STRUCTURE – FUNCTION RELATIONSHIPS OF TYPE II RESTRICTION ENDONUCLEASES

by

Arzu Uyar

B.S. in Chemical Engineering, Istanbul University, 2004

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

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my thesis advisor, Prof. Pemra Doruker Turgut for her enthusiastic guidance and support. Her understanding and also encouragement helped me throughout the research and writing of this thesis. It has been a great honor for me to work with her.

I am also grateful to the members of my thesis committee, Prof. Türkan Haliloglu and Assist. Prof. Demet Akten Akdoğan, for devoting their precious time to read and comment on my thesis.

I gratefully thank to Özge Kürkçüoğlu for her guidance, advices and kindly answering to my endless questions.

I would like to thank to all my former and present friends in Polymer Research Center.

I owe very special thanks to my best friend, Elif Dereli. Thank you for all things you have done for me. You always encouraged me and also helped whenever I needed. I hope our friendship will go on forever. Also I want to thank my friend, Selin Taşdöğen, because every time she called and encouraged me.

The last but not the least, my gratitude is to my parents. In my life, none of these would be possible without their endless support, understanding and patience. This thesis is dedicated to my family.

ABSTRACT

STRUCTURE – FUNCTION RELATIONSHIPS OF TYPE II RESTRICTION ENDONUCLEASES

Restriction endonucleases (REs) constitute an enzyme family, which can recognize and cleave specific foreign DNA sequences to protect prokaryotic cells. The vibrational dynamics of Type II REs are investigated by elastic network models, namely Gaussian Network Model (GNM) and Anisotropic Network Model (ANM), in order to analyze their structure-function relationships. Collective deformations of both apo REs and their specific complexes with DNA indicate scissor-like (ratchet-like) rotations and twisting-type motions in the slowest modes of dimeric and tetrameric REs. These motions can give insights about the sliding mechanism for target site location and the oscillation between closed and open states of REs, respectively. These modes are shown to be present in the non-specific complexes of EcoRV and BamHI, which further indicates their functional importance in terms of DNA recognition and sliding mechanism. Some important differences between cognate and non-cognate complexes of EcoRV and BamHI in terms of orientational cross-correlations are detected in functionally important two modes. Hot spot residues determined from high-frequency modes are clustered around the DNA binding region and at the DNA nodes, which indicates a possible communication pathway between cleavage sites.

ÖZET

TİP II RESTRİKSİYON ENDONÜKLEAZLARININ YAPI-İŞLEV İLİŞKİLERİ

Restriksiyon endonükleazları prokaryotik hücrelere karşı korumak için özgül yabancı DNA dizilimini tanıyabilen ve kesebilen enzim ailesini oluştururlar. Tip II restriksiyon endonükleazlarının titreşimsel dinamiği, yapı-işlev ilişkilerini incelemek için Gaussian Ağyapı Modeli (GNM) ve Anizotropik Ağyapı Modeli (ANM) isimli elastik ağyapı modelleri ile araştırılmıştır. Hem apo hem de DNA ile özgül komplekslerinin toplu deformasyonları, dimerik ve tetramerik restriksiyon endonükleazlarının en yavaş modlarında makas benzeri (dişli çark benzeri) dönme ve bükülme türü hareketleri göstermektedir. Bu hareketler sırasıyla hedef konum yeri için kayma mekanizması ve restriksiyon endonükleazlarının kapalı ve açık durumları arasındaki salınımları hakkında bilgiler verebilir. Bu modların EcoRV ve BamHI'in özgül olmayan komplekslerinde bulunduğu gösterilmiştir ki, ayrıca bu, DNA'yı tanıma ve kayma mekanizması bakımından bu hareketlerin işlevsel önemini göstermektedir. İşlevsel olarak önemli olan iki modda, yönelimsel karşılıklı ilişkiler bakımından EcoRV ve BamHI'in özgül ve özgül olmayan kompleksleri arasında bazı önemli farklılılar bulunmuştur. Yüksek frekanslı modellerden belirlenen hareketli rezidüler DNA'ya bağlanma bölgesinin etrafında ve DNA düğümlerinde toplanmaktadır ki bu, kesme yerleri arasındaki muhtemel bir iletişim yolunu göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

3-D	Three-dimensional
B_i	B-factor
bp	Base-pair
C(i,j)	Orientational cross-correlations
C_α	Alpha-carbon
D	Asp
E	Glu
H	Hessian matrix
h(x)	Heavyside step function
K	Lys
k	mode
k_B	Boltzman constant
ms	Mean-square
N	Total number of nodes
P	Pro
Pu	Purine
Py	Prymidine
Q-loop	Glutamine-rich loop
r_{cutt}	Cutoff distance
R_i	Position vector of residue i
R-loop	Recognition loop
T	Temperature
u_j^A	Eigenvector set of A with j^{th} mode
u_k	Eigenvector
V_{tot}	Overall potential energy
$\Delta G^{\#}_{dis}$	Activation barrier for dissociation
$\Delta G^{\#}_{slide}$	Activation barrier for sliding
ΔG^0	Increment in the energy of unbound state
ΔR	Matrix of instantaneous fluctuations of ΔR_i

ΔR_i	Displacement vector for R_i
ΔR_{ij}	Fluctuation in the distance vector
Å	Angstrom
α	Alpha helix
β	Beta strand
γ	Uniform Hookean force constant
λ_k	Eigenvalue
Γ	Kirchhoff matrix
ANM	Anisotropic Network Model
GNM	Gaussian Network Model
MD	Molecular Dynamics
PDB	Protein Data Bank
REs	Restriction endonucleases
RMSD	Root mean square deviation
TIM	Triosephosphate isomerase

1. INTRODUCTION

Proteins are large and complex molecules that are made up of amino acids linked by peptide bonds. The linear sequence of amino acids determines each protein's unique 3-dimensional (3-D) structure and its specific function. Proteins play essential roles in the function and regulation of cells.

Enzymes are specialized proteins that catalyze biochemical reactions in the cell. And restriction endonucleases (REs) constitute an enzyme family, which can recognize and cleave specific foreign DNA sequences to protect prokaryotic cells (Roberts and Halford, 1993; Raleigh and Brooks, 1998). Because they cut DNA into small fragments, REs are crucial for the recombinant DNA technology in genomic studies. With the aid of REs, recombinant DNA is now used for many applications such as cloning, recombinant vaccines, prevention and cure for diseases, production of recombinant pharmaceuticals, Southern hybridization analysis, DNA sequencing and global gene expression analysis (Roberts, 2005).

Because of the wide range of usage areas of REs it is important to understand their interactions with DNA and hence their functions. In the present computational study, the vibrational dynamics of Type II REs, which is an RE family widely used in recombinant DNA technology, are investigated in order to understand their recognition and cleavage of DNA.

Normal mode technique is a harmonic vibrational analysis of protein motions around the native folded state. Elastic network model is an efficient coarse-grained normal mode approach, where the alpha-carbons (C_α) of the protein stand for the nodes of the network. Each node is connected to all other neighboring nodes by identical harmonic springs that have a uniform force constant (Bahar *et al.*, 1997).

The Gaussian network model (GNM) is an elastic network model that is commonly used to analyze the vibrational dynamics of proteins and their complexes. Because all residue fluctuations are implicitly assumed to be isotropic in the GNM, this method is

limited to the evaluation of the mean-square displacements and cross-correlations between residue fluctuations. However, the residue fluctuations are in general anisotropic. In this respect, the anisotropic network model (ANM) has proven to be useful to include the anisotropic effects in vibrational protein dynamics. The ANM approach provides three dimensional descriptions of the 3N-6 internal modes, where N represents the total number of nodes (Atilgan *et al.*, 2001). Thus, the directionalities of the fluctuations can be determined because ANM calculates the x, y and z components of the fluctuations. The collective modes of motion leading to global conformational changes and the critical sites controlling these modes can be identified by the ANM (Atilgan *et al.*, 2001). The low-frequency modes extracted from ANM give information about the mechanism of the cooperative motions of the native protein.

The low-frequency modes make the major contributions to thermal conformational fluctuations. And higher frequency fluctuations that are more localized involve only a few residues and identify kinetically hot regions that are important for stability (Bahar *et al.*, 1998; Bahar and Jernigan, 1998). The deformation from the native state coordinates in the slow modes obtained by elastic network models can be used to determine the dynamic domains in the structure. Domain motions, such as hinge bending or closure, are known to play an important role in the function of many proteins (Hayward, 1999).

In the present study, vibrational dynamics of Type II REs are investigated by GNM and ANM approaches in order to understand their recognition and cleavage of cognate DNA. Six REs are used, four of them from EcoRI family and two of them from EcoRV family. Collective deformations are analyzed to have better understanding of similarities and differences among the REs. Moreover, non-cognate DNA complexes of EcoRV and BamHI are studied in order to differentiate their dynamics from their cognate complexes with DNA. Finally, ANM and GNM results of REs are compared with those of another homo-dimeric enzyme triosephosphate isomerase (TIM).

2. RESTRICTION ENDONUCLEASES

Restriction endonucleases (REs) are elements of restriction modification (RM) systems that take place ubiquitously among prokaryotic organisms (Roberts and Halford, 1993). A restriction endonuclease produced by the bacteria is an enzyme that cuts double-stranded foreign DNA only at a particular sequence of nucleotides called the "recognition sequence", to protect itself against the attack of viruses. Each bacterium protects its own DNA at its specific recognition sequence by adding a bulky methyl group to the bases of N4 or C5 at cytosine or N6 at adenine. However, invading viral DNA does not have the protective methyl group, therefore is cleaved wherever RE encounters the recognition sequence (Pingoud *et al.*, 2005).

REs have been named according to their origin (Smith and Nathans, 1973). The first letter of the name indicates the genus while the second two letters indicate the species of the prokaryotic cell from which they were isolated. As an example, for EcoRI, which is a well known RE, "Eco" comes from Escherichia (genus) and coli (species) while "R" comes from RY13 (strain) and "I" comes from the order of discovery from Escherichia coli. Similarly, BamHI is isolated from Bacillus amyloliquefaciens strain H and order of discovery I.

The recognition sequences and cleavage sites of REs are different from each other. For example, EcoRI recognizes the sequence of GAATTC and cleaves between G and A while EcoRV recognizes GATATC and cleaves between T and A. In addition, the lengths of recognition sites are different. The enzymes EcoRI, SacI and SstI each recognize a 6 base-pair (bp) sequence of DNA, whereas NotI recognizes a sequence of 8 bp in length, and the recognition site for Sau3AI is only 4 bp in length. Length of the recognition sequence points to how frequently the enzyme will cut in a random sequence of DNA. Enzymes with a 6 bp recognition site will cut, on average, every 4^6 or 4096 bp; a 4 bp recognition site will occur roughly every 256 bp (Welsh *et al.*, 2004).

REs having the same recognition site are called "isoschizomers" as for EcoRI and RsrI having the identical recognition sites (G/AATTC, where / denotes the cleavage site)

(Stephenson *et al.*, 1989). On the other hand, isoschizomers recognizing the same sequence but cleaving from different positions are called “neoschizomers”. SmaI and XmaI recognize the sequence (CCCGGG), but SmaI cleaves from C and G (CCC/GGG) while XmaI from C and C (C/CCGGG). And, in some cases, dissimilar structures may have equivalent functions in terms of cleavage site and recognition sequence (Bujnicki, 2004).

Some of REs can recognize indefinite recognition sequences while some of them only recognize a definite sequence. For example, the enzyme BamHI has definite recognition sequence GGATCC. In contrast, HinfI and XhoII have indefinite recognition sites. HinfI recognizes sequence starting with GA ending in TC with any base between these nucleotides. Similarly, XhoII recognizes and cleaves the sequences of AGATCT, AGATCC, GGATCT and GGATCC (Rena and Houslay, 1998). In addition, the recognition site for one enzyme may contain the recognition site for another as in the case of BamHI containing the recognition site for Sau3AI (Friedhoff *et al.*, 2001).

Generally REs recognize palindromic sequences (Roberts *et al.*, 2003). They read DNA the same forward and backward sites, as in the case of NgoMIV that cleaves between nucleotides G and C in 5'G/CCGGC and in 3'CGGCC/G. Also, there are some REs which have non-palindromic sequence such as AccBSI can recognize and cleave in the middle of the 5'GAG/CGG-3'CTC/GCC sequence (Abdurashitov *et al.*, 1997). In Figure 2.1.(a), examples for palindromic and non-palindromic sequences are given.

When the REs cleave their recognition sites, they generate one of three different types of ends, which are “5' or 3' overhangs, called sticky ends” and “blunt ends”. In Figure 2.1.(b), representation for sticky and blunt ends is given. Sticky ends are sometimes called “cohesive ends”, because they will readily join with their partner by base pairing (Pingoud *et al.*, 2005). Blunt ends are without overhangs and generated by the enzymes that cut at precisely opposite sites in the two strands of DNA.

REs are divided into four families, namely Type I, Type II, Type III, and Type IV families, according to subunit composition and cofactor requirement (Roberts *et al.*, 2003). Type I, Type III and Type IV families are composed of multiple different subunits separately responsible for DNA recognition and cleavage. These families require ATP

(Type I and Type III) or GTP (Type IV) during DNA cleavage with Mg^{2+} cofactor, and S-adenosylmethionine as the donor of the methyl group for activity (Type I and Type III) (Dryden *et al.*, 2001; Roberts *et al.*, 2003; Roberts *et al.*, 2004; Pingoud *et al.*, 2005; Mücke *et al.*, 2001; Janscak *et al.*, 2001; McCleland and Sczcelkun, 2004). Type II restriction endonucleases differ from the other families of REs by having simple subunit organization. They are usually homodimeric or homotetrameric enzymes that cleave DNA within or close to their recognition site and do not require ATP or GTP but only Mg^{2+} as cofactor (Pingoud *et al.*, 2005).

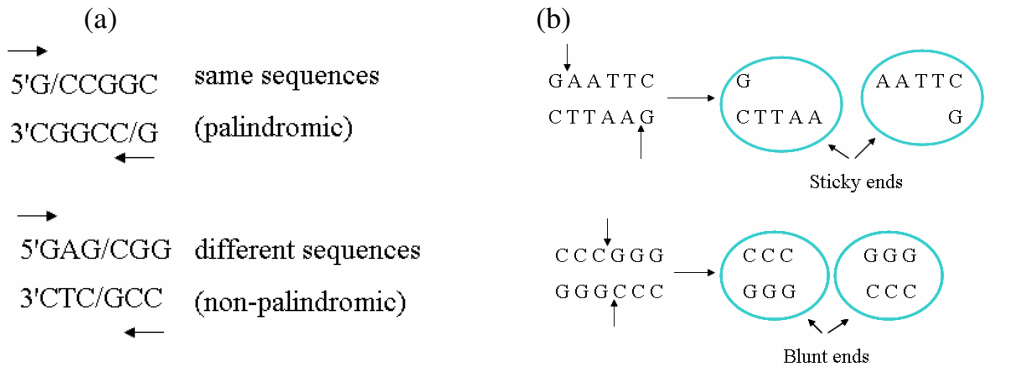


Figure 2.1. (a) Palindromic and non-palindromic sequences, and (b) sticky and blunt ends

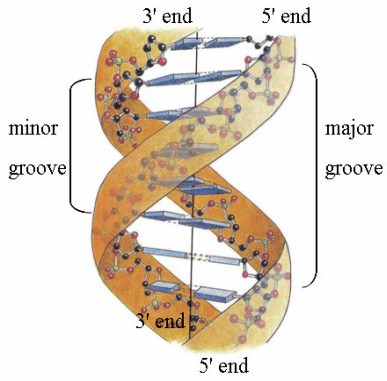


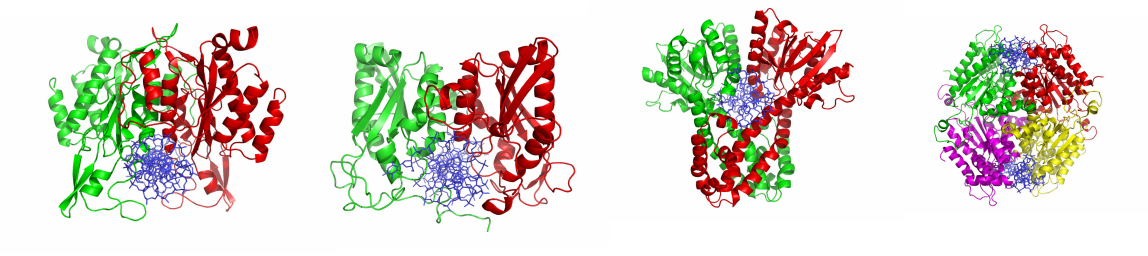
Figure 2.2. Minor and major grooves of DNA

More than 3500 Type II enzymes have been identified (Roberts *et al.*, 2007), over 50 have been sequenced and biochemically characterized, and they are classified according to their mechanism of recognizing and cleaving DNA (for a comprehensive review, see Pingoud *et al.*, 2005). Since 1990, REs have been crystallographically analyzed and crystal structure information for 16 Type II REs is available from Protein Data Bank (PDB)

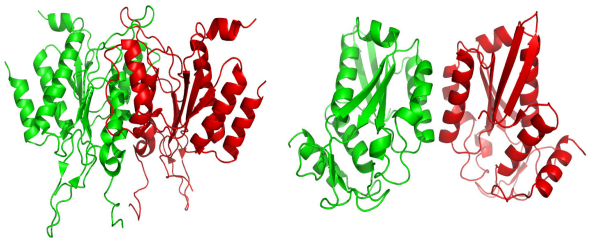
(Berman *et al.*, 2000). EcoRI is the first crystallographically analyzed enzyme (McClarin *et al.*, 1986), which has been revised and published in 1990 (Kim *et al.*, 1990). Three years later, EcoRV structure was published (Winkler *et al.*, 1993). When the crystal structures of REs were compared, it is revealed that they share a common structural core composed of five-stranded mixed beta-sheet flanked by two alpha-helices (Venclovas *et al.*, 1994). This core which serves as an important structure stabilization center (Fuxreiter and Simon, 2002) involves two or three acidic residues (Asp or Glu) and one Lys residue as catalytic residues (Aggarwal, 1995). In some of REs, the catalytic residue Lys is replaced by Glu or Gln (Newman *et al.*, 1994; Lukacs *et al.*, 2000).

REs are originated from at least four different folds. One of the most common sequence pattern of REs is (PD-x_{10-30}(D/E)xK) (Bujnicki, 2004), which is found in Type II REs, where P: Pro; D: Asp; E: Glu; K: Lys, and x denotes any amino acid. Type II REs belonging to the (PD…(D/E)xK) superfamily can be divided into EcoRI and EcoRV families in terms of secondary structure elements' topology and the subunit arrangement (Huai *et al.*, 2000; Anderson, 1993). In EcoRI family, the fifth beta-strand is parallel, while in EcoRV family it is antiparallel to the fourth strand (Huai *et al.*, 2000). EcoRI and EcoRV family enzymes approach DNA from different direction. EcoRI family enzymes (i.e., BamHI, BglII, BsoBI, EcoRI, EcoRII, FokI, NgoMIV) usually approach the DNA from the major groove, whereas enzymes of the EcoRV family enzymes (i.e., EcoRV, HincII, NaeI, PvuII) from the minor groove (Anderson, 1993). In Figure 2.2 (adapted from Branden and Tooze, 1999), major and minor grooves of DNA are given. Enyzmes of EcoRI family recognize the DNA mostly using an alpha-helix and a loop, and in most cases they produce 5´ overhanged ends. Enzymes of EcoRV family recognize DNA by means of a beta-strand and a beta-like turn and usually they produce blunt or 3´ overhanged ends (Pingoud *et al.*, 2005). Figure 2.3 shows four X-ray structures of free restriction endonucleases and six X-ray structures of enzyme-cognate DNA complexes of EcoRI and EcoRV families, which are the material used in this study. All figures are drawn by PyMol v0.99 program (DeLano, 2002). Identical chains are shown in different colors with ribbon representation and DNA is shown atomistically. NgoMIV is a tetrameric enzyme and the rest are homo-dimers.

EcoRI Family

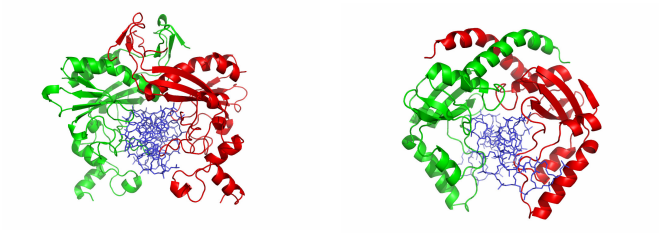


EcoRI-DNA complex BamHI-DNA complex BsoBI-DNA complex NgoMIV-DNA complex

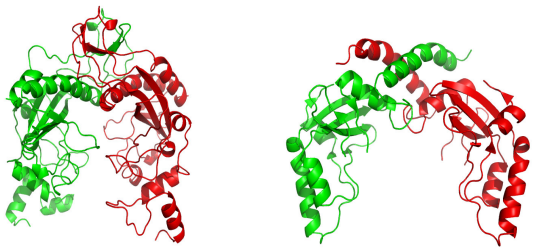


EcoRI-apo form BamHI-apo form

EcoRV Family



EcoRV-DNA complex PvuII-DNA complex



EcoRV-apo form PvuII-apo form

Figure 2.3. Crystal structures of Type II family REs

EcoRI, member of Type IIP family REs (P for palindromic), is a homodimeric enzyme with 550 residues in total. Its high-resolution crystal structure at 2.5 Å can be found with the PDB code 1ERI, where the coordinates for the first 16 residues in each monomer have not been reported (Kim *et al.*, 1990). Catalytic residues of EcoRI located in each monomer are Asp91, Glu111, and Lys113 and its 3-D structure is shown in Figure 2.4. Its recognition sequence with cleavage sites is (G/AATTC-CTTAA/G).

Structure analyses show that DNA is kinked by EcoRI in the middle of the recognition sequence (Kim *et al.*, 1990; Kim *et al.*, 1994) shown in the side view (Figure 2.4.(b)). DNA contacts involve several non-adjacent structural components: (a) a bundle of two helices from each subunit which enter the widened major groove and carry residues involved in base and backbone interactions at their amino termini; (b) an extended chain which passes through the major groove of the recognition site; (c) a β-strand that is parallel to the DNA backbone, which contains active site residues crucial for catalysis as well as residues involved in phosphate contacts; (d) two arms (inner and outer loops as shown in Figure 2.4.(a)) that reach around the DNA and are responsible for backbone contacts outside of the recognition sequence (McClarin *et al.*, 1986; Kim *et al.*, 1990). There are 18 protein-base hydrogen bonds and 10 van der Waals contacts to all pyrimidines (Pingoud and Jeltsch, 1997).

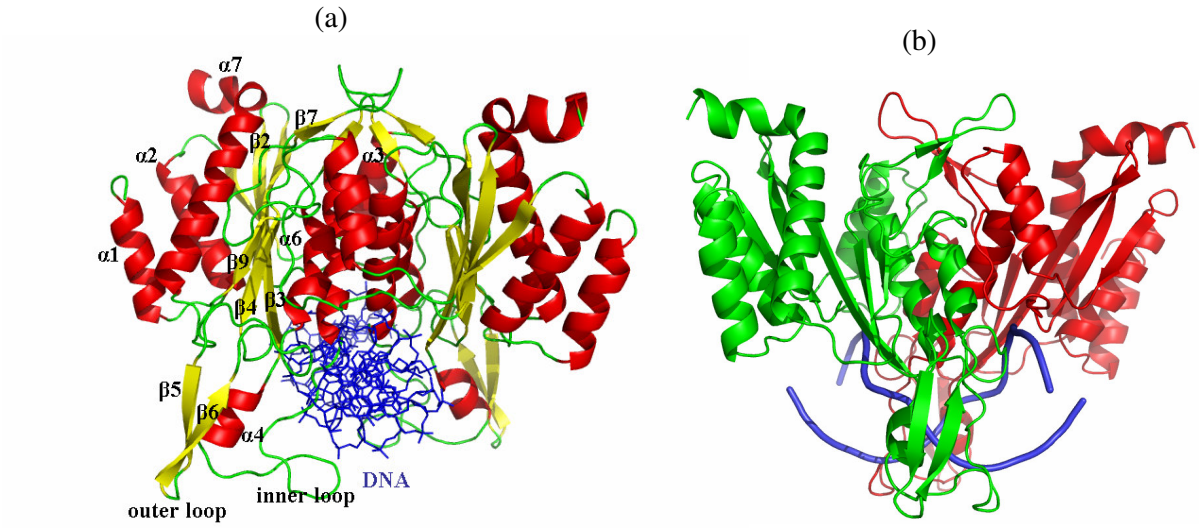


Figure 2.4. (a) EcoRI restriction enzyme complex with cognate DNA, and (b) Side view of EcoRI showing kinked DNA

BamHI, member of Type IIP family REs, is a homodimeric enzyme with each monomer having 213 residues. Its PDB code is 2BAM at 2.0 Å resolution (Viadiu and Aggarwal, 1998) with missing chain end residues (208-213) in monomer A and (211-213) in momoner B. Catalytic residues in each monomer of BamHI with a recognition sequence (G/GATCC-CCTAG/G) are Asp94, Glu111, and Glu113 and its 3-D structures in complex with cognate and a non-cognate DNA (1ESG.pdb) (Viadiu and Aggarwal, 2000) are given in Figure 2.5.(a) and (b), respectively.

Significant DNA distortion is not observed for this enzyme which approaches DNA from the major groove. Here, the amino-termini of two α-helices from each subunit and adjacent loops compose the major groove contacts. On the other hand, minor groove contacts involve the C-terminal α-helix from only one subunit, which in the specific complex is unwound and adopts an extended conformation (indicated as structured loop on Figure 2.5). The C-terminal α-helix of the other subunit does not enter to the minor groove and folds back to follow the sugar-phosphate backbone. One subunit of BamHI forms the majority of its phosphate contacts to one DNA half site (GGA), but nearly all of its base contacts to the other DNA half site (TCC). The C-terminal arm of only one subunit is asymmetrically related to both DNA half-sites (Pingoud and Jeltsch, 1997).

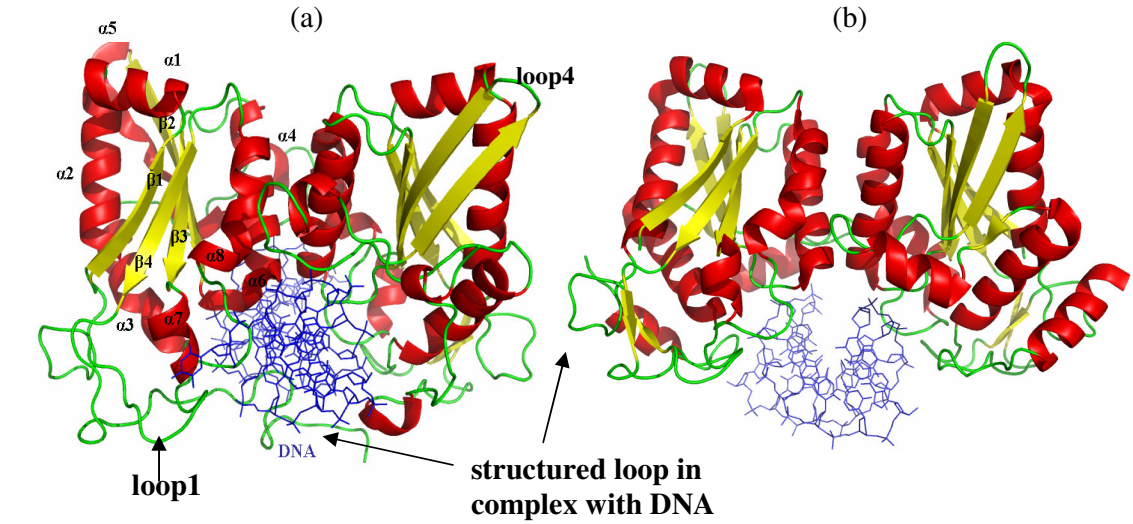


Figure 2.5. BamHI restriction enzyme complex with (a) cognate DNA, and (b) non-cognate DNA

BsoBI is a homodimeric enzyme of 646 residues in the Type IIP family REs. In the crystal structure (PDB code: 1DC1.pdb) at 1.70 Å resolution (van der Woerd *et al.*, 2001) the coordinates for the residues (1-4, 221-229) in monomer A and (1-3, 221-229) in monomer B have not been reported (Figure 2.6.(a)). For better visualization of each monomer, side view of BsoBI-DNA complex is given in Figure 2.6.(b). Catalytic residues of BsoBI in each monomer are Asp212, Glu240, and Lys242, with recognition sequence (C/PyCGPuG-GPuGCPy/C) where Py denotes prymidine (Cytosine, Thymine or Uracil) and Pu denotes purine (Adenine or Guanine).

BsoBI approaches the DNA from major groove and holds DNA by forming a 20 Å long tunnel. Each subunit is in interaction with each recognition half site, and base-specific contacts are made in the major and minor grooves. Several hydrogen bonds can be observed between protein and the outer and inner CG base pairs. PyPu base pair is involved in only one direct hydrogen bond between Lys81 and the N7 of the purine, and in one water-mediated bidentate hydrogen bond between Asp246 and N7 purine as well as the substituent in position 6 of the purine (Pingoud and Jeltsch, 2001).

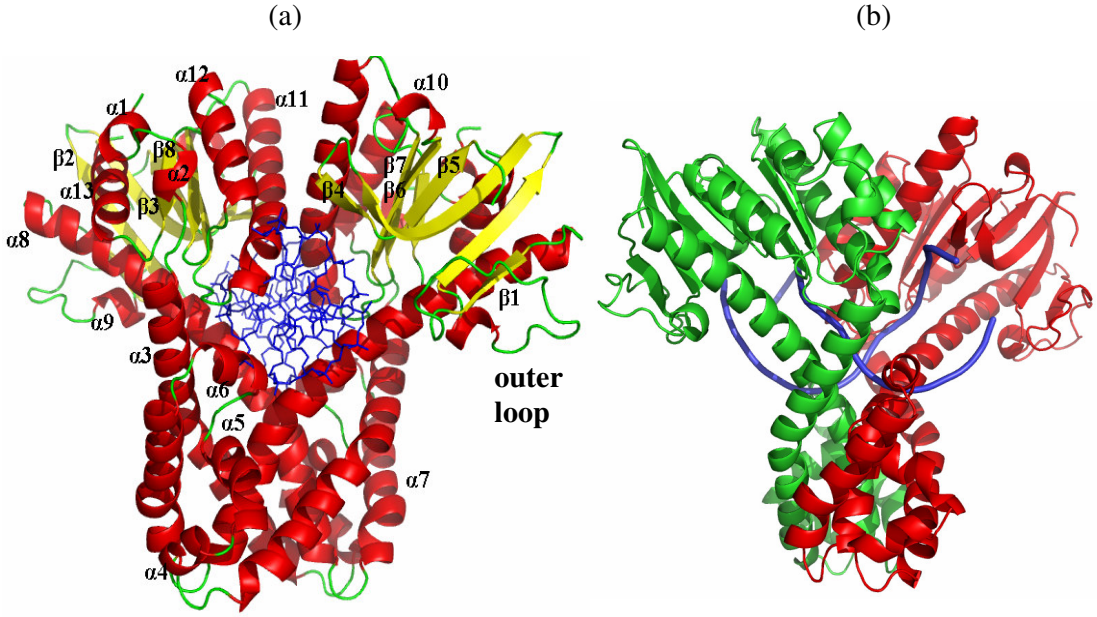


Figure 2.6. (a) BsoBI restriction enzyme complex with cognate DNA, and (b) side view of BsoBI-DNA complex

NgoMIV, a member of Type IIF family REs, is a homo-tetrameric enzyme with each monomer having 286 residues. Its catalytic residues in each monomer are Asp140, Glu201, and Lys187 and its recognition sequence is (G/CCGGC-CGGCC/G). For NgoMIV, only crystal structure of the enzyme-product complex is available, which can be found with the PDB code 1FIU at 1.60 Å resolution (Deibert *et al.*, 2000) (Figure 2.7). The tetrameric structure of NgoMIV is formed by side-by-side contacts (between monomers 1/4 and 2/3) and cross contacts (between monomers 1/3 and 2/4) as shown in Figure 2.7.(b)) (Deibert *et al.*, 2000).

NgoMIV approaches DNA from the major groove where it makes most of the base-specific contacts with the target sequence. One subunit makes hydrogen bonds to the GCC half site in the major groove, whereas other subunit makes a hydrogen bond to the C of the outer GC base pair in the miner groove. Three structural elements, loops preceding three alpha-helices, forms base-specific contacts. Three adjacent amino acids (Arg191, Asp193, and Arg194) make all possible hydrogen bonds to the two adjacent GC base pairs in the major groove. Altogether there are 20 hydrogen bonds to the bases of recognition sequence (Pingoud and Jeltsch, 2001).

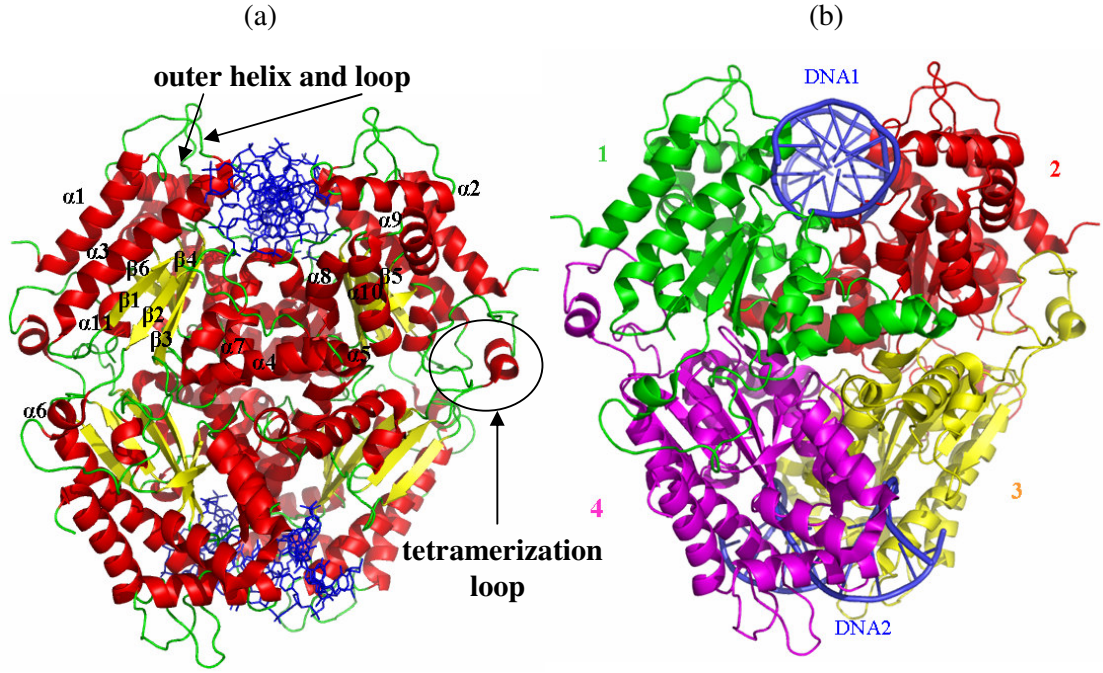


Figure 2.7. (a) NgoMIV restriction enzyme complex with cognate DNA, and (b) monomers of NgoMIV-DNA complex

EcoRV, which is a homodimeric enzyme with each monomer of 244 residues, is a member of Type IIP family REs. Its PDB code is 1BGB at 2.0 Å resolution (Horton and Perona, 1998). Coordinates for the residues (141-147) in monomer A and (11-17, 97-100, 141-145) in monomer B have not been reported. Catalytic residues of EcoRV in each monomer are Asp74, Asp90, and Lys92 and its 3-dimensional structure is shown for cognate and non-cognate (2B0E.pdb) (Hiller *et al.*, 2005) complexes in Figure 2.8. Its recognition sequence is (GAT/ATC-CTA/TAG).

EcoRV is the only RE for which structure analyses exist for the free enzyme (i.e. apo form), a non-specific and two specific complexes, as well as of an enzyme-product complex. When the non-specific and specific complex structures are compared, the most obvious change is the distortion of DNA by a 55° central kink which unwinds DNA. Two recognition elements form the interactions of protein with DNA. The primary recognition element is recognition loop (R-loop) (Figure 2.8), responsible for hydrogen bonding and van der Waals contacts. The other recognition element glutamine-rich loop (Q-loop) (Figure 2.8) which involves catalytic residue Asp74, is responsible for hydrogen bonds in the minor groove. There are 18 hydrogen bonds and two van der Waals contacts. (Pingoud and Jeltsch, 1997).

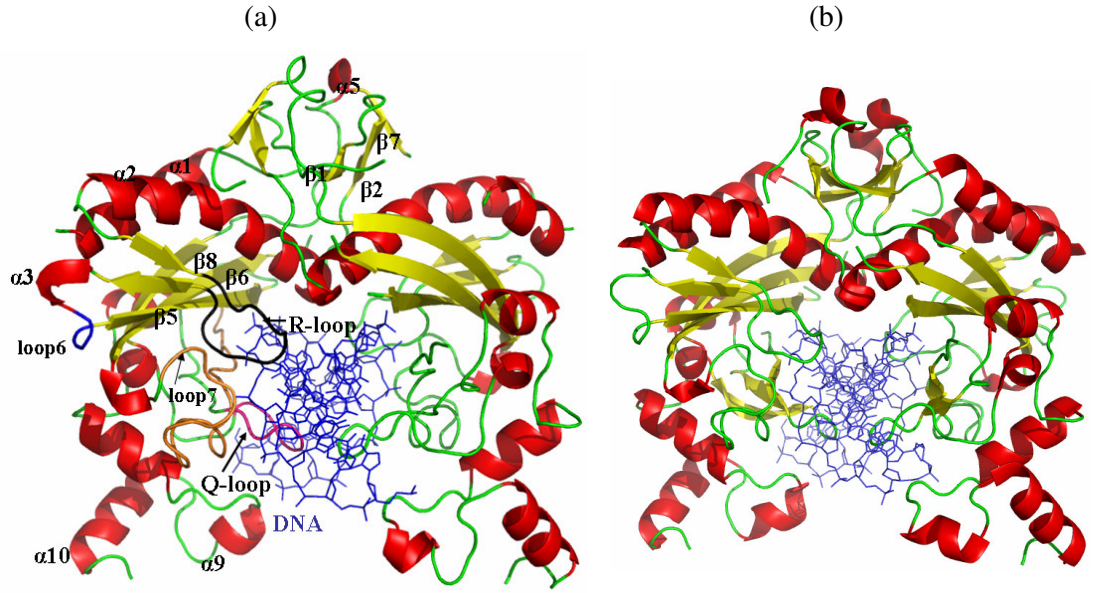


Figure 2.8. EcoRV restriction enzyme complex with (a) cognate DNA, and (b) non-cognate DNA

PvuII is a member of Type IIP family REs, and a homodimeric enzyme with each monomer of 157 residues. Its crystal structure has PDB code 1EYU at 1.78 Å resolution (Horton and Cheng, 2000) with missing coordinates of the residues (1) in monomer A and (1, 53-55) in monomer B. Catalytic residues of PvuII in each monomer are Asp58, Glu68, and Lys70 and its recognition sequence is (CAG/CTG-GTC/GAC) (Figure 2.9). It generates blunt ends.

PvuII is similar to EcoRV in terms of structure (Athanasiadis *et al.*, 1994; Cheng *et al.*, 1994) but it does not distort DNA while EcoRV does (Cheng *et al.*, 1994). As in the EcoRV-DNA complex, the DNA lies in a gap formed by the two identical subunits. The DNA binding regions of each subunit includes two substructures from different locations in the primary sequence: binding and catalytic regions are mutually combined. A two-stranded antiparallel beta-sheet is essential in recognition. Also it makes interactions with bases and phosphates outside of the recognition site. There are 12 hydrogen bonds and several van der Waals contacts (Pingoud and Jeltsch, 1997).

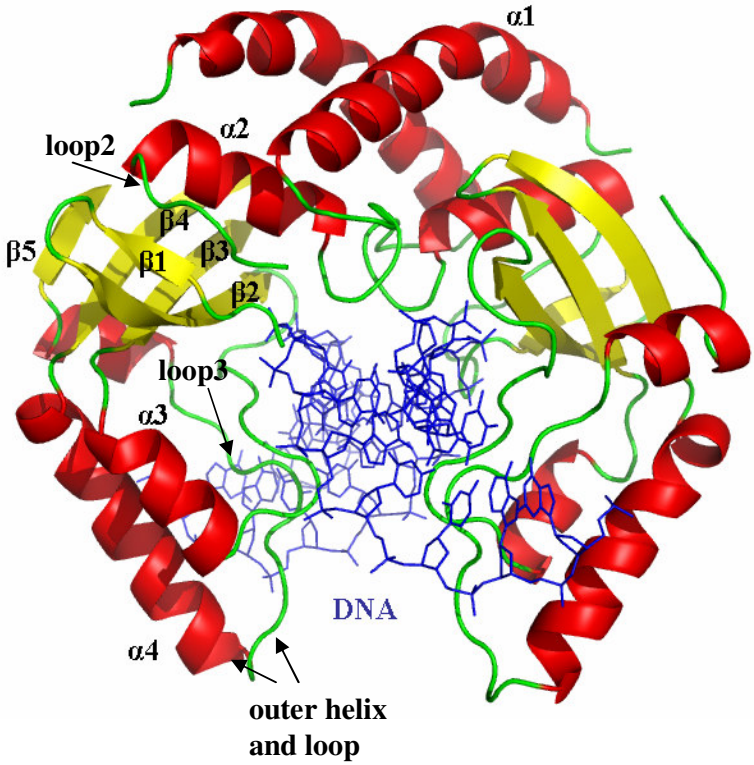


Figure 2.9. PvuII restriction enzyme complex with cognate DNA

3. METHODS

3.1. Gaussian Network Model

Dynamics characteristics of biomolecular structures can be modeled as those of elastic network models. The Gaussian Network Model (GNM) proposed by Bahar and coworkers (Bahar *et al.*, 1997) is one of the simpler models introduced at residue level which can elucidate the dynamics of proteins and their complexes. In this model, residues are assumed to undergo Gaussian-distributed fluctuations about their mean positions defined by their native conformation. Neighboring nodes that are located within a cutoff distance (r_{cut}) are linked by elastic springs with a uniform harmonic force constant to form a perfectly elastic network.

The overall potential energy for this network can be calculated by summing up the harmonic potential between neighboring residue pairs.

$$V_{tot} = \sum_i \sum_j V(\mathbf{R}_i, \mathbf{R}_j) = \sum_i \sum_j \frac{1}{2}\gamma(\Delta\mathbf{R}_{ij} \cdot \Delta\mathbf{R}_{ij}) \tag{3.1}$$

Here γ is the uniform Hookean force constant between the interacting residues, $\mathbf{R}_i$ is the position vector of i^{th} residue, $\Delta\mathbf{R}_{ij}$ is the fluctuation in the distance vector $\mathbf{R}_{ij}=\mathbf{R}_j-\mathbf{R}_i$ from the equilibrium position. The equation for the total potential can also be expressed as

$$V_{tot} = (\frac{\gamma}{2})tr[\Delta\mathbf{R}^{\mathbf{T}}\mathbf{\Gamma}\Delta\mathbf{R}] \tag{3.2}$$

where $\mathbf{\Gamma}$ is the Kirchhoff matrix of contacts, for which the ij^{th} off-diagonal element is -1 if i^{th} and j^{th} residues are in contact, and 0 otherwise, and $\Delta\mathbf{R}$ is the matrix of instantaneous fluctuations of $\Delta\mathbf{R}_i$.

The diagonal elements of the inverse Kirchhoff matrix Γ^{-1} gives the mean-square fluctuations of the nodes explicitly calculated from eigenvalues and eigenvectors that represent the frequencies and the shapes of individual modes, respectively. The number of

eigenvalues is equal to the number of residues N. Because of the degrees of freedom of the system is one, the smallest eigenvalue will be zero, therefore N-1 eigenvectors and eigenvalues will be available for the system.

The cross correlation of fluctuations of the i^{th} and j^{th} residues is calculated as:

$$< \Delta \mathbf{R}_i \Delta \mathbf{R}_j >= (3k_B T / \gamma)[\mathbf{\Gamma}^{-1}] = \sum_{k=1}^{N} (3k_B T / \gamma)[\lambda_k{}^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \qquad (3.3)$$

where λ_k and $\mathbf{u}_k$ are the eigenvalue and eigenvector set associated with the k^{th} mode of the Kirchhoff matrix **Γ,** and N is the total number of modes of the structure. The last term in the equations explains how the harmonic fluctuations around the equilibrium positions can be decomposed into series of modes from slowest to fastest for a structure with known C-alpha coordinates. The fluctuations indicated in the slowest modes (with low eigenvalues) usually describe the global motion, whereas the fast modes (with high eigenvalues) describe the high frequency local fluctuations. Moreover, it was found that the residues contributing the high frequency fluctuations contributed to the stability and function of the molecule (Bahar *et al.*, 1998).

The X-ray crystallographic Debye-Waller factors (Frauenfelder *et al.*, 1979) called B-factors are related to mean-square fluctuations of each node and can be calculated as:

$$B_i = \frac{8.\pi^2 < \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i >}{3} \qquad (3.4)$$

where B_i denotes the temperature factor or B-factor of a residue.

3.2. Anisotropic Network Model

An extension of the GNM for elucidating the fluctuation dynamics of proteins is Anisotropic Network Model (ANM) (Atilgan *et al.*, 2001) which is a simplified version of normal mode analysis (NMA) (Brooks *et al.*, 1988). In this model, anisotropy of fluctuations is taken into account with 3-D character of motion, whereas they are assumed

to be isotropic in GNM. In reality the fluctuations are anisotropic in general, and it is important to assess the directions of collective motions, as these can be directly relevant to biological function and mechanisms. Similar to GNM, nodes of the elastic network correspond to the C-alphas of residues and close neighboring nodes are linked by harmonic springs having a uniform force constant γ. The total potential energy of the system of N residues is expressed as

$$V = (\frac{\gamma}{2})\Delta\mathbf{R}^{\mathbf{T}}\mathbf{H}\Delta\mathbf{R} \tag{3.5}$$

where $\Delta\mathbf{R}$ is a 3N-dimensional vector of the fluctuations $\Delta\mathbf{R}_i$ in the position vectors $\mathbf{R}_i$ of the individual sites ($1 \leq i \leq N$), $\Delta\mathbf{R}^T$ is its transpose, and $\mathbf{H}$ is the 3N x 3N Hessian matrix composed of the second derivatives of the potential for over all interactions sites

$$V_{tot} = \frac{\gamma}{2}\sum_i\sum_j h(r_{cut} - R_{ij})(\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \tag{3.6}$$

where $h(x)$ is the Heavyside step function ($h(x) = 1$ if $x \geq 0$, and zero otherwise), R_{ij} is the distance between the i^{th} and j^{th} centers, and r_{cut} is the cutoff distance defining the range of direct interactions. $\mathbf{H}$ is composed of N x N super-elements of size 3 x 3 of H_{ij}

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & & & H_{2N} \\ \vdots & & & \vdots \\ H_{N1} & & & H_{NN} \end{bmatrix} \tag{3.7}$$

The ij^{th} super-element for $i \neq j$ H_{ij} of $\mathbf{H}$ is

$$H_{ij} = \begin{bmatrix} \partial^2 V/\partial X_i \partial X_j & \partial^2 V/\partial X_i \partial Y_j & \partial^2 V/\partial X_i \partial Z_j \\ \partial^2 V/\partial Y_i \partial X_j & \partial^2 V/\partial Y_i \partial Y_j & \partial^2 V/\partial Y_i \partial Z_j \\ \partial^2 V/\partial Z_i \partial X_j & \partial^2 V/\partial Z_i \partial Y_j & \partial^2 V/\partial Z_i \partial Z_j \end{bmatrix} \tag{3.8}$$

Here X_i, Y_i, and Z_i are the components of the distance vector $\mathbf{R}_i$. The elements of **H** are calculated by using coordinates of the C-alphas of the structure taken from PDB. The counterpart of the Kirchhoff matrix **Γ** of the GNM is similar to $(1/\gamma)\mathbf{H}$ in the ANM. Singular value decomposition of **H** yields 3N-6 non-zero eigenvalues, and 3N-6 eigenvectors that represent the respective frequencies and shapes of individual modes, respectively. Six eigenvalues are zero because of the degrees of freedom for the set of the equations are six.

The inverse of **H** is composed of N x N super-elements, each of which scales with the 3 x 3 matrix of correlations between the components of pairs of fluctuation vectors. The cross correlations between residue fluctuations are found from

$$< \Delta\mathbf{R}_i \Delta\mathbf{R}_j >= (k_B T/\gamma) tr[\mathbf{H}^{-1}]_{ij} = \sum_{k=1}^{N} (k_B T/\gamma)[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \qquad (3.9)$$

where λ_k is the k^{th} nonzero eigenvalue and $\mathbf{u}_k$ is the corresponding eigenvector. B-factors can be calculated from the mean-square fluctuations of each node by Equation 3.4.

In the GNM and ANM, each amino acid is represented by a single node consisting of its C-alpha atom, and each nucleotide by three nodes at P (phosphate group), C4* (sugar), and C2 (base) coordinates (Delarue and Sanejouand, 2002) (see Figure 3.1. atoms denoted by red circles). For coarse-grained calculations, r_{cut} value is taken as 13 and 17 Å in ANM, whereas 7 and 10 Å in GNM. In fully atomistic ANM calculations, r_{cut} value is taken as 9 Å.

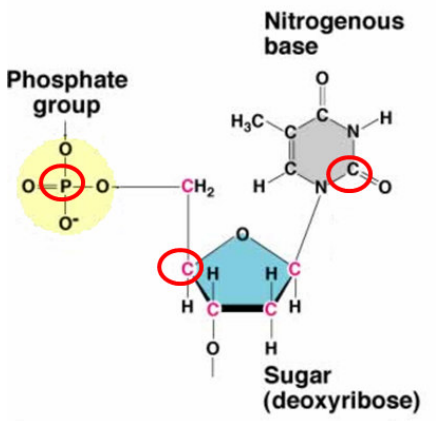


Figure 3.1. Nodes in a nucleotide for construction of elastic network model

4. RESULTS AND DISCUSSION

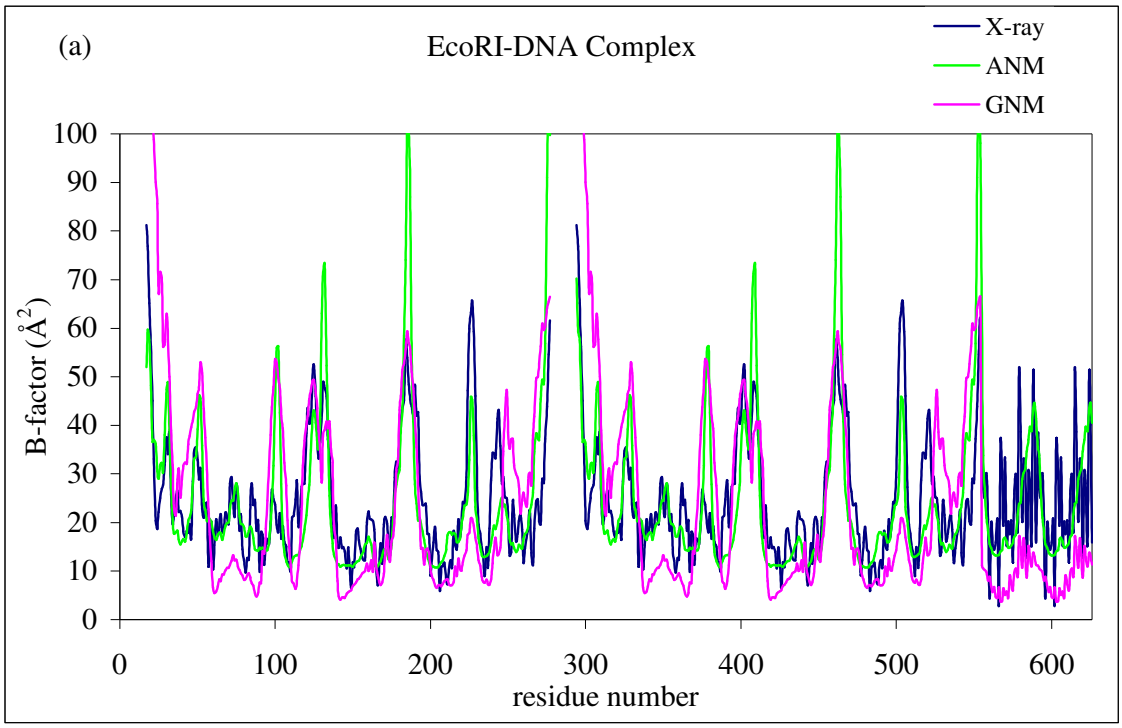
4.1. Equilibrium Fluctuations

In GNM and ANM, the connectivity matrices are constructed using the coarse-grained native structure represented by C-alpha (amino acids), and P, C4* and C2 atoms (nucleotides) (see Figure 3.1). The cutoff distance, which defines the range of coarse-grained interactions, is taken as r_{cut} = 7 Å for GNM, and 13 Å for ANM calculations. After Kirchhoff and Hessian matrices are diagonalized in GNM and ANM calculations, respectively, the predicted B-factor of residues can be found by adding the mean-square (ms) fluctuations over all modes and using Equation 3.4 (Bahar *et al.*, 1997; Atilgan *et al.*, 2001). The theoretical B-factors are compared with the experimental values from the X-ray structures in order to see the validity of each elastic network model for the REs.

In Figures 4.1 to 4.6, X-ray temperature factors and theoretical results from both ANM and GNM are displayed for the complex, DNA-removed (if apo form is not available) and apo forms of EcoRI, BamHI, BsoBI, NgoMIV, EcoRV and PvuII, respectively. The missing residues in each monomer are indicated with gaps on the plots. There is generally good agreement between the theoretical values of each elastic network model and the X-ray results. However, for BsoBI and NgoMIV the theoretical B-factor results obtained from GNM are not as satisfactory as the other cases. In ANM, certain peaks corresponding to either chain ends or solvent-exposed loops exhibit very high mobility in the absence of crystal contacts. Still these consistent results in general justify further analysis of these systems.

4.2. Global Dynamics

The low-frequency (slow) modes give insights about the cooperative conformational motions of proteins around their native state. These motions are relevant to biological functions of many enzymes and proteins (Bahar *et al.*, 1998).



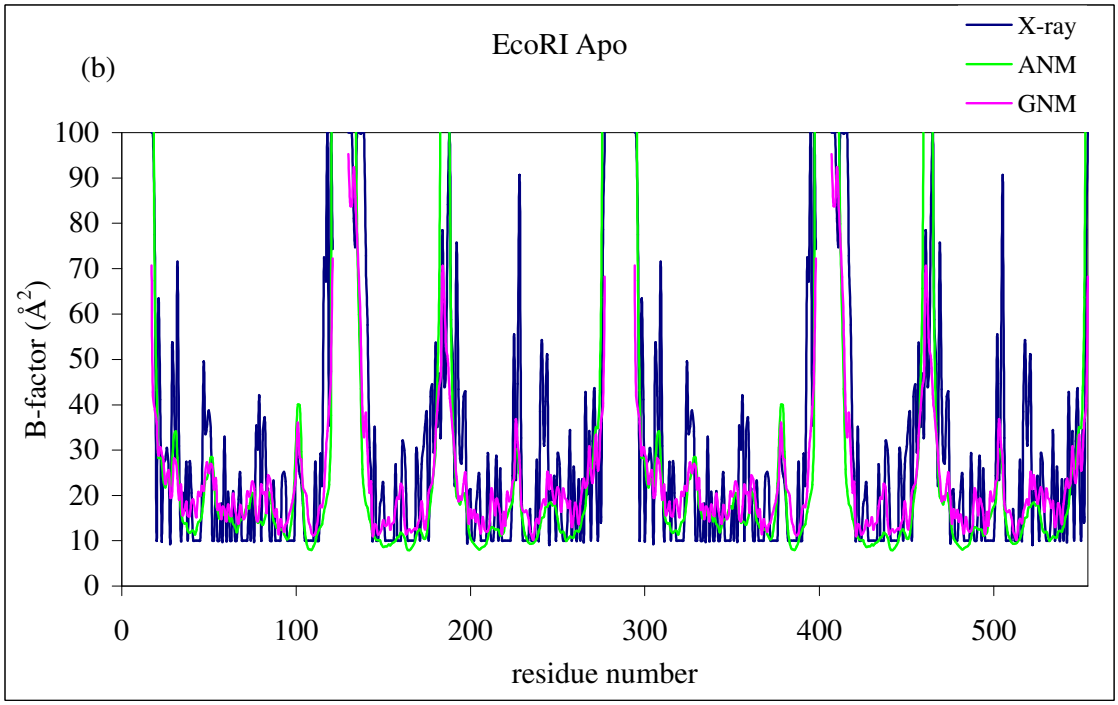
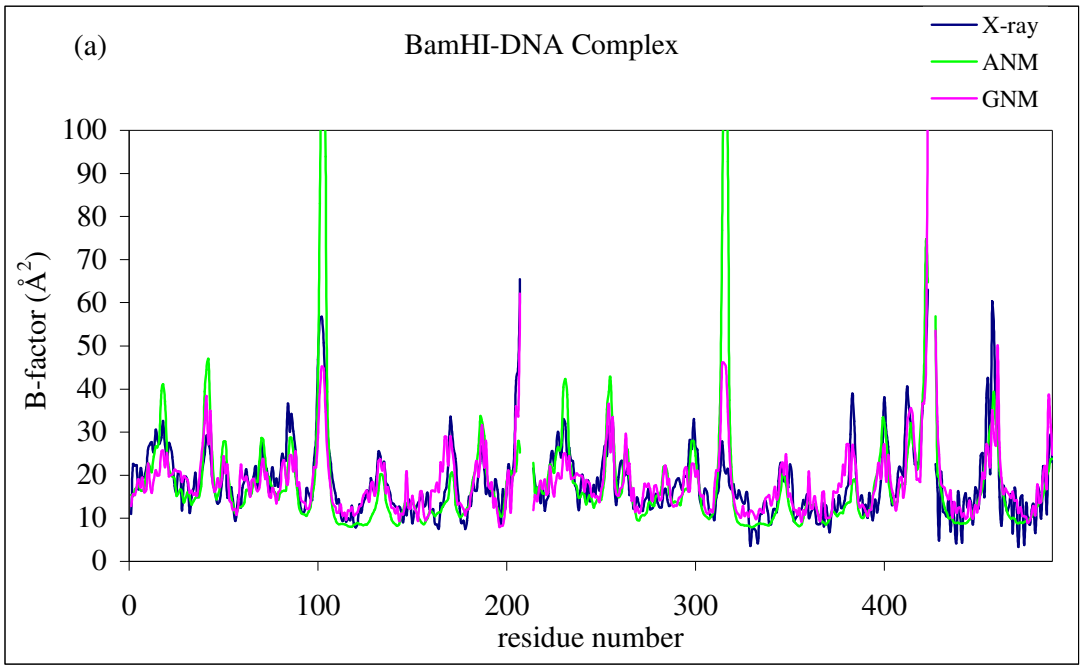


Figure 4.1. B-factors of EcoRI comparing X-ray, ANM and GNM results for (a) complex (b) apo forms



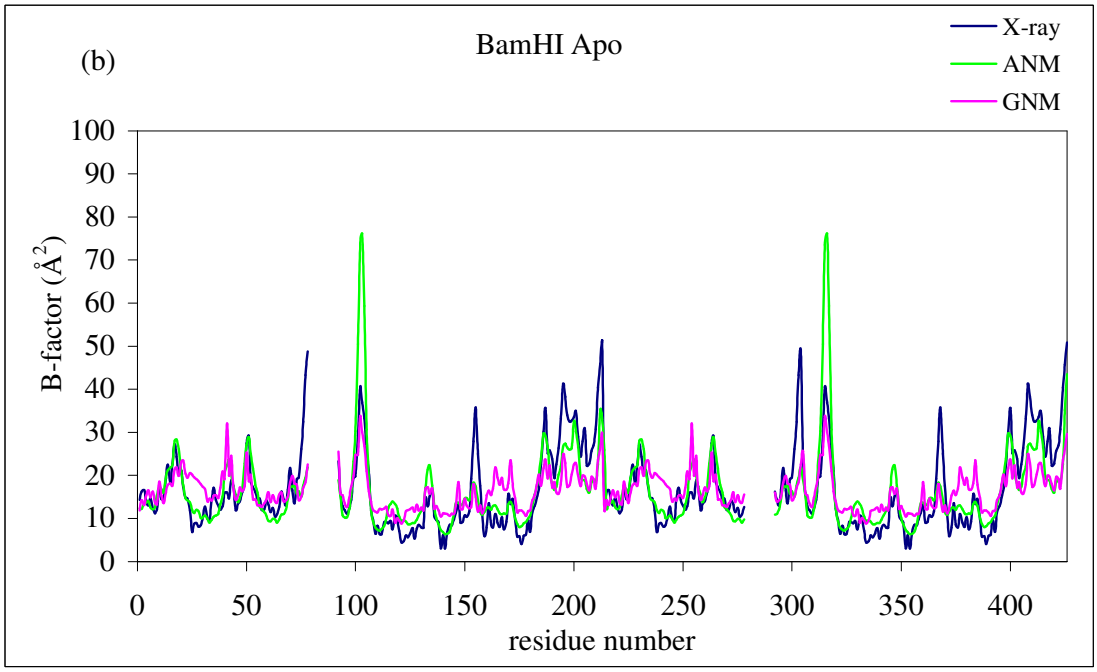
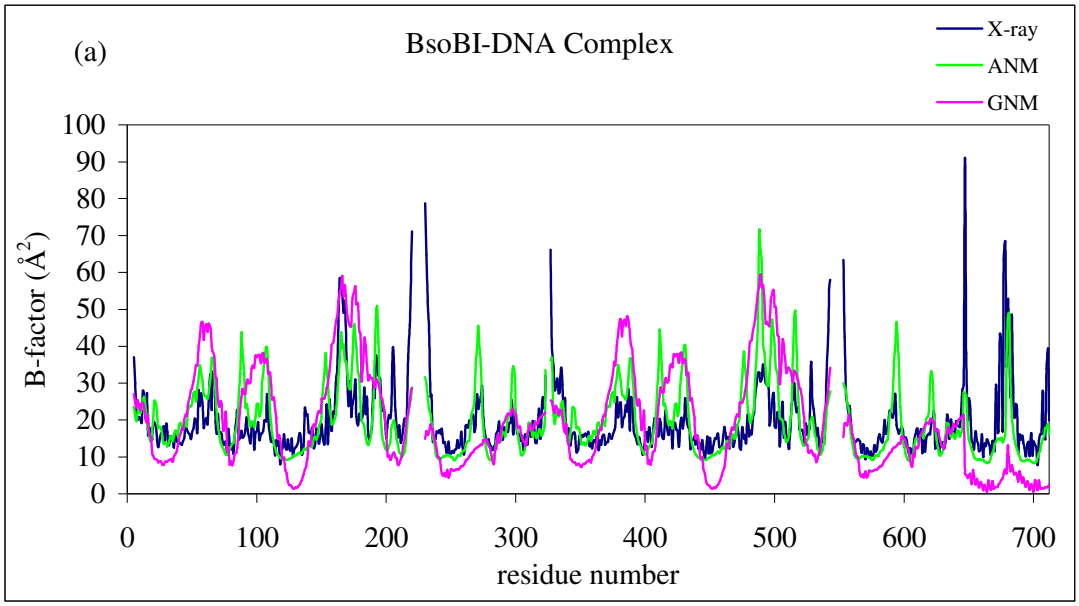


Figure 4.2. B-factors of BamHI comparing X-ray, ANM and GNM results for (a) complex, and (b) apo forms



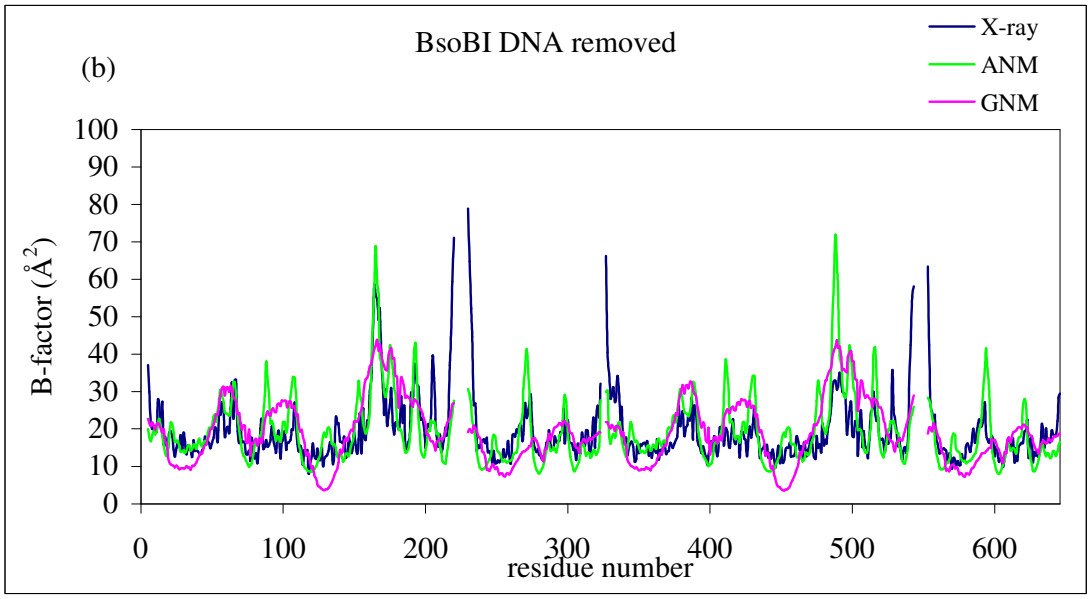
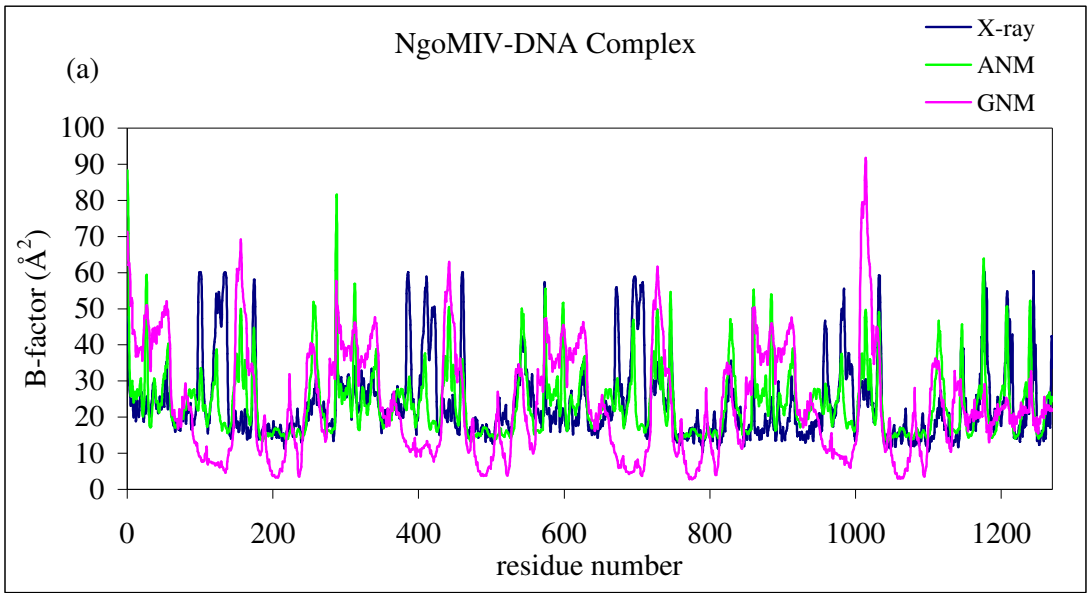


Figure 4.3. B-factors of BsoBI comparing X-ray, ANM and GNM results for (a) complex, and (b) free forms



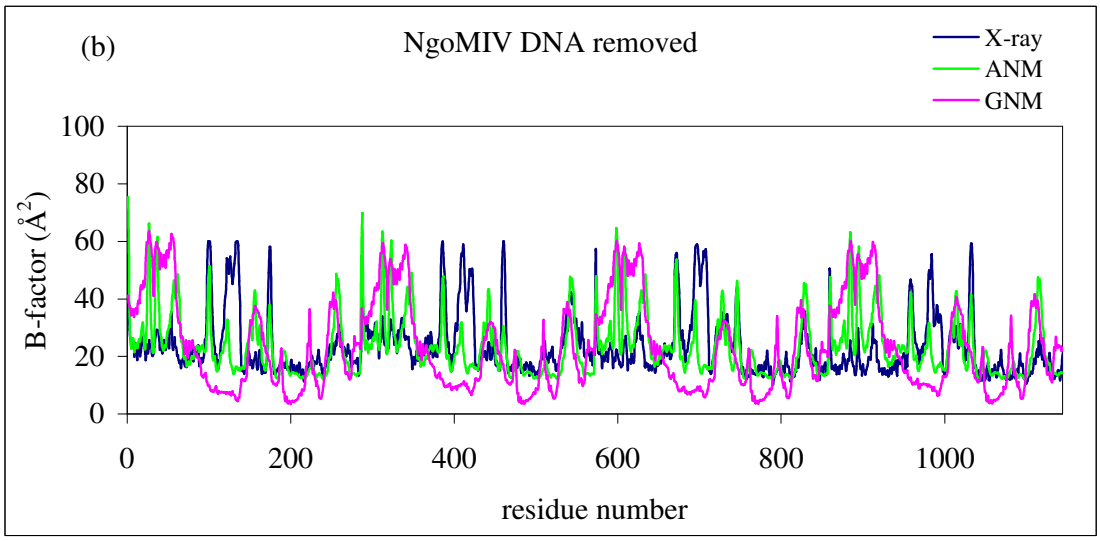


Figure 4.4. B-factors of NgoMIV comparing X-ray and ANM results for (a) complex, and (b) free forms

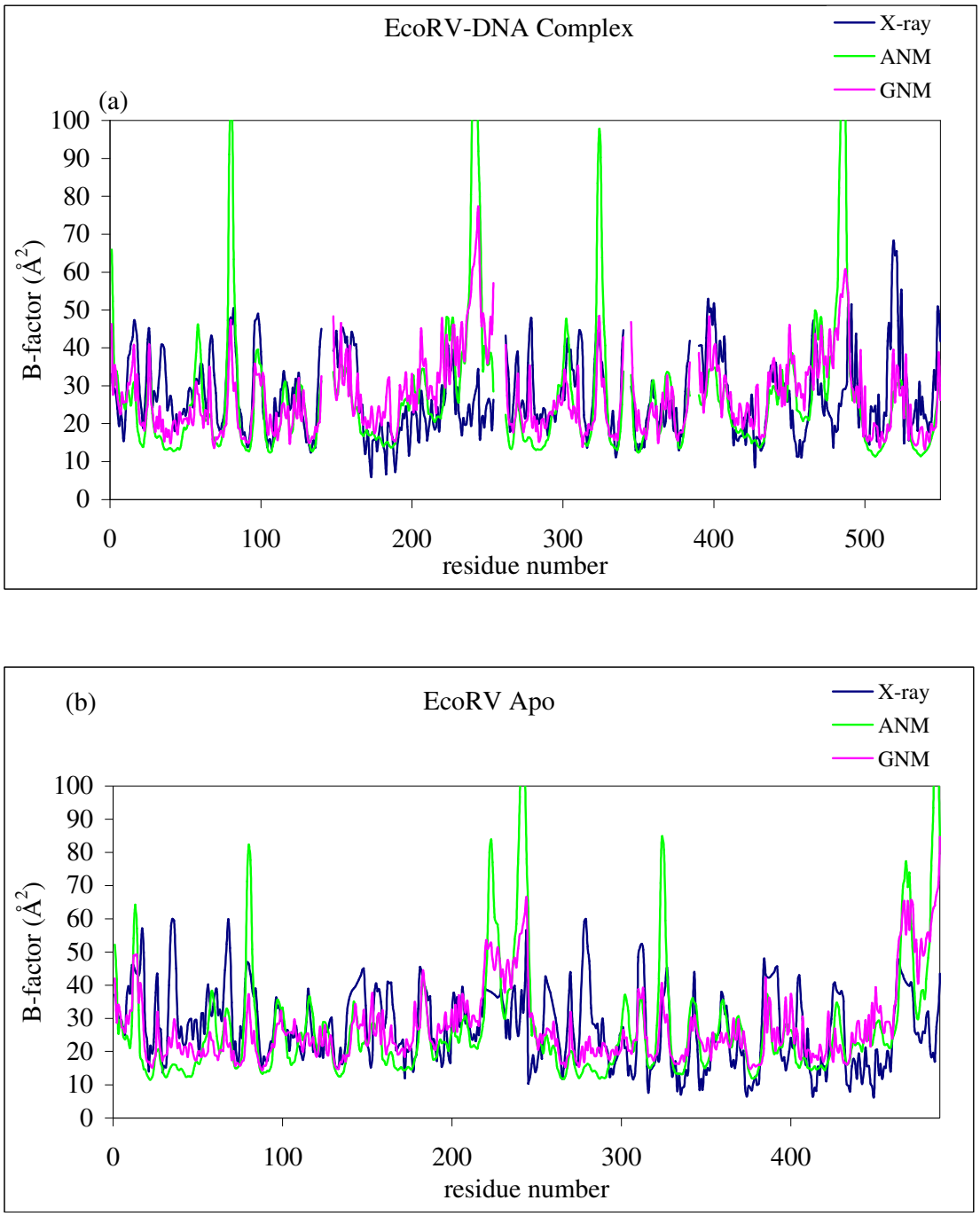


Figure 4.5. B-factors of EcoRV comparing X-ray, ANM and GNM results for (a) complex, and (b) apo forms

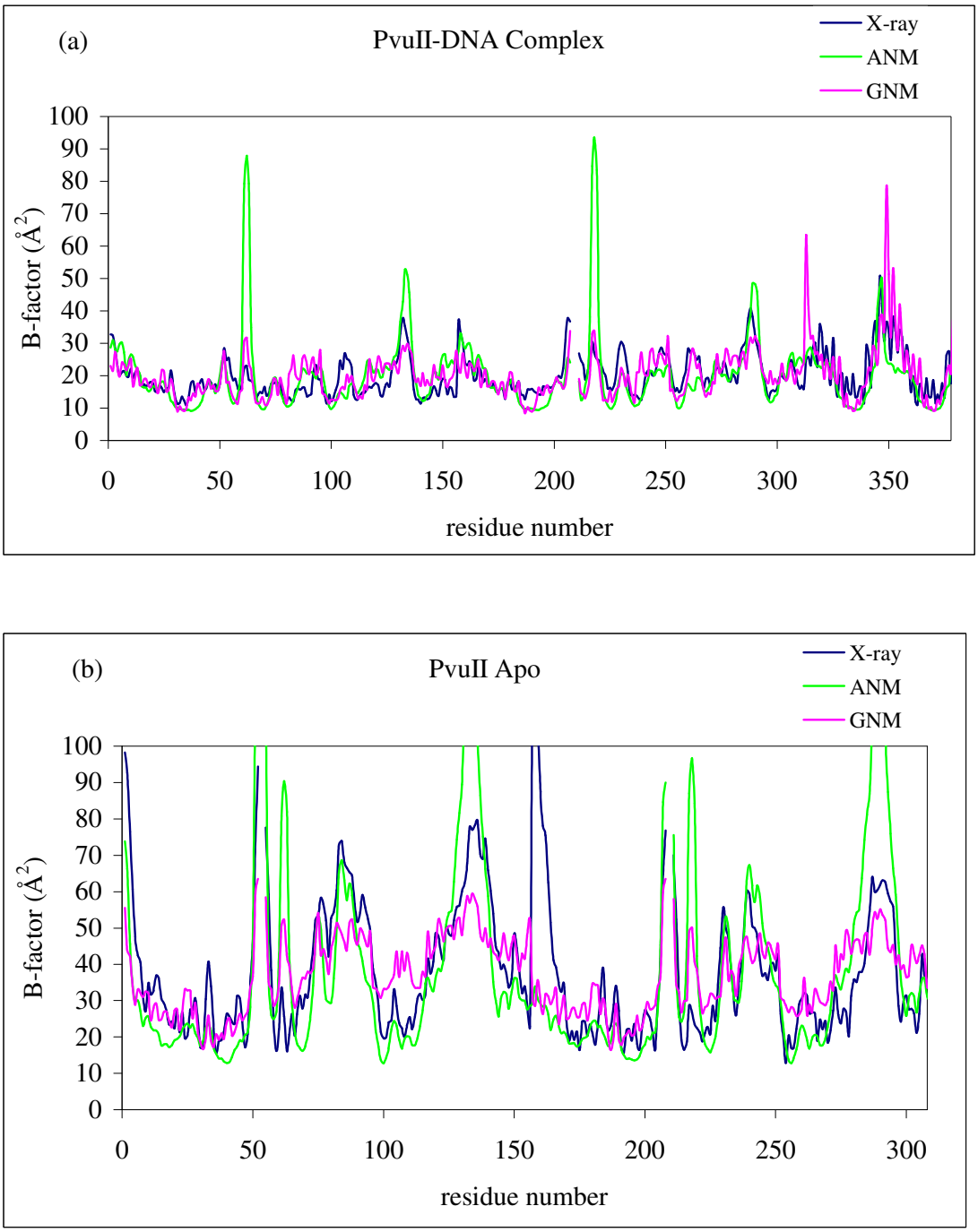


Figure 4.6. B-factors of PvuII comparing X-ray, ANM and GNM results for (a) complex, and (b) apo forms

4.2.1. Collective Deformations

Ms fluctuations showing the cumulative action of the slowest 10 modes (comprising ~15% of the overall motion) from ANM are plotted in Figures 4.7 to 4.9. The missing residues in each monomer are indicated with gaps. Dashed lines indicate the apo or DNA removed form, and the solid lines indicate the complex with DNA form.

ANM calculations are performed for apo structures of BamHI (1BAM.pdb) (Newman *et al.*, 1994) and apo form of EcoRI (1QC9.pdb) (Chandrasekhar *et al.*, (*unpublished*)) to observe the changes in dynamics due to DNA binding. In Figure 4.7.(a-b), ms fluctuations can be seen for the complex and apo form of BamHI in (a) and for the complex and apo form of EcoRI in (b). Slow mode behavior of the apo form of BamHI is quite different from the complex form. The difference lies in the enhanced mobility of the loop1 and α6-7-8 surrounding DNA and also the loop4, which connects β2 and β3, in both monomers, which are expected in the absence of DNA. Structured loop is not mobile in complex form, because in this form the structured loop forms contacts with the DNA. For BamHI-DNA complex, the loop4 in both monomers is very mobile in the first 10 collective modes because they are at the surface of the protein and free to move.

The inner and outer loops near to DNA as shown in the Figures 4.7.(b) are extremely mobile in the first 10 collective modes in EcoRI apo form. Extreme mobility of these loops is due to missing residues in the loops. As a result, the details of the collective motions in apo EcoRI cannot be clearly observed.

In Figure 4.8.(a), the cumulative ms fluctuations of apo (1RVE.pdb, Winkler *et al.*, 1993) and complex forms is compared for EcoRV. The loops (R-loop, Q-loop, and loop7) near to DNA, as indicated in Figure 4.8.(a), show high mobility in apo form of EcoRV. The loop6 is at the surface of the protein, therefore this loop has extreme mobility in both apo and complex forms.

In Figure 4.8.(b), it can be seen that the residues (forming loop3, outer helix, and outer loop) surrounding DNA and near to DNA cleavage site are mobile in apo form of PvuII similar to EcoRV. Residue in loop2 in monomer 1 denoted by circle show high

fluctuations because of the missing residues of this loop in apo form of PvuII. However, same behavior is not observed for the same residue in monomer 2.

ANM calculations are performed to observe dynamics differences between DNA bound and removed forms of BsoBI and NgoMIV, since apo structures are not available. By the removal of DNA from native structure of BsoBI, the outer loop located on monomer 1 increases pointing that this area stabilized upon DNA binding, and this behavior is not observed on the symmetric region at monomer 2 (Figure 4.9.(a)). Similarly in NgoMIV, DNA interacting regions (forming outer loop and helix) and several protein surfaces (forming α9) seem to be stabilized in the presence of DNA (Figure 4.9.(b)).

4.2.2. Scissor - like Motion

For each RE, one of the four slowest modes from ANM exhibits a scissor-like motion. In this particular motion, two monomers undergo a ratchet-like rotation like scissors. Displacement vectors for these modes are shown in Figures 4.10 to 4.12 with two monomers represented in red and green and DNA in blue. In order to better visualize the motion, an arbitrary scaling factor is used to exaggerate the displacements. On these figures, arrows show one alternative direction for the complex and apo forms of the enzymes. Cartoon presentations of RE structures on the top of the figure indicate different views of enzymes for the panels and coloring of the monomers.

The first slowest harmonic modes of motion are shown in Figure 4.10 for DNA complex (a-b), and for apo forms (c-d) of BamHI and EcoRI. In the first mode of BamHI-DNA complex, the outer loops exhibit a scissor-like motion along the DNA strands. These loops move parallel to DNA strands, whereas the core domain and the DNA remain immobile (Figure 4.10.(a)). In the first mode of its apo form, the enhanced mobility of the core domain, the outer loops and α5 helices can be observed (Figure 4.10.(c)). Similar results are observed in slowest mode of EcoRI, as in a previous study (Doruker *et al.*, 2006) (Figure 4.10.(b), (d)). Scissor-like motion of two monomers in the presence of DNA dramatically change to an immobilization of core domain and perpendicular motion of inner and outer loops into the DNA cavity in apo-form.

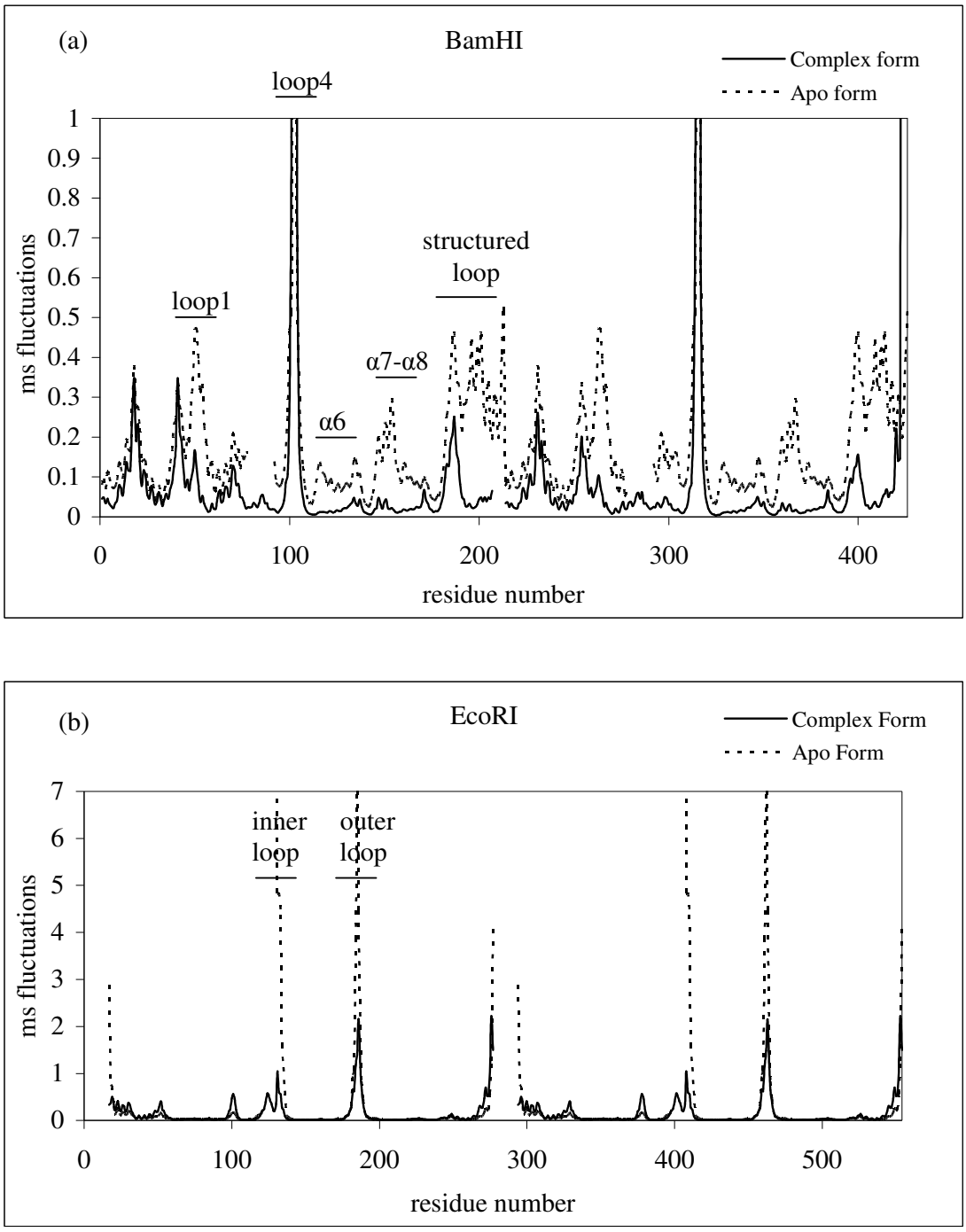


Figure 4.7. Mean-square fluctuations including 10 low-frequency modes for the apo and complex forms of (a) BamHI, and (b) EcoRI

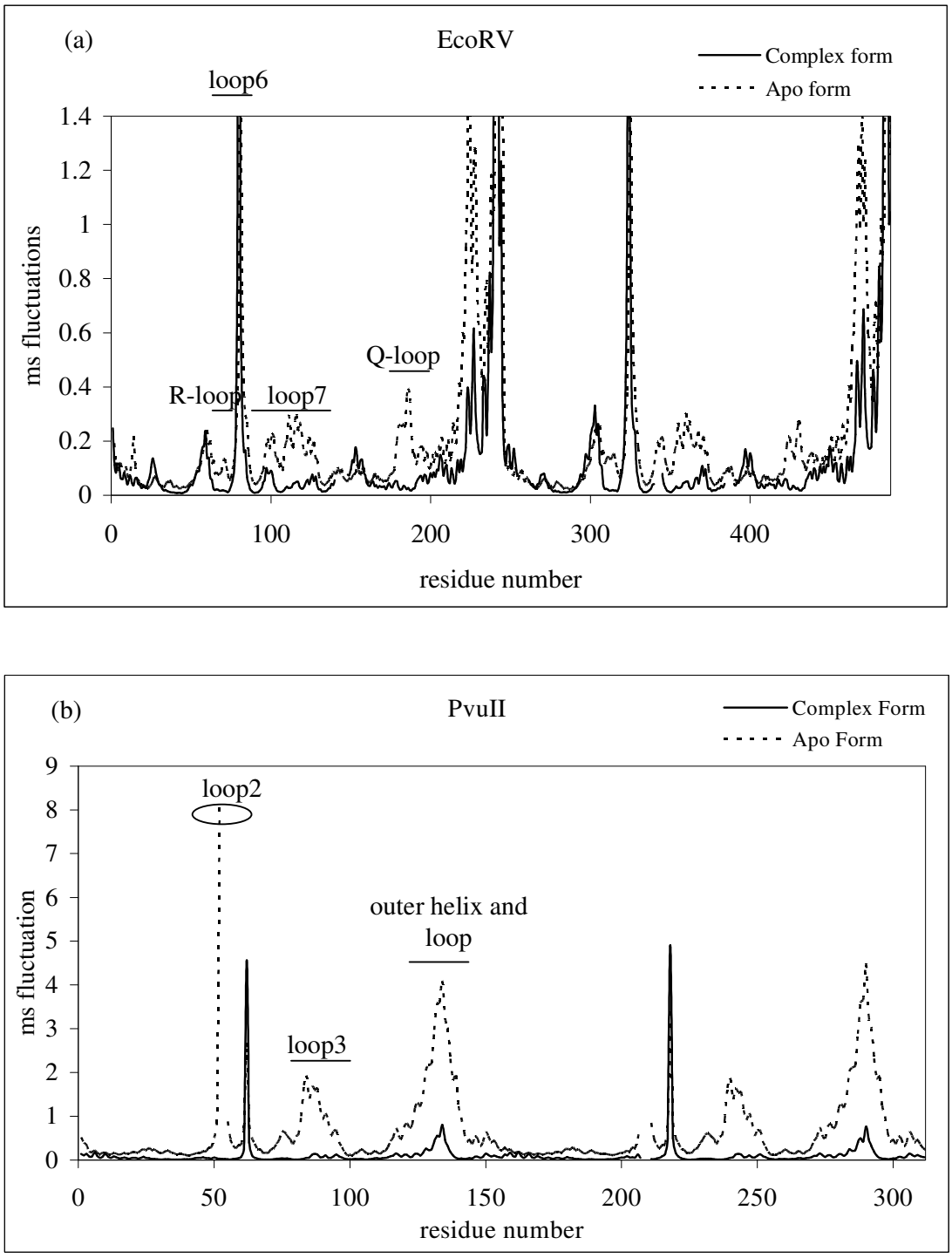


Figure 4.8. Mean-square fluctuations including 10 low-frequency modes for the apo and complex forms of (a) EcoRV, and (b) PvuII

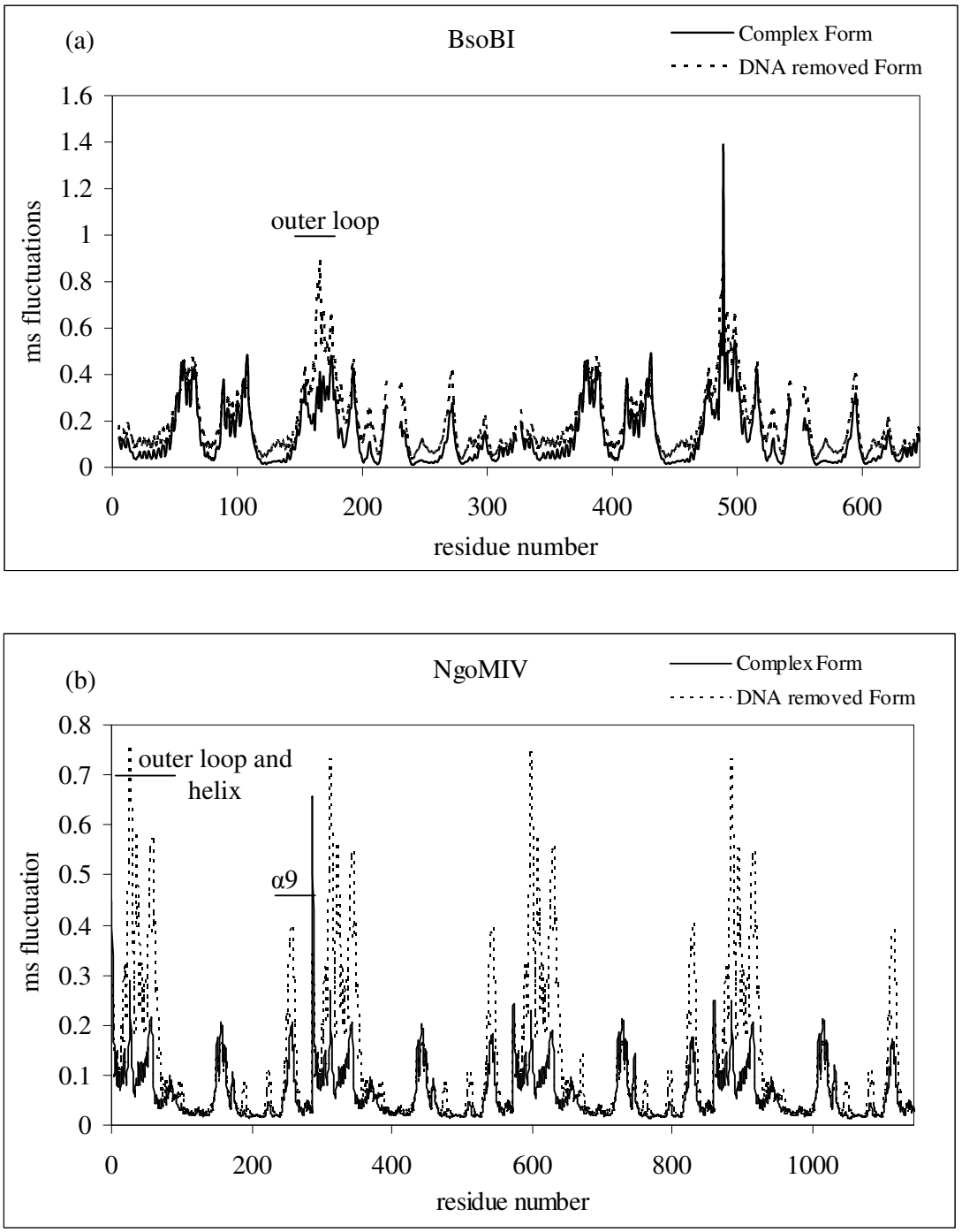


Figure 4.9. Mean-square fluctuations including 10 low-frequency modes for the DNA removed and complex forms of (a) BsoBI, and (b) NgoMIV

In the Figure 4.11.(a-d), the displacement vectors for the modes of the PvuII and EcoRV are displayed with their DNA-bound and apo-forms. As in the collective motions observed for EcoRI family members, the helices (α3-α4 in PvuII and α9-α10 in EcoRV)

and inner loops surrounding DNA binding pocket are moving parallel to DNA in the first mode of PvuII-DNA and EcoRV-DNA complexes (Figure 4.11.(a)-(b)). Also the same behavior can be observed in the second mode of PvuII and the first mode of EcoRV apo forms (Figure 4.11.(c)-(d)). On the other hand, the residues (forming α5 helix and β1, β2, and β7 strands) on the top of EcoRV monomers and DNA form a relatively immobile region (Figure 4.11.(b)) while the core domain seems to become more mobile in the first mode of its apo form (Figure 4.11.(d)).

In the second mode of BsoBI-DNA complex, scissor-like motion of monomers can be observed (Figure 4.12.(a)). Interestingly in this motion, upper part of one monomer moves with the lower part of the other monomer (green with red and vice versa). Same behavior is observed in the first mode of BsoBI DNA removed form (Figure 4.12.(c)). In the first mode of NgoMIV-DNA complex and the second mode of its DNA removed form, monomers in side-by-side contacts (monomer 1/3 and 2/4) move together and again form scissor-like motion (Figure 4.12.(b)-(d)).

This particular motion that is observed in apo and complex forms may give an idea about the mechanism for REs' efficient target site location on DNA. Facilitated diffusion has been shown to be effective in target site location (Jack *et al.*, 1982; Ehbrecht *et al.*, 1985; Terry *et al.*, 1985). There are three different proposed mechanisms: sliding, jumping or hopping, and intersegment transfer (Berg and von Hippel, 1985; Jeltsch and Urbanke, 2004; Halford and Marko, 2004; Coppey *et al.*, 2004).

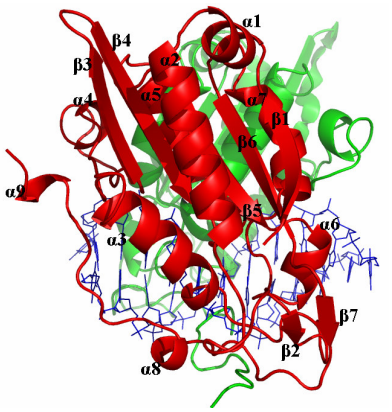
Sliding is the random movement of an enzyme along the DNA, following the pitch of the double helix, when the enzyme is in a nonspecifically bound state (Jeltsch *et al.*, 1994; Sun *et al.*, 2003). It is also called linear or 1-D diffusion. Schematic representation of this mechanism can be seen in Figure 4.13.(a), which is adapted from the study of Jeltsch and Urbanke (Jeltsch and Urbanke, 2004). This implies that specific sites tend not be overlooked by an RE sliding along the DNA. During linear diffusion, the non-specific binding mode is not given up and the water layer around DNA and enzyme, characteristic for the non-specific binding mode, remains largely intact (Pingoud *et al.*, 2005).

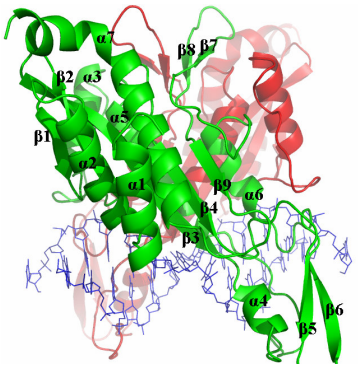
In contrast, jumping or hopping is normal diffusion that takes into account that the chance of reassociation of a DNA binding protein to the DNA molecule close to the site it has dissociated from is much greater than associating with another DNA molecule or a distant site on the same molecule. Schematic representation can be seen in Figure 4.13.(b). During this mechanism, the non-specific binding mode is given up and the water layer characteristic for the free DNA and the free protein is re-formed. Jumping or hopping does not follow the pitch of the double helix, meaning that specific sites along the DNA can be overlooked during hopping, depending on the step size.

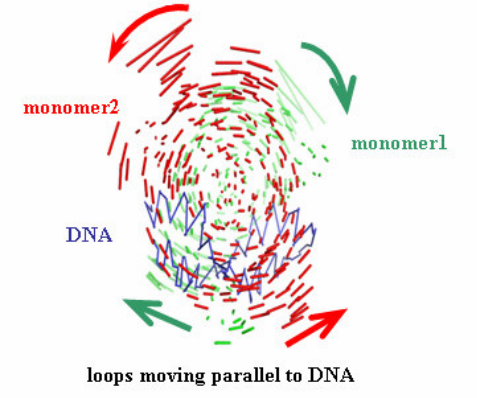
Moreover, the comparison of sliding and jumping mechanisms in terms of their activation barrier requirements may give insights. The sliding rate depends on the size of an activation barrier ($\Delta G^{\#}_{slide}$) between the enzyme being bound to the DNA in one register and in the next (Figure 4.14., adapted from the study of Jeltsch and Urbanke (Jeltsch and Urbanke *et al.*, 2004)). Non-specific DNA binding is mediated by electrostatics and by hydrogen bonds to the phosphate groups. Since electrostatic forces are not directional, an attractive electrostatic potential will appear smooth with $\Delta G^{\#}_{slide}$ being very small. The protein has to overcome the $\Delta G^{\#}_{slide}$ in order to slide on the DNA by one base pair denoted by filled arrows on Figure 4.14. For a hopping movement (dashed arrows), all molecular interactions of protein must be broken to dissociate from the DNA, which requires a larger activation barrier than $\Delta G^{\#}_{slide}$ for dissociation $\Delta G^{\#}_{dis}$. Thus, thermodynamically the energy of the unbound state is increased by ΔG^{0}_{dis}. Thus, sliding mechanism seems more probable in terms of activation barrier requirement. Similarly, in scissor-like motion, it is observed that not only the dimeric but also the tetrameric REs move along the DNA in such a way that they are trying to find their target site location. In this respect scissor-like motion may give an insight about the sliding mechanism for target site location of REs.

4.2.3. Twisting Motion

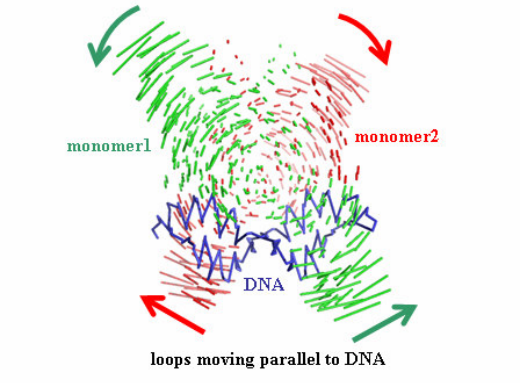
Twisting motion of subunits is another behavior observed in the slowest modes for REs. Rescaled displacement vectors of residues in the slow mode with twisting behavior are shown in Figures 4.15 to 4.17.





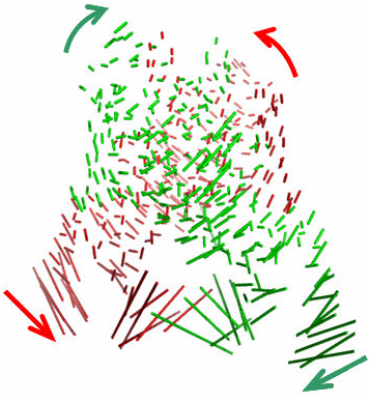


(a) BamHI-DNA complex, first mode



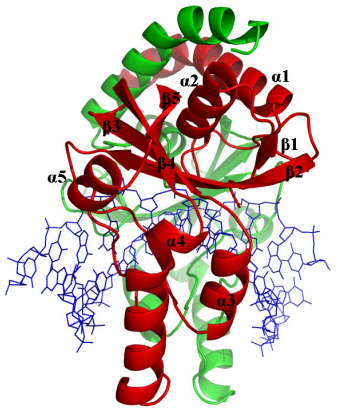
(b) EcoRI-DNA complex, first mode

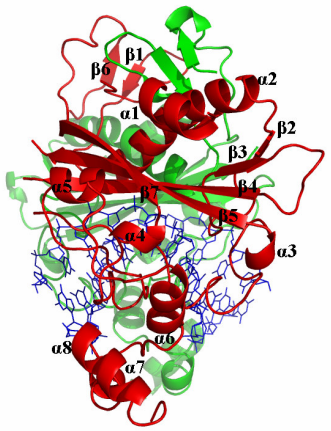
(c) BamHI apo, first mode

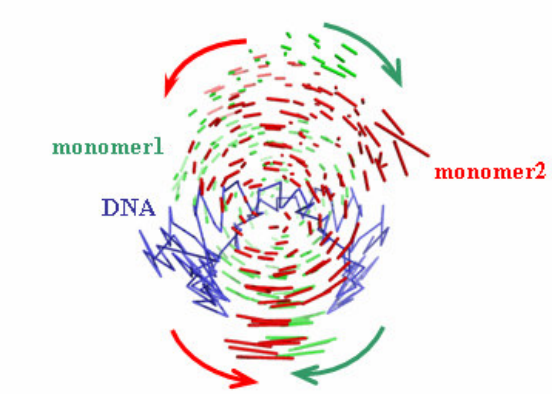


(d) EcoRI apo, first mode

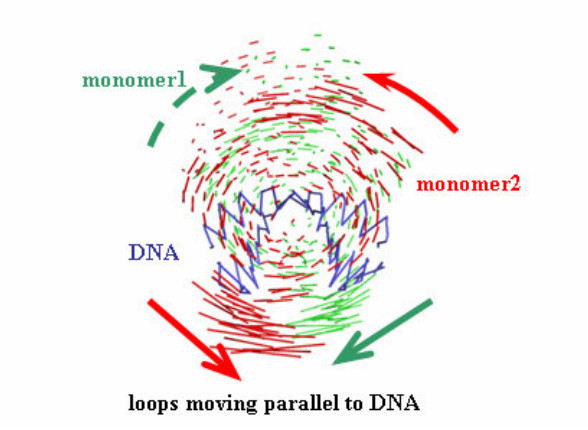
Figure 4.10. Displacement vector representations showing scissor-like motion for BamHI and EcoRI



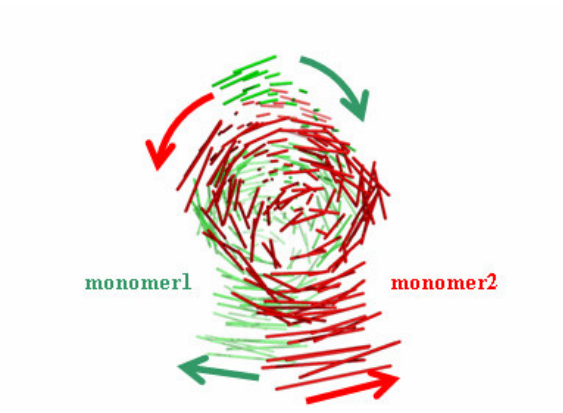




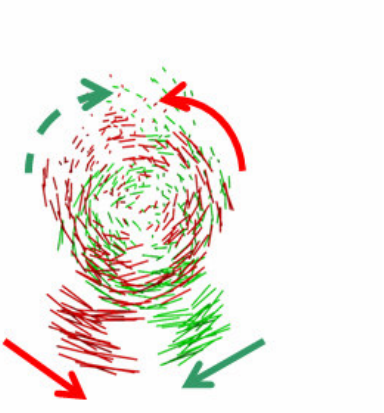
(a) PvuII-DNA complex, first mode



(b) EcoRV-DNA complex, first mode

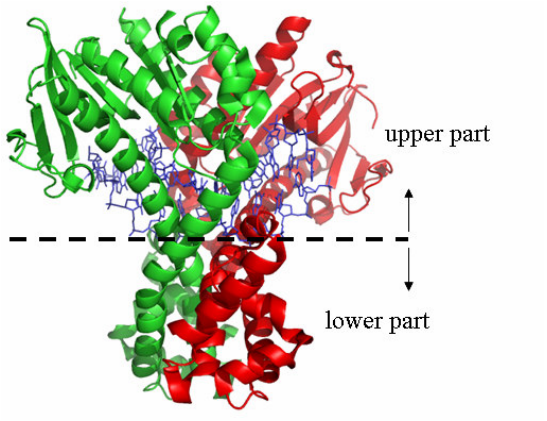


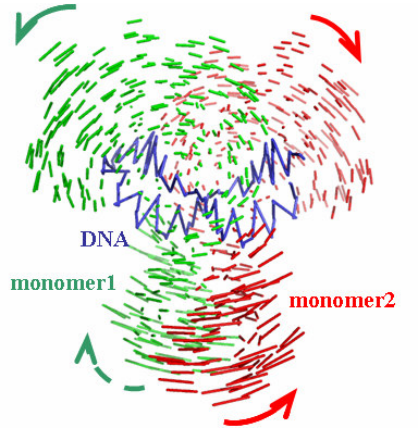
(c) PvuII apo, second mode



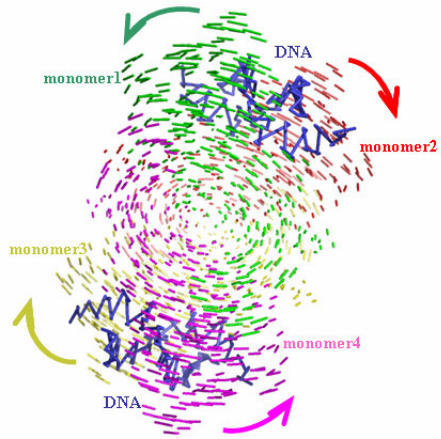
(d) EcoRV apo, first mode

Figure 4.11. Displacement vector representations showing scissor-like motion for PvuII and EcoRV

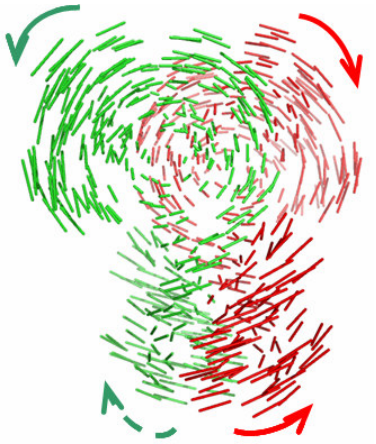




(a) BsoBI-DNA complex, second mode



(b) NgoMIV-DNA complex, first mode



(c) BsoBI DNA removed, first mode

(d) NgoMIV DNA removed, second mode

Figure 4.12. Displacement vector representations showing scissor-like motion for BsoBI and NgoMIV

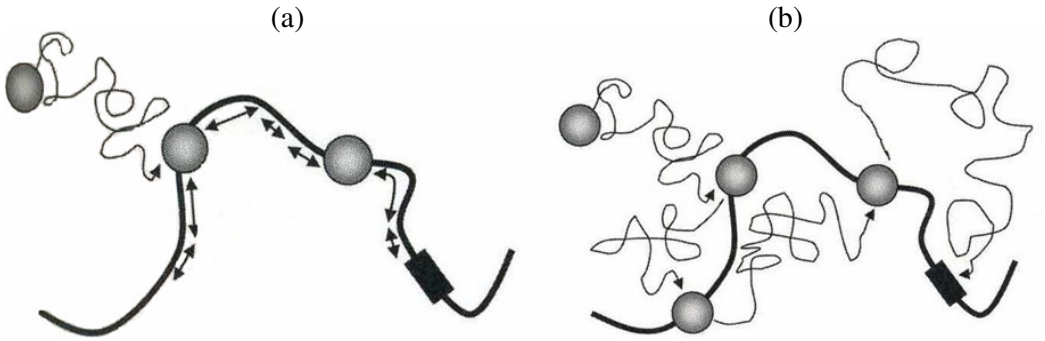


Figure 4.13. Target site location by (a) sliding, and (b) jumping or hopping

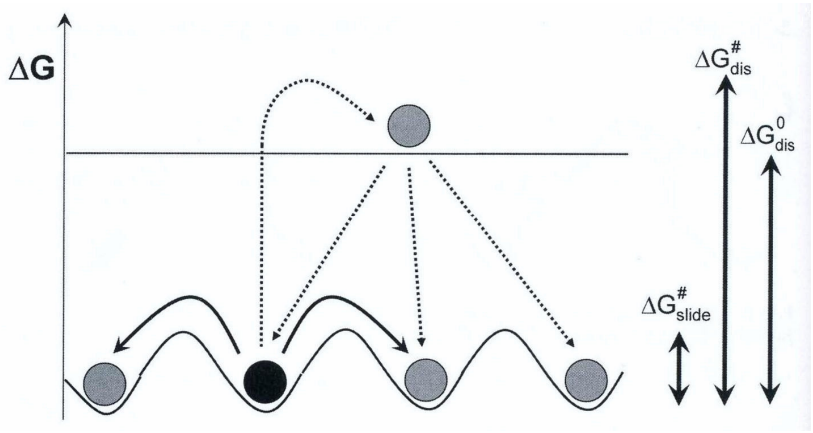


Figure 4.14. Comparison of the changes in Gibbs energy associated with a sliding and jumping movement of a protein on DNA

Similarly, arrows show one of the alternative oscillation directions for the complex, apo and DNA removed forms of the REs. The red and green colors represent different monomers and DNA is shown in blue. The cartoon representations of REs on the top of the figure indicate the perspective shown in each panel.

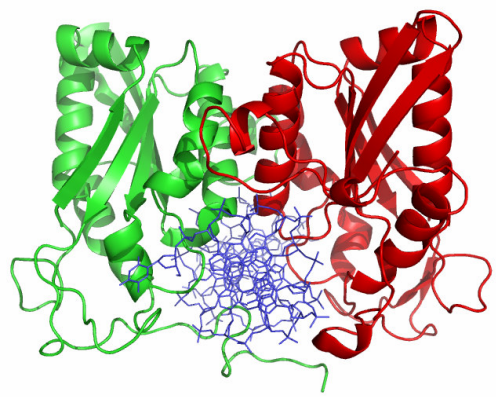
Figure 4.15 displays the second slowest harmonic modes of motion for DNA complex and apo forms of BamHI (a, c) and of EcoRI (b, d), respectively. In the second mode of BamHI-DNA complex, the residues surrounding the DNA are immobile and motion of two monomers describe a counter-rotation of the two monomers named as twisting motion (Figure 4.15.(a)). The loop4 of both monomers are relatively mobile. The same twisting motion but with higher magnitude is observed in the second mode of apo form (Figure 4.15.(c)). Similarly, in the second mode of EcoRI-DNA complex and apo

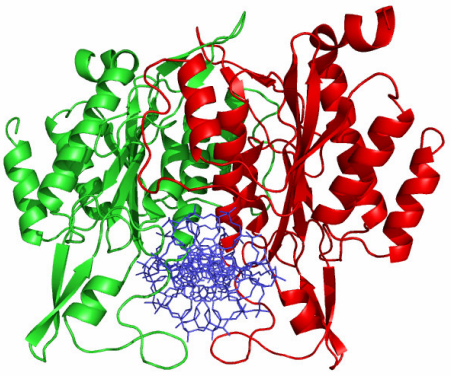
form, the inner and outer loops move parallel to DNA and both monomers twist around the immobile core (Figure 4.15.(b)) as shown in a previous study (Doruker *et al.*, 2006). Twisting motion with high magnitude of EcoRI apo form is not observed because of the high fluctuations of inner loops.

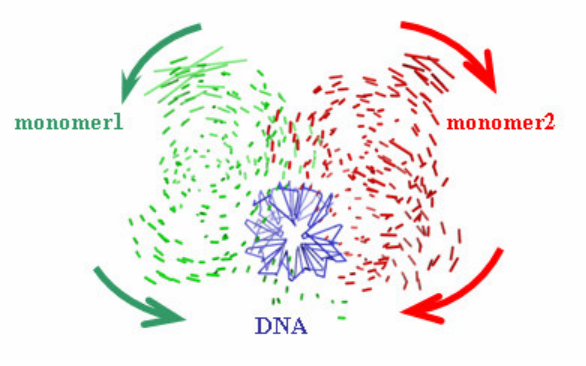
In the second mode of PvuII-DNA complex, the motion of two monomers describe a counter-rotation similar to previous cases (Figure 4.16.(a)). The loops and helices surrounding the DNA have more mobility while the α1 helix on the top of each monomer, which is parallel to DNA, is not highly mobile. However, in the second mode of apo form, the core domain and the residues (forming α3-4 and loops surrounding DNA) are highly mobile (Figure 4.16.(c)). Similar twisting motion is observed in the third mode of EcoRV-DNA complex and the second mode of apo form (Figure 4.16.(b-d)).

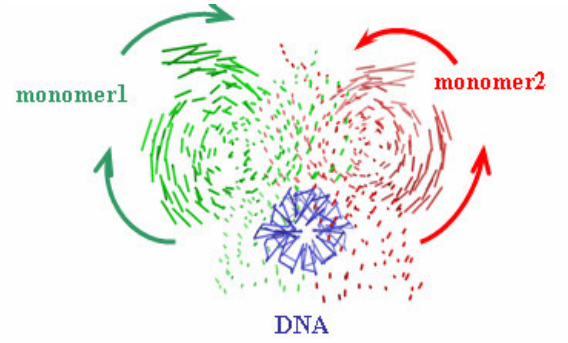
In Figure 4.17.(a)-(c), eigenvector representations of BsoBI-DNA complex and DNA removed forms are shown. In the third mode of BsoBI-DNA complex and the fourth mode of DNA removed form, both upper parts and lower parts of each monomer are in a reverse rotation describing a twisting motion (Figure 4.17.(a)-(c)). For NgoMIV-DNA complex and DNA removed form, eigenvector representations are shown in Figure 4.17.(b)-(d). Each monomer is denoted by different colors as green, red, dark yellow and magenta, respectively. Twisting motion with low magnitudes is observed in all monomers of DNA complex (Figure 4.17.(b)-(d)). In the absence of DNA, these fluctuations increase specifically in the loops 2-3 (Figure 4.17.(b)-(d)).

REs in apo form have to open their DNA binding site in order to bind to DNA. In most instances the enzymes make their binding sites accessible in a tong-like motion (Scheuring-Vanamee *et al.*, 2004). There is a question about oscillation of REs between their closed and open states (Pingoud *et al.*, 2005). In twisting motion, it is observed that the enzymes oscillate between their closed and open states. This result found in this study may be an important clue for the answer to the question on oscillation between closed and open states.

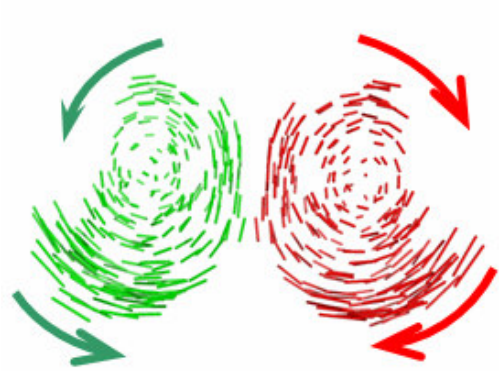


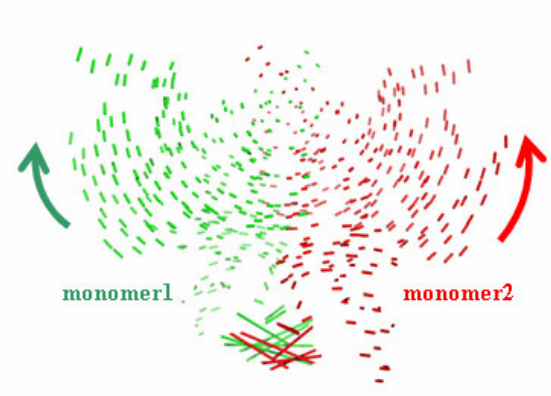






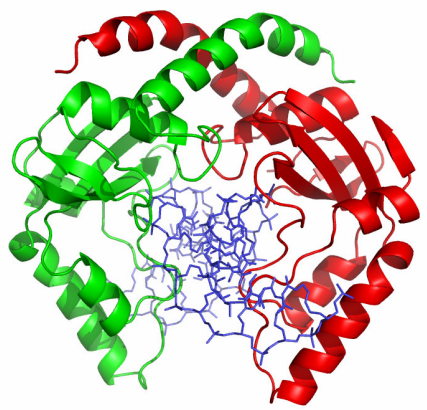
(a) BamHI-DNA complex, second mode (b) EcoRI-DNA complex, second mode

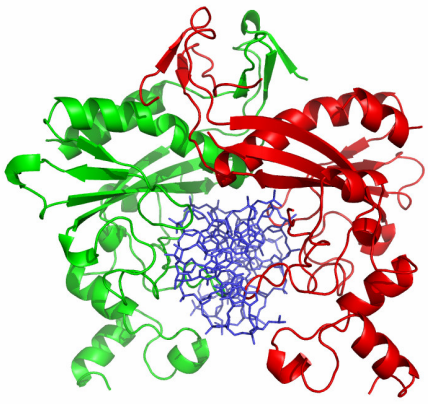


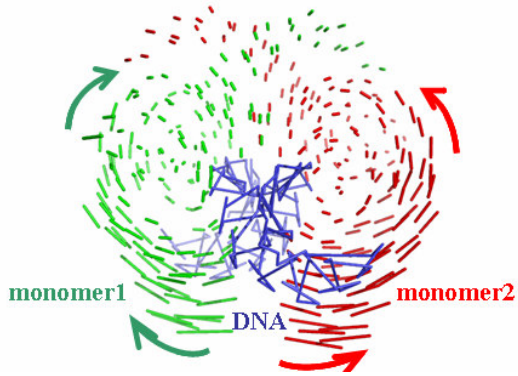


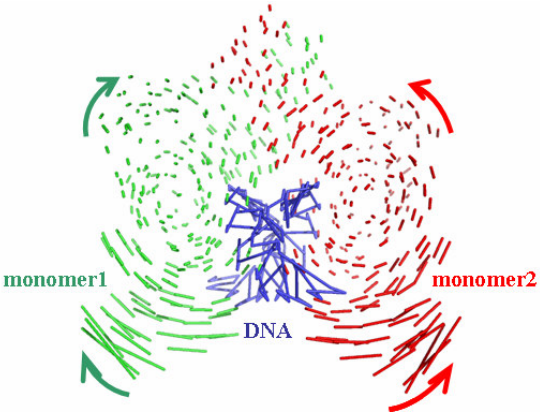
(c) BamHI apo, second mode (d) EcoRI apo, second mode

Figure 4.15. Displacement vectors showing twisting motion of BamHI and EcoRI









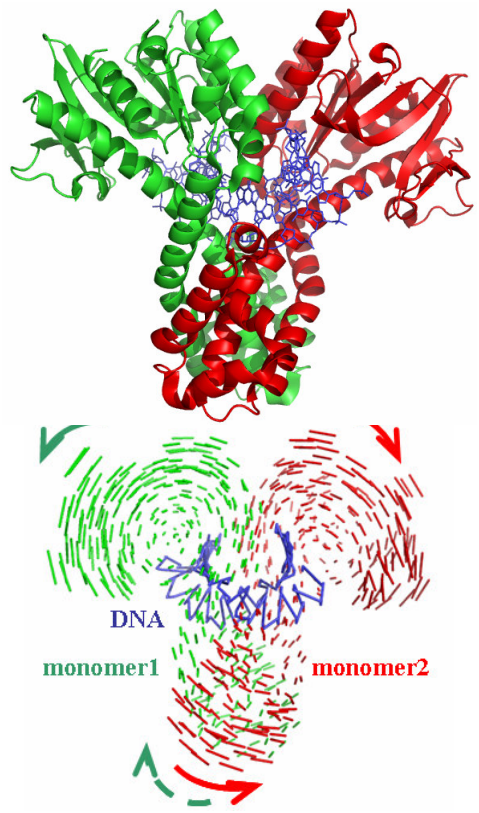
(a) PvuII-DNA complex, second mode

(b) EcoRV-DNA complex, third mode

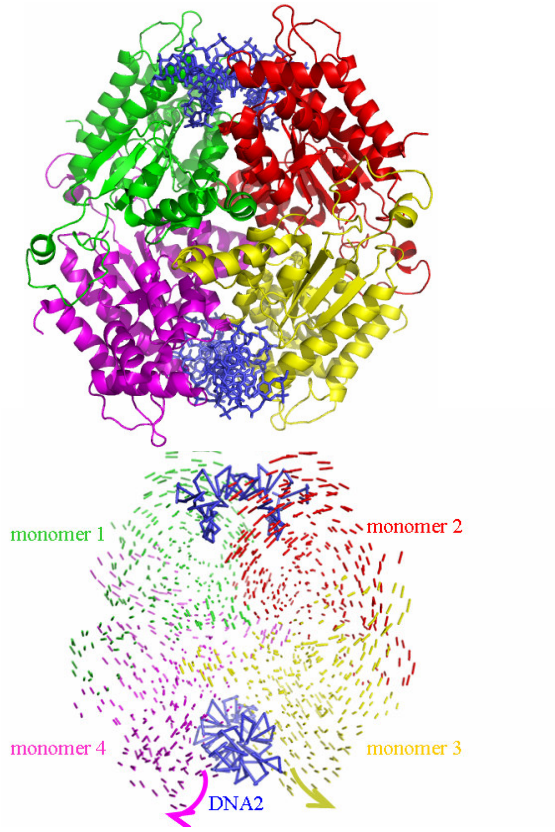
(c)PvuII apo, first mode

(d) EcoRV apo, second mode

Figure 4.16. Displacement vectors showing twisting motion of PvuII and EcoRV

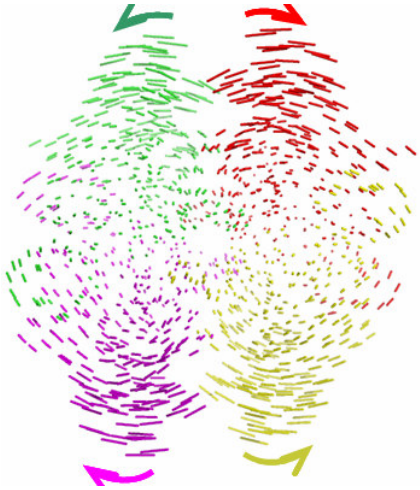


(a) BsoBI-DNA complex, third mode



(b) NgoMIV-DNA complex, fourth mode

(c) BsoBI DNA removed, fourth mode



(d) NgoMIV DNA removed, fourth mode

Figure 4.17. Displacement vectors showing twisting motion of BsoBI and NgoMIV

4.2.4. Orientational cross-correlations

The normalized orientational cross-correlations, *C(i,j)*, between residue fluctuations are calculated by the following equation.

$$C(i,j) = \frac{\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle}{\left[\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle \langle \Delta \mathbf{R}_j \cdot \Delta \mathbf{R}_j \rangle \right]^{1/2}} \tag{4.1}$$

The cross-correlations vary within the range [-1, 1] with the lower and upper limits indicating respective fully anticorrelated and correlated fluctuations, whereas *C(i,j)* = 0 indicates uncorrelated fluctuations in terms of orientation. First five slowest modes calculated with ANM and GNM are used to obtain the cumulative effect on orientational cross-correlations. In all cross-correlation figures, the positively correlated (red) fluctuations are given in upper left-hand triangle, being separated from the negative correlations (blue) shown in the lower right-hand triangle for clarity. The missing residues are indicated with gaps that are uncorrelated with the rest of the residues.

EcoRI: Orientational cross-correlations are calculated from cumulative first five normal modes of ANM. The results for EcoRI-DNA complex is represented in Figure 4.18.(a). These results are in conformity with the cross-correlations from previous study calculated for the first ten modes of motion (Doruker *et al.*, 2006). Left top square shaped area indicates that two monomers within the core region are positively correlated pointing to one of the distinct domains within the structure. The upper diagonal triangles indicate the positive correlation of the core domain residues within each monomer.

Each inner loop is strongly correlated with the closest DNA strand and not correlated with the other strand (Figure 4.18.(a)). For DNA and inner loops interactions, the strongest cross-correlation values ~0.90 are observed for the kinked region in the middle of DNA sequence. For the nucleotides G and A that are at the cleavage site (G/**AATT**C-C**TTAA**/G), it is observed that the strongest cross-correlation values ~0.9 are with the active site residues in both monomers (Asp91, Glu111, Lys113). The catalytic residues within each monomer are also positively correlated with each other. In addition, some positive correlations are observed between two catalytic sites. It indicates that a

possible way of communication between the distal active sites that are known to have concerted catalytic activity.

Residues Val128 to Lys130 that forms the tip of inner loop are strongly correlated with the nucleotides thymine and cytosine (TTC), which can be significant in the catalysis. Residues 142-170 and 443-471 are moving positively correlated with the DNA strands (Doruker *et al.,* 2006).

Cross-correlations obtained from ANM calculations for apo form of EcoRI is represented in Figure 4.18.(b). Left top square shaped area indicates correlation of two monomers within the core domain. The behavior is similar, but the intra-monomer positive correlations are enhanced whereas inter-monomer positive correlations are less pronounced.

Cross-correlations obtained from GNM (r_{cut}= 7 Å) calculations for EcoRI-DNA complex is represented in Figure 4.18.(c). Again, left top square shaped area shows correlation of two monomers forming the core domain. Positive correlations are weaker than those obtained from ANM calculations for EcoRI-DNA complex, whereas enhanced negative correlation values exist between monomers. The upper diagonal triangles indicate positive strong correlation of the core domain residues in each monomer. In Figure 4.18.(d), cross-correlations obtained from GNM calculations for the apo form of EcoRI is represented. Positive correlations of the core domain residues in each monomer increase, whereas positive correlations of two monomers within the core domain decrease in the apo form of EcoRI.

BamHI: As in the EcoRI-DNA complex, the strong positively correlated correlation regions of two monomers forming the core domain at the interface can be seen in Figure 4.19.(a) for BamHI-DNA complex from ANM calculations. Left top square shaped area shows correlation of two monomers within the core domain. The loop (residue indices 79-92) is correlated with the same loop (residue indices 79-91) of the other monomer in DNA complex. This structured loop in cognate DNA complex is at different position in non-cognate DNA complex form (see Figure 2.5) (Viadiu and Aggarwal, 2000).

From ANM calculations, it is observed that the cleavage site is strongly correlated (~0.90) with Asp94, Glu111, Glu113, which are the active site residues of BamHI located symmetrically on both monomers. These catalytic residues are positively correlated with each other and also with the catalytic site on the other monomer indicating a possible way of communication between the distal active sites as in the EcoRI case. When ANM and GNM cross-correlation results are compared it is seen that the results are not in perfect conformity within each other for both DNA complex and apo forms. In Figures 4.19.(c)-(d), inter-monomer positive correlation values are replaced mainly by inter-monomer negative correlations and intra-monomer positive correlations, for both DNA complex and apo forms of BamHI contrary to ANM results.

PvuII: In Figure 4.20.(a)-(b), cross-correlation plots can be seen for PvuII-DNA complex and apo form obtained from ANM calculations, respectively. Residues above DNA of each monomer denoted by black square are positively correlated in both DNA complex and apo form of PvuII (Figure 4.20.(a)-(b)). Interestingly, enhanced correlation of outer helix α1 (see Figure 2.9) with the same helix of the other monomer can be seen in apo form denoted by black square (Figure 4.20.(b)). In apo form, positive correlations between two monomers decreased. As for ANM results of PvuII, it is observed that residues above DNA of each monomer are strongly positive correlated in both DNA complex and apo forms of PvuII from GNM calculations (Figure 4.20.(c)-(d)). However, positive correlation (only ~ 0.20) of outer helix with the residues of other monomer above DNA can be seen in apo form (Figure 4.20.(d)). In apo form, it is observed that two monomers are almost negatively correlated.

EcoRV: Cross-correlations obtained from ANM calculations for EcoRV-DNA complex and apo form is represented in Figure 4.21.(a) and 4.21.(b), respectively. From the calculations, domains of EcoRV cannot be clearly seen. Only four regions can be distinguished as strongly positive correlated denoted by black squares in both DNA complex and apo forms. Results obtained from GNM calculations indicate that the same four regions are strongly positive correlated in both DNA complex and apo forms (Figure (4.21.(c)-(d)).

BsoBI: Cross-correlations obtained from ANM calculations for BsoBI-DNA complex and apo form is represented in Figure 4.22.(a) and 4.22.(b), respectively. Domains of BsoBI cannot be clearly seen from ANM calculations. Only a square shaped region can be distinguished as strongly positive correlated region denoted by black circle in both DNA complex and apo forms. This region can be more clearly seen in Figures 4.22.(c)-(d) obtained from GNM calculations.

NgoMIV: For NgoMIV-DNA complex and DNA removed form, cross-correlations obtained from ANM calculations are represented in Figures 4.23.(a)-(b), respectively. From Figure 4.23.(a), positive side-by-side contacts between monomers 1-4 and 2-3 and positive cross contacts between monomers 1-3 and 2-4 (see Figure 2.7.(b)) (Deibert *et al.*, 2000) can be clearly seen. Tetramerization loop, as shown in Figure 2.7.(a) (residues 147-176), between monomers 1-4 and 2-3 is positively correlated with each other as suggested Deibert *et al.* (Deibert *et al.*, 2000). However, cross-correlations obtained from GNM calculations for NgoMIV-DNA complex indicate that cross contacts between monomers 1-3 and 2-4 are mostly negatively correlated (Figure 4.23.(c)). For NgoMIV DNA removed form, ANM results indicate that cross-correlation values within each monomer increases (Figure 4.23.(b)). From GNM calculations, it is observed that side-by-side contacts increases in NgoMIV DNA removed form (Figure 4.23.(d)).

In summary, in the absence of DNA, the connectivity throughout the network changes and results in relatively different local harmonic motions in slow modes. Specifically orientational cross-correlations obtained from ANM results indicate that positive inter-subunit correlations generally weaken in the absence of DNA. However, this confirmation surprisingly can not be observed from GNM results. GNM locates only some key positive interactions at the interface. It could be used to determine native contacts from crystal structures (Kantarci *et al.*, 2006).

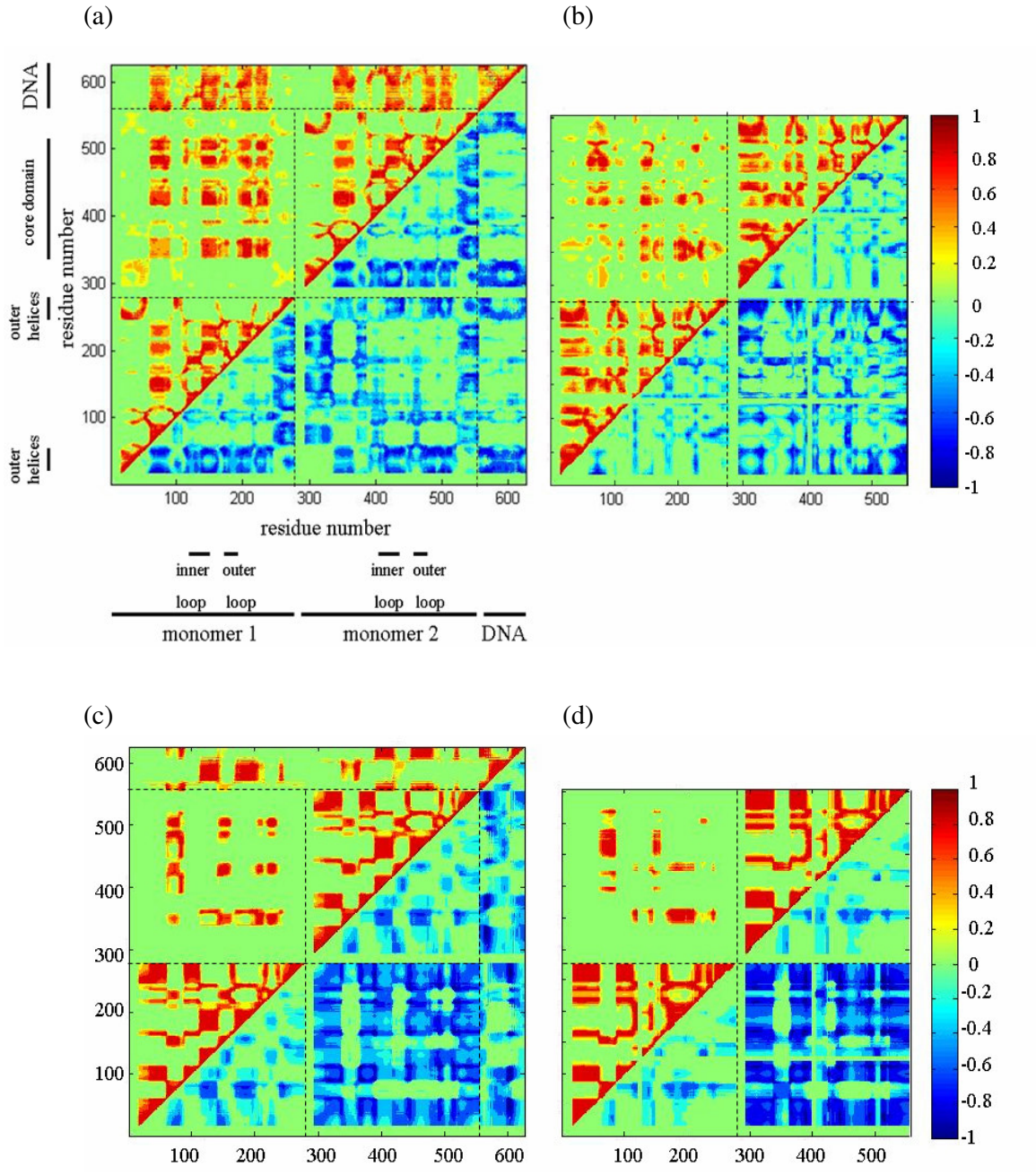


Figure 4.18. Orientational cross-correlations between residue fluctuations of EcoRI (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

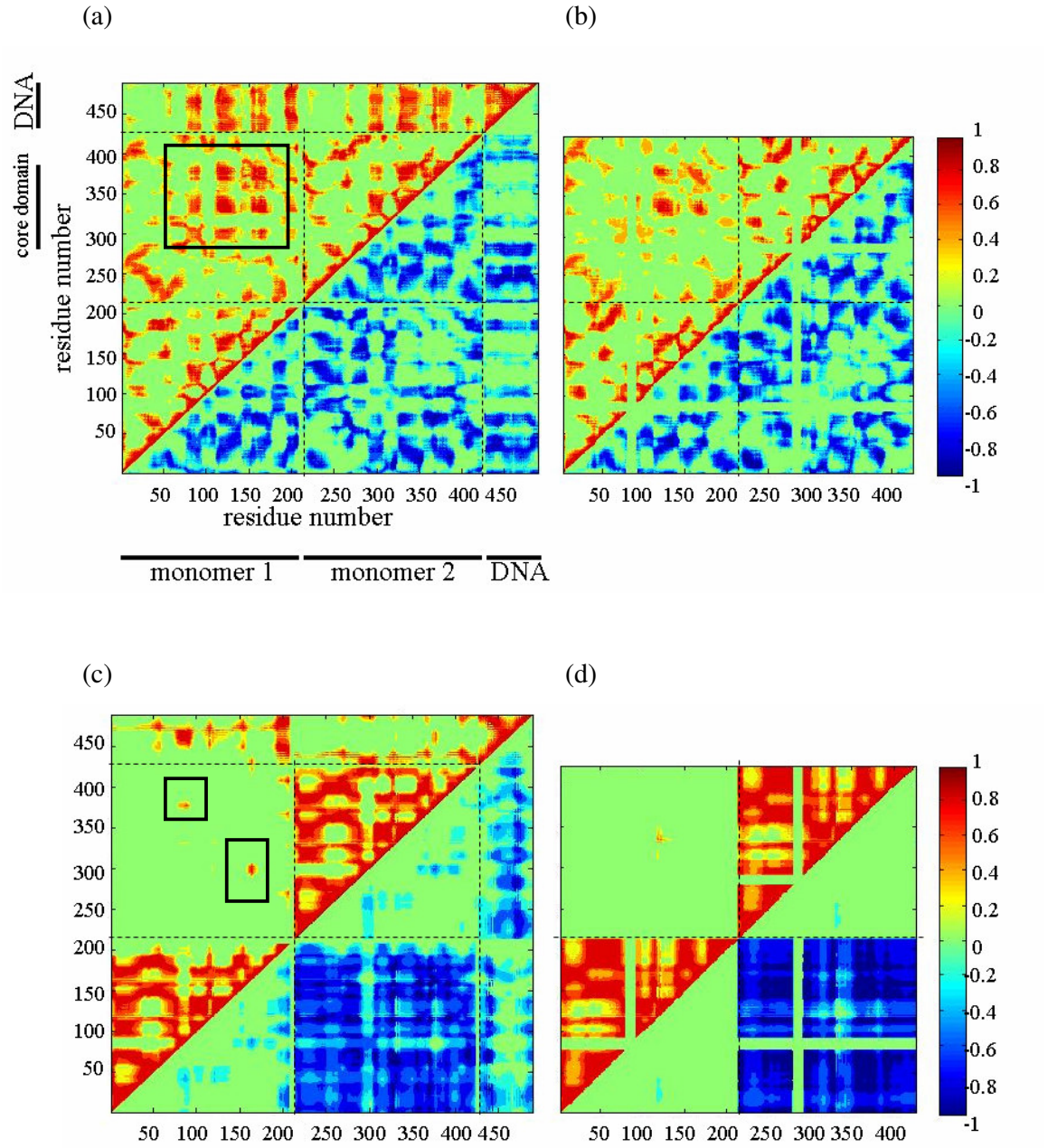


Figure 4.19. Orientational cross-correlations between residue fluctuations of BamHI (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

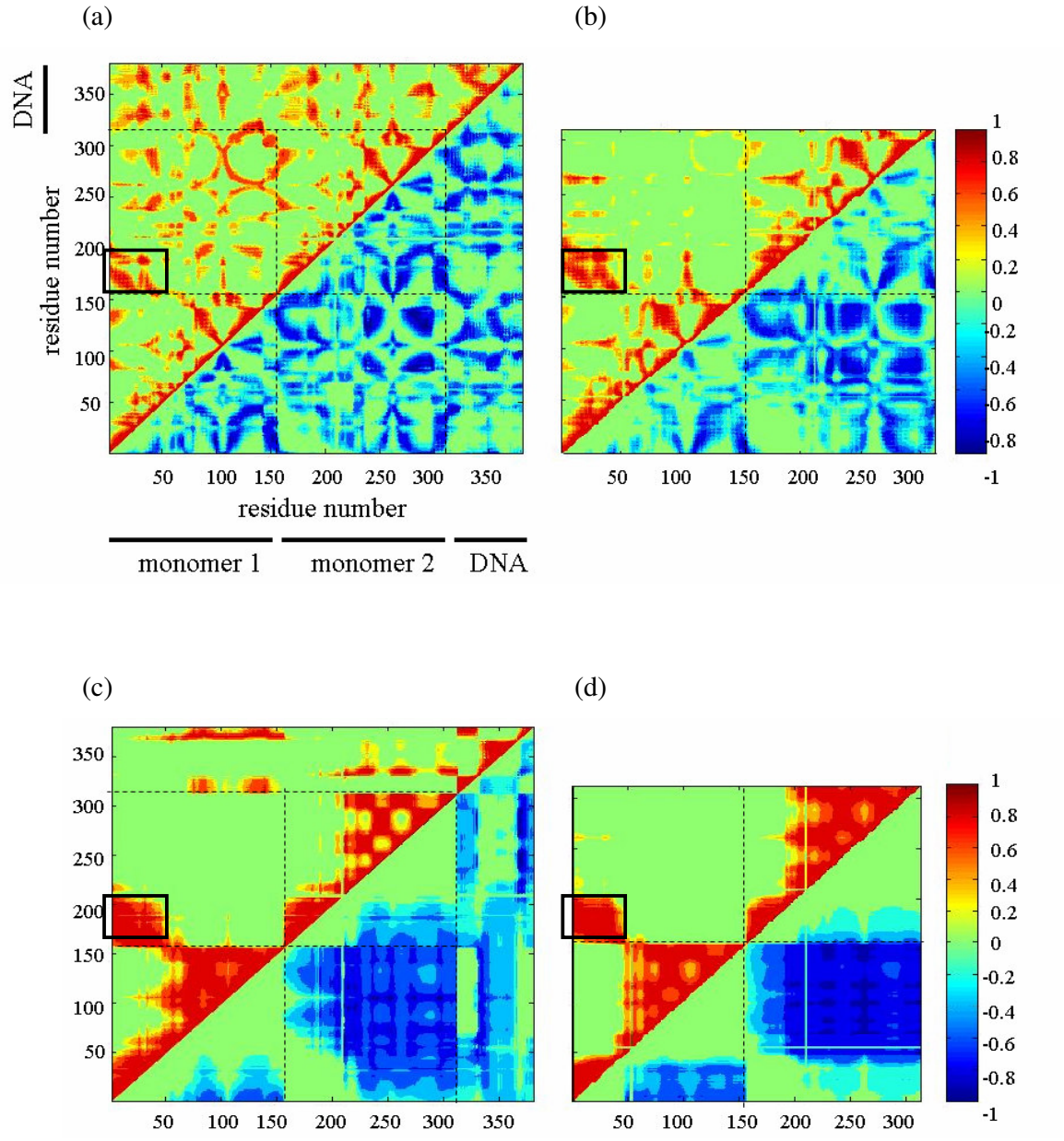


Figure 4.20. Orientational cross-correlations between residue fluctuations of PvuII (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

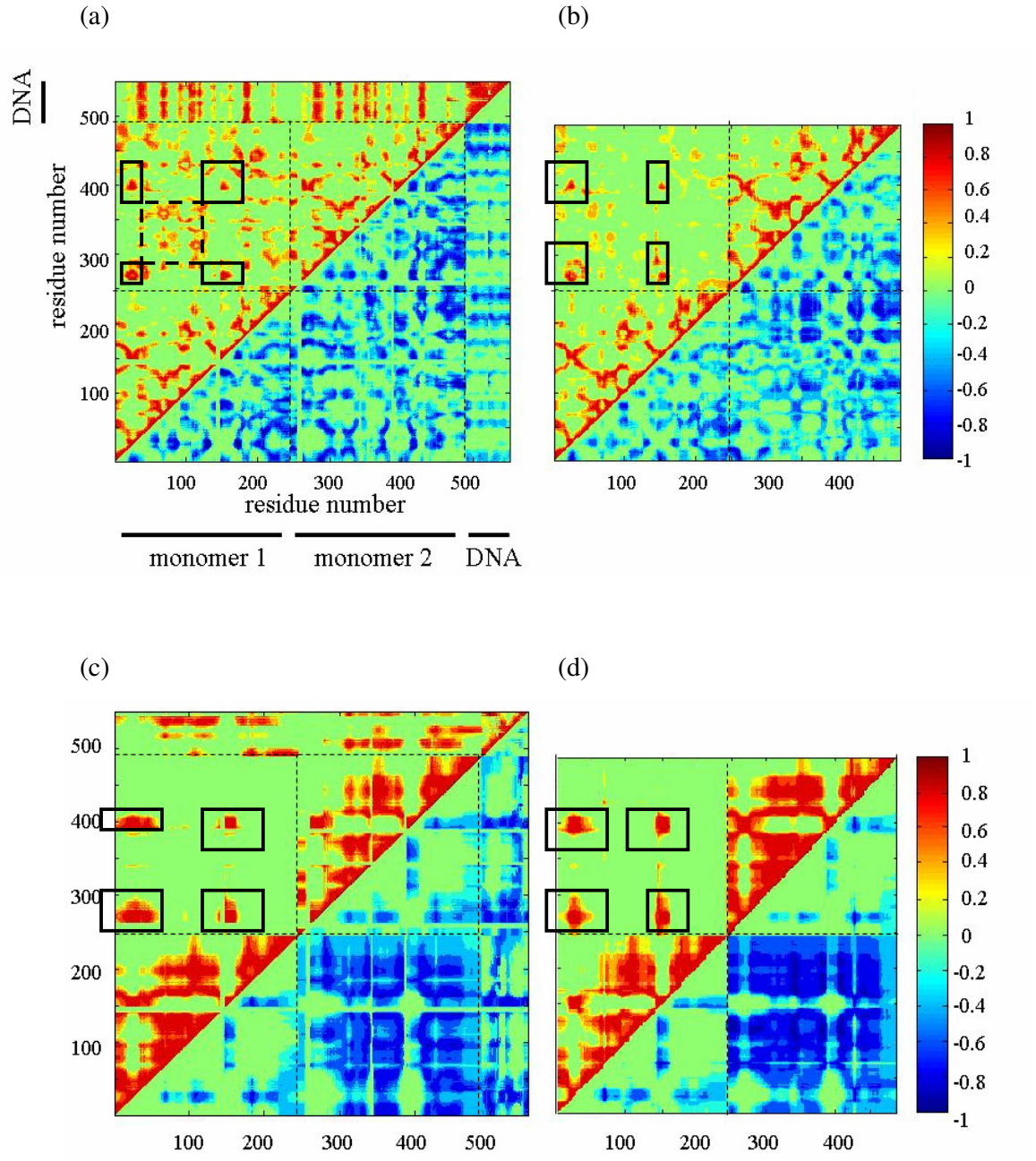


Figure 4.21. Orientational cross-correlations between residue fluctuations of EcoRV (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

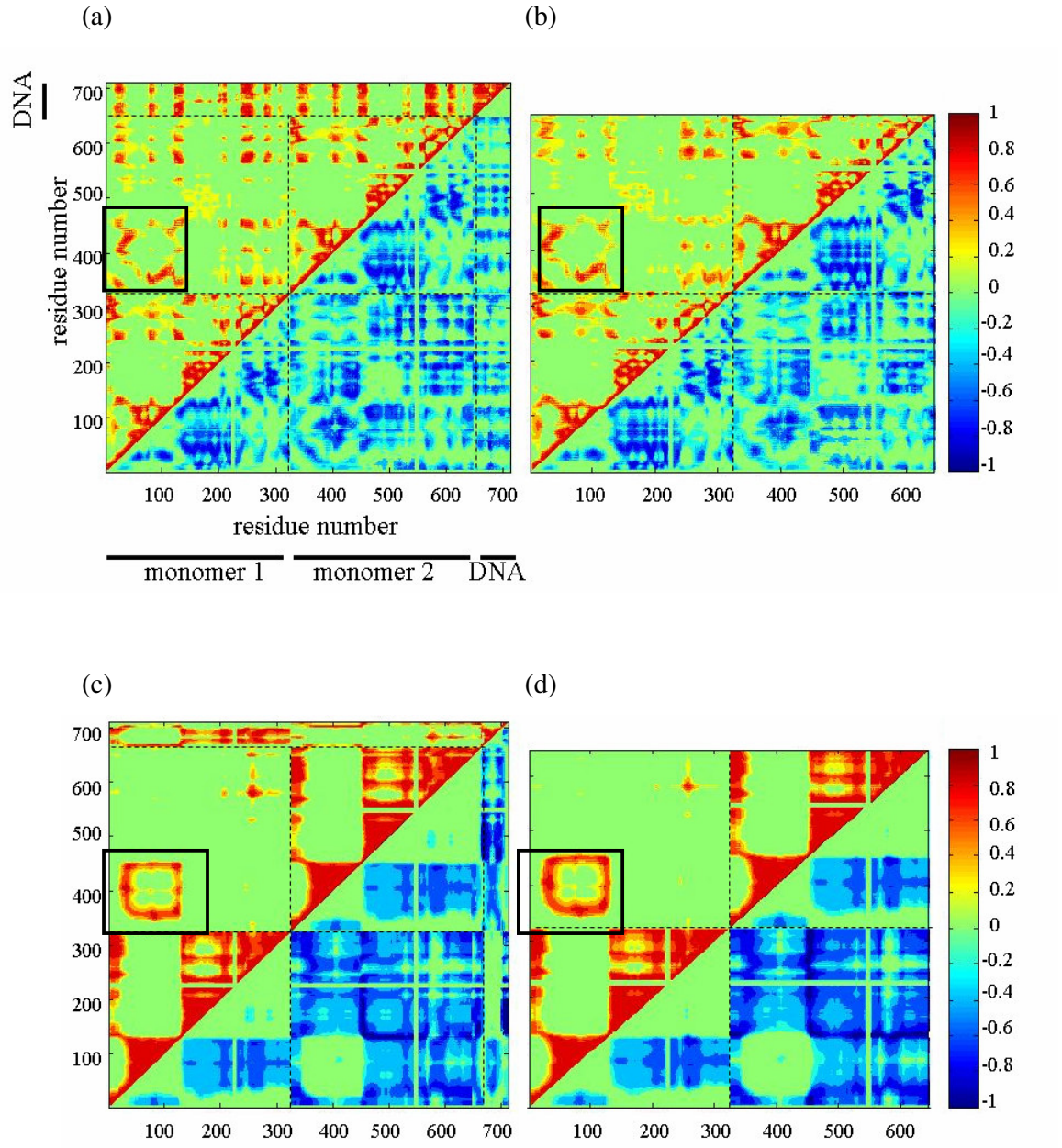


Figure 4.22. Orientational cross-correlations between residue fluctuations of BsoBI (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

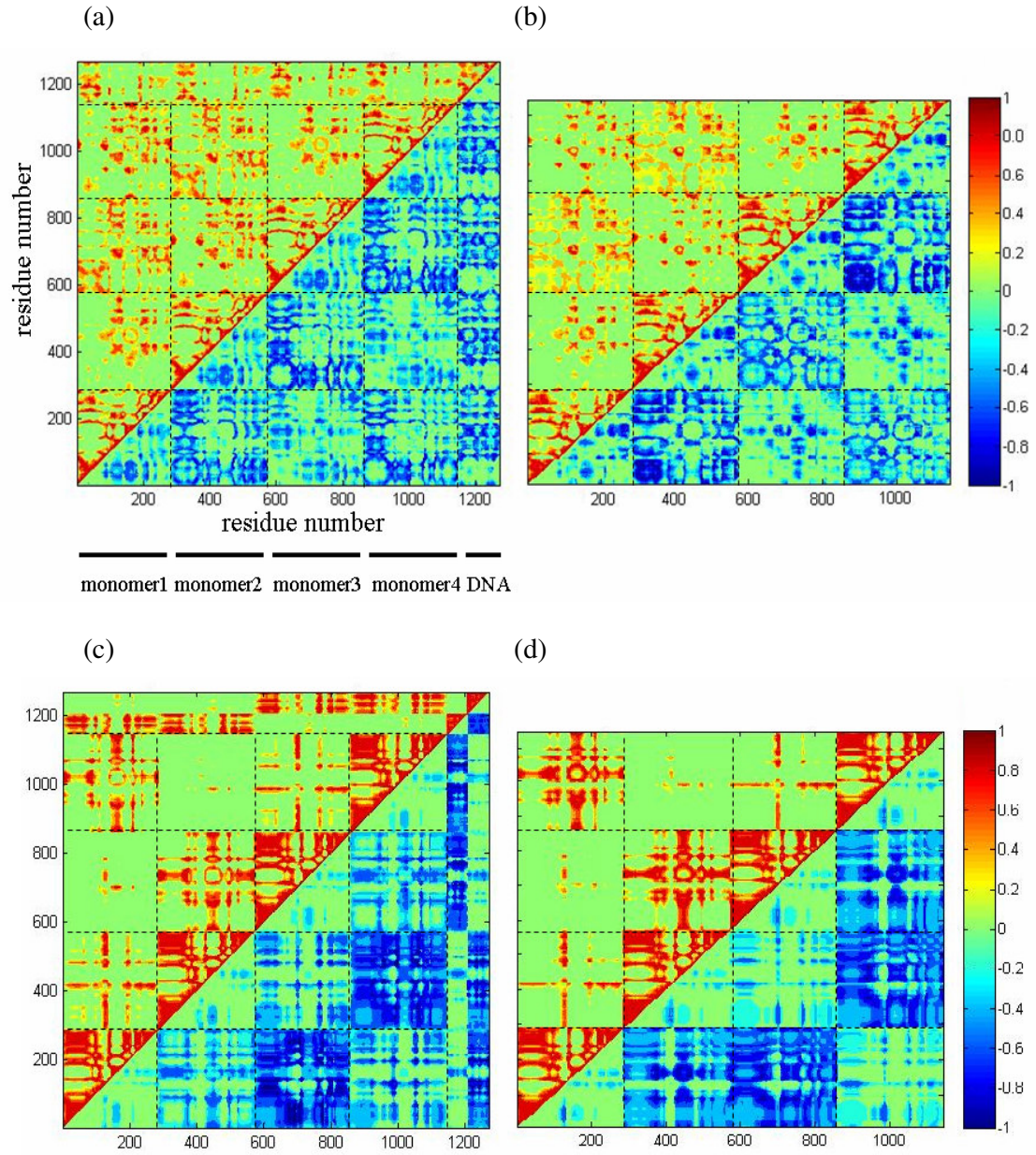


Figure 4.23. Orientational cross-correlations between residue fluctuations of NgoMIV (including first five modes) (a) DNA complex (ANM), (b) apo form (ANM), (c) DNA complex (GNM), and (d) apo form (GNM).

4.2.5. Comparison of Cognate and Non-cognate DNA Complexes

REs recognize their specific DNA sequences and cleave only at their recognition sites. However, they can bind to nonspecific DNA but they cannot cleave this non-cognate DNA. Various experimental and computational studies have been performed to understand this selectivity of REs, but recognition and cleaving mechanisms of cognate DNA are still obscure. For REs, a single base pair change in DNA sequence can reduce their cleavage activity by over a million-fold (Lesser *et al.*, 1990; Thielking *et al.*, 1990). In this study, crystal structures of BamHI and EcoRV complexed with cognate and non-cognate DNA are used to observe the differences in their harmonic dynamics. BamHI, complex with cognate DNA (2BAM.pdb), has GGATCC sequence, while it has GGATTC sequence in its complex with non-cognate DNA (1ESG.pdb) (see Figure 2.5). EcoRV recognizes and cleaves the sequence of GATATC in its complex with cognate DNA (1BGB.pdb), whereas it nonspecifically binds to the sequence of GAAUTC, a complex with non-cognate DNA (2B0E.pdb) (see Figure 2.8).

Equilibrium fluctuations and global dynamics of these REs are analyzed by ANM (r_{cut}: 13 Å) in order to see whether there is any difference in these properties for complexes with cognate and non-cognate DNA of BamHI and EcoRV.

The predicted B-factor values and collective deformations from ANM are separately compared for cognate and non-cognate DNA complexes of BamHI. It is observed that the predicted B-factor results, slow and fast modes are similar for the complexes (not shown). Also, the same results are obtained for EcoRV complexes with cognate and non-cognate DNA.

The level of agreement between the collective motions of cognate and non-cognate DNA complexes is calculated using overlap definition in Equation 4.1 (Amadei *et al.*, 1993; Hess, 2000).

$$u_{jk}^{AB} = \mathbf{u}_j^A \cdot \mathbf{u}_k^B \tag{4.1}$$

where A and B indicate the different eigenvector sets from ANM calculations, and subscripts j and k indicate the slow mode indices of each eigenvector. The limits of u_{AB} are between 0 and 1 that indicate random directions and perfect overlap, respectively.

In Figure 4.24.(a), overlap matrix of EcoRV cognate and non-cognate DNA complexes for first 10 modes is displayed. Slowest modes are in conformity with each other for EcoRV complexes (with the overlap value ~0.98) as expected because root mean-square deviation (RMSD) value between EcoRV-DNA complexes is 0.68 Å. For BamHI cognate and non-cognate DNA complexes, overlap matrix is displayed in Figure 4.24.(b). First modes are in well conformity (with the overlap value ~0.82) and also the second mode of cognate DNA complex is in conformity with the third mode of non-cognate DNA complex of BamHI (with the overlap value ~0.75) although the RMSD value between these structures is 2.57 Å. The first and second modes giving sliding and opening up of DNA binding site, respectively, are observed both in cognate and non-cognate complexes. It is important that we detect differences between cognate and non-cognate complexes in these two functionally important modes.

(a) (b)

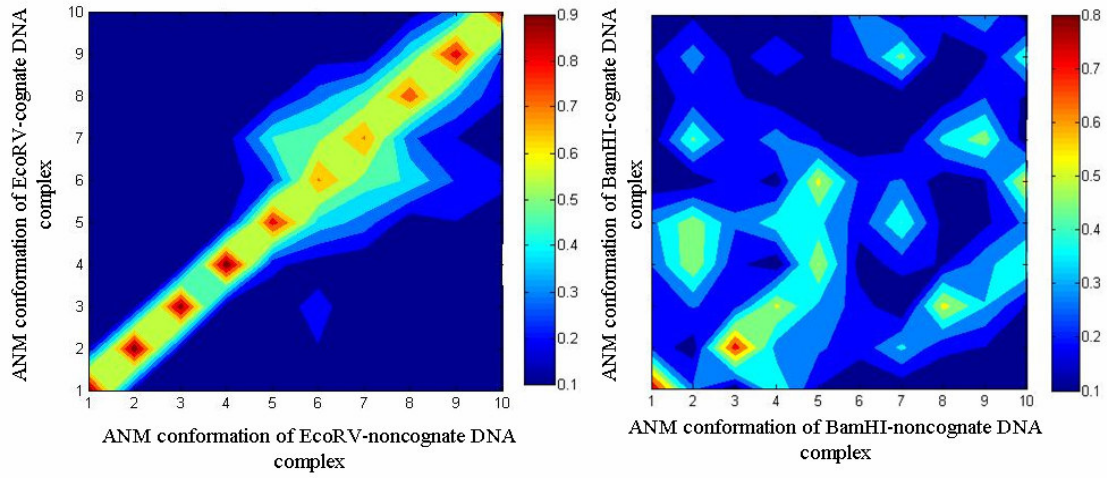


Figure 4.24. Overlap matrix for the first10 eigenvectors from cognate and noncognate DNA complexes (a) of EcoRV, and (b) of BamHI

Cross correlations for slowest first 5 modes, the second mode, and the first mode are calculated for cognate and non-cognate complexes of EcoRV and also BamHI. The cross-correlation matrix elements of non-cognate complex for all residue pairs (i,j) are subtracted from the corresponding (i,j) elements for the cognate complex. The resulting differences are plotted in matrix form. The difference values vary in the range [-2, 2] in these cross correlation difference plots. The lower limit indicates that two residues are fully anticorrelated (~ -1) in cognate complex, while they are fully correlated (~1) in non-

cognate complex. And the opposite situation is valid for the upper limit. In all plots red and dark blue regions indicate large differences, meaning that orientational cross-correlations reverse directions.

Cross-correlation difference plots for first 5 modes, second mode and first mode for EcoRV complexes are displayed. From first five modes difference plot (Figure 4.25), it is observed that there is no significant difference between EcoRV complexes with cognate and non-cognate DNA. However, looking at the first and second modes separately, some differences can be distinguished. For the second slowest mode, the significant differences between complexes are observed in both monomers of protein (labeled as I in core domain, II-IV symmetric residues, and III-V chain ends in Figure 4.26 and displayed in 3-D structure of EcoRV). When first mode cross-correlation values are compared between cognate and non-cognate complexes of EcoRV, it is observed that there are significant differences on DNA nucleotides (labeled as I-II residues at the top of monomer 1, and III-IV nucleotides at recognition site) (Figure 4.27). Observation of significant difference on DNA in only first mode supports that the first mode can be important for the recognition and cleavage of cognate DNA.

In fact, protein-DNA interactions may be more clearly observed by fully atomistic calculations for complexes with cognate and non-cognate DNA. From atomistic ANM calculations of EcoRV (r_{cut}= 9 Å), a similar difference plot to Figure 4.25 is obtained for first five slowest mode difference (not shown). For the second slowest mode, the significant differences between complexes are observed in both monomers of protein (labeled as I-II-V in core domain, III-VI-VII-VIII residues close to DNA binding region, and IV chain end in Figure 4.28 and also displayed in 3-D structure of EcoRV). These regions are similar to Figure 4.26. From Figure 4.29, it can be seen that significant difference on DNA nucleotides for complexes with cognate and non-cognate DNA is obtained again in first mode (labeled as II-V residues at the top of monomer 1, I-IV residues in core domain, III in Q-loop, and VI nucleotides at recognition site in the figure and also displayed in 3-D structure of EcoRV). Results of cross-correlation difference for coarse-grained and full atom calculations point to similar regions. However, the importance of side chain interactions between DNA and amino acids point to the first mode in terms of correlation differences in the recognition site.

In Figure 4.30, including the first five modes of BamHI complexes, differences between complexes can be seen only for the chain end residues. For the second slowest mode, the significant differences between complexes are observed in both monomers of protein (labeled as I-IV residues at α1-2, II-VI residues close to DNA binding region, and III-V chain end residues in Figure 4.31 and also displayed in 3-D structure of BamHI). Each region labeled as III and VI has different cross-correlation values in complexes. This result is expected because these regions are unstructured in non-cognate DNA complex of BamHI. For the first mode, significant differences can be clearly observed on DNA nucleotides (labeled as I-V residues at α2, II-III-VI-VII residues in core domain, IV-VIII chain end residues, and IX nucleotides at recognition site)(Figure 4.32). Thus, first mode can have importance for the function of BamHI-cognate DNA complex.

In summary, differences between cognate and non-cognate complexes are revealed by both coarse-grained and fully atomistic ANM calculations. Although cognate and non-cognate DNA complexes have similar scissor-like and twisting motions in their first and second modes, respectively, large differences in the orientational cross-correlations are observed in residues specifically at the recognition site in the first mode. Also, in the second mode, some important differences are obtained on residues near to DNA binding region. The existance of the same first and second modes in non-specifically bound complexes indicate that these modes, as suggested before, could be related to the sliding and opening of binding site, respectively.

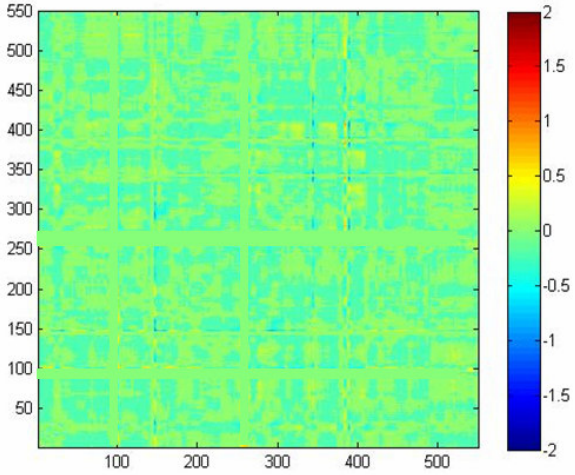


Figure 4.25. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of EcoRV (including first five slowest modes)

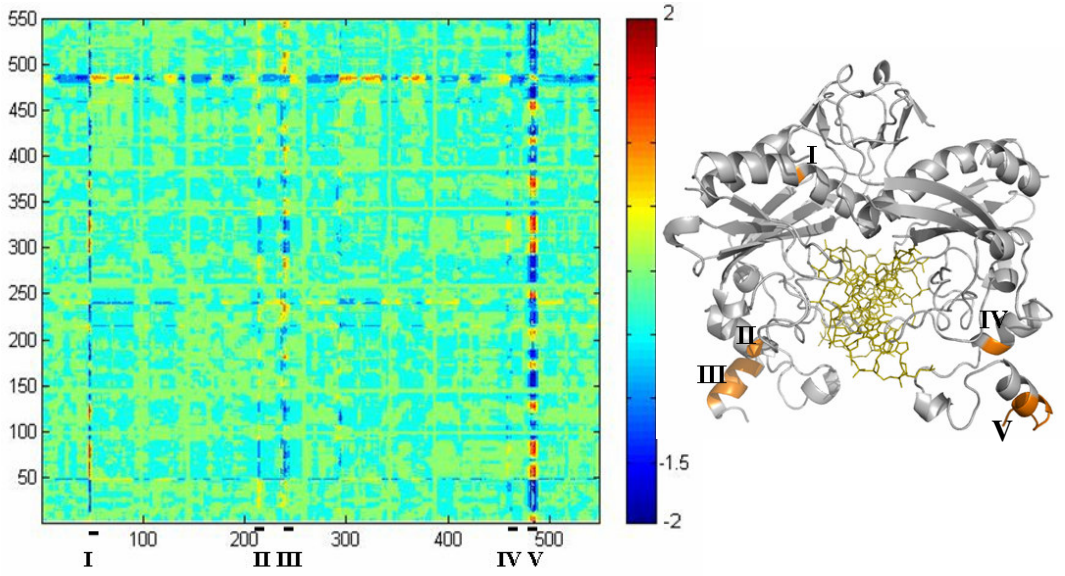


Figure 4.26. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of EcoRV (including second slowest mode)

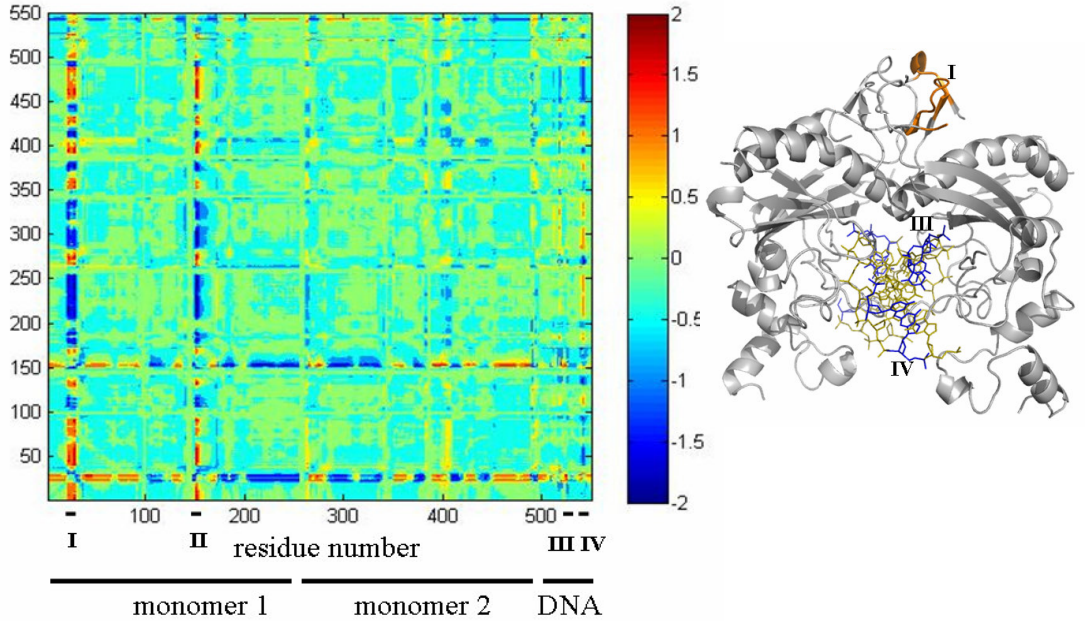


Figure 4.27. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of EcoRV (including first slowest mode)

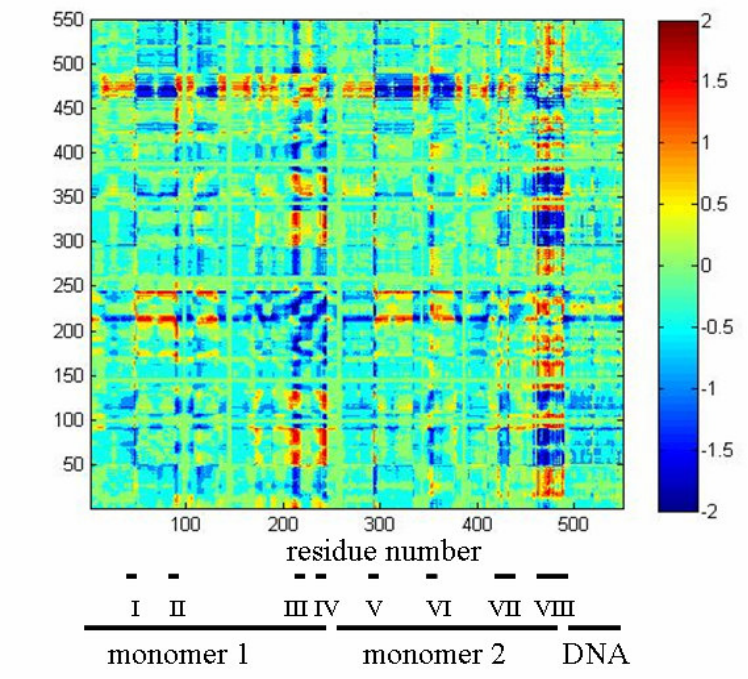


Figure 4.28. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of EcoRV from full atom run (including second slowest mode)

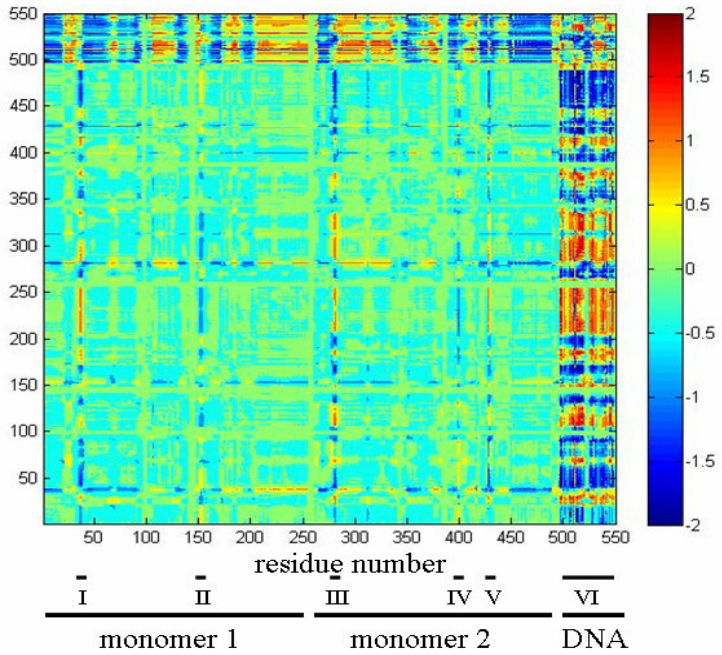


Figure 4.29. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of EcoRV from full atom run (including first slowest mode)

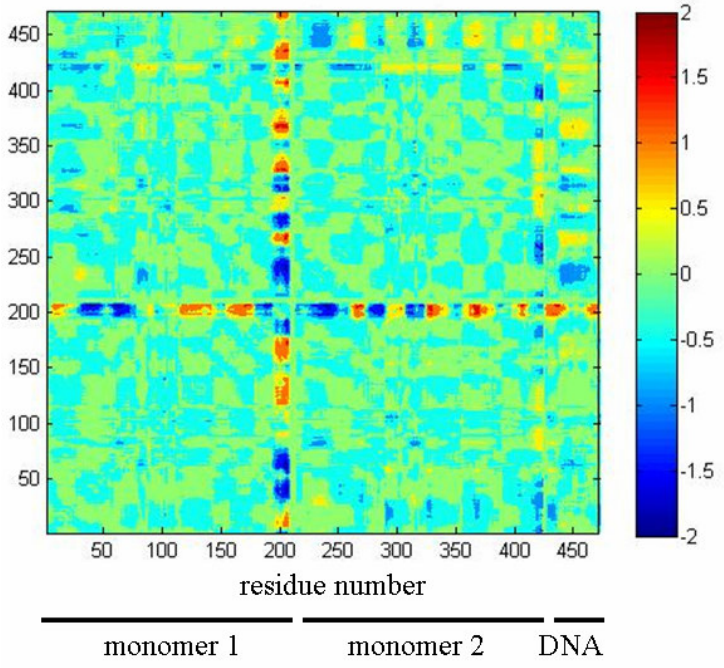


Figure 4.30. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of BamHI (including first five slowest modes)

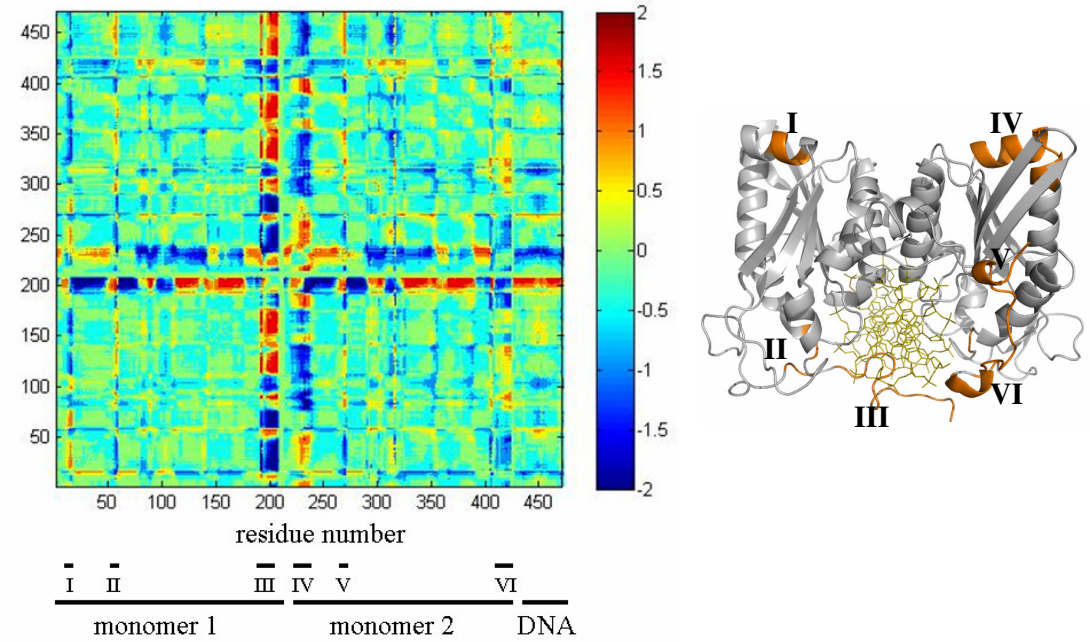


Figure 4.31. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of BamHI (including second slowest mode)

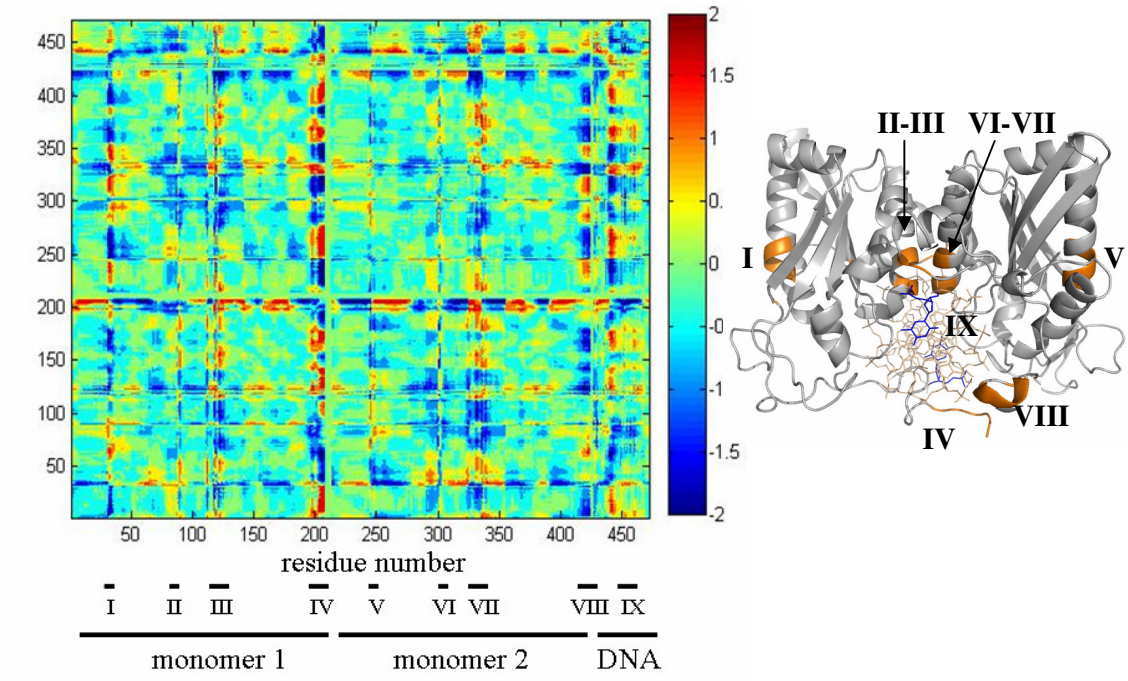


Figure 4.32. Difference plot of cross-correlation values for cognate and non-cognate DNA complexes of BamHI (including first slowest mode)

4.3. Fast Modes: Hot Spots

If a residue when mutated to alanine lead to an increase in the absolute binding energy by more than 2 kcal/mol, it can be defined as a "hot spot" (Bogan and Thorn, 1998; Clackson and Wells, 1995). In elastic network models, the residues with high ms fluctuations in the fast modes are also termed as hot spot residues (Bahar *et al.*, 1998; Haliloglu *et al.*, 2005). These highly fluctuating residues in the fast modes, which are important for the stability of the protein (Bahar *et al.*, 1998), correlate with the conserved residues and also they are important for mutation studies (Hu *et al.*, 2000; Ma *et al.*, 2003; Keskin *et al.*, 2004; Haliloglu *et al.*, 2005).

To identify the residues that contribute to the stability of the RE models, the high-frequency modes obtained from ANM (r_{cut}= 13 Å) and GNM (r_{cut}= 7 Å) calculations are analyzed after normalization of the areas under the curves. Different number of cumulative fast modes are compared for ANM and GNM in Figure 4.33.(a)-(b), respectively. Comparison of fast mode shapes for cumulative 5, 10, and 20 modes do not indicate

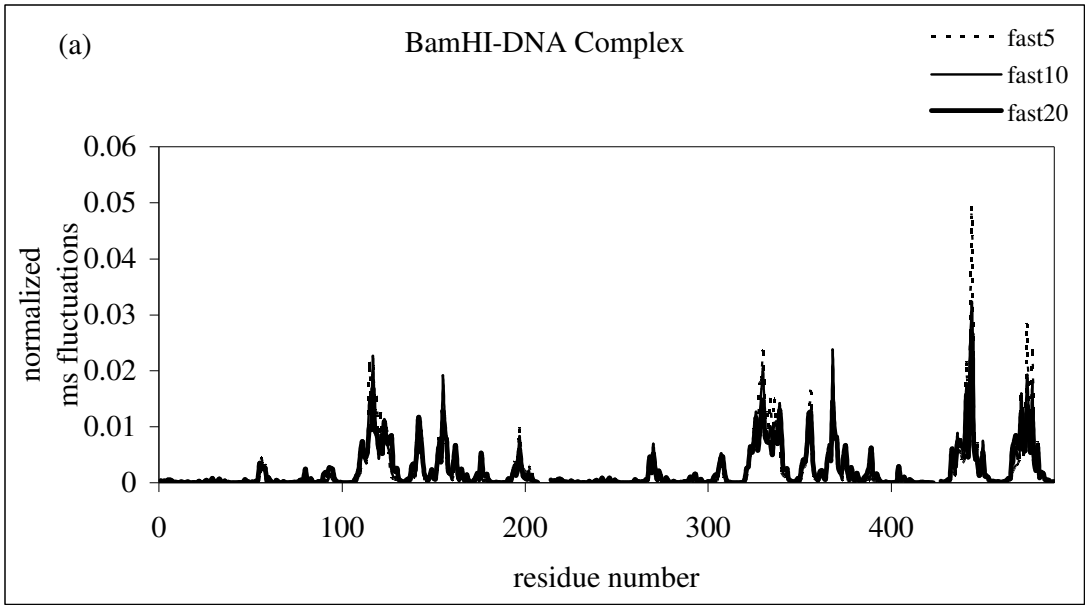
significant differences in the peaks. As a result, the hot spot residues are defined according to cumulative 20 high-frequency modes. Similar to the work of Haliloglu *et al.* (Haliloglu *et al.*, 2005), high amplitude ms fluctuations, which correspond to high frequency local vibrations of hot spot residues, are selected as nodes with a normalized ms fluctuation above 0.005 $Å^2$. For NgoMIV, EcoRI and PvuII cases, threshold value is taken as 0.002, 0.004, and 0.008 $Å^2$. On average, the number of hot spots in peaks above the threshold should not be more than 15% of the protein size for all the structures.

For BamHI-DNA complex and apo form, cumulative effect of 20 high-frequency modes are normalized and compared for GNM (in magenta) and ANM (in green) in the Figure 4.34.(a)-(b). As it can be seen from the figure, the GNM and ANM results are not in close conformity. Similar results are observed for the complex, apo, and DNA removed forms (if apo form is not available) of other RE models. Cumulative fast modes ms fluctuations of these REs are given for ANM and GNM calculations in Appendix Figures 6.1. to 6.5.

The choice of the cutoff radius may have an effect on the fast modes specifically because of the presence of DNA in complex or the DNA binding cavity in apo form. For this reason, ANM and GNM calculations are repeated for new sets of cutoffs to see whether the discrepancy between ANM and GNM results could be corrected. In ANM calculations, the cutoff value is increased to 17 Å and in GNM to 10 Å as suggested by Sen *et al.* (Sen *et al.*, 2006). In Figure 4.35, ANM results are compared for r_{cut}= 13 Å and 17 Å and GNM calculations are compared for r_{cut}= 7 Å and 10 Å. Increasing the cutoff value in the construction of the connectivity matrix does not change the fast modes for ANM calculations as clearly seen in Figure 4.35.(a) for BamHI-DNA complex. However, in GNM calculations new high frequency peaks emerge with increasing cutoff value (Figure 4.35.(b)). And the results point that hot spots observed from GNM (r_{cut}: 10 Å) are in more conformity with the ANM (r_{cut}: 13 Å) calculations (Figure 4.35.(c)).

For better visualization, the hot spot residues are shown as spheres on native structure of BamHI-DNA complex for each elastic network model. In Figure 4.36.(a), the hot spot residues obtained from ANM (r_{cut}= 13 Å) are shown by green spheres.

They form an inner core at DNA binding region. For cutoff radius of 7Å in GNM, the hot spot residues (magenta) are scattered away from the core domain and DNA (Figure 4.36.(b)).



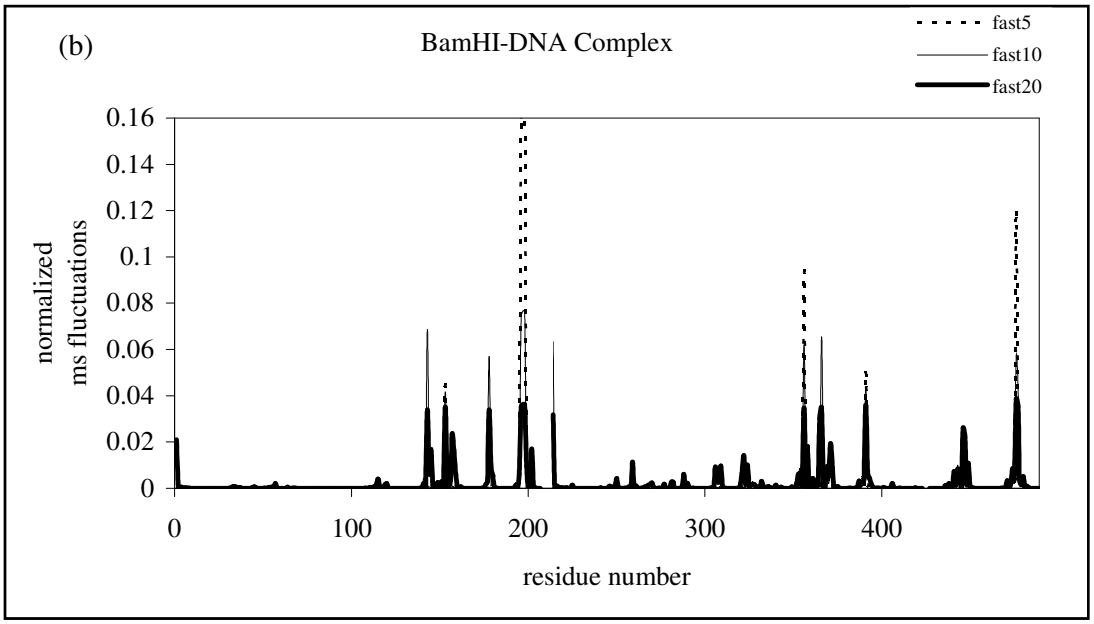


Figure 4.33. Comparison of different cumulative fast modes (a) ANM (r_{cut}= 13 Å), and (b) GNM (r_{cut}= 7 Å)

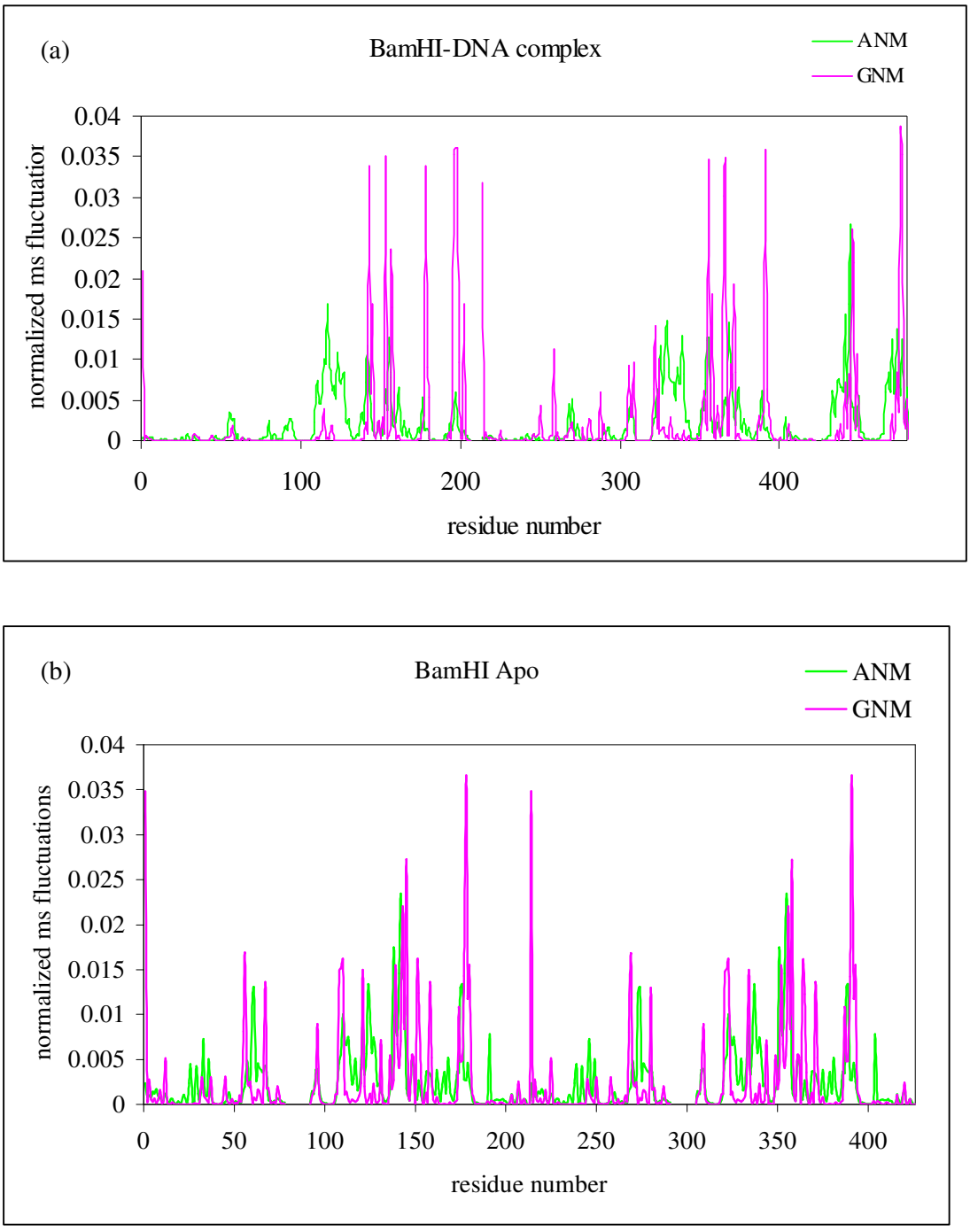


Figure 4.34. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}= 13 Å) and GNM (r_{cut}= 7 Å) (a) for BamHI-DNA complex, and (b) BamHI apo form

Hot spot residues obtained from GNM with r_{cut}= 10 Å are closer to DNA binding region as in the case of ANM (Figure 4.36.(c)). Similarities of ANM and GNM calculations in terms of hot spot residues are given in Figure 4.36.(d)-(e) looking from a side view. Green spheres indicate the hot spot residues obtained from ANM (r_{cut}= 13 Å), while dark blue spheres indicate the hot spot residues obtained from GNM (r_{cut}= 10 Å), that coincide with ANM calculations. Same representations for comparison of ANM and GNM results are given for other RE-DNA complex models in Figure 4.37. Again, the hot spot residues form an inner core at DNA binding region. For NgoMIV-DNA complex, only ANM results are given in Figure 4.37.

Number and percentage of hot spots for the RE models are tabulated with respect to cutoff value and network model in Table 4.1. Also, for the recognition sites on DNA, hot spots are determined and tabulated for each elastic network model in Table 4.2. For NgoMIV-DNA complex and its apo form, GNM (r_{cut}= 10 Å) calculations could not be performed, therefore hot spot data of these structures could not be given for 10 Å cutoff value in Table 4.1 and 4.2. ANM gives more hot spots with the same threshold value. And hot spots at recognition site are revealed by ANM. Increasing cutoff value in GNM leads to more hot spots at recognition site.

In summary, increasing the cutoff value affects the high-frequency modes in GNM calculations. Looking at Figure 4.37 both in ANM (r_{cut}= 13 Å) and GNM (r_{cut}= 10 Å) calculations, it is important that the hot spot residues form an inner core that can act as a pathway for their communications between the DNA cleavage sites, because it is known that the both DNA sides should be cleaved at the same time for REs that used in this study.

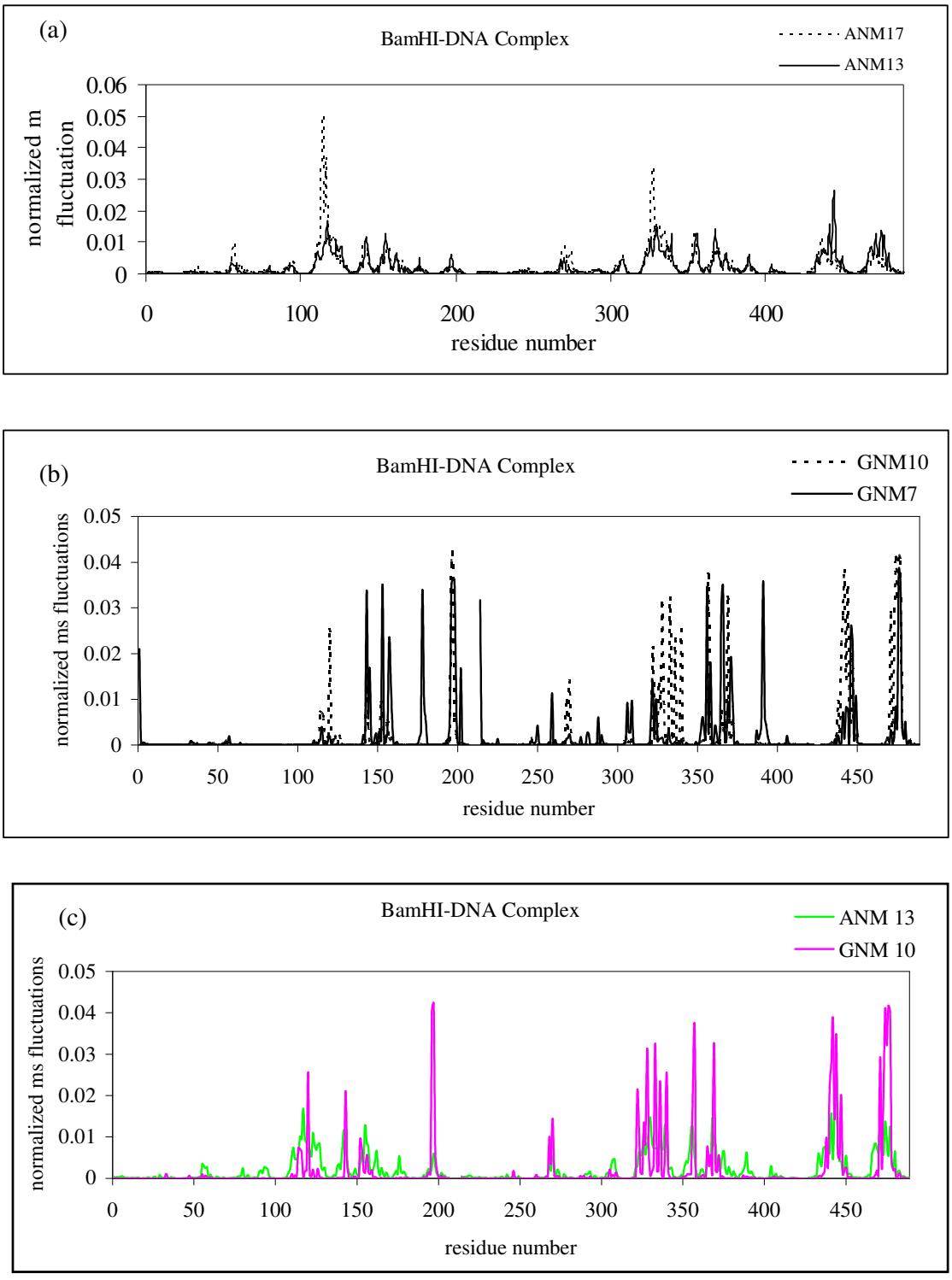


Figure 4.35. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes for different cutoff values of (a) ANM, (b) GNM for BamHI-DNA complex, and (c) Comparison of ANM (13 Å) and GNM (10 Å)

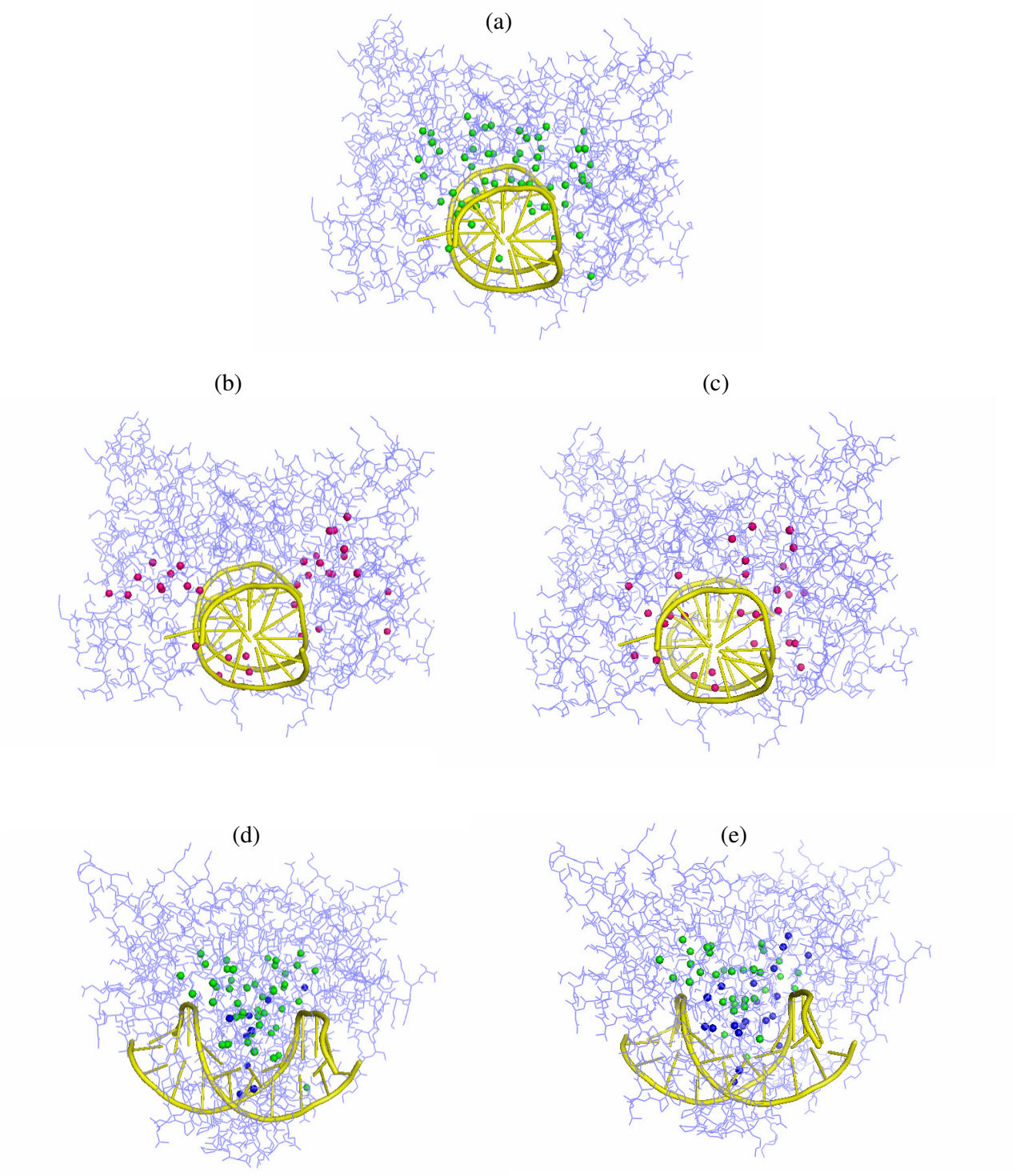


Figure 4.36. Hot spots of (a) ANM r_{cut}: 13 Å, (b) GNM r_{cut}: 7 Å, (c) GNM r_{cut}: 10 Å, (d) similarity of ANM r_{cut}: 13 Å and GNM r_{cut}: 7 Å, and (e) similarity of ANM r_{cut}: 13 Å and GNM r_{cut}: 10 Å

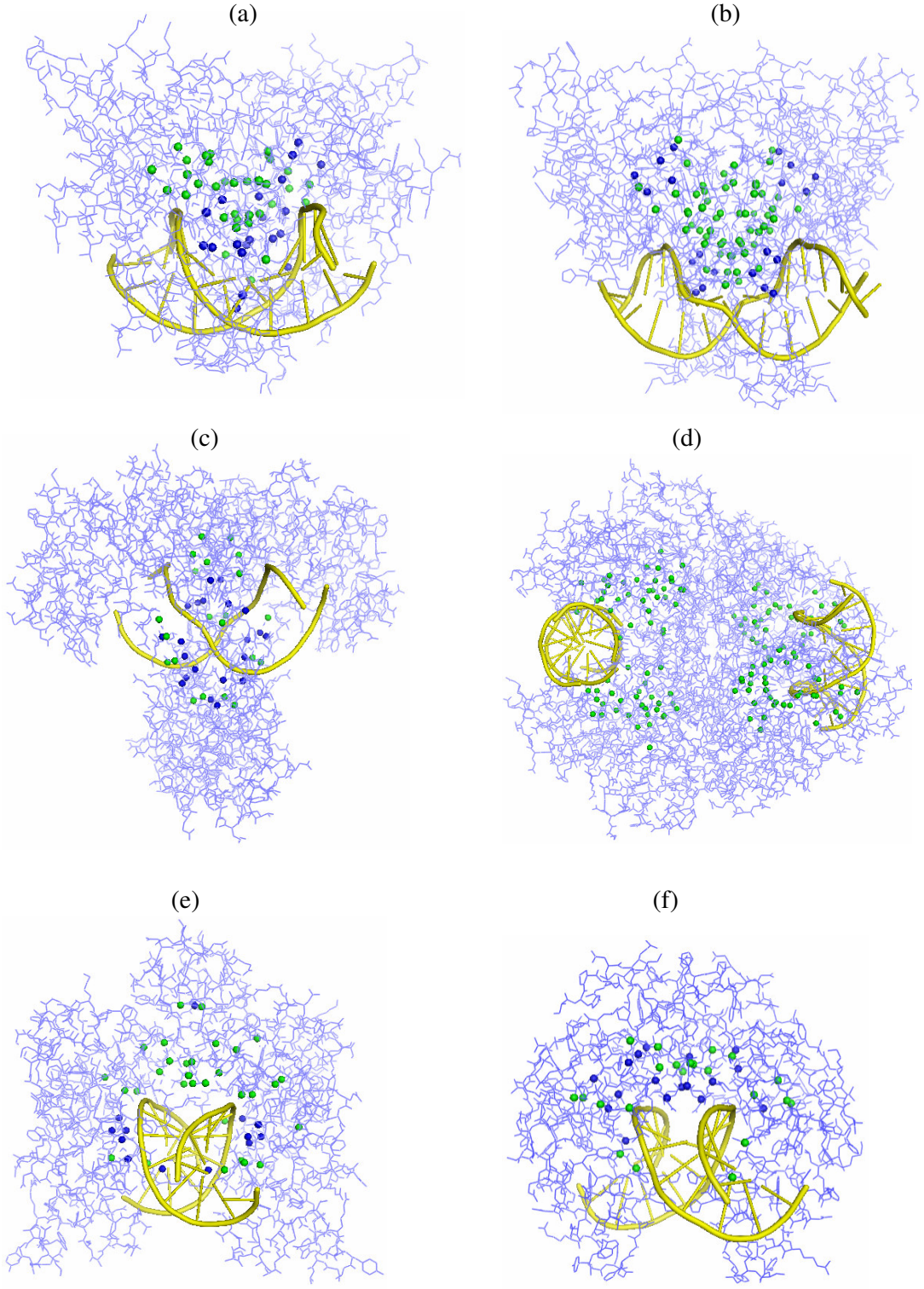


Figure 4.37. Hot spots for (a) BamHI-DNA complex, (b) EcoRI-DNA complex, (c) BsoBI-DNA complex, (d) NgoMIV-DNA complex, (e) EcoRV-DNA complex, and (f) PvuII-DNA complex

Table 4.1. Number and percentage of hot spots for the RE models

Enzyme	Total number of nodes	Number / Percentage of hot spots		
		ANM 13 Å	GNM 7 Å	GNM 10 Å
BamHI-DNA	480	79 / 16.5	47 / 9.8	43 / 9
EcoRI-DNA	594	78 / 13.1	40 / 6.7	32 / 5.3
BsoBI-DNA	687	71 / 10.3	42 / 6.1	50 / 7.3
NgoMIV-DNA	1270	177 / 14	71 / 5.6	-
EcoRV-DNA	465	70 / 15	45 / 9.7	47 / 10.1
PvuII-DNA	378	51 / 13.5	32 / 8.5	35 / 9.3
BamHI apo	400	56 / 14	58 / 14.5	34 / 8.5
EcoRI apo	506	68 / 13.4	55 / 10.9	36 / 7.1
BsoBI DNA removed	621	57 / 9.2	44 / 7.1	32 / 5.2
NgoMIV DNA removed	1144	99 / 8.7	71 / 6.2	-
EcoRV apo	488	62 / 12.7	50 / 10.2	37 / 7.6
PvuII apo	308	42 / 13.6	33 / 10.7	29 / 9.4

Table 4.2. Hot spots on recognition sites for the RE models

Enzyme	Hot spots on recognition site		
	ANM 13 Å	GNM 7 Å	GNM 10 Å
BamHI-DNA	**G/GATCC**	G/G**ATCC**	G/**GATC**C
	G/GATCC	G/G**ATC**C	G/**GAT**CC
EcoRI-DNA	**G/AA**TTC	G/AA**T**TC	G/A**A**TTC
	G/AATTC	G/AA**T**TC	G/A**A**TTC
BsoBI-DNA	**C/TCGAG**	C/**TCG**AG	C/**TCGA**G
	C/TCGAG	C/**TCG**AG	C/**TCGA**G
NgoMIV-DNA	G/**CCGG**C	G/**CC**GGC	-
	G/**CCGG**C	G/**CC**GGC	-
	G/**CCGG**C	G/**CC**GGC	-
	G/CCGGC	G/**CC**GGC	-
EcoRV-DNA	G**AT/ATC**	GAT/ATC	**GAT/ATC**
	G**AT/ATC**	GA**T**/ATC	**GAT/ATC**
PvuII-DNA	C**AG/CTG**	CAG/C**TG**	CA**G/CTG**
	C**AG/CTG**	CAG/C**T**G	CA**G/CTG**

4.4. Comparison of Restriction Endonucleases with Another Dimeric Enzyme: TIM

Results in section 4.3 point that the cutoff radius may affect the fast modes of REs. In order to understand whether this result is specific to RE-DNA complexes or not, calculations are performed for another homo-dimeric enzyme triosephosphate isomerase (TIM).

The crystal structure of TIM (PDB code: 1NEY.pdb) at 1.20 Å resolution (Jogl *et al.*, 2003) is used. TIM is active as a dimer and each identical monomer has 248 residues. Conserved and interface residues (Wierenga *et al.*, 1991) of TIM are displayed on its 3-D structure in Figure 4.38.

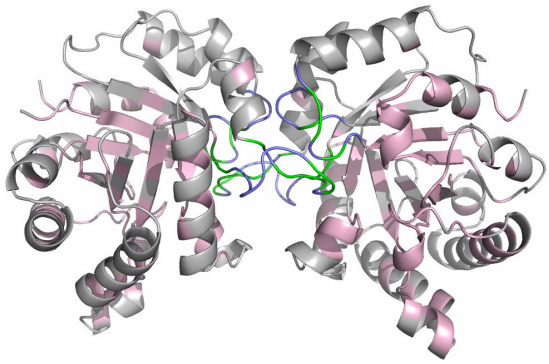


Figure 4.38. Conserved and interface residues of TIM (conserved: green and pink; interface: green and blue)

The cumulative ms fluctuations averaged over different number of fast modes of TIM are compared for GNM (r_{cut}= 7 Å) in Figure 4.39. Same peaks are observed for all cases and only the ms fluctuation values increase with the decreasing cumulative highest mode. Cumulative ms fluctuations including the 20 highest modes for TIM is shown in Figure 4.40.(a) for both ANM r_{cut}= 13 Å (solid line) and r_{cut}= 17 Å (dashed line). It is observed that the results are in conformity for the different cutoff values of ANM. Also, interface residues shown with the lines on the figure can be seen as peaks in both cutoff values of ANM calculations. In Figure 4.40.(b), cumulative ms fluctuations averaged over 20 highest modes obtained from GNM calculations are displayed for r_{cut}= 7 Å (solid line) and r_{cut}= 10 Å (dashed line). Some interface residues can not be seen as peaks for GNM (r_{cut}= 7 Å) emerge as peaks for GNM r_{cut}= 10 Å. In Figure 4.41, ANM and GNM results are compared. ANM (r_{cut}= 13 Å) results are in conformity with GNM (r_{cut}= 7 Å) results.

Increasing the cutoff value in GNM calculations leads to even closer conformity in terms of hot spots of interface residues.

For better visualization, the hot spot residues obtained from the high frequency local vibrations of TIM are displayed as green spheres for ANM (r_{cut}= 13 Å), and magenta spheres for GNM (r_{cut}= 7 Å) and GNM (r_{cut}= 10 Å) calculations in Figure 4.42.(a-b-c), respectively. They are selected as nodes with a normalized ms fluctuation above 0.005 $Å^2$. Hot spot residues obtained from ANM (r_{cut}= 13 Å) calculation includes interface residues, whereas a few interface residues are included in GNM (r_{cut}= 7 Å) calculation. And more interface residues are revealed as hot spots with increasing cutoff value to 10 Å in GNM. Number of hot spots and their percentage values with respect to total structure for each elastic network model are tabulated in Table 4.3. Total number of interface residues (Wierenga *et al.*, 1991) is 54 as shown in Figure 4.38.

All methods are satisfactory in terms of predicting a high percentage of conserved residues. ANM and GNM predict similar hot spots for TIM. The cutoff radius in GNM is only important for the location of hot spot residues at the interface. The percentages of conserved residues predicted are also quite similar. In view of the results on TIM, we can state that the discrepancy between GNM and ANM in REs results from the presence of the DNA binding site at the dimeric interface, as a result of which hot spots near the binding site are not satisfactorily predicted by GNM (r_{cut}= 7 Å).

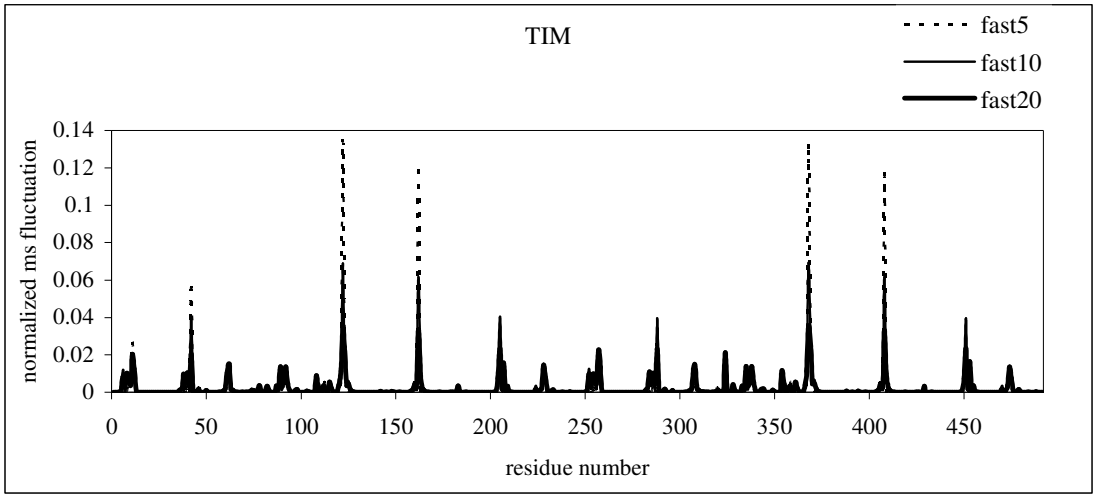


Figure 4.39. Comparison of cumulative ms fluctuations averaged over different highest modes of TIM obtained from GNM r_{cut}= 7 Å calculation

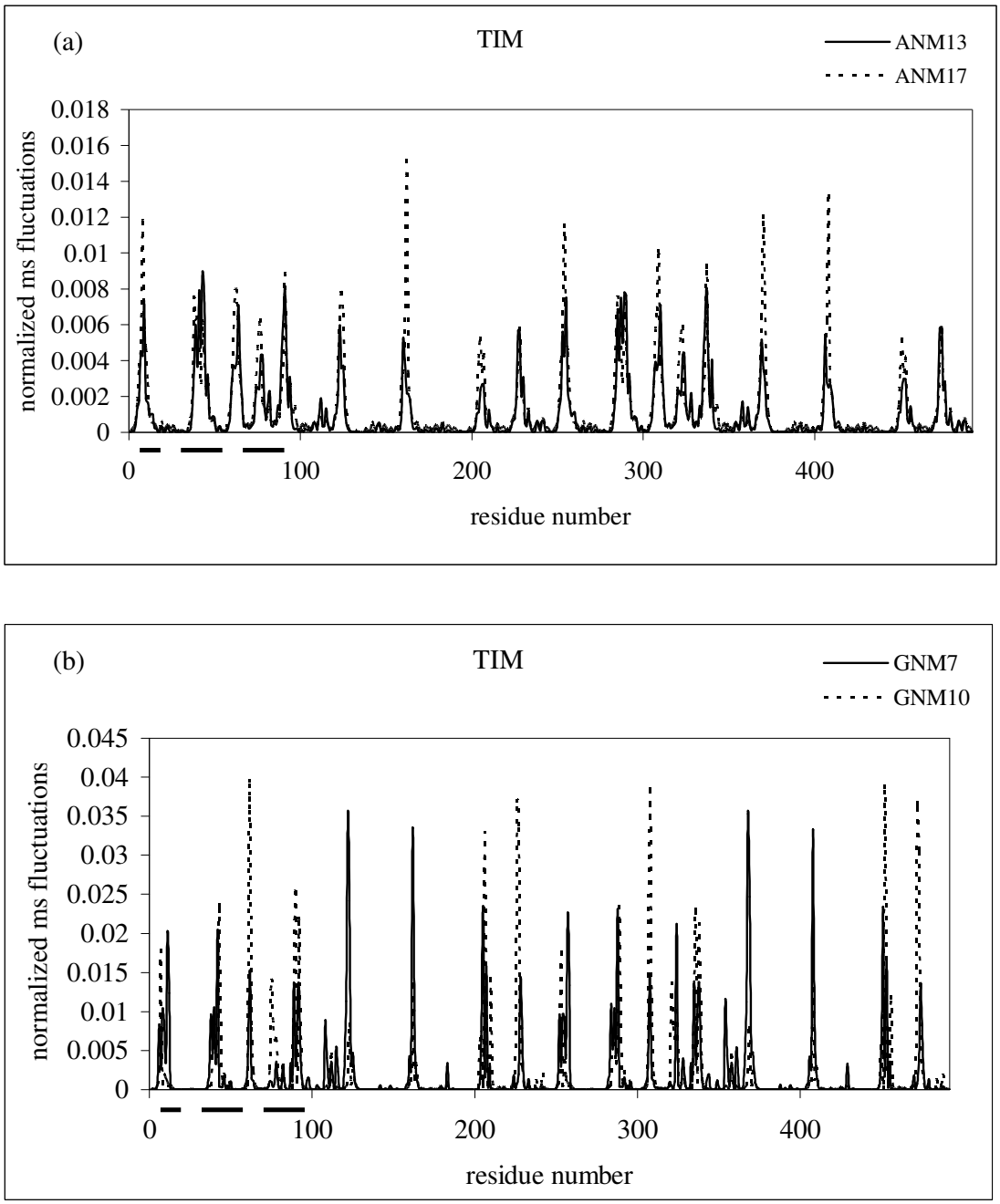


Figure 4.40. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes for different cutoff values of (a) ANM, and (b) GNM for TIM

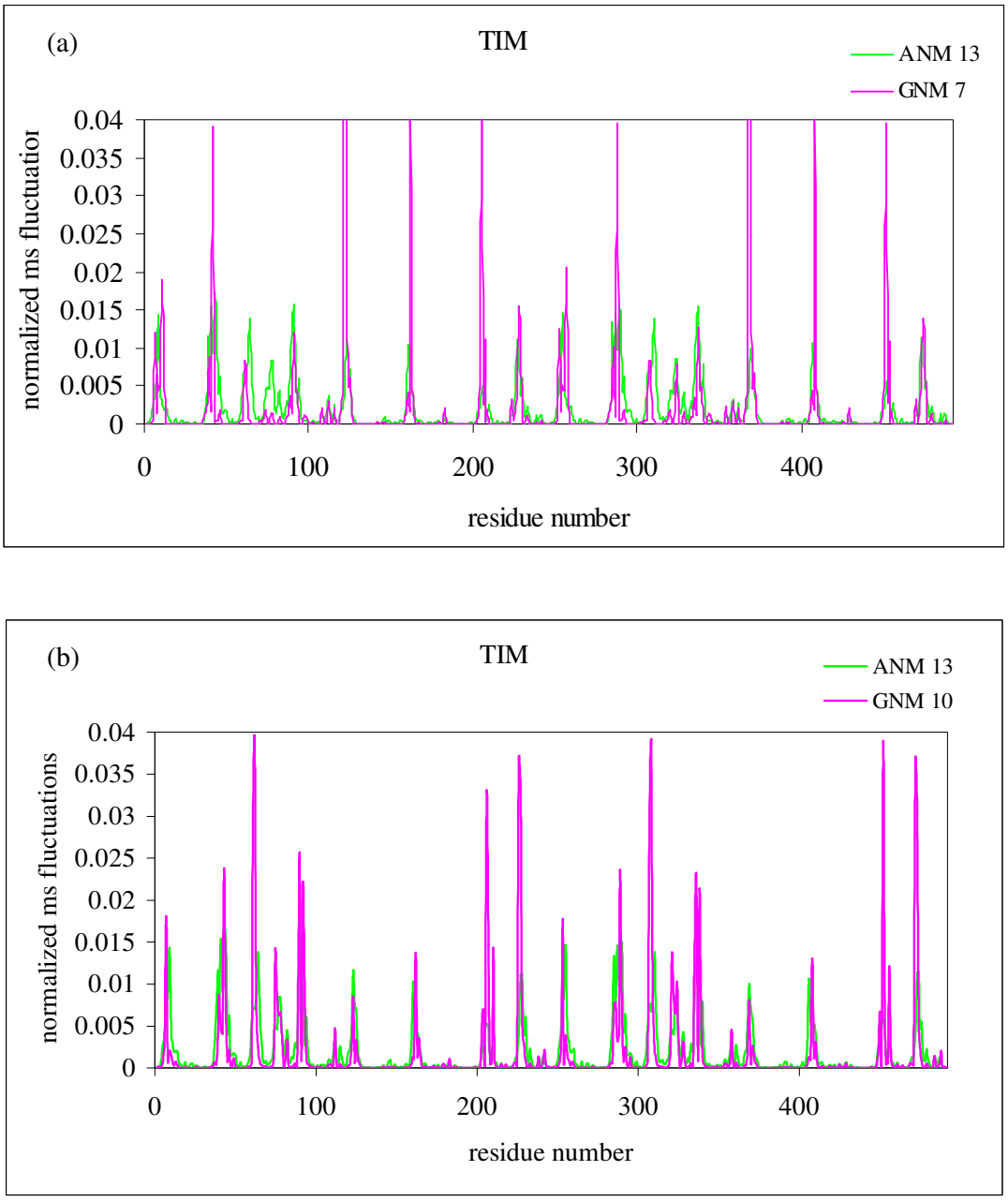


Figure 4.41. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes of TIM from (a) ANM (r_{cut}= 13 Å) and GNM (r_{cut}= 7 Å), and (b) ANM (r_{cut}= 13 Å) and GNM (r_{cut}= 10 Å)

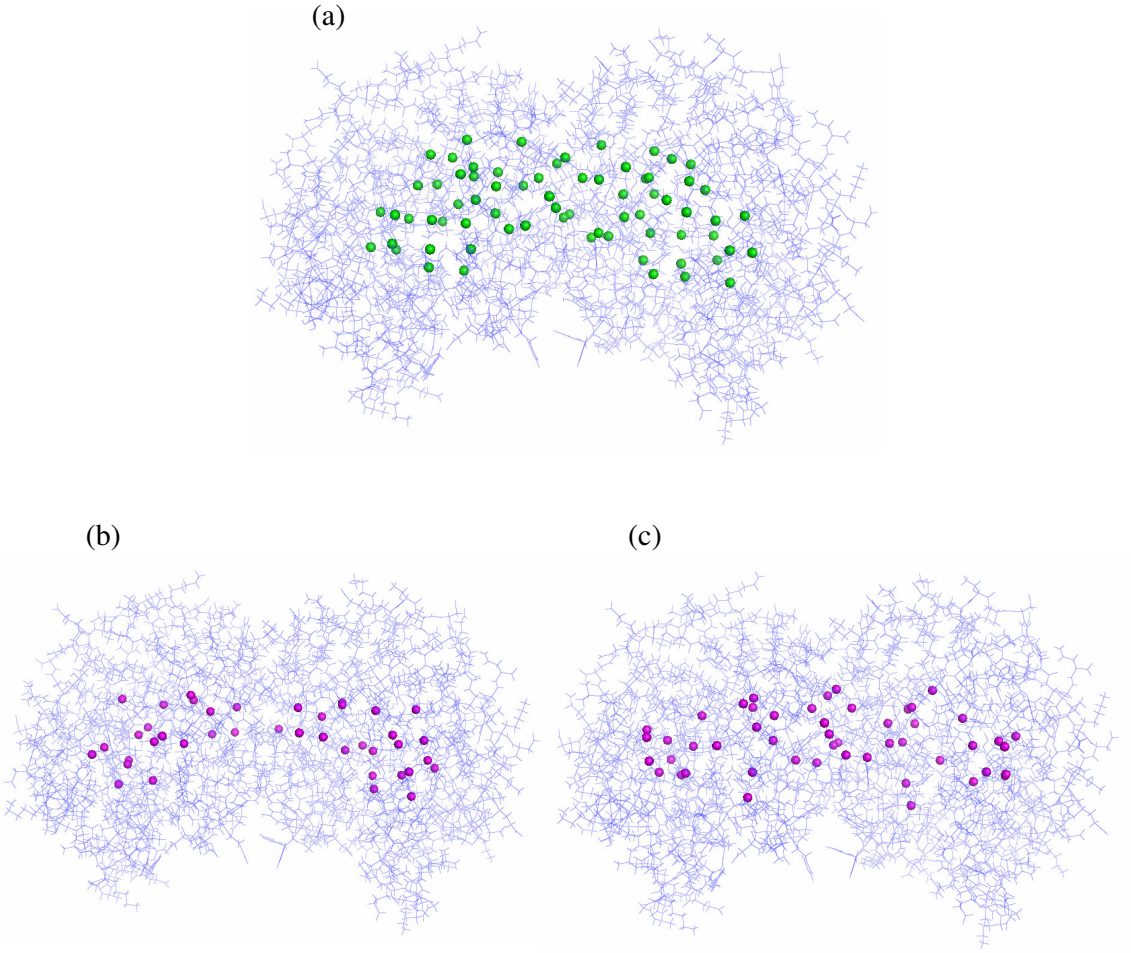


Figure 4.42. Hot spot residues of TIM (a) ANM r_{cut}= 13 Å, (b) GNM r_{cut}= 7 Å, and (c) GNM r_{cut}= 10 Å

Table 4.3. Number of hot spots for TIM

Model	Total number of hot spots	Number of hot spots at interface	Number of hot spots as conserved residues	Percent value of hot spots with respect to total structure
ANM 13 Å	70	12	41	14.1
GNM 7 Å	41	5	32	8.3
GNM 10 Å	50	10	34	10.1

When ANM and GNM results of TIM are compared in terms of orientational cross-correlations, the results are different for the correlations between two monomers (Figure 4.43.(a)-(c)). ANM (r_{cut}= 13 Å) correlation results for TIM are in well conformity with the correlations obtained from Molecular Dynamics (MD) study of TIM (Cansu and Doruker, 2007, (*accepted)*) (Figure 4.43.(b)). Positive inter-subunit correlations are dominant in ANM calculations. In contrast, GNM shows positive correlations only for some of key interface residues (71-76), as in the case in REs. However, these key interactions obtained by GNM are important in terms of predicting native interfaces (Kantarci *et al.*, 2006).

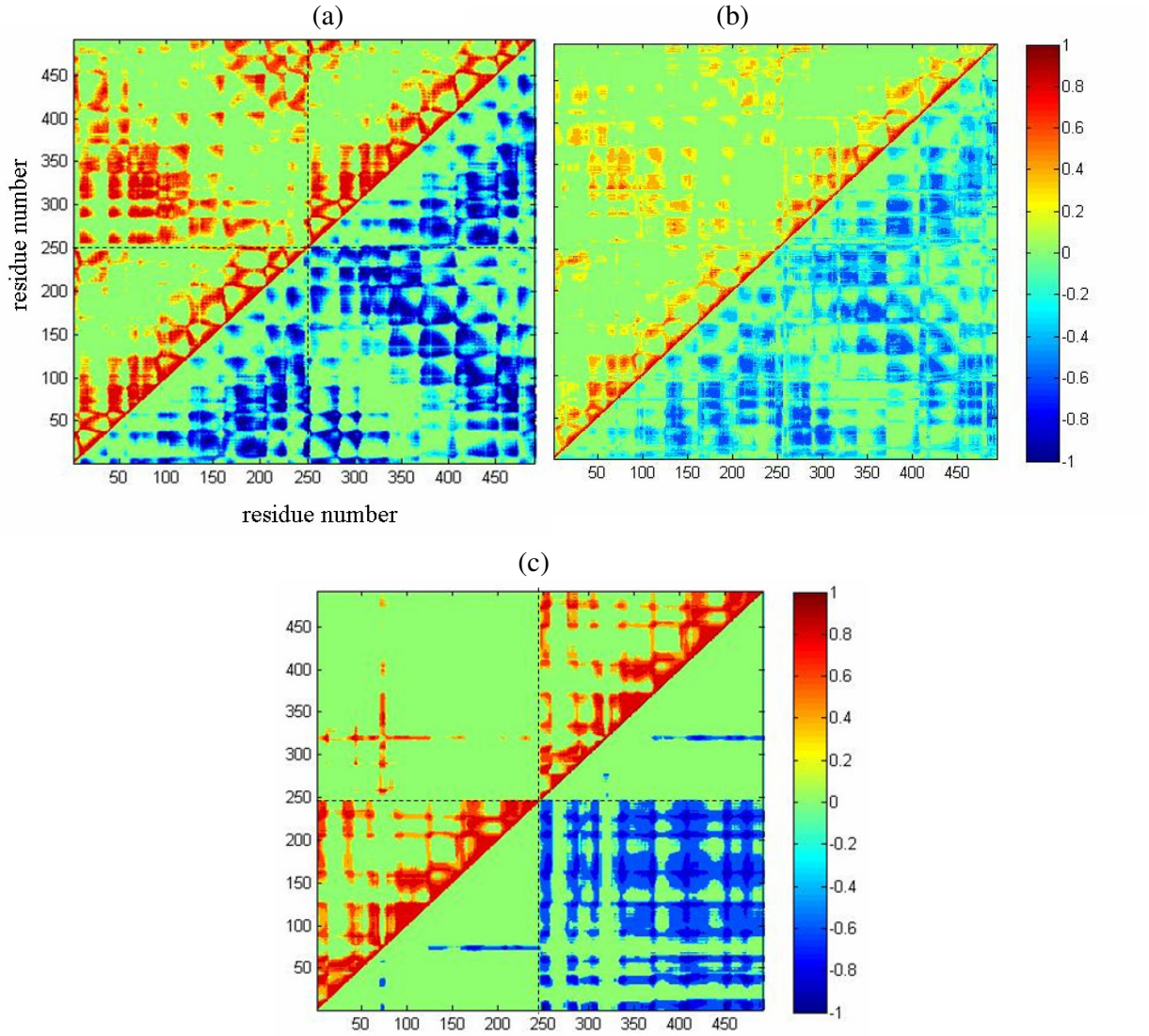


Figure 4.43. Orientational cross-correlations of TIM obtained from (a) ANM (r_{cut}: 13Å) and (b) MD, and (c) GNM (r_{cut}: 7 Å)

5. CONCLUSIONS and RECOMMENDATIONS

REs are widely used in genomic studies as they can recognize and cleave specific (cognate) DNA sequences. Their interactions with DNA strands have been studied for decades in order to understand their cleavage mechanism and also specificity on DNA sequences. The present study demonstrates the collective motions of six Type II RE family members analyzed with the elastic network models, namely ANM and GNM for their complex structures with DNA, apo forms, and DNA removed forms (if apo form is not available).

Collective deformations observed in the slowest modes of REs are mainly composed of a scissor-like (a ratchet-like) rotation and twisting motions. These motions are observed both in cognate complex and apo forms. Scissor-like motion can give an insight about the sliding mechanism for target site location of REs because in this particular motion enzymes move along the DNA in such a way that they are trying to find their target site. Also, in twisting motion, it is observed that the enzymes oscillate between their closed and open states. This result found in this study may explain the opening up of the DNA binding site in apo form and in the complex after the cleavage. The slowest modes giving sliding and opening up of DNA binding site are observed both in cognate and non-cognate complexes.

The orientational cross-correlations calculated from ANM indicate that positive inter-subunit correlations generally weaken in the absence of DNA. However, this confirmation can not be observed from GNM results, because GNM locates only some key positive interactions at the interface. We detect differences in the orientational correlations of residues between cognate and non-cognate complexes in these two functionally important modes. Large differences in the orientational cross-correlations are observed in residues specifically at the recognition site in the first mode. Also, in the second mode, some important differences are obtained on residues near to DNA binding region.

Hot spots are clustered around DNA binding region according to ANM results. Also, hot spots at recognition site are revealed by ANM. Increasing the cutoff value affects

the high-frequency modes in GNM calculations and leads to more hot spots at recognition site. It is important that the hot spot residues form an inner core that can act as a pathway for their communications between the DNA cleavage sites, because it is known that the both DNA sides should be cleaved at the same time for REs that used in this study.

For comparison with REs, another homo-dimeric enzyme TIM has been studied. ANM and GNM predict similar hot spots for TIM. All methods are satisfactory in terms of predicting a high percentage of conserved residues. The cutoff radius in GNM is only important for the location of hot spot residues at the interface. The percentages of conserved residues predicted are also quite similar. In view of the results on TIM, we can state that the discrepency between GNM and ANM in REs results from the presence of the DNA binding site at the dimeric interface, as a result of which hot spots near the binding site are not satisfactorily predicted by GNM (r_{cut}= 7 Å).

Besides elastic network models used in this study, conformation change analysis from BamHI apo form to BamHI-DNA (cognate) complex can be performed by targeted MD studies. The findings in this analysis may give insights about the mechanism of recognition and cleavage of BamHI.

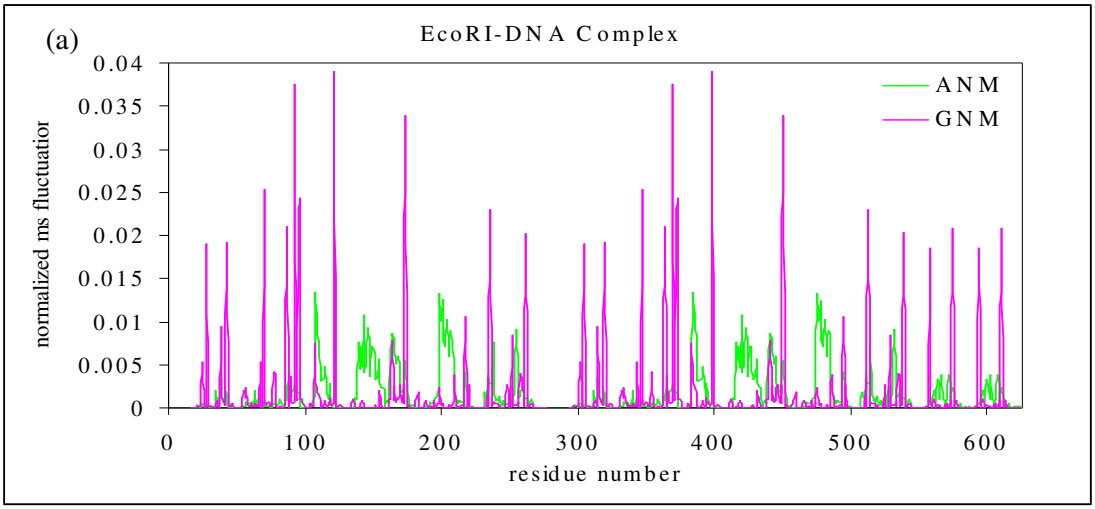
Modes obtained from atomistic calculations for complexes with cognate and non-cognate DNA can be analyzed in detail in order to locate differences between these complexes in terms of changes in inter-atomistic distances, which may give insights about the cleavage mechanism of REs.

Scissor-like motion giving sliding along DNA may have an importance for DNA recognition and binding, therefore other DNA binding enzymes can be analyzed by elastic network models in order to see whether they have this particular motion in their slowest modes.

6. APPENDIX

6.1. Additional Results for Hot spots

Figures 6.1 to 6.5 show the cumulative effect of 20 high-frequency modes comparison for each elastic network model.



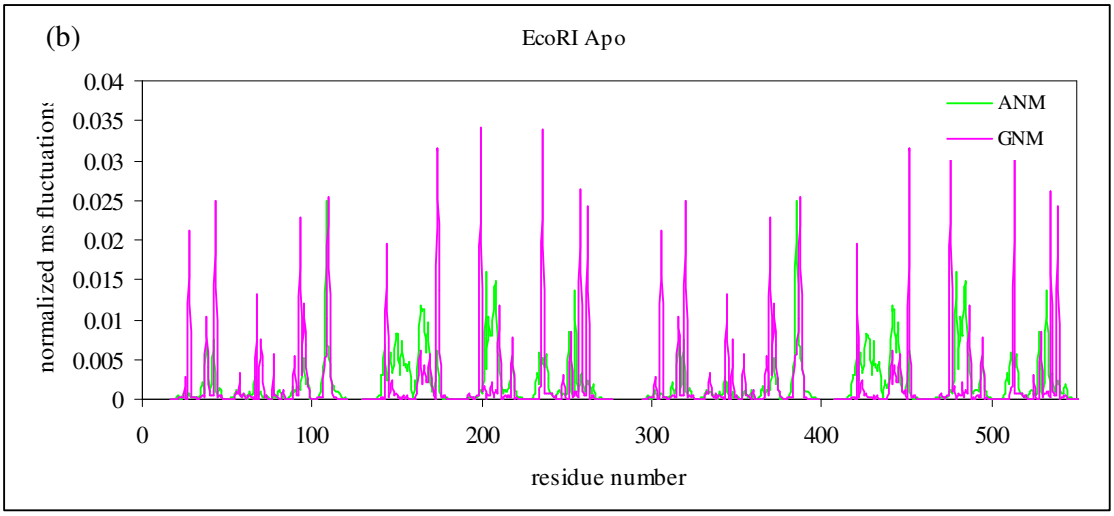


Figure 6.1. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}: 13 Å) and GNM (r_{cut}: 7 Å) for (a) EcoRI-DNA complex, and (b) EcoRI apo form

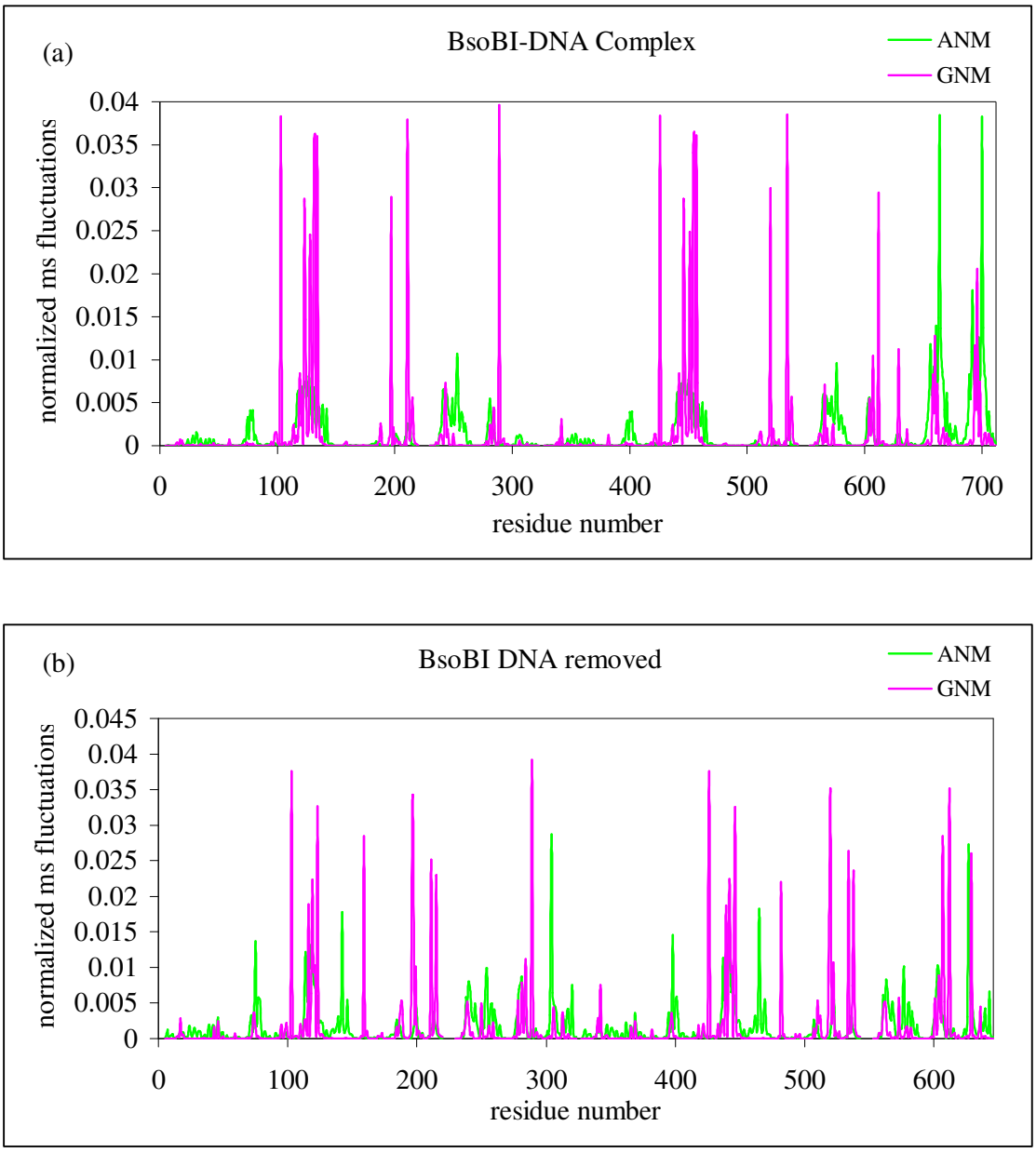


Figure 6.2. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}: 13 Å) and GNM (r_{cut}: 7 Å) for (a) BsoBI-DNA complex, and (b) BsoBI DNA removed form

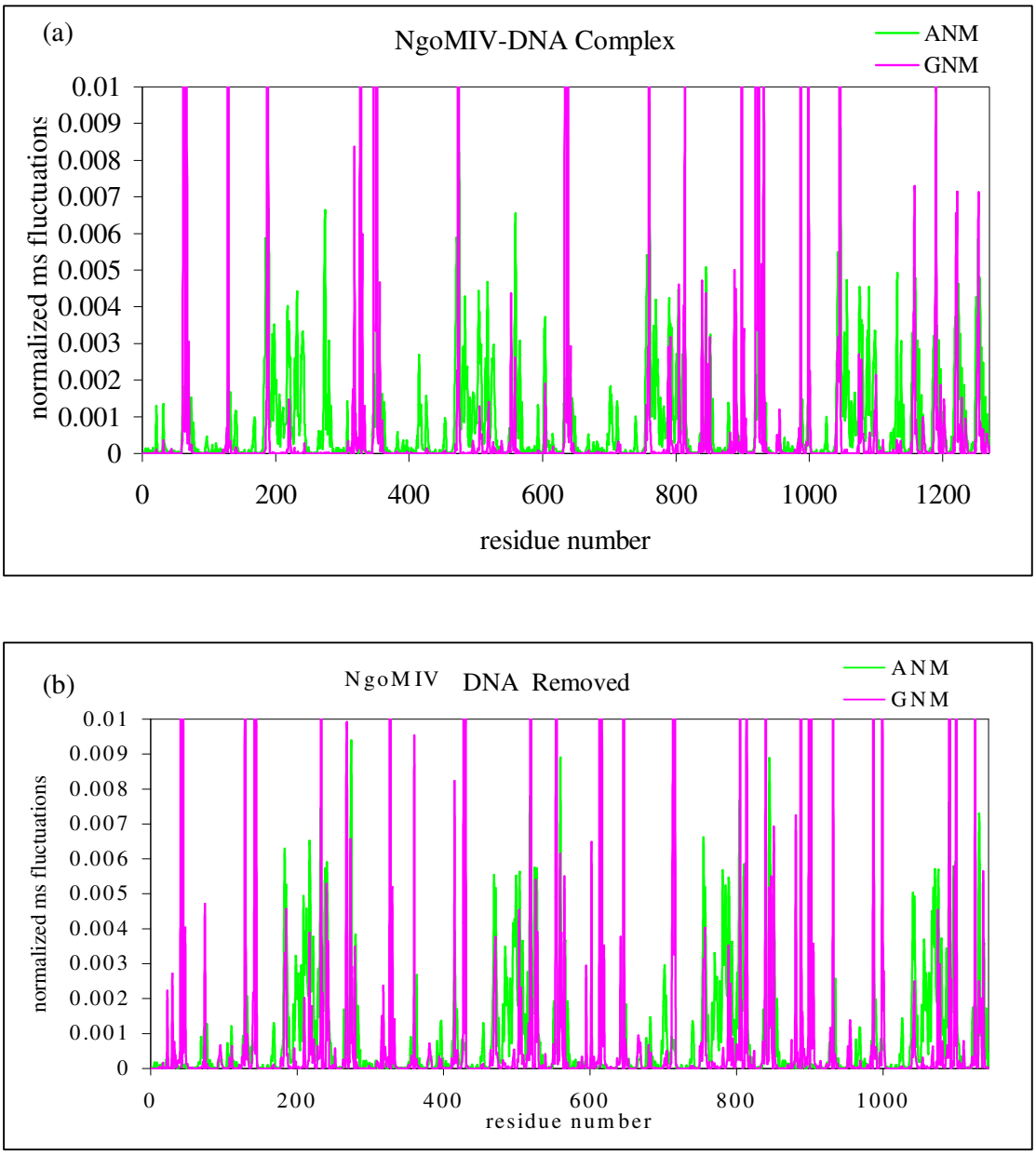
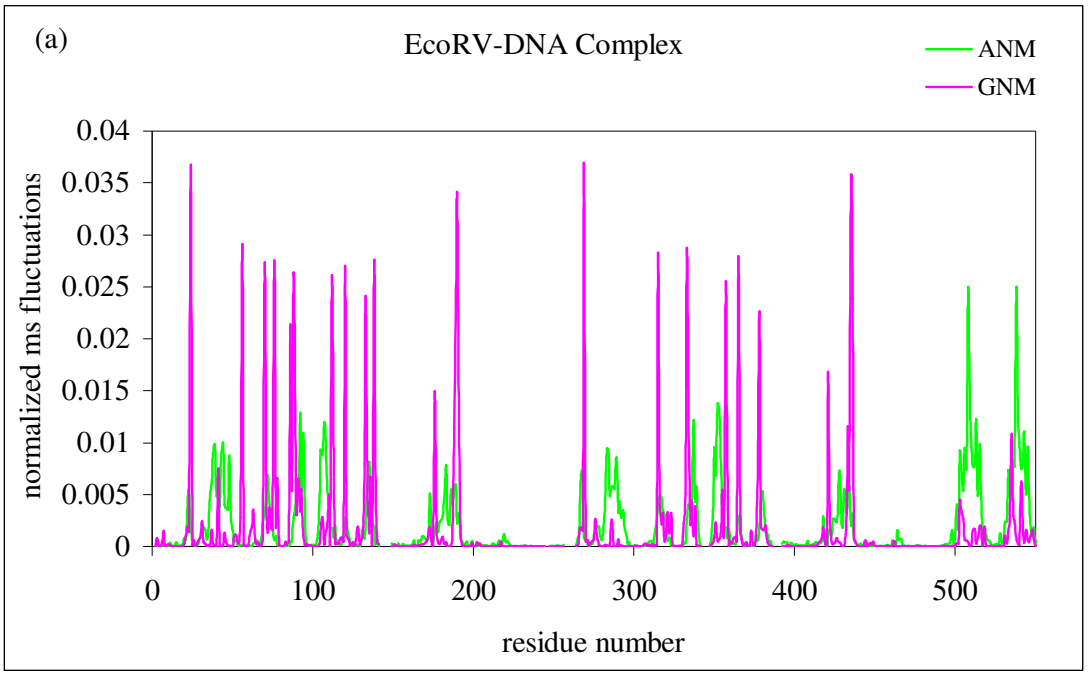


Figure 6.3. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}: 13 Å) and GNM (r_{cut}: 7 Å) for (a) NgoMIV-DNA complex, and (b) NgoMIV DNA removed form



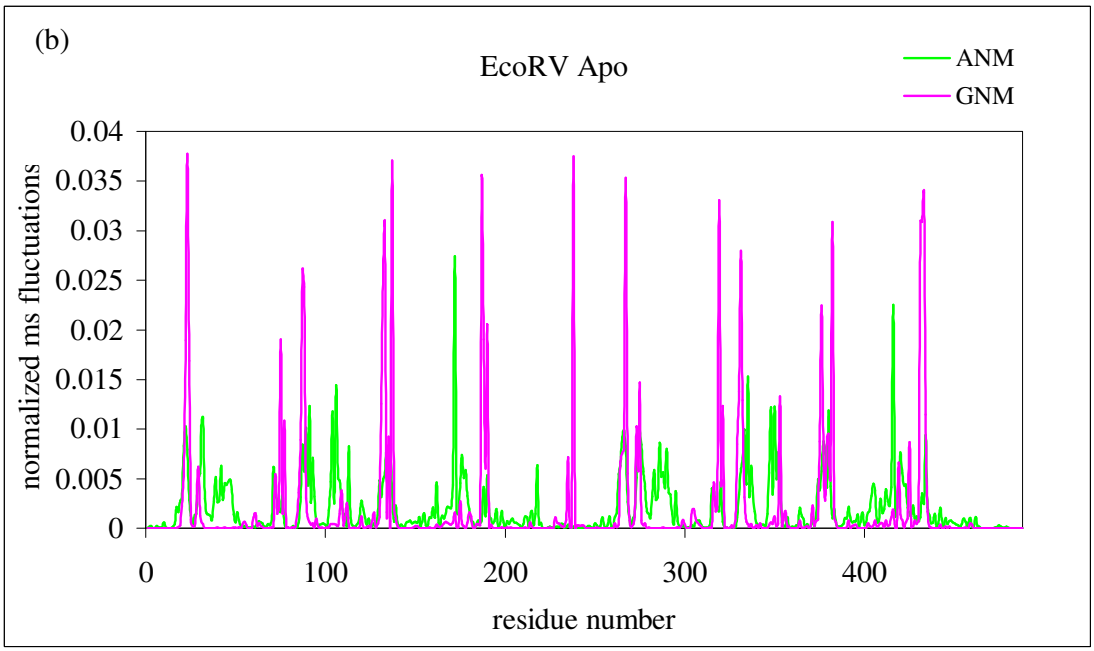


Figure 6.4. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}: 13 Å) and GNM (r_{cut}: 7 Å) for (a) EcoRV-DNA complex, and (b) EcoRV apo form

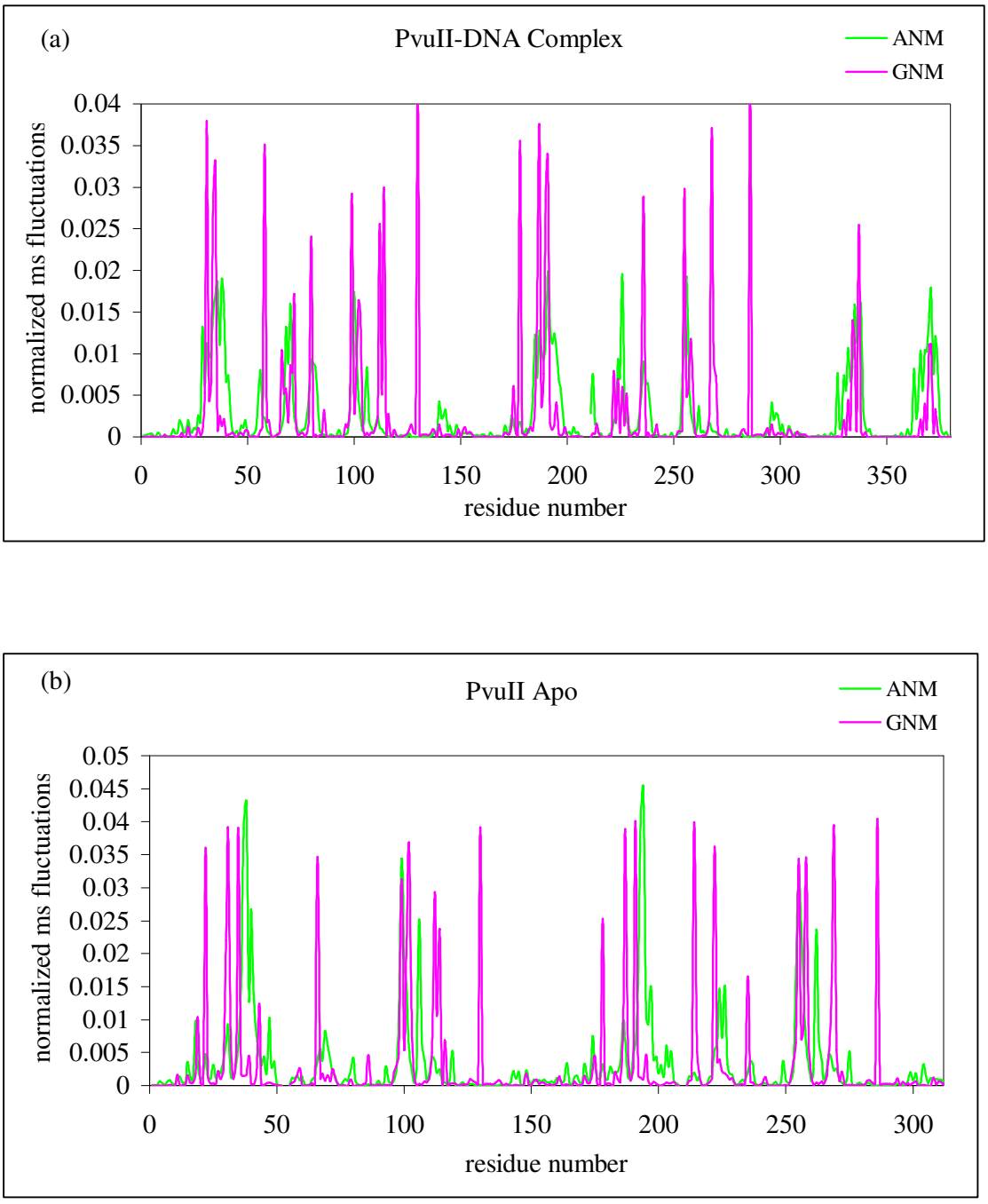


Figure 6.5. Mean-square fluctuations showing cumulative effect of 20 high-frequency modes from ANM (r_{cut}: 13 Å) and GNM (r_{cut}: 7 Å) for (a) PvuII-DNA complex, and (b) PvuII apo form

REFERENCES

Abdurashitov, M. A., O. A. Belichenko, A. V. Shevchenko, V. S. Dedkov, S. K. Degtyarev, 1997, "BstAPI, an ApaBI isoschizomer, cleaves DNA at 5'-GCANNNN NTGC-3'", *Nucleic Acids Res.*, Vol. 25, No. 12, pp. 2301-2302.

Aggarwal, A. K., 1995, "Structure and function of restriction endonucleases", *Curr. Opin. Struct. Biol.*, Vol. 5, pp. 11-19.

Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen, 1993, "Essential dynamics of proteins", *Proteins*, Vol. 17, pp. 412-425.

Anderson, J. E., 1993, "Restriction endonucleases and modification methylases", *Curr. Opin. Struct. Biol.*, Vol. 3, pp. 24-30.

Athanasiadis, A., M. Vlassi, D. Kotsifaki, P. A. Tucker, K. S. Wilson, and M. Kokkinidis, 1994, "Crystal structure of PvuII endonucleases reveals extensive structural homologies to EcoRV", *Nat. Struct. Biol.*, Vol. 1, pp. 469-475.

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, 2001, "Anisotropic of fluctuation dynamics of in proteins with an elastic network model", *Biophysical Journal*, Vol. 80, pp. 505-515.

Bahar, I., A. R. Atilgan and B. Erman, 1997, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential", *Folding and Design*, Vol. 2, pp. 173-181.

Bahar, I., A. R. Atilgan and M. C. Demirel, 1998, "Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability", *Physical Review Letters*, Vol. 80, No. 12, pp. 2733-2736.

Bahar, I., and R. L. Jernigan, 1998, "Vibrational dynamics of transfer RNAs: comparison of the free and synthetase bound forms", *J. Mol. Biol.*, Vol. 281, pp. 871-885.

Berg, O. G., and P. H. von Hippel, 1985, "Diffusion-controlled macromolecular interactions", *Annu. Rev. Biophys, Chem*., Vol. 14, pp. 131-160.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, 2000, "The Protein Data Bank", *Nucleic Acids Research*, Vol. 28, pp. 235-242.

Bogan, A. A., and K. S. Thorn, 1998, "Anatomy of hot spots in protein interfaces", *J. Mol. Biol.*, Vol. 280, pp. 1-9.

Branden C. and Tooze J., 1999, *Introduction to Protein Structure*, Garland Publishing, Inc., New York.

Brooks, C. L., M. Karplus, B. M. Pettitt, 1988, "Proteins: A theoretical perspective on Structure, Dynamics and Thermodynamics", *Adv. Chem. Phys.*, Vol. 71. pp. 1-259.

Bujnicki, J. M., 2004, "Molecular phylogenetics of restriction endonucleases", in *Restriction Endonucleases*, pp. 63-93, Pingoud A. (ed.), Springer, Berlin.

Cansu, S. and P. Doruker, 2007, *accepted.*

Chandrasekhar, K., M. M. Horvath, C. Samudzi, J. Choi, J. M. Rosenberg, 1999, *to be published.*

Cheng, X., K. Balendiran, I. Schildkraut, and J. E. Anderson, 1994, "Structure of PvuII endonuclease with cognate DNA", *EMBO J*., Vol. 13, pp. 3927-3935.

Clackson, T. and J. A. Wells, 1995, "A hot spot of binding energy in a hormone-receptor interface", *Science*, Vol. 267, No. 5196, pp. 383-386.

Coppey, M., O. Benichou, R. Voituriez and M. Moreau, 2004, "Kinetics of target site localization of a protein on DNA: a stochastic approach", *Biophys. J*., Vol. 87, pp. 1640-1649.

Deibert, G., S. Grazulis, G. Sasnauskas, V. Siksnys, and R. Huber, 2000, "Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA", *Nat. Struct. Biol.*, Vol. 7, pp. 792-799.

DeLano, W.L., 2002, "The PyMOL Molecular Graphics System", *DeLano Scientific*, Palo Alto, CA, USA.

Delarue, M. and Y. H. Sanejouand, 2002, "Simplified normal modes analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model", *J. Mol. Biol.*, Vol. 320, pp. 1011-1024.

Doruker, P., L. Nilsson, and O. Kurkcuoglu, 2006, "Collective Dynamics of EcoRI-DNA Complex by Elastic Network Model and Molecular Dynamics Simulations", *Journal of Biomolecular Structure & Dynamics*, Vol. 24, No. 1, pp. 1-15.

Dryden, D. T., N. E. Murray, D. N. Rao, 2001, "Nucleoside triphosphate-dependent restriction enzymes", *Nucleic Acids Research*, Vol. 29, pp. 3728-3741.

Ehbrecht H. J., A. Pingoud, C. Urbanke, G. Maass, and C. Gualerzi, 1985, "Linear diffusion of restriction endonucleases on DNA", *J. Biol. Chem.*, Vol. 260, pp. 6160-6166.

Frauenfelder, H., G. A. Petsko and D. Tsernoglou, 1979, "Temperature-dependent X-ray diffraction as a probe of protein structural dynamics", *Nature,* Vol. 280, pp. 558 – 563.

Friedhoff, P., R. Lurz, G. Lüder, and A. Pingoud, 2001, "Sau3AI, a Monomeric Type II Restriction Endonuclease That Dimerizes on the DNA and Thereby Induces DNA Loops", *The Journal of Biological Chemistry*, Vol. 276, No. 26, pp. 23581-23588.

Fuxreiter¸ M., and I. Simon, 2002, "Protein stability indicates divergent evolution of PD-(D/E)XK type II restriction endonucleases", *Protein Science*, Vol. 11, pp. 1978-1983.

Halford, S. E., and J. F. Marko, 2004, "How do site-specific DNA-binding proteins find their targets?", *Nucleic Acids Res.*, Vol. 32, pp. 3040-3052.

Haliloglu, T., O. Keskin, M., B. Ma, and R. Nussinov, 2005, "How Similar Are Protein Folding and Protein Binding Nuclei? Examination of Vibrational Motions of Hot Spots and Conserved Residues", *Biophysical Journal*, Vol. 88, pp. 1552-1559.

Hayward, S., 1999, "Structural principles governing domain motions in proteins", *Proteins*, Vol. 36, pp. 425-435.

Hess, B., 2000, "Similarities between principal components of protein dynamics and random diffusion", *Phys. Rev. E*, Vol. 62, pp. 8438-8448.

Hiller, D.A., A. M. Rodriguez, J. J. Perona, 2005, "Non-cognate Enzyme-DNA Complex: Structural and Kinetic Analysis of EcoRV Endonuclease Bound to the EcoRI Recognition Site GAATTC", *J.Mol.Biol*., Vol. 354, pp. 121-136.

Horton, J. R., and X. Cheng, 2000, "PvuII endonuclease contains two calcium ions in active sites", *J. Mol. Biol*., Vol. 300, pp. 1049-1056.

Horton, N. C., and J. J. Perona, 1998, "Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts", *J. Biol. Chem*., Vol. 273, pp. 21721-21729.

Hu, Z., B. Ma, H. Wolfson, and R. Nussinov, 2000, "Conservation of polar residues as hot spots at protein interfaces", *Proteins*, Vol. 39, pp. 331-342.

Huai, Q., J. D. Colandene, Y. Chen, F. Luo, Y. Zhao, M. D. Topal, and H. Ke, 2000, "Crystal structure of NaeI- An evolutionary bridge between DNA endonuclease and topoisomerase", *EMBO J.,* Vol. 19, pp. 3110-3118.

Jack, W. E., B J. Terry, and P. Modrich, 1982, "Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence", *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp. 4010-4014.

Janscak, P., U. Sandmeier, M. D. Szczelkun, T. A. Bickle, 2001, "Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I", *J. Mol. Biol.*, Vol. 306, pp. 417-431.

Jeltsch, A. and C. Urbanke, 2004, "Sliding or Hopping? How Restriction Enzymes Find Their Way on DNA", in *Restriction Endonucleases*, pp. 95-110, Pingoud A. (ed.), Springer, Berlin.

Jeltsch, A., J. Alves, H. Wolfes, G. Maass, and A. Pingoud, 1994, "Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA", *Biochemistry,* Vol. 33, pp. 10215-10219.

Jogl, G., S. Rozovsky, A. E. McDermott, L. Tong, 2003, "Optimal alignment for enzymatic proton transfer: Structure of the Michaelis complex of triosephosphate isomerase at 1.2-A resolution", *Proc. Natl. Acad. Sci. USA*, Vol. 100, pp. 50-55.

Kantarci, N., P. Doruker, and T. Haliloglu, 2006, "Cooperative fluctuations point to the dimerization interface of p53 core domain", *Biophys. J.*, Vol. 91, pp. 421-432.

Keskin, O., C. J. Tsai, H. Wolfson, and R. Nussinov, 2004, "Anew, structurally nonredundant, diverse data set of protein-protein interfaces and its implications", *Protein Sci.*, Vol. 13, pp. 1043-1055.

Kim, Y. C., J. C. Grable, R. Love, P. J. Greene, and J. M. Rosenberg, 1990, "Refinement of EcoRI endonuclease crystal structure: A revised protein chain tracing", *Science*, Vol. 249, pp. 1307-1309.

Kim, Y., J. Choi, J. C. Grable, P. J. Greene, P. Hager, and J. M. Rosenberg, 1994, "Studies on the canonical DNA-EcoRI endonuclease complex and the EcoRI kink", *Structural Biology: The state of the art*, Vol. 1 (Sarma, R. H., and M. H. Sarma, eds), Adenine Press, Schenectady NY, pp. 225-246.

Lesser, D. R., M. R. Kurpiewski, and L. Jen-Jacobson, 1990, "The energetic bases of specificity in the EcoRI endonuclease-DNA interaction", *Science*, Vol. 250, pp. 776-786.

Lucaks, C. M., R. Kucera, I. Schildkraut, and A. K. Aggarwal, 2000, "Understanding the immutability of restriction enzymes: Crystal structure of BglII and its DNA substrate at 1.5 Å resolution", *Nat. Struct. Biol.*, Vol. 7, pp. 134-140.

Ma, B., T. Elkayam, H. Wolfson, and R. Nussinov, 2003, "Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces", *Proc. Natl. Acad. Sci. USA*, Vol. 100, pp. 5772-5777.

McClarin, J. A., C. A. Frederick, B. C. Wang, P. Greene, H. W. Boyer, J. Grable, and J. M. Rosenberg, 1986, "Structure of the DNA-EcoRI endonuclease recognition complex at 3 Å resolution", *Science*, Vol. 234, pp. 1526-1541.

McCleland, S. E., and M. D. Sczcelkun, 2004, "The type I and III restriction endonucleases: structural elements in molecular motors that process DNA", in *Restriction Endonucleases*, pp. 111-135, Pingoud A. (ed.), Springer, Berlin.

Mücke, M., S. Reich, E.. Moncke-Buchner, M. Reuter and D. H. Kruger, 2001, "DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites", *J. Mol. Biol.*, Vol. 312, pp. 687-698.

Newman, M., T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal, 1994, "Structure of restriction endonuclease BamIII and its relationship to EcoRI", *Nature*, Vol. 368, pp. 660-664.

Pingoud, A., and A. Jeltsch, 1997, "Recognition and cleavage of DNA by type-II restriction endonucleases", *Eur. J. Biochem*, Vol. 246, pp. 1-22.

Pingoud, A., and A. Jeltsch, 2001, "Structure and function of type II restriction endonucleases", *Nucleic Acids Research*, Vol. 29, No. 18, pp. 3705-3727.

Pingoud, A., M. Fuxreiter, V. Pingoud and W. Wende, 2005, "Type II restriction endonucleases: structure and mechanism", *Cell. Mol. Life Sci*., Vol. 62, pp. 685-707.

Raleigh, E. A. and J. E. Brooks, 1998, in De Bruijn, F. J., J. R. Lupski and G. M. Weinstock (eds), *Bacterial Genomes*. Chapman & Hall, New York, pp. 78-92.

Rena, G., and M. D. Houslay, 1998, "A simple method for sequencing small DNAs by introducing precise overlapping ends into restriction digestion fragments", *Nucleic Acids Research,* Vol. 26, No. 16, pp. 3867-3868.

Roberts, R. J. and S. E. Halford, 1993, "Type II restriction endonucleases", in Linn, S. M., R. S. Lloyd and R. J. Roberts (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY, pp. 35-88.

Roberts, R. J., 2005, "How restriction enzymes became the workhorses of molecular biology", *PNAS*, Vol. 102, No. 17, pp. 5905-5908.

Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. F. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Krüger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P.-C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. B. Vitor, G. G. Wilson and S.-Y. Xu, 2003, "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes", *Nucleic Acids Res*., vol. 31, pp. 1805-1812.

Roberts, R. J., T. Vincze, J. Posfai, D. Macelis, 2007, "REBASE - enzymes and genes for DNA restriction and modification", *Nucleic Acids Research*, Vol. 35, pp. 269-270.

Scheuring-Vanamee, E., H. Viadiu, C. M. Lukacs and A. K. Aggarwal, 2004, "Two of a Kind: BamHI and Bg1II", in *Restriction Endonucleases*, pp. 215-236, Pingoud A. (ed.), Springer, Berlin.

Sen, T. Z., Y. Feng, J. V. Garcia, A. Kloczkowski, and R. L. Jernigan, 2006, "The Extent of Cooperativity of Protein Motions Observed with Elastic Network Models Is Similar for Atomic and Coarser-Grained Models", *J. Chem. Theory Comput.*, Vol. 2, No. 3, pp. 696-704.

Smith, H. O., and D. Nathans, 1973, "A suggested nomenclature for bacterial host modification and restriction systems and their enzymes", *J. Mol. Biol*., Vol. 81, pp. 419-423.

Stephenson, F. H., B. T. Ballard, H. W. Boyer, J. M. Rosenberg, and P. J. Greene, 1989, "Comparison of the nucleotide and amino acid sequences of the RsrI and EcoRI restriction endonucleases", *Gene*, Vol. 85, pp. 1-13.

Sun, J, H. Viadiu, A. K. Aggarwal, and H. Weinstein, 2003, "Energetic and Structural Considerations for the Mechanism of Protein Sliding along DNA in the Nonspecific BamHI-DNA Complex", *Biophysical Journal*, Vol. 84, pp. 3317-3325.

Terry, B. J., W. E. Jack, and P. Modrich, 1983, "Facilitated diffusion during catalysis by EcoRI endonuclease: nonspecific interactions in EcoRI catalysis", *J. Biol. Chem*., Vol. 260, pp. 13130-13137.

Thielking, V., J. Alves, A. Fliess, G. Maass, and A. Pingoud, 1990, "Accuracy of the EcoRI restriction endonuclease: binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences", *Biochemistry,* Vol. 29, pp. 4682-4691.

van der Woerd, M., J. Pelletier, S. Xu, and A. Friedman, 2001, "Restriction Enzyme BsoBI-DNA Complex: A tunnel for recognition of degenerate DNA sequences and potential histidine catalysis", *Structure*, Vol. 9, pp. 133-144.

Venclovas, C., A. Timinskas, and V. Siksnys, 1994, "Five-stranded β-sheet sandwiched with two α-helices: a structural link between restriction endonucleases EcoRI and EcoRV", *Proteins*, Vol. 20, pp. 279-282.

Viadiu, H., and A. K. Aggarwal, 1998, "The role of metals in catalysis by the restriction endonucleases BamHI", *Nat. Struct. Biol.*, Vol. 5, pp. 910-916.

Viadiu, H., and A. K. Aggarwal, 2000, "Structure of BamHI Bound to Nonspecific DNA: A Model for DNA sliding", *Molecular Cell*, Vol. 5, pp. 889-895.

Welsch, A. J., S. E. Halford, and D. J. Scott, 2004, "Analysis of type II restriction endonucleases that interact with two recognition sites", in *Restriction Endonucleases*, pp. 297-317, Pingoud A. (ed.), Springer, Berlin.

Wierenga, R. K., M.E. Noble, J.P. Postma, H. Groendijk, K.H. Kalk, W.G. Hol and F.R. Opperdoes, 1991, "The crystal structure of the open and the closed conformation of the flexible loop of trypanosomal triosephosphate isomerase", *Proteins*, Vol. 10, pp. 33–49.

Winkler, F. K., D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos, and K. S. Wilson, 1993, "The crystal structure of EcoRV endonuclease and of its complexes with cognate- and non-cognate DNA fragments", *EMBO J.*, Vol. 12, pp. 1781-1795.