Model.*, vol. 22(5), pp. 377-395, 2004.
8. Alberts et al., *Essential Cell Biology* (2nd Edition), Garland Science, 2004.
9. Murray et al., *Harper’s Illustrated Biochemistry* (27th Edition), McGraw-Hill, 2006.
10. Bhagavan, N.V., *Medical Biochemistry* (4th Edition), Jones and Bartlett, 2002.

11. Humphrey, W., A. Dalke and K. Schulten, "VMD - Visual Molecular Dynamics", J. Molec. Graphics, vol. 14, pp. 33-38, 1996.
12. Schultz, S., G. Shields and T.A. Steitz, "Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees (PDB code: 1CGP)", Science, vol. 253, pp. 1001-1007, 1991.
13. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne, "The Protein Data Bank", Nucleic Acids Research, vol. 28, pp. 235-242, 2000.
14. Harman, J.G., "Allosteric regulation of the cAMP receptor protein", BBA, vol. 1547, pp. 1-17, 2001.
15. Garcia, A. and J.G. Harman, "Simulations of CRP:(cAMP)₂ in noncrystalline environments show a subunit transition from the open to the closed conformation", Protein Science, vol. 5, 62-71, 1996.
16. Kolb, A., S. Busby, H. Buc, S. Garges and S. Adhya, "Transcriptional regulation by cAMP and its receptor protein", Annu.Rev.Biochem., vol. 62, pp. 749-795, 1993.
17. Berrera, M., S. Pantano and P. Carloni, "Catabolite Activator Protein in Aqueous Solution : A Molecular Simulation Study", J. Phys. Chem., vol. 111, pp. 1496-1501, 2007.
18. Weber, I.T. and T.A. Steitz, "Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution", J. Mol. Biol., vol. 198, pp. 311-326, 1987.
19. Passner, J.M., S.C. Schultz and T.A. Steitz, "Modeling the cAMP induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution (PDB code: 1G6N)", J. Mol. Biol., vol. 304, pp. 847-859, 2000.

20. Case, D.A., T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Jr. Merz, A. Onufriev, C. Simmerling, B. Wang and R. Woods, "The Amber biomolecular simulation programs", *J. Comput. Chem.*, vol. 26, pp. 1668-1688, 2005.
21. Karplus, M. and G.A. Petsko, "Molecular dynamics simulations in biology", *Nature*, vol. 347, pp. 631-639, 1990.
22. Case, D.A., T.A. Darden, T.E.I. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman et al., "AMBER 8", University of California, San Francisco, 2004.
23. Duan, Y., C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins", *J. Comput. Chem.*, vol. 24, pp. 1999-2012, 2003.
24. Leach, A.R., *Molecular Modeling: Principles and Application* (2nd Edition), Prentice Hall, 2001.
25. Verlet, L., "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules", *Physical Review*, vol. 159, pp. 98-106, 1967.
26. Jorgensen, W. L., J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein, "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys.*, vol. 79, pp. 926-935, 1983.
27. "H++: A web-based server for computational prediction of protonation states and pK of ionizable groups in macromolecules", <http://biophysics.cs.vt.edu/H++>.
28. Gordon, J.C., J.B. Myers, T. Folta, V. Shoja, L.S. Heath and A. Onufriev, "H++: A server for estimating pKas and adding missing hydrogens to macromolecules.", *Nucleic Acids Res.*, Jul 1; 33 (Web Server issue): W368-71, 2005.

29. Bashford, D. and M. Karplus, "pKa of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model", *Biochemistry*, vol. 29, pp. 10219-10225, 1990.
30. Ryckaert, J. P., G. Ciccotti and H.J.C. Berendsen, "Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes", *J. Comput. Phys.*, vol. 23, pp. 327-341, 1977.
31. Essman, U., L. Perera, M.L. Berkowitz, T.A. Darden, H. Lee and L.G. Pedersen, "A smooth Particle Mesh Ewald method", *J. Chem. Phys.*, vol. 103, pp. 8577-8593, 1995.
32. Berendsen, H.J.C., J.P.M. Postma, W.F. Van Gunsteren, A. DiNola and J.R. Haak, "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.*, vol. 81, pp. 3684-3690, 1984.
33. Karpen, M.E., D.J. Tobias and C.L. Brooks, "Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV", *Biochemistry*, vol. 32(2), pp. 412-420, 1993.
34. Torda, A.E. and W.F. van Gunsteren, "Algorithms for clustering molecular dynamics configurations", *J. of Comp. Chem.*, vol. 15(12), pp. 1331-1340, 1994.
35. Bui, J.M., Z. Radic, P. Taylor and J.A. McCammon, "Conformational Transitions in Protein-Protein Association: Binding of Fasciculin-2 to Acetylcholinesterase", *Biophysical Journal*, vol. 90, pp. 3280-3287, 2006.
36. Best, C. and H.C. Hege, "Visualizing and Identifying Conformational Ensembles in Molecular Dynamics Trajectories", *Computing in Science and Engineering*, vol. 04(3), pp. 68-75, 2002.
37. Cui, Q. and I. Bahar (Editors), *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, CRC Press, Boca Raton, FL, 2005.
38. Flory, P.J., "Statistical thermodynamics of random networks", *Proc. R. Soc. Lond. A*, vol. 351, pp. 351-380, 1976.

39. DeLano, W.L., "The PyMOL Molecular Graphics System", DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>, 2002.
40. Kumarevel, T.S. et al., "Crystal structure of activated crp protein from E. coli. (PDB code: 2GZW)", to be published.
41. Passner, J.M. and T.A. Steitz, "The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer (PDB code: 2CGP)", PNAS, vol. 94, pp. 2843-2847, 1997.
42. Scott, S.P. and S. Jarjous, "Proposed structural mechanism of Escherichia coli cAMP receptor protein cAMP-dependent proteolytic cleavage protection and selective and nonselective DNA binding", Biochemistry, vol. 44, pp. 8370-8748, 2005.
43. Busby, S. and R.H. Ebright, "Transcription Activation by Catabolite Activator Protein (CAP)", J. Mol. Biol., vol. 293, pp. 199-213, 1999.
44. McKay D.B., I.T. Weber and T.A. Steitz, "Structure of Catabolite Gene Activator Protein at 2.9-Å Resolution: Incorporation of amino acid sequence and interactions with cyclic AMP", J. of Biol. Chem, vol. 257 (16), pp. 9518-9524, 1982.
45. Youn H., R.L. Kerby, J. Koh and G.P. Roberts, "A C-helix residue, Arg123, has important roles in both the active and inactive forms of the cAMP receptor protein", J. of Biol. Chem, vol. 282(6), pp. 3632-3639, 2007.
46. Tutar, Y. and J.G. Harman, "Effect of salt bridge on transcription activation of CRP-dependent lactose operon in Escherichia coli", Archives of Biochemistry and Biophysics, vol. 453, pp. 217-223, 2006.
47. Walker, R., "TUTORIAL B3: Case Study - Folding TRP Cage (Advanced analysis and clustering)", <http://amber.scripps.edu/tutorials/basic/tutorial3/>.

48. Frishman, D. and P. Argos, "Knowledge-Based Protein Secondary Structure Assignment", *PROTEINS: Structure, Function, and Genetics*, vol. 23, pp. 566-579, 1995.
49. Won H.Y., T. Yamazaki, T.W. Lee, M.K. Yoon, S.H. Park, Y. Kyogoku and B.J. Lee, "Structural Understanding of the Allosteric Conformational Change of Cyclic AMP Receptor Protein by Cyclic AMP Binding", *Biochemistry*, vol. 39, pp. 13953-13962, 2000.
50. Mongan, J., "Interactive Essential Dynamics", *Journal of Computer-Aided Molecular Design*, vol. 18(6), pp. 433-436, 2004.
51. Amadei, A., A.B.M. Linssen and H.J.C. Berendsen, "Essential Dynamics of Proteins", *PROTEINS: Structure, Function, and Genetics*, vol. 17, pp. 412-425, 1993.
52. Armon A., D. Graur and N. Ben-Tal, "ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information", *J. Mol. Biol.*, vol. 307, pp. 447-463, 2001.
53. Doruker, P., A.R. Atilgan and I. Bahar, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to α -Amylase Inhibitor", *PROTEINS: Structure, Function, and Genetics*, vol. 40, pp. 512-524, 2000.
54. Lawson, C.L., D. Swigon, K.S. Murakami, A.A. Darst, H.M. Berman and R.H. Ebright, "Catabolite activator protein: DNA binding and transcription", *Current Opinion in Structural Biology*, vol. 14, pp. 10-20, 2004.
55. Baker, C.H., S.R. Tomlinson, A.E. Garcia and J.G. Harman, "Amino Acid Substitution at Position 99 Affects the Rate of CRP Subunit Exchange", *Biochemistry*, vol. 40(41), pp. 12329-12338, 2001.