

COMPUTATIONAL STUDIES OF THE DNA MISMATCH RECOGNITION PROTEIN

By

Sean Ming-Yin Law

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Biochemistry and Molecular Biology

2011

ABSTRACT

COMPUTATIONAL STUDIES OF THE DNA MISMATCH RECOGNITION PROTEIN

By

Sean Ming-Yin Law

The MutS DNA mismatch recognition protein was studied by using a combination of molecular dynamics simulations and normal mode analysis. Both methods revealed uniquely different structural conformations that were collectively used to characterize a new functional cycle for mismatch recognition.

The DNA dynamics from the MutS simulations were also assessed. The G·T mismatch contained within the DNA was found to be relatively stable, whereas the 5' adjacent base next to the mispaired thymine was highly dynamic. In one simulation, the 5' adjacent base opened up via the major groove and stayed flipped-out for the entire duration of the 200 ns simulation. The energetics of base-flipping in the MutS-DNA system were examined and the relevance and importance of these observations were discussed.

The development of a new path-based restraint is presented and applied to study DNA translocation in the Hin recombinase test system. Using multiple path-base restraints, the DNA was successfully translocated by one full base pair in both the forward and backward directions. The method for calculating the corresponding free energy profile along a single DNA translocation reaction coordinate was also reformulated and explained.

ACKNOWLEDGEMENTS

Six productive, challenging, and revealing years at Michigan State University have finally come to a dénouement. Each passing day as a graduate student has taught me more about myself than I could have ever imagined. However, my initial interest in research began well before my days at Michigan State University and, therefore, I must acknowledge Dr. Walter J. Whiteley at York University (Toronto, Canada). The opportunities that were given to me as an undergraduate researcher really opened up my eyes, and Walter's guidance and advice have played a vital role in my scientific journey.

As an amateur structural biologist with an interest in applied mathematics, I became a member of Dr. Michael Feig's group at Michigan State University during the spring of 2006. Michael's exacting nature, endless patience, passion for science, and insistence of "quality over quantity" provided the perfect environment for my professional growth and development. As a mentor and advisor, Michael challenged and encouraged me to step beyond my boundaries and both inspired and motivated me to work harder. Over the years, I have truly come to understand and appreciate all of the things that Michael has done and I consider myself exceptionally privileged and forever indebted. I will always remember the one piece of timely advice that Michael had shared with me near the end of my degree which accurately defines what I have learned while under his supervision – "Gambate!" (*Japanese*: Work hard/don't give up!).

My path through graduate school would not have been possible without the support from many past and current members of the Feig lab. First, I would like to thank Dr. Shayantani Mukherjee, Dr. Katarzyna Maksimiak, Hugh Crosmun, and Taraz Buck for their indispensable contributions to the MutS project. Next, I would like to acknowledge the first generation of

group members, namely Andrew Stumpff-Kane, Dr. Seiichiro Tanizaki, Dr. Kitiyaporn Wittayanarakul, Brian Connelly, and Jacob Clifford with whom I have had the pleasure of working with. I am also grateful for my other colleagues, Afra Panahi, Maryam Sayadi, Dr. Yi-Ming Cheng, Dr. Srinivasa Gopal, Dr. Kahsay Gebreyohannes, Dr. Alexander Predeus, Dr. Liang Fang, Dr. Kangasabai Vadivel, Vahid Mirjalili, Seref Gul, and Nan Liu. Thank you all for the great memories and for always making the next day more enjoyable than the last.

I would like to express my sincerest gratitude to all members of my guidance committee, Dr. Robert Cukier, Dr. Robert Hausinger, Dr. Charles Hoogstraten, Dr. Honggao Yan, and Dr. William Wedemeyer. Their expert advice, critical support, and careful tutelage were crucial for allowing me to reach my research goals. I would also like to extend my appreciation to Dr. Leslie Kuhn and former members of the Kuhn Lab for adopting me as their computational half-son and I would also like to acknowledge Dr. Zachary Burton for his critical feedback and helpful discussions. A special thanks goes out to Edward Kabara as I would have never survived my first year of graduate school without his honest friendship. For fear of omission, I would like to thank all past and present members of the Department of Biochemistry and Molecular Biology as well as all of the admirable administrative personnel. My days as a graduate student have finally reached the end of a long chapter but I promise you that the next page will be filled with new and exciting adventures.

Finally, to my mom, dad, Stella, Sacha, and to the rest of my supportive friends and family - 多謝.

TABLE OF CONTENTS

List of Tables	vii
List of Figures.....	viii
Chapter 1 Introduction	1
1.1 Overview.....	2
1.2 Replication Fidelity and Methyl-directed DNA Mismatch Repair.....	2
1.3 Molecular Structure of the DNA Mismatch Recognition Protein	4
1.4 Investigating the Structural Dynamics of the DNA Mismatch Recognition Complex.....	6
1.5 Computer Simulations	8
Chapter 2 Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal Mode Analysis.....	23
2.1 Abstract.....	24
2.2 Introduction.....	25
2.3 Methods.....	30
2.4 Results and Discussion	32
2.5 Conclusions.....	48
2.6 Acknowledgements.....	49
Chapter 3 An Analysis of <i>E. coli</i> MutS Dynamics from Molecular Dynamics Simulations.....	67
3.1 Abstract.....	68
3.2 Introduction.....	69
3.3 Methods.....	71
3.4 Results.....	73
3.5 Discussion.....	76
3.6 Conclusions.....	80
Chapter 4 Base-Flipping Mechanism in Post-Mismatch Recognition by MutS.....	96
4.1 Abstract.....	97
4.2 Introduction.....	98
4.3 Materials and Methods.....	100
4.4 Results and Discussion	106
4.5 Conclusions.....	116
4.6 Acknowledgements.....	117
Chapter 5 A Path-Based Reaction Coordinate for Biased Sampling of Nucleic Acid Translocation	158
5.1 Introduction.....	159
5.2 Methods.....	162
5.3 Results.....	173
5.4 Discussion.....	175

5.5 Conclusion	177
Chapter 6 Conclusions and Perspectives	183
6.1.....	184
References.....	191

LIST OF TABLES

Table 2.1	Overlap Index for a Pair of Modes, Each From Two Different Sets.....	50
Table 3.1	Pair-wise Overlap of the First Eigenvector From the Nine Simulations and the Combined Trajectory	81
Table 4.1	DNA Sequence Used in All MutS Simulations.....	118
Table 5.1	DNA Sequence Used in All Hin Recombinase Simulations	178

LIST OF FIGURES

Figure 1.1 Schematic representation of the methyl-directed DNA mismatch repair in <i>E. coli</i>	17
Figure 1.2 DNA mismatch recognition proteins	19
Figure 1.3 Schematic representation of the Hamiltonian Replica Exchange Method	21
Figure 2.1 Crystal structure of MSH2-MSH6.....	51
Figure 2.2 The dynamic behavior of the ATPase site in both chains	53
Figure 2.3 Root mean square fluctuation of C α atoms	55
Figure 2.4 Protein backbone showing thermal fluctuations color coded by B-factors	57
Figure 2.5 Average covariance from the first 10 modes in MSH2-MSH6 and MutS	59
Figure 2.6 Mode motions of MSH-free projected on to the minimized crystal structure.....	61
Figure 2.7 Close-up views of the motions in the nucleotide-binding domain of MSH-free	63
Figure 2.8 Schematic diagram representing distinct conformational states during the functional cycle of MSH2-MSH6 or MutS	65
Figure 3.1 The MutS-DNA structure with different nucleotide-bound conformations	82
Figure 3.2 Secondary structure propensity	84
Figure 3.3 The percent contribution of the first 20 eigenvectors.....	86
Figure 3.4 Schematic “Porcupine plot” for the ATP:ATP simulation.....	88
Figure 3.5 MutS ATPase domain.....	90
Figure 3.6 Schematic “Porcupine plot” for the ATP:ADP simulation	92
Figure 3.7 The RMSF for the S1 and S2 monomers.....	94
Figure 4.1 X-ray crystal structure of <i>E. coli</i> MutS	119
Figure 4.2 Pseudodihedral angle definition	121
Figure 4.3 C α protein RMSD and heavy atom DNA RMSD	123

Figure 4.4 K-means clustering of the nine simulations using a 2.5 Å radius	125
Figure 4.5 DNA base pair hydrogen bonding.....	127
Figure 4.6 Comparison of the G·T and G/C(-1) major groove widths	129
Figure 4.7 Comparison of T22 glycosyl rotation angle	131
Figure 4.8 Correlation of C21 base-flipping in NONE:NONE simulation with various structural quantities.....	133
Figure 4.9 Snapshots from the NONE:NONE simulation of base flipping.....	135
Figure 4.10 Free energy profiles from the HREM simulation.....	137
Figure 4.11 Comparison of the C21 backbone ζ torsion angle.....	139
Figure 4.12 Free energy profiles from the same HREM simulation.....	141
Figure 4.13 HREM sampling overlap.....	143
Figure 4.14 Water residence time calculations from the NONE:NONE simulation	145
Figure 4.15 Solvent-accessible surface area (SASA) calculations	147
Figure 4.16 Protein domain motions.....	149
Figure 4.17 S1 DNA binding domain movement from unbiased and HREM simulations along Y and Z directions.....	152
Figure 4.18 Allosteric signaling from the DNA binding domain to the ATPase domains	154
Figure 4.19 Visualization of the effects of base flipping on the ATPase domains.....	156
Figure 5.1 Snapshots from forward and backward DNA translocation in Hin-recombinase	179
Figure 5.2 Free energy profile for DNA translocation	181

Chapter 1

Introduction

1.1 Overview

In this dissertation the structures of the *Escherichia coli* (*E. coli*) and human DNA mismatch recognition proteins, MutS and MSH2-MSH6 (MutS α), respectively, are studied using modern computational techniques. This dissertation is composed of six chapters followed by a combined reference section for all chapters. Chapter 1 offers a basic introduction to the DNA mismatch repair process and presents some of the experimental and theoretical methods that are referenced within this dissertation. Chapter 2 discusses the identification of distinct MutS and MutS α conformational states using normal mode analysis accompanied by the structural characterization of a complete functional cycle for mismatch recognition. Chapters 3 and 4 describe the work from several 200 nanosecond (ns) molecular dynamics computer simulations of *E. coli* MutS. Chapter 3 presents an analysis of the overall protein dynamics and Chapter 4 details the proposed role and effects of DNA base-flipping in MutS. Chapter 5 describes the development of a novel path-based biasing potential that can be used for studying DNA translocation. Finally, a summary of the conclusions is given in Chapter 6.

1.2 Replication Fidelity and Methyl-directed DNA Mismatch Repair

In 1953, James Watson and Francis Crick presented their classical paper which accurately described the double helical structure of DNA (1). Because DNA serves as the genetic blueprint, maintaining the integrity of DNA is essential and requires mechanisms to prevent errors due to attack from external factors and due to infidelity during improper replication or homologous recombination. During DNA synthesis, the error frequency in base misincorporation alone is estimated to be about $10^{-1} - 10^{-2}$ (2-3) but this error is greatly reduced to $10^{-5} - 10^{-7}$ by the

nucleotide selectivity of DNA polymerase and the replisomal 3'→5' proofreading exonuclease (along with a small reduction by accessory proteins such as the single-stranded DNA-binding proteins (SSB)) (2-3). However, this cumulative error rate is still far too high considering that the human genome is made up of roughly three billion base pairs and, on average, the replication machinery only commits three base pair errors per replication cycle (3). Fortunately, the last line of defense is the DNA mismatch repair (MMR) system which is capable of increasing the accuracy by a factor of 10^3 and thereby improving the cumulative error frequency to $\sim 10^{-10}$ (3). The MMR pathway is conserved from prokaryotes to eukaryotes and MMR deficiencies in humans have been linked to an increase risk of colorectal cancer as well as other forms of cancer (4).

The methyl-directed MMR pathway in *E. coli* (Figure 1.1), which has been reconstituted from purified proteins (5), is the most well-studied bacterial mismatch repair pathway and therefore also serves as an excellent prototype for understanding eukaryotic mismatch repair. In *E. coli*, the MutS DNA mismatch recognition protein is responsible for scanning the DNA in search of short insertion/deletion loops (IDLs) and base-base mismatches resulting from occasional polymerization errors that have eluded proofreading (6). After binding to a mismatch, MutS, in the presence of ATP, recruits MutL and forms a MutS-DNA-MutL ternary complex (7-9). Next, MutH, a protein that is bound within ~ 1 kb of the mismatch at a hemimethylated d(GATC) site located either 3' or 5' to the lesion and which is activated by the assembly of the ternary complex, gets recruited. Then, the latent endonuclease found in MutH cleaves the newly synthesized (unmethylated) DNA strand and not the Dam methylated parental strand (10). This incision acts as a point of entry for binding of SSB and for MutL-facilitated loading of DNA helicase II. Dependent upon the location of the nick with respect to the mismatch, removal of the

damaged DNA strand can occur in either direction (11-12) and is completed by either 3'→5' exonucleases (ExoI and ExoX) or 5'→3' exonucleases (RecJ and ExoVII) (13-15). Finally, the excised DNA is re-synthesized and sealed by DNA polymerase III and DNA ligase, respectively. For a more thorough review of DNA MMR, the reader is directed to several excellent and comprehensive reviews (4-5, 16-18).

1.3 Molecular Structure of the DNA Mismatch Recognition Protein

The earliest documented effort for predicting the 3-dimensional structure of the human MutS homolog 2 (hMSH2) DNA mismatch recognition protein was published in 1998 by de las Alas *et al.* (19). In that work, a then novel prediction-based threading method was used to identify structural homologs of hMSH2 and coordinates were manually assigned based on matches between the predicted secondary structure of hMSH2 and the secondary structure of the structural homologs found in the Protein Data Bank (PDB). However, two years later, two independent groups published the first ever high-resolution X-ray crystal structures of MutS from *E. coli* (20) and *Thermus aquaticus* (*T. aquaticus*) (21), neither of which bore any resemblance to the previously predicted structure. Over the years, these seminal structures have played an important role in improving our overall understanding of mismatch recognition. As suggested by their highly conserved amino acid sequences, the pair of crystal structures was shown to be remarkably similar (Figure 1.2A and Figure 1.2B) with each structure being made up of two individual homodimeric subunits that form the shape of a θ symbol (or, alternatively, a pair of praying hands). Each monomer consists of five structural domains, each of which is found to match the fold of a previously determined protein structure (21). Domain I is the DNA

binding domain and even though both subunits (S1 and S2) share the same sequence, only the S1 DNA binding domain interacts directly with the mismatch via a highly conserved Phe-X-Glu motif (20-30) (located near the N-terminus of the protein) while the S2 subunit makes non-specific DNA contacts. Domain II is referred to as the connector and has been implicated in ATP-dependent interactions with MutL (9, 31). Domain III, the core domain, is largely responsible for providing structural support and is poised for transmitting long-range allosteric signals between both ends of the large protein (21, 32). Domain IV resides at the tip of the protein forming a clamp around the DNA with the help of domain I and, in both structures, the DNA is bent by about 60° near the site of the mismatch. However, in the absence of DNA, both domains I and IV are found to be highly mobile (21). Finally, domain V is the location for the well conserved nucleotide-binding site (20-21) (Figure 1.2D) which belongs to the ATP binding cassette (ABC) superfamily. Biochemical studies of *E. coli* and *T. aquaticus* MutS have demonstrated that the two chemically identical ATPase subunits act asymmetrically, each showing different affinities for ADP, ATP, and non-hydrolyzable ATP analogues (33-34). This domain also contains a conserved helix-turn-helix (HTH) motif which is crucial for mismatch binding, ATPase activity, and protein dimerization (35) (see Figure 1.2D).

From single-molecule total internal reflection fluorescence microscopy (TIRFM), it has been clearly demonstrated that the DNA mismatch protein scans the DNA diffusively (in the presence or absence of ADP) (36) and, after recognizing a mismatch, ATP binds to MutS and causes the protein to dissociate from the mismatch and to form a stable clamp around the DNA (37-42). This is the so-called “sliding clamp” conformation. There has been an ongoing debate as to whether or not MutS sliding (not scanning) is dependent on the hydrolysis of ATP (as supported by the *active translocation model*) (39, 42) or whether the purpose of ATP hydrolysis

is to allow the instantaneous recovery of MutS back to a scanning mode (as supported by the *molecular switch model*) (38, 40). A more comprehensive discussion of these models is provided in the following review articles (16, 18, 43).

Since 2000, several more structures of *E. coli* MutS containing a point mutation (44), bound to various mismatches (45), and doubly bound with ATP (instead of ADP as in the original structure) (29) have been published. However, these new structures only capture small local changes in the protein compared to the original and it has been suggested that: 1) different mismatches are recognized using the same binding mode (45); and 2) the MutS structure observed in the various crystals is possibly a trapped intermediate that is incapable of hydrolyzing ATP (29). More recently, several crystal structures of the hMSH2-hMSH6 (MutS α) human homolog (bound to different mismatches) were determined (30) (see Figure 1.2C) and were found to preserve many of the structural features first identified in the prokaryotic MutS homolog.

1.4 Investigating the Structural Dynamics of the DNA Mismatch Recognition Complex

The structure and dynamics of the MutS-DNA complex (and its homologs) have been studied by using a wide range of experimental and theoretical techniques. As discussed above, the high-resolution structural data from X-ray crystallography has been important for our understanding of the overall MutS structure, but it also provides some additional insight into the mobility of the protein as reflected by thermal B-factors. In principle, B-factors indicate the spread of electron density around a specific position in the map, and so parts of the structure that are disordered are

reflected by high B-factor values while low B-factors represent low mobility (46). However, the meaning of B-factors must be interpreted cautiously because crystal packing forces could have an adverse effect on the protein motions. At a much lower resolution, small angle X-ray scattering (SAXS) has also been helpful in identifying three different nucleotide-dependent conformations of *Thermus thermophilus* MutS (47). In that study, the size and general shape were measured in solution using SAXS, and it was found that the MutS structure (in the absence of DNA) was stretched out in the presence of ADP, more compact in the presence of ATP, and existed as an intermediate between the two when nucleotides were absent. While new observations were made using this method, the lack of atomic resolution makes it impossible to determine the precise nucleotides that are bound in the asymmetrical ATPases. This is an important point because each ATP binding site can be independently occupied by either ADP, ATP, or nothing at all, which means that there are a total of nine different possible nucleotide-bound combinations to consider (two subunits and three nucleotide configurations).

The dynamic nature of the MutS-DNA complex has been best characterized by using atomic force microscopy (AFM) (48). This method utilizes a flexible probe/stylus that scans the contours of a surface that is deposited with protein-DNA complexes and is capable of producing images with nanometer (nm) resolution (48-49). AFM images showed that MutS bound to homoduplex DNA belonged to a single population where the DNA took on a bent conformation while both bent and unbent DNA conformations were observed when MutS was bound at a mismatch site (48). These results led to the proposition that the DNA is kinked by 60° upon recognition of a mismatch (i.e., as in the crystal structure) but, ultimately, the protein undergoes a conformational change as a result of DNA unbending (16). These ideas were later expanded based on single molecule fluorescence resonance energy transfer (smFRET) experiments (50).

Since MutS was found to be more stable at an unbent mismatch site than at a homoduplex site, it was hypothesized that the unbent state bound by MutS may be stabilized by flipping out one of the mismatched bases (16, 48, 51). Interestingly, while there has been no direct evidence of the mismatch base (or any other base) flipping out of the helical stack, the dynamics of the bases surrounding the mismatch have been probed using 2-aminopurine, a fluorescent adenine analog that is commonly used to study DNA base flipping, and it was found that the 5' adjacent base next to the mismatch experienced enhanced dynamics when bound by MutS (52). The relevance of this study is discussed in more detail in Chapter 4.

The experimental methods discussed in the above section have all been vital in contributing to the current understanding of the MutS-DNA complex. However, most of the methods described above lack the level of detail required to accurately define the different structural conformations expected in the MutS system. Atomic detail computer simulations are capable of filling this void and can play an important complementary role in clarifying experimental data. As well, molecular dynamics (MD) has been used extensively to investigate other protein-DNA complexes (53-56) and is therefore well suited for studying the MutS-DNA system. In the next section, some of the key computational techniques referenced in this dissertation are introduced.

1.5 Computer Simulations

In 1958, Kendrew *et al.* published the first high-resolution protein structure of myoglobin (57) which, to some extent, contributed to the initial view that proteins were rigid rather than dynamic structures (58-59). Nearly 20 years later, McCammon *et al.* broke new ground by being the first

to capture the protein dynamics of the bovine pancreatic trypsin inhibitor (BPTI) at atomic resolution from an 8.8 picosecond (ps) MD computer simulation (60). Since then, MD simulations have played a valuable role in complementing experiment by providing molecular level insight into the dynamical motions of individual macromolecules (59, 61). Modern MD simulations, which treat atoms as being the smallest particle in the system, begin with defining the potential energy function, $U(\vec{R})$, which is traditionally made up of bonding terms and non-bonding terms and is written as a function of the Cartesian coordinates, \vec{R} (61-64):

$$\begin{aligned}
 U(\vec{R}) = & \sum_{\substack{\text{bond} \\ \text{lengths}}} K_b (b - b_0)^2 + \sum_{\substack{\text{bond} \\ \text{angles}}} K_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2 \quad (1.1) \\
 & + \sum_{\text{elec}} \frac{q_i q_j}{D r_{ij}} + \sum_{\text{vdW}} \left[\left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) \right]
 \end{aligned}$$

Bonding terms typically consist of bond lengths (b), bond angles (θ), dihedral angles (ϕ), and improper dihedral angles (ω) (Eq. (1.1)). All constants with the subscript 0 represent equilibrium values and K_b , K_θ , K_ϕ , and K_ω denote the respective force constants. The dihedral angle term is modeled as a sinusoidal function where n and δ are the periodicity and phase shift, respectively. The non-bonding terms consist of electrostatic and van der Waals interactions (Eq. (1.1)) where r_{ij} corresponds to the distance between atoms i and j . The electrostatic interactions are calculated between point charges q_i and q_j using a Coulombic

potential where D represents the effective dielectric function for the medium. The combined parts within the van der Waals term is often referred to as the Lennard-Jones 6-12 potential and accounts for the repulsion of atomic cores at short distances and for the attractive London dispersion forces. A more thorough explanation of these and other new or missing terms can be found in the following references (61-64).

Once the potential energy function is established, a molecular dynamics simulation can be initiated by solving Newton's equation of motion:

$$F_i = m_i a_i \quad (1.2)$$

where m_i and a_i are the mass and acceleration of atom i , respectively. F_i is the force acting on atom i and is computed from the gradient of the potential energy function, $U(\vec{R})$:

$$F_i = -\nabla_i U(\vec{R}) \quad (1.3)$$

Starting with the coordinates of a high-resolution crystal structure, the standard procedure for running an MD simulation typically involves first minimizing the structure in order to remove steric clashes and to relieve local strains within the structure. Next, initial velocities are randomly assigned to each atom from the Maxwellian distribution starting from a low temperature, T , and a_i is computed from the force as described above. Then, Newton's equation of motion (which is an ordinary differential equation with no analytical solution) can be solved through numerical integration using discrete steps in order to determine the new position of each atom, \vec{r}_i , at some time $t + \Delta t$. The Taylor expansion of the coordinate for a particular atom around time t can be written as:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i \Delta t + \frac{1}{2} \vec{a}_i \Delta t^2 + \dots \quad (1.4)$$

where $\vec{r}_i(t)$ is given and Δt should be some short time step (typically between 1-2 fs) that is smaller than the period of the highest frequency motion and chosen to ensure stability in the potential energy between each simulation step (61, 65). Several integrators exist for continuously integrating Newton's equations of motion (*e.g.* Verlet, leap-frog, velocity Verlet, etc.), but discussions of the pros and cons are beyond the scope of this dissertation and the reader is referred to the following resources (58, 65). Normally, to avoid creating local "hot spots" with high velocities, the simulation is gradually heated by applying new velocities from a corresponding Maxwellian distribution at a given temperature up to the target temperature. Once the system is fully equilibrated, the simulation is ready for its production run. The resulting production simulation trajectory will contain snapshots of the system collected from the full trajectory and any dynamic variable (*e.g.* angles, distances, energies, etc) can be measured and plotted as a function of the simulation time. More importantly, average values can also be calculated from the time series plots for comparison with experiment. Some of the most popular MD simulation programs currently available include CHARMM (63-64), AMBER (66), GROMOS (67), and NAMD (68).

Since the inception of computer simulations, there have been significant advances in computer hardware and software. In fact, in 2010, using a specially constructed, state-of-the-art machine called *Anton* (designed for producing extensively long simulation trajectories) (69), Shaw *et al.* revisited the historical BPTI protein and became the first group ever to simulate a protein in explicit solvent for a full millisecond (ms) (70), nearly 100 times longer than what was previously possible and more than 10^8 times longer than the original BPTI simulation. Extending

the simulations from microsecond (μs) to ms time scales for much larger systems is not trivial, but it opens the door for studying more complex biological phenomena. In this dissertation, both the CHARMM (63-64) and NAMD (68) simulation programs have been used, along with the CHARMM27/CMAP force field (71-73), in order to study the conformational dynamics of the *E. coli* MutS protein on sub- μs time scales (see Chapter 3 and 4).

Often times large barriers exist between two conformational states (*e.g.* open and closed states) which may not be effectively sampled by using straight MD due to the long simulation times required to observe these transitions. Thus, it may be necessary to employ enhanced sampling techniques such as Umbrella Sampling (US) (74) or the Hamiltonian Replica Exchange Method (HREM) (75) in order to overcome these barriers. In US, a carefully chosen restraining potential (or umbrella potential), $U_{umbrella}(\xi)$, is added to the potential energy function, $U(\vec{R})$, in order to bias the sampling towards a particular region along the reaction coordinate of interest, ξ , that otherwise would be rarely visited. The resulting biased probability distribution, $P_{biased}(\xi)$, can be unbiased according to:

$$P_{unbiased}(\xi) = e^{\beta U_{umbrella}(\xi)} \times P_{biased}(\xi) \times e^{-\beta f} \quad (1.5)$$

where, $\beta = 1/k_B T$ (k_B : Boltzmann constant and T : temperature) and f is a constant that comes from adding $U_{umbrella}(\xi)$ to $U(\vec{R})$. The corresponding free energy profile (or sometimes referred to as the potential of mean force (PMF)), $w_{unbiased}(\xi)$, can then be obtained from:

$$w_{unbiased}(\xi) = -k_B T \ln P_{unbiased}(\xi) \quad (1.6)$$

$U_{umbrella}(\xi)$ can take on any functional form but is usually chosen as a quadratic function:

$$U_{umbrella}(\xi) = K_{\xi}(\xi - \xi_0)^2 \quad (1.7)$$

where K_{ξ} is the force constant that controls the width of the umbrella potential and ξ_0 is the target equilibrium value along ξ . In this case, a sufficiently large value of K_{ξ} would lead to small deviations away from ξ_0 and consecutive overlapping simulations that progress ξ_0 incrementally along ξ between two states, where the structural output of the last simulation is used as input in the next simulation in a daisy chain fashion, ultimately leading to barrier crossing. These individually biased simulations (often called “windows”) can be easily unbiased according to Eq. (1.5) and a relative free energy profile (or PMF) along ξ can be constructed using the Weighted Histogram Analysis Method (WHAM) (76).

HREM is a method where multiple non-interacting copies of the same system, called replicas, each using a restraining potential with a different value of ξ_0 are generated and independently simulated for a given time. Periodically, replicas are compared with neighboring conditions (or neighboring values along ξ) and may be swapped based on a specific energetic criteria (75) (see Figure 1.3). The probability of accepting or rejecting an exchange follows the Monte Carlo Metropolis criterion, W :

$$\begin{aligned} W(X_i, E_m; X_j, E_n) &= 1 && \text{for } \Delta \leq 0; \\ W(X_i, E_m; X_j, E_n) &= \exp(-\Delta) && \text{for } \Delta > 0 \end{aligned} \quad (1.8)$$

where $\Delta = \beta \left\{ \left[E_m(X_j) + E_n(X_i) \right] - \left[E_m(X_i) + E_n(X_j) \right] \right\}$ and $E(X)$ is the potential energy of a system for a given configuration, X . After a number of exchange cycles, the HREM technique essentially promotes multiple instances of barrier crossing by multiple replicas and, similar to US, allows accurate relative free energies to be calculated. Both the US and HREM methods are applied to study DNA base flipping in MutS (see Chapter 4). Also, the general lack of understanding about how MutS distinguishes between homoduplex and mismatch DNA as it scans duplex DNA has motivated the development of a novel multidimensional path-based restraining potential. In Chapter 5, this new umbrella potential for studying DNA translocation is presented along with a discussion on generating PMFs along arbitrary reaction coordinates from multidimensional WHAM.

Normal mode analysis (NMA) is an effective computational method for deducing large-amplitude conformational dynamics. Unlike MD simulations where the computational cost of simulating slow conformational changes in large macromolecular assemblies becomes prohibitive as the number of atoms within the system increases significantly, NMA extracts biologically relevant motions (often represented by the lowest frequency vibrational modes) by approximating the potential energy surface of the system as a parabolic function around the potential energy minimum (65, 77-78). If we let X^0 be the equilibrium configuration comprised of N atoms and whose potential energy resides at a minimum, then the Taylor expansion of the potential energy function $E(X)$ around X^0 (truncated after the quadratic term for small displacements) can be written as:

$$E(X) = E(X^0) + \sum_{i=1}^{3N} \left(\frac{\partial E}{\partial X_i} \right)_0 (X_i - X_i^0) + \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 E}{\partial X_i \partial X_j} \right)_0 (X_i - X_i^0)(X_j - X_j^0) \quad (1.9)$$

However, since X^0 is the minimum of the potential energy function and the potential energy can be defined relative to X^0 , then both $\left(\frac{\partial E}{\partial X_i}\right)_0$ and $E(X^0)$ can be set to zero and Eq. (1.9)

reduces to:

$$E(X) = \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 E}{\partial X_i \partial X_j} \right)_0 (X_i - X_i^0)(X_j - X_j^0) \quad (1.10)$$

Thus, the potential energy surface is approximated by a harmonic function centered around the energy minimum and which is governed by the second derivatives in Eq. (1.10). The substitution of Eq. (1.10) into Newton's equation of motion yields:

$$F(X) = -\nabla E(X) \quad (1.11)$$

which can be rearranged and expressed in matrix form as:

$$m \frac{d}{dt^2} \begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_{3N} \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 E}{\partial X_1 \partial X_1} & \cdots & \frac{\partial^2 E}{\partial X_1 \partial X_{3N}} \\ \vdots & & \vdots \\ \frac{\partial^2 E}{\partial X_{3N} \partial X_1} & \cdots & \frac{\partial^2 E}{\partial X_{3N} \partial X_{3N}} \end{bmatrix} \begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_{3N} \end{bmatrix} = 0 \quad (1.12)$$

where $\Delta X_i = X_i - X_i^0$ and the $3N \times 3N$ matrix of second-order partial derivatives is commonly referred to as the Hessian. Diagonalization of the Hessian matrix produces a set of $3N$ eigenvalues and $3N$ eigenvectors which are directly related to the frequencies and normal modes, respectively. A more detailed derivation of Eq. (1.10) and a discussion of the methods for diagonalizing the Hessian is provided in reference (78).

Normally, for a system with N atoms, standard NMA is carried out in three steps: 1) the starting structure is extensively energy minimized using an appropriate force field; 2) The Hessian is calculated; and 3) The eigenvalues and eigenvectors (or normal modes) are computed by diagonalizing the Hessian. Of course, as with all approximations extra care must always be taken to make use of experimental information in order to properly pinpoint and assign the functionally relevant modes (79). The method of NMA is used in Chapter 2 to examine the biologically important motions in the human and *E. coli* MutS proteins.

In the following chapters, the results from several extensive computer simulations will be presented and carefully compared to experimental data. In addition, a newly developed path-based umbrella potential used for studying DNA translocation will be proffered.

Figure 1.1 Schematic representation of the methyl-directed DNA mismatch repair in *E. coli*. Post-replicative DNA mismatch repair begins when MutS scans the newly synthesized DNA in search of base-base mismatches or insertion/deletion loops. Upon mismatch binding, additional downstream repair proteins (MutL, MutH, DNA helicase, exonuclease, SSB) are recruited and the mismatch containing strand is excised. Following, DNA polymerase III and ligase resynthesize the missing DNA strand and the DNA is repaired. SSB has been omitted for clarity. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Figure 1.1

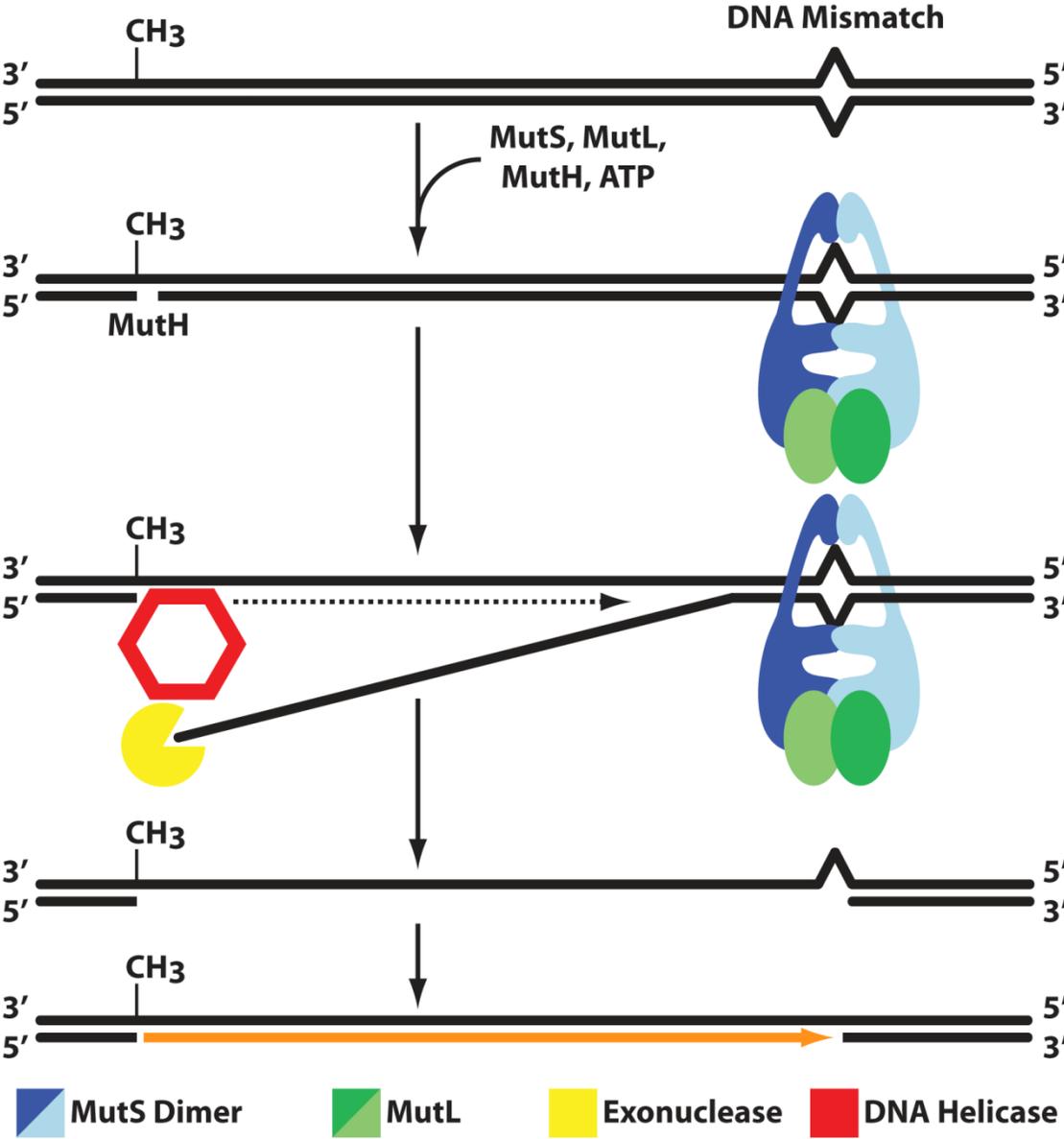


Figure 1.2 DNA mismatch recognition proteins (front and side views) bound to mismatch DNA. The DNA binding domains are colored red/pink, the connector domains are colored orange/pale orange, the core domains are colored yellow/pale yellow, the clamp domains are colored green/pale green, the ATPases are colored blue/pale blue, and the DNA is colored brown/tan. A) *E. coli* MutS (20). B) *T. aquaticus* MutS (21). C) Human MutSa (30). D) *E. coli* MutS dual ATPases bound by two ATP molecules (purple spheres) (20). The conserved helix-turn-helix is denoted by HTH.

Figure 1.2

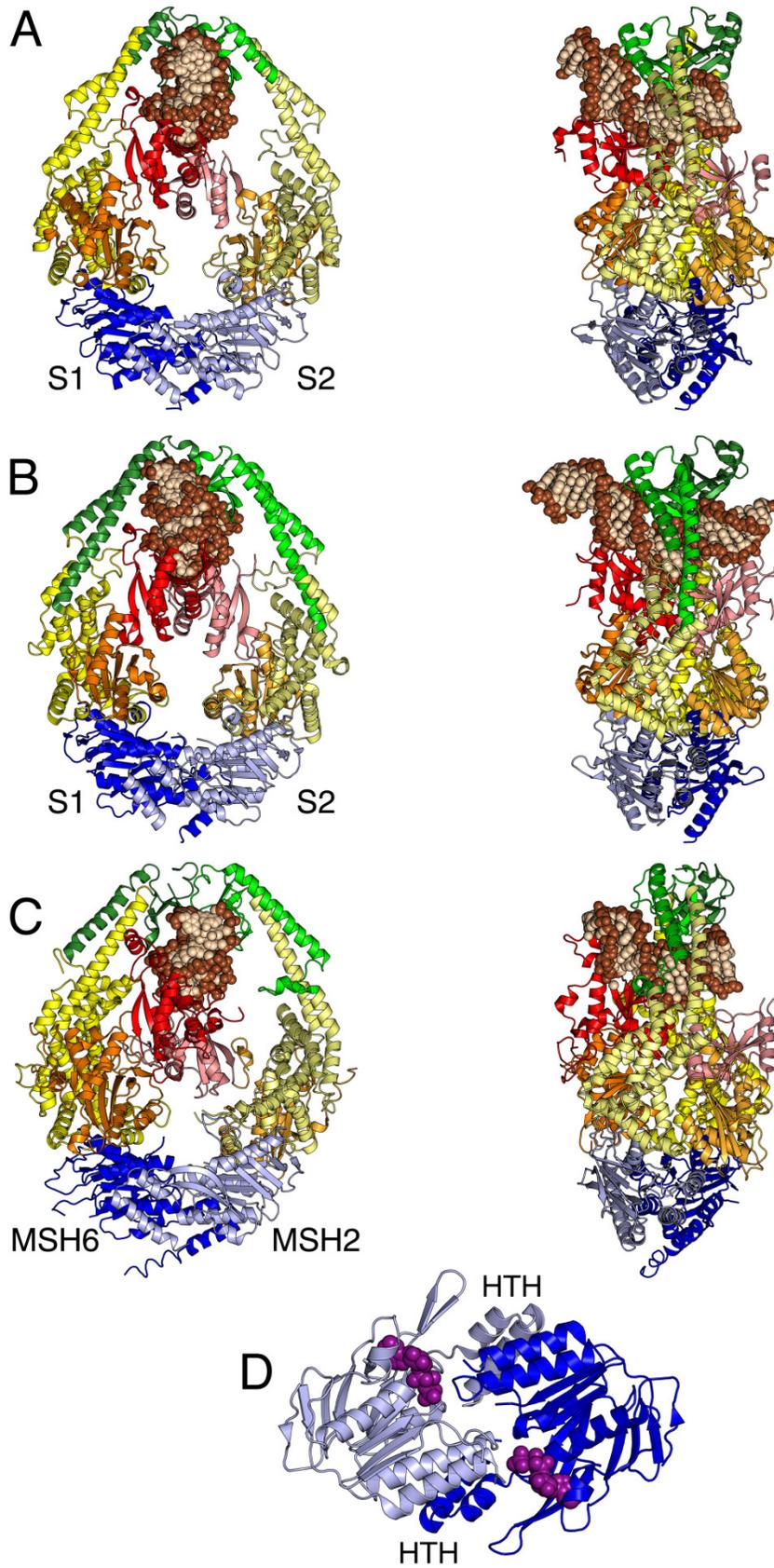
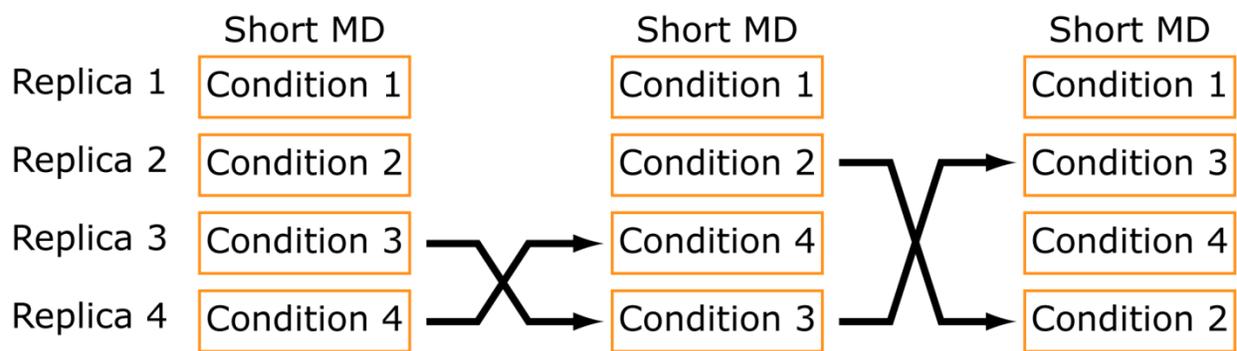


Figure 1.3 Schematic representation of the Hamiltonian Replica Exchange Method. Multiple simulations are coupled and conditions (or replicas) are exchanged in periodic intervals according to a Monte Carlo Metropolis criterion in order to enhance sampling along a specific reaction coordinate (75).

Figure 1.3



Chapter 2

Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal Mode Analysis

Shayantani Mukherjee, Sean M. Law, and Michael Feig

Adapted from

Biophys. J., 2009, **96**, 1707-1720

Sean Law contributed significantly to the analysis of the data from normal mode analysis and also produced several figures for the publication.

2.1 Abstract

Post-replication DNA mismatch repair is essential in maintaining the integrity of genomic information in prokaryotes and eukaryotes. The first step in mismatch repair is the recognition of base-base mismatches and insertions/deletions by bacterial MutS or eukaryotic MSH2-MSH6. Crystal structures of both proteins bound to mismatch DNA reveal a similar molecular architecture, but provide limited insight into the detailed molecular mechanism of long-range allostery involved in mismatch recognition and repair initiation. This study describes normal mode calculations of MutS and MSH2-MSH6 with and without DNA. The results reveal similar protein flexibility and suggest common dynamic and functional characteristics. A strongly correlated motion is present between the lever domain and ATPase domains, which proposes a pathway for long-range allostery from the N-terminal DNA binding domain to the C-terminal ATPase domains, as suggested from experimental studies. Detailed analysis of individual low frequency modes of both MutS and MSH2-MSH6 shows changes in the DNA binding domains coupled to the ATPase sites, which are interpreted in the context of experimental data to arrive at a complete molecular-level mismatch recognition cycle. Distinct conformational states are proposed for DNA scanning, mismatch recognition, repair initiation, and sliding along DNA after mismatch recognition. Hypotheses based on the results presented here form the basis for further experimental and computational studies.

2.2 Introduction

DNA mismatch repair (MMR) pathways maintain the integrity of genomic DNA by eliminating errors incorporated during replication and recombination. The initial steps of DNA-mismatch recognition and repair initiation in the post-replication MMR pathway are mostly conserved from bacteria to human with MutS in prokaryotes and MutS homologs (MSH) in eukaryotes recognizing defective DNA and initiating repair (16-17, 80). A functional MSH protein leading to correct mismatch recognition and subsequent deletion is especially important in humans for the avoidance of cancer phenotypes (81).

Prokaryotic MutS is comprised of monomers with identical sequence, termed S1 and S2, although it forms a structural heterodimer when bound to DNA (20-21). MutS is known to recognize base-base mismatches and short base insertions or deletions leading to their successful repair. In eukaryotes, at least 7 variants of MSH have been identified. They form a number of heterodimers, of which MSH2-MSH6 corresponds most closely to MutS (with MSH2 corresponding to S2 and MSH6 corresponding to S1) (80). Like MutS, the MSH2-MSH6 complex also recognizes base pair mismatches with high efficiency and single base insertions or deletions but does not efficiently recognize longer base insertions or deletions (16, 80).

After successful association of MutS or MSH2-MSH6 with a mismatch, a complex is formed in the presence of ATP with MutL in prokaryotes (7) or MutL homologs (MLH) in eukaryotes (82) in order to promote downstream repair events. Crystal structures of prokaryotic MutS from *Escherichia coli* (*E. coli*) (20), *Thermus aquaticus* (21) and human MSH2-MSH6 (30) bound to different base pair mismatches or a single thymine insertion/deletion have become available. The structures all show the same architecture with two main functional sites at

opposite ends of the dimer: a DNA binding site and an ATPase site. As evidenced by the crystal structures, the clamp and DNA binding domains (domains IV and I, respectively) from both chains (S1 and S2 of MutS or MSH6 and MSH2 of human) encircle the mismatched DNA (Figure 2.1). However, only the DNA-binding domain of one of the chains is in direct contact with the mismatch, giving rise to structural and functional asymmetry between the dimer moieties. Specific contacts with the mismatch base are made through a conserved ‘Phe-X-Glu’ motif in the DNA binding domain of chain S1 in MutS and MSH6. Insertion of this motif into the minor groove of the DNA is coincident with significant DNA bending ($\sim 60^\circ$) and minor groove widening at and around the mismatch site compared to canonical DNA. The bent conformation of the DNA is further stabilized through non-specific contacts from the clamp domain.

The nucleotide binding domains (domain V) reside on the opposite end of the protein with the ATP binding sites (ATPase sites) lying close to the dimerization interface. Biochemical studies have provided evidence for functional coupling between DNA scanning, mismatch recognition, repair initiation and ATPase activity (34, 44, 83-84), which suggests allosteric signaling within the MutS or MSH dimers. Each MutS ATPase domain, belonging to the ATP binding cassette (ABC) superfamily (85), is comprised of functionally important residues from both chains as shown in Figure 2.1C (29). The nucleotide binding site residing in each particular chain consists of Walker A and Walker B loops that are important for nucleotide phosphate binding and phosphate catalysis, respectively. Another loop containing a conserved phenylalanine residue (596 in MutS, 650 in MSH2, and 1108 in MSH6) stacks with the nucleotide adenine ring and the cavity is completed by the signature loop of the opposite monomer, which has been suggested to play an important role in catalysis. Several studies

suggest that the ATPase activities of the two chains are strongly correlated with each other and that they follow a sequential, rather than a simultaneous, pattern of ATP hydrolysis (44). Moreover, both sites show intrinsic asymmetry in the ATPase activity with nucleotide binding affinities changing significantly for each ATPase site during the recognition cycle (33-34, 44, 83-84). In free enzyme or when bound to regular DNA, the chain that contacts the DNA mismatch (S1 or MSH6) has a higher affinity for ATP compared to the other chain, while chain S2 or MSH2 binds mostly ADP (33, 84). It is further known that ATP hydrolysis is fast in S1/MSH6 when the protein is bound to regular DNA and that ADP release is the rate-limiting step (34). The ATPase site of the other chain has a much slower hydrolysis rate (84). These results highlight a differential behavior of the two ATPase sites when the protein is bound to regular DNA as depicted schematically in Figure 2.2. During scanning of regular DNA, the nucleotide-binding domain of chain S1/MSH6 binds ATP followed by fast hydrolysis to ADP. However, since exchange of ADP for ATP is not as fast as hydrolysis, ADP would be bound to this site for the majority of the time. At the same time, ADP is also bound predominantly to the other nucleotide-binding domain of chain S2/MSH2.

Experimental data suggests that mismatch binding promotes the exchange of ADP for ATP while stalling ATP hydrolysis of S1/MSH6 (34, 83). The resulting prolonged ATP bound state at S1/MSH6 'authorizes' recognition of a mismatch by the DNA binding domain, whereas ATP is readily hydrolyzed when the DNA binding domain is bound to regular DNA (24). Furthermore, stable ATP binding by S1/MSH6 ultimately leads to reduced ADP binding affinity in the ATPase site of S2/MSH2. This presumably enhances the ATP binding affinity of S2/MSH2 (84). The dual ATP bound state is believed to trigger a conformational change to a sliding clamp conformation where the mismatch is released by the DNA binding domain and

rebinding of mismatched DNA is inhibited (83-84). Interestingly, a recent single molecule study on MSH2-MSH6 has demonstrated that the sliding motion along DNA after mismatch recognition is independent of ATP hydrolysis (36).

While there is a general understanding of the long-range allostery of MutS and its homologs involved in recognition and repair initiation, the molecular-level events leading to the functional correlation between N-terminal DNA mismatch recognition and C-terminal nucleotide binding and hydrolysis have remained elusive. Further advances in this respect have been hindered by the fact that the available crystal structures only show the mismatch-bound state and do not provide information about the different nucleotide bound combinations in the two ATPase domains. Until now, the complex interplay between functional states of the two ATPase and DNA binding sites is mostly understood from biochemical kinetic studies that, on the other hand, fail to offer a molecular-level understanding of the process. While experiments may continue to reveal additional information for different functional states, conformational sampling of proteins can also be studied by theoretical means. Molecular dynamics simulations that often offer insights in this regard are not easily applicable to MutS because of the long time scales of the mismatch recognition process and the large system size of the MutS-DNA complex. Normal mode analysis (NMA) is an alternative strategy for studying large-scale conformational changes in biomolecules. NMA relies on a harmonic approximation of the potential energy surface around a minimum energy structure and the resulting lowest frequency dynamic modes often resemble biologically relevant functional motions (27-28). Here, we have applied NMA to study the conformational dynamics of MutS and its eukaryotic homolog MSH2-MSH6. The results suggest a new molecular-level understanding of the long-range allosteric pathway in the functional interplay between DNA mismatch recognition, nucleotide binding activity, and repair

initiation. Structural characterization of distinct conformational states along with the elucidation of a complete functional cycle offers possible avenues of validating the proposed cycle through experiments.

2.3 Methods

Normal mode calculations were performed on *E. coli* MutS and its human homolog, MSH2-MSH6, both in the absence of any bound nucleotides. Calculations were performed on each protein in the presence or absence of DNA, resulting in a total of 4 sets of normal mode calculations. They are referred to as MSH-DNA, MSH-free, MutS-DNA, and MutS-free for MSH2-MSH6 with DNA, MSH2-MSH6 without DNA, MutS with DNA, and MutS without DNA, respectively. Initial structures of *E. coli* and human protein were obtained from PDB IDs 1E3M (20) and 2O8B (30), respectively, and missing loops were constructed using Modeller (86). The structures were then extensively energy minimized using the CHARMM22/CMAP force-field (87) and distance dependent dielectric ($\epsilon = 4$). The root mean square deviation (RMSD) of the minimized structures with respect to the crystal structures were 2.08 Å, 2.42 Å, 1.28 Å and 2.06 Å for MSH-DNA, MSH-free, MutS-DNA, and MutS-free, respectively. Low RMSD values indicate that extensive minimization in absence of explicit water or DNA does not lead to significant structural deviations from the crystal structure. Normal modes were calculated using the block-normal mode approach using the VIBRAN module in CHARMM (88-89), version c33a2, and with the same force field as used for minimization. Only low frequency modes were analyzed in both proteins as those are the most relevant for describing functional motions involving the entire complex. In order to calculate the similarity between individual modes among MutS and MSH2-MSH6, we defined the overlap index for each pair of modes (i, j)

as: $\left| \left(\sum_k S_{ik} \cdot H_{jk} \right) / k \right|$; where k is the number of aligned residues of MutS and MSH, S_{ik} is the

k^{th} component unit vector of i^{th} mode of MutS and H_{jk} is the k^{th} component unit vector of

j^{th} mode of MSH. Each dot product contributing to the sum is between unit vectors and can possess a maximum value of 1 for residue pairs moving in exactly the same direction, or a value of -1 for residue pairs moving in exactly the opposite direction. The value of the overlap index can thus reach a maximum value of 1 for an ideal case where all aligned residues of two proteins are moving in exactly same or opposite direction. The sequence alignment between MSH2-MSH6 and MutS was taken from previous work (30). Molecular graphics were generated using PyMOL (90).

2.4 Results and Discussion

Normal mode calculations were carried out to explore the possible conformational dynamics of MutS and MSH2-MSH6 from the perspective of the crystal structures. Apart from conducting NMA calculations on both proteins with bound DNA, we have also considered proteins without DNA in order to allow dynamics beyond the DNA mismatch bound form. Results from the analysis of MutS and MSH2-MSH6 in the presence and absence of DNA are discussed below. Data from 4 different sets of normal mode calculations are referred to as MSH-DNA, MSH-free, MutS-DNA, and MutS-free for MSH2-MSH6 with DNA, MSH2-MSH6 without DNA, MutS with DNA, and MutS without DNA, respectively.

Flexibility of MutS and MSH2-MSH6 from Normal Modes and X-ray Data

Root mean square fluctuations (RMSF) provide information about inherent protein flexibility. They can be deduced from experimental B-factors or can be calculated from normal modes (91). The results in Figure 2.3 (*A-D*) show that the RMSFs calculated from experimental B-factors are uniformly high due to the limited resolution of the MSH2-MSH6 crystal structure (2.75 Å) and do not provide significant information about relative domain fluctuations. RMSFs from B-factors of the MutS crystal structure (with a resolution of 2.2 Å) are still high but indicate increased flexibility in MutS domains I, IV and parts of III, in particular, for chain S2. On the contrary, RMSFs calculated from the first 10 normal modes show significant differences in the domain movements. MutS-free and MSH-free exhibit large flexibility in the DNA binding and clamp domains (I and IV) and to a lesser degree between the lever domains (III). In contrast, the ATPase domains (V) show comparably low structural fluctuations. Mode calculations for MutS-

DNA and MSH-DNA provide qualitatively similar results, but with damped flexibility in the clamp domains and in the DNA binding domains of chain MSH6 and MutS S1. Figure 2.4 (A-D) show both proteins colored according to the B-factors calculated from normal mode RMSF values.

The NMA-based dynamics of MutS and MSH2-MSH6 are remarkably similar between chains as well as between the prokaryotic and eukaryotic enzymes. MSH6 and domain I of MSH2 appear to be slightly more rigid compared to MutS which may be related to the functional specialization of MSH2-MSH6. One may speculate that MutS requires increased flexibility to recognize both mismatches and longer insertions/deletions in contrast to MSH2/MSH6 which only recognizes mismatches and single base insertions/deletions. While an absolute comparison of RMSF values from a small number of normal modes between proteins of different size may be problematic, very similar results were obtained when motions of the first 100 modes were accumulated (data not shown).

Furthermore, it appears that MSH2 and MutS S2 are slightly more flexible than MSH6 and MutS S1, respectively, with the exception of the clamp domain for MSH-DNA and MutS-DNA systems. The increased flexibility in domain I of MSH2 or MutS S2 and decreased flexibility of the clamp domain compared to the other chain correspond to the structural asymmetry of MutS and MSH2-MSH6. Increased flexibility of domain I of MSH2/MutS S2 is probably due to the fact that they do not make considerable contacts with the DNA, while extensive DNA contacts of the clamp domains of MSH2/MutS S2 compared to the other chain accounts for its decreased flexibility. It was observed that the clamp domain of MSH2 and MutS S2 make 83 and 86 atomic contacts with the DNA, respectively, while that of MSH6 and MutS S1 make only 62 and 42 contacts, respectively. It was further observed that clamp and lever of

MutS S1 are more flexible compared to that of MSH6, which again is a result of less atomic contacts made by the clamp of S1 with the DNA than that of MSH6. The number of atomic contacts was calculated by considering protein heavy atoms around 5 Å of the DNA in the minimized structure of both proteins. Finally, Figure 2.4 highlights that the clamp and lever domains of MSH2 in MSH-DNA are slightly more flexible than the corresponding domains in MutS S2 of MutS-DNA. These differences are probably the result of the proteins being bound to DNA segments of varying lengths. MutS-DNA has a longer DNA (3 base pair steps more than MSH-DNA) which topologically constrains the mobility of MutS S2, giving rise to a comparatively rigid S2 clamp than that of MSH2. The portion of the lever domain of S2 tightly connected to the clamp also undergoes some degree of rigidification. The rigidification of MutS S2 may be more close to reality, as DNA undergoing repair in the cell is much longer than those observed in the crystal structures. It should also be mentioned that a substantial longer DNA can alter the extent of flexibility observed in the clamp of the other chain; namely S1 and MSH6. Hence, different level of flexibility of clamp and lever due to a much larger DNA cannot be directly inferred from these studies using fragmented DNA, except for the fact that an overall decreased flexibility of the clamps and levers will result in both chains when compared to DNA free systems.

Finally, it was observed that the RMSF for protein with DNA spikes at residues 1275 to 1281 in MSH6 and residues 663-666 in MutS S2. This is likely a manifestation of the tip effect (92) and is considered physically meaningless.

Correlated Motions in MutS and MSH2-MSH6

Covariance plots averaged over the ten lowest NMA modes were calculated to examine correlated motions in all 4 systems under investigation. The results shown in Figure 2.5 indicate similar overall correlations in MutS and MSH2-MSH6 in the absence and presence of DNA. Furthermore, both chains of MutS and MSH2-MSH6 show similar average correlation patterns, with only minor variations, despite the structural asymmetry of the complex. Common to all chains are correlations within each domain, reflecting rigid body domain motions, like correlations between adjacent domains II (connector domain) and III (lever domain), between III and V (ATPase domain), and between I (DNA binding domain) and II. While correlations within the same subunit are generally positive, correlations between dimer moieties are mostly negative with the exception of a strong positive correlation between the two ATPase domains and the two clamp domains as a result of dimerization.

The plots indicate high positive correlation between the lever domains and parts of the ATPase domains immediately adjacent to the lever including the ATPase binding sites. Experiments suggest the presence of long range allostery between the N-terminal DNA binding domain and the C-terminal ATPase domains, although a clear understanding of the allosteric pathway is missing. Strong correlations between the ATPase sites and lever domains highlight the propagation of signals within the two functional sites via the levers. Furthermore, the DNA binding domain in MSH6 and MutS S1 has a strong negative correlation with the ATPase domain in MSH2 and MutS S2, respectively, in particular for MSH-DNA and MutS-DNA, again suggesting conserved domain motions important for allostery.

Correspondence Between MutS and MSH2-MSH6 Modes

The analysis of RMSFs and motional correlations indicates that MutS and MSH2-MSH6 exhibit similar dynamic characteristics, both in the presence and absence of DNA. Furthermore, the correlation analyses from the first ten modes suggests dynamic coupling between DNA binding and ATPase activity. In order to explore this point in more detail, the ten lowest-frequency modes were individually compared between the same protein, i.e., MSH-free and MSH-DNA or MutS-free and MutS-DNA. The same comparison was also performed between different proteins, i.e., MSH-free and MutS-free and MSH-DNA and MutS-DNA. Table 2.1 shows the overlap indices calculated between any pair of modes from 4 different systems as described in the methods section. An overlap index value of 1.0 means that atoms move in identical directions in the two modes that are compared, a value of 0.0 means that motions are entirely orthogonal or that atom motions have zero amplitude. While a value of 1.0 or close to it is unlikely even for very similar structures, visual inspection of the MutS and MSH modes indicate that the motions are qualitatively similar when overlap indices are at 0.6 and above and, to a lesser but still substantial extent, when values are between 0.5 and 0.6, especially when MutS is compared to MSH. Relatively low overlap indices despite visually similar motions are due to uncertainties in the alignment between the two proteins with a sequence identity of only 21% and 24% for MSH2 and MSH6 (30), differences in structure, and significant overall flexibility due to the multi-domain nature of both MutS and MSH.

Table 2.1 shows that the highest degree of overlap on a mode-by-mode basis exists between MSH-free and MSH-DNA and also between MSH-free and MutS-free systems. There is a lesser degree of one-to-one correspondence between MutS-free and MutS-DNA and also between MSH-DNA and MutS-DNA, with individual modes being reordered more significantly

according to frequencies in these systems. High one-to-one overlap is found between modes 1, 2, 3, 4 of MSH-free and modes 1, 5, 3, 4 of MutS-free, between modes 1, 3, 4, 9, 10 of MSH-DNA and modes 2, 5, 4, 9, 10 of MutS-DNA, between modes 1, 2, 4, 5, 8, 9 of MSH-DNA and modes 2, 3, 4, 5 and 9, 10, 7 of MSH-free, and between modes 1, 2, 3, 5 of MutS-DNA and modes 2, 5, 3 and 9, 7 of MutS-free. It is apparent that many modes do not match on a one-to-one basis, but share common features with multiple modes (e.g. mode 2 of MSH-free matches modes 2, 5, and 8 of MutS-free). It is known from previous studies that complex domain motions responsible for altered functional states in a large protein are often better represented as a combination of low frequency modes. Our studies also suggest that the low frequency modes of both MSH and MutS exhibit almost similar domain flexibilities, while specific protein dynamics are often seen to occur as a combination of multiple modes showing different degrees of mode mixing in both proteins. The degree of mode mixing observed in this study will likely change as a result of using different force fields or coarse-grained models, but low frequency normal mode space will likely be conserved in all normal mode analyses, provided the starting structure remains the same. Thus, the main aim of this study is to highlight the conserved nature of domain motions in both proteins, rather than highlighting any specific mode or modes responsible for the protein function.

The presence of DNA alters the structural flexibility to some extent, as evident from the differences between modes in the presence and absence of DNA. For example, mode 1 is present in MSH-free and MutS-free, but not in MSH-DNA or MutS-DNA. As described in more detail below, the mode involves large motions of the clamp domains that are not possible in the presence of DNA. Visual inspection further reveals that altered motions of the clamp and DNA binding domains for structures in the presence and absence of DNA is a major factor in reduced

mode overlap indices between the two protein systems, despite otherwise similar overall motion. Interestingly, MSH modes are much more conserved in MSH-free and MSH-DNA systems than for the two systems in MutS. This suggests that protein flexibility is altered more in MutS than in MSH2-MSH6 through specific DNA interactions, especially near the DNA binding domain and clamps. This further reflects a more rigid overall structure in MSH2-MSH6 that is optimized to interact with mismatched DNA while MutS requires more structural flexibility to interact both with mismatched DNA or significantly distorted DNA structures with insertions or deletions.

In this study, we will focus more on modes from MSH-free and MutS-free since they are more likely to indicate motions from the known mismatch bound crystallographic structures towards alternate states during DNA scanning and mismatch repair. Comparison of the modes between MSH-free and MutS-free indicate that modes 1, 3, 4, and 9 from both complexes significantly overlap and may be considered equivalent. Modes 2 and 5 overlap significantly, suggesting that these modes are simply reordered with respect to their frequencies. However, there is also overlap between modes 2 of both complexes suggesting common features in both modes. Otherwise, there is significant mode overlap along the diagonal for modes 7 through 9 and additional limited off-diagonal overlap for modes 6 through 10. Overall, mode overlaps between MSH-DNA and MutS-DNA are lower but high overlap indices are again mostly limited to diagonal or near-by off-diagonal elements, with modes 1, 2, 3, 4, 5, 6, 9, and 10 of MSH-DNA corresponding to modes 1 and 2, 2, 5, 4, 6, 6, 9, and 10 of MutS-DNA.

Low-frequency Modes in MSH2/MSH6

Based on our analyses, the dynamic characteristics are largely conserved between MutS and MSH2-MSH6, both in the presence or absence of DNA. This is not surprising but it is also not trivial given the structural differences between the eukaryotic and prokaryotic enzymes and the slightly different biological functions. In the following, we will describe the lowest frequency modes in more detail with a focus on the modes of MSH-free.

The protein motion during each of the first 5 modes of MSH-free is shown in Figure 2.6. Close-up views of the ATPase domains of selected modes are shown in Figure 2.7. Initial visual inspection suggests the following general conclusions about the nature of domain motions in both MutS and MSH2-MSH6: 1) Most of the modes show an overall breathing motion of the DNA binding cavity involving the clamp and DNA binding domains. The parts of the clamp domains directly bound to the DNA backbone always show damped motion in proteins with DNA, although movements of other parts of the DNA binding cavity show a similar kind of breathing motion. Such an opening/closing motion of the DNA binding cavity corresponds to conformational transitions between a mismatch-bound state and scanning/sliding conformations where the interaction with DNA is presumed to be weaker. 2) Many modes show a correlation between opening/closing of the DNA binding cavity and alterations in the ATPase domain, in particular, the nucleotide-binding cleft. This finding establishes that MutS or MSH2-MSH6 is capable of allosteric communication between DNA binding and ATPase activity. The correlation between motions of the DNA binding domains and the ATPase domains varies as it may involve the MSH6, MSH2, or both ATPase domains in an alternating fashion. 3) A mode that affects the nucleotide-binding cleft in both ATPase domains in the same manner and at the same time is not observed in any of the 4 cases studied. This finding agrees with the experimental evidence that

ATPase activity in MSH2-MSH6 involves the two domains only in a sequential rather than simultaneous fashion (44).

The individual modes are described in detail in the following:

Mode 1 involves a wagging motion of the clamp domain along the direction of the DNA. The rotating motion around the core, apparent in the rest of the enzyme, results from a fixed center of mass. If the protein is aligned at domains I, II, III, and V, only the clamp domain IV moves in this mode. Mode 1 involves both chains to the same extent. The exact functional role of this mode is unclear but it may be related to the translocation of MSH2-MSH6 along DNA in the absence of mismatch when the clamps do not establish strong contacts with the DNA backbone. This mode is absent in both proteins bound to DNA mismatch, presumably due to residue contacts with the bent DNA.

Mode 2 consists of a partial opening/closing motion of the DNA binding site that is less pronounced than in some of the other modes. The unique aspect of this mode is an alternating opening/closing of the nucleotide binding clefts between the MSH2 and MSH6 ATPase domains (see Figure 2.6). It appears likely that this mode is involved in coupling MSH6 and MSH2 ATPase activity in a sequential fashion. As mentioned above, mode 2 in MSH-free has high overlap with mode 5 of MutS-free. However, this is due to similar motions in the DNA binding, clamp, and core domains. The alternating opening/closing of the two nucleotide binding sites is not present in mode 5 of MutS-free but is seen instead in mode 2 of MutS-free. An alternating ATPase movement correlated similarly to motions in the DNA binding cavity is also observed in mode 1 of MutS-DNA and MSH-DNA, suggesting that the inter-domain correlation is conserved in all systems.

Mode 3 couples opening of the DNA binding site with closing of the nucleotide binding cleft in MSH2. The opening of the DNA binding site is achieved by the movement of the clamp domains away from the DNA as well as the movement of the DNA binding domain in MSH6 out of the plane of the MSH2-MSH6 complex. Relative to the DNA, this motion moves domain I out of the DNA groove rather than along its helical axis. In the open form of this mode, most DNA contacts of MSH6 near the mismatch site are lost and the DNA can essentially slide freely relative to MSH2-MSH6. This mode is highly conserved in all other systems of MutS and MSH and is thus expected to play an important role in the protein's functional cycle. An almost identical mode is observed for mode 2 and 3 for MSH-DNA and MutS-DNA. We note, though, that the overlap index between the two modes in MSH-DNA and MutS-DNA is small due to altered clamp movements in MutS-DNA but otherwise they show similar domain motions.

Mode 4 is comprised mainly of a sideways motion of the clamp and part of the lever domains towards either the MSH2 or MSH6 side of the enzyme. This mode is asymmetric with respect to the overall complex. A symmetric version of this mode would result in clamp domain separation and lead to an open dimer where the clamp domains are far away from each other as proposed for the DNA-free complex from small-angle X-ray scattering (47). The symmetric mode is not observed, presumably due to limitations of the harmonic approximation in normal mode analysis.

Mode 5 involves closing of the DNA binding cavity that is coupled with opening of the nucleotide binding cleft in MSH6. The closing of the DNA binding cavity is achieved primarily by the motion of the clamp domains directly towards the DNA. A similar overall motion is also found in mode 2 of MutS-free, although the coupling between opening and closing of the DNA binding cavity with changes in the ATPase domain of MutS S1 is more pronounced in mode 6 of

MutS-free. It is likely that MutS-free achieves a motion equivalent to the MSH-free mode 5 through a combination of modes 2 and 6. A similar correlated motion between DNA binding cavity and ATPase domains is further observed in mode 5 of MSH-DNA and mode 4 of MutS-DNA.

Functional Cycle of MSH2-MSH6 and MutS from Normal Modes

The crystal structures of MutS and the MSH2-MSH6 complex only show the mismatch bound conformation. It is clear, however, that other functional states are involved during scanning of regular DNA, authorization of mismatch repair, and sliding of the enzyme along DNA during and immediately after repair before DNA scanning is resumed. On the molecular level, these different states are likely reflected in altered conformations of MSH2-MSH6 and MutS. X-ray crystallographic approaches have not identified alternate states of MSH2-MSH6, but there is evidence of alternating ATP and ADP bound states from small-angle X-ray scattering (47), where ATP binding has resulted in more compact protein conformations. The normal mode analysis presented here offers first insights into the functional dynamics of MSH2-MSH6 and MutS beyond the known DNA-mismatch bound crystal structures. Through a combination of the conserved low-frequency modes in both proteins, it is possible to propose, for the first time, a complete functional cycle of MSH2-MSH6 and MutS that is in full agreement with experimental observations. The proposed molecular-level picture of the cycle is illustrated in Figure 2.8 and described in detail in the following:

DNA binding: The functional cycle of MSH2-MSH6 and MutS begins with binding to newly replicated DNA. Experimental data suggests that DNA-free MutS is present in an open

form. Upon association with DNA the clamp domains are presumed to close. The asymmetric mode 4 indicates how the clamp domains might separate starting from the DNA-bound form without significantly affecting the structure of the rest of the enzyme.

DNA scanning and mismatch recognition: Once MSH2-MSH6 or MutS is bound to DNA, it will begin scanning for base mismatches. According to single molecule experiments, MSH2-MSH6 moves along regular DNA via one-dimensional diffusion (36), while DNA binding kinetics indicate that the protein is not bound strongly to DNA in the absence of a mismatch (16, 23, 93). In contrast, MSH2-MSH6 and MutS interact closely with mismatched DNA in a highly bent form as evidenced by the crystal structures (20-21, 30). The formation of highly bent DNA is greatly facilitated by the presence of base mismatches or base insertions/deletions (94) and is believed to be the main feature by which mismatch DNA base pairing is recognized (16). The transition from scanning to mismatch recognition is therefore expected to involve a significant change in the DNA binding domain from a relaxed conformation with relatively weak protein-DNA interactions to a tightened conformation where the enzyme holds on to highly bent DNA. The opening/closing motion of the DNA binding cavity in mode 5 of MSH2-MSH6 describes such a transition in molecular detail.

The transition from DNA scanning to mismatch recognition is coupled to the fast exchange of ADP to ATP and subsequent stalling of ATP hydrolysis in MSH6 according to kinetic experiments (34, 44, 83-84). Mode 5 couples closing of the DNA binding cavity to opening of the MSH6 or MutS S1 nucleotide binding cleft, and vice versa. The nucleotide binding cleft is sandwiched between the Walker A motif and a loop, which acts as a flap over the adenine moiety. This loop contains a conserved Phe residue (Phe596 in MutS, Phe650 in MSH2 and Phe1108 in MSH6) that stacks with the adenine ring in all available crystal structures.

Previous studies of ATP binding in some ATPases have revealed that binding often induces tightening of the site that is required for ATP hydrolysis as suggested by an increase in the hydrophobicity of the binding pocket (95) or closing of specific loops in the presence of the nucleotide resulting in a tightened cavity (96). We hypothesize that an open nucleotide binding cleft in the MSH2-MSH6 and MutS ATPase domains encourages ATP binding but inhibits hydrolysis. On the contrary, closing of the ATPase cavity predominantly involves movement of the loop bound to the adenine ring towards the catalytic center (Walker B motif), thereby ensuring successful ATP catalysis. Mode 5 therefore provides a molecular level picture of how mismatch recognition through deformation of DNA at the mismatch site might be coupled to the experimentally observed changes in MSH6 ATP activity.

In most of the crystal structures of MSH2-MSH6 and MutS, ADP or a non-hydrolyzable ATP analog is present in both or any one of the chains, while only one MutS structure with bound ATP on both chains has been reported so far (29). The difficulties of observing the protein with stable ATP bound MSH6/S1 may be attributed to crystal packing that does not allow the formation of ATPase domains that are truly catalytically inactive (29). Thus, the crystal structure with ATP at S1 may not be fully representative of the true non-hydrolysable state of the protein, but rather represent a different trapped intermediate state.

Initiation of repair: The next step after base mismatch recognition is initiation of the repair process. This involves binding of MutL/MLH to MutS/MSH2-MSH6 (16-17, 80) which then signals further downstream events. Furthermore, kinetic studies indicate that ADP exchanges for ATP in the MSH2 ATPase domain subsequent to ATP binding in the MSH6 ATPase domain (84). The sequential coupling of ATP binding to the two ATPase domains mirrors alternating ATP hydrolysis activity in other dimerized ATPase domains as in ABC

transporters (97) and can be understood in terms of the alternating opening/closing of the nucleotide binding clefts seen in mode 2. We hypothesize that the initially very open ATP binding site in MSH6 following mismatch recognition partially closes upon ATP binding which in turn leads to opening of the MSH2 ATP binding site according to mode 2 and subsequent exchange of ADP for ATP in MSH2. Mode 2 also involves structural rearrangement outside the ATPase domain indicating that the enzyme assumes a distinct conformation at this step of the functional cycle, possibly to facilitate MutL/MLH binding.

Repair and mismatch release: Recent experiments suggest that after the initiation of repair, MSH2-MSH6 and MutS form a mobile clamp state that slides along the DNA in search of downstream repair proteins (36, 40, 83-84, 98). A transition from the mismatch bound state to a sliding conformation requires re-opening of the DNA binding cavity to a form that still holds the DNA but is not competent to rebind mismatched DNA (83). Moreover, this sliding activity is not powered by ATP hydrolysis. Current understanding of this transition is unclear from any experimental studies. We propose that the most conserved mode in both proteins, i.e., mode 3, describes the molecular events involved in the formation of sliding clamp conformation. In mode 3, the DNA binding cavity is opened by the release of the clamps from the DNA coupled to a large motion of the DNA binding domain perpendicular to the DNA helix. As a result, intimate interactions with the mismatch through the DNA groove become impossible, in particular, interactions involving the highly conserved Phe-X-Glu motif that is known to interact specifically at the mismatch site (20-21, 29-30). The protein is capable of sliding along the DNA in this state. The release of DNA mismatch binding and sliding according to mode 3 is coupled to a tightening of the MSH2 ATP binding site which would facilitate eventual ATP hydrolysis in MSH2 and allow recovery of the DNA scanning mode.

Although, specific normal modes have been mentioned while describing the molecular events during scanning, mismatch binding and sliding clamp formation, it is likely that opening and closing of the DNA binding cavity is actually occurring as a result of multiple low frequency modes. This is even more likely as almost all of the low frequency modes studied, except mode 1 in MSH-free and MutS-free, exhibit some kind of breathing motion of the DNA binding cavity involving different domain motions like that of the clamps, levers and DNA binding domain. The specific modes used to describe the conformational changes in the functional cycle only show the necessary synchronization between the opening/closing of the DNA binding cavity and the ATPase cleft, and are thus used to describe the experimentally observed allosteric effects.

Validation Through Experiments and Further Simulation

The results presented from normal mode calculations make a number of predictions about the functional dynamics of MutS and MSH2-MSH6 and the existence of additional functional states that have not been characterized on a molecular level to date. In particular, this study proposes molecular level details of long range allosteric coupling between the N-terminal DNA binding domains and the C-terminal ATPase sites as well as coupling between the two adjacent ATPase sites, which are known to exhibit a sequential pattern of action. In addition, the results presented here provide an atomic level characterization of distinct states in the functional cycle of MutS and MSH2-MSH6. A more open DNA scanning conformation is proposed and a sliding clamp state is predicted where the DNA binding domain is rotated out of the enzyme to result in structures that are significantly different from the crystal structure. These predictions should stimulate further experimental and computational studies to validate the predictions made here.

In particular, structural experiments could probe the nature of the DNA scanning and sliding conformations based on the predictions presented here while biochemical studies may test mutations that would disrupt the proposed domain movements. Furthermore, the proposed structures for alternate functional states can be subjected to more extensive computational studies to examine their stability and transitions between those states.

2.5 Conclusions

Results from the normal mode calculations of MSH2-MSH6 and MutS are presented to develop a molecular level picture of distinct conformational states involved in their functional cycles. A comparison of the modes between MSH2-MSH6 and MutS reveals striking similarities, indicating that the two enzymes are not just structurally, but also dynamically and functionally equivalent on the molecular level. The most important result indicates the presence of a strong motional correlation between the ATPase domains and the lever domains in all low frequency modes analyzed, while individual modes highlight the specific nature of the correlation between the N-terminal DNA binding domains and the ATPase domains. This indicates that both MutS and MSH2-MSH6 are structurally capable of establishing long-range allostery during their functional cycle. Based on a detailed analysis of the lowest-frequency modes in the context of the available experimental data, a detailed mechanism is proposed that involves DNA scanning, mismatch recognition, repair initiation, and sliding of MSH2-MSH6/MutS along DNA before scanning is resumed.

Normal mode calculations can provide an approximate view of biologically relevant dynamics in biomolecules but are limited by the theoretical nature of the methodology. The ideas presented here suggest a number of experiments that could validate and extend the proposed mechanism of DNA mismatch recognition by MSH2-MSH6 and MutS. Furthermore, the normal mode results can serve as starting points for additional computational studies that may investigate the proposed functional states and transitions between them in more detail.

2.6 Acknowledgements

The authors thank Dr. Katarzyna Maksimiak and Mr. Hugh Crosmun for valuable contributions during the early stages of the project. Financial support from NSF CAREER grant 0447799 and the Alfred P. Sloan Foundation is acknowledged as well as access to computational resources at the High Performance Computing Center at Michigan State University.

Table 2.1 Overlap Index for a Pair of Modes, Each From Two Different Sets

Mode No.	1	2	3	4	5	6	7	8	9	10	
1	0.8	0.1	0.2	0.2	0.0	0.1	0.2	0.3	0.2	0.1	MSH-free (rows) vs. MutS-free (columns)
2	0.2	0.5	0.1	0.0	0.6	0.1	0.3	0.5	0.0	0.1	
3	0.0	0.3	0.8	0.3	0.2	0.2	0.1	0.0	0.3	0.1	
4	0.2	0.1	0.2	0.6	0.3	0.5	0.3	0.3	0.2	0.1	
5	0.1	0.5	0.1	0.1	0.5	0.4	0.3	0.2	0.1	0.2	
6	0.1	0.0	0.1	0.5	0.2	0.3	0.3	0.3	0.4	0.1	
7	0.2	0.2	0.1	0.1	0.3	0.0	0.5	0.3	0.3	0.5	
8	0.3	0.2	0.1	0.1	0.0	0.4	0.5	0.5	0.5	0.0	
9	0.1	0.2	0.2	0.2	0.2	0.4	0.1	0.0	0.6	0.3	
10	0.1	0.3	0.1	0.3	0.0	0.4	0.2	0.1	0.2	0.3	
1	0.5	0.6	0.4	0.0	0.0	0.0	0.0	0.2	0.0	0.1	MSH-DNA (rows) vs. MutS-DNA (columns)
2	0.3	0.5	0.4	0.1	0.1	0.2	0.1	0.3	0.1	0.2	
3	0.3	0.2	0.2	0.2	0.8	0.0	0.4	0.0	0.2	0.0	
4	0.3	0.2	0.5	0.6	0.2	0.3	0.0	0.1	0.2	0.0	
5	0.0	0.1	0.2	0.3	0.2	0.5	0.3	0.1	0.3	0.1	
6	0.3	0.2	0.0	0.2	0.3	0.5	0.0	0.1	0.2	0.4	
7	0.3	0.0	0.1	0.4	0.0	0.4	0.0	0.3	0.1	0.2	
8	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.0	0.2	0.4	
9	0.1	0.2	0.0	0.2	0.2	0.0	0.1	0.0	0.7	0.1	
10	0.1	0.0	0.2	0.1	0.1	0.2	0.4	0.4	0.1	0.6	
1	0.5	0.7	0.0	0.1	0.3	0.0	0.2	0.1	0.0	0.2	MSH-DNA (rows) vs. MSH-free (columns)
2	0.1	0.0	0.9	0.1	0.2	0.0	0.2	0.2	0.1	0.2	
3	0.5	0.4	0.1	0.2	0.4	0.4	0.3	0.4	0.2	0.2	
4	0.2	0.1	0.1	0.9	0.2	0.3	0.2	0.3	0.3	0.0	
5	0.2	0.0	0.2	0.0	0.6	0.2	0.4	0.1	0.7	0.0	
6	0.2	0.2	0.2	0.1	0.0	0.4	0.3	0.0	0.2	0.2	
7	0.0	0.0	0.2	0.0	0.1	0.5	0.2	0.2	0.4	0.1	
8	0.0	0.2	0.0	0.1	0.3	0.3	0.1	0.1	0.0	0.8	
9	0.0	0.1	0.3	0.1	0.1	0.2	0.7	0.2	0.2	0.0	
10	0.0	0.0	0.1	0.0	0.2	0.4	0.2	0.4	0.1	0.1	
1	0.5	0.6	0.4	0.3	0.2	0.3	0.3	0.4	0.1	0.0	MutS-DNA (rows) vs. MutS-free (columns)
2	0.4	0.3	0.2	0.3	0.6	0.1	0.2	0.4	0.1	0.2	
3	0.2	0.2	0.7	0.3	0.1	0.2	0.1	0.1	0.6	0.1	
4	0.1	0.2	0.4	0.2	0.0	0.5	0.3	0.5	0.3	0.4	
5	0.2	0.2	0.0	0.3	0.5	0.4	0.6	0.0	0.1	0.1	
6	0.3	0.1	0.2	0.5	0.1	0.1	0.2	0.3	0.0	0.2	
7	0.4	0.3	0.1	0.0	0.4	0.0	0.2	0.1	0.4	0.3	
8	0.1	0.1	0.2	0.1	0.3	0.2	0.3	0.3	0.2	0.2	
9	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.4	0.1	0.3	
10	0.2	0.2	0.2	0.1	0.1	0.2	0.0	0.1	0.3	0.3	

Values ≥ 0.5 are in bold; values ≥ 0.6 are also in italic.

Figure 2.1 Crystal structure of MSH2-MSH6 in front (A) and sideways (B) orientation. Protein domains are indicated in red (I, DNA binding), orange (II, connector), yellow (III, lever), green (IV, clamp) and blue (V, ATPase). DNA is indicated in light (base pairs) and dark (backbone) brown, while bound ADP molecules are magenta. Darker shades refer to MSH6; lighter shades refer to MSH2. Close-up view of the nucleotide-binding domain (C) highlights the Walker A motif in yellow, the Walker B motif in orange, and the signature loop in green.

Figure 2.1

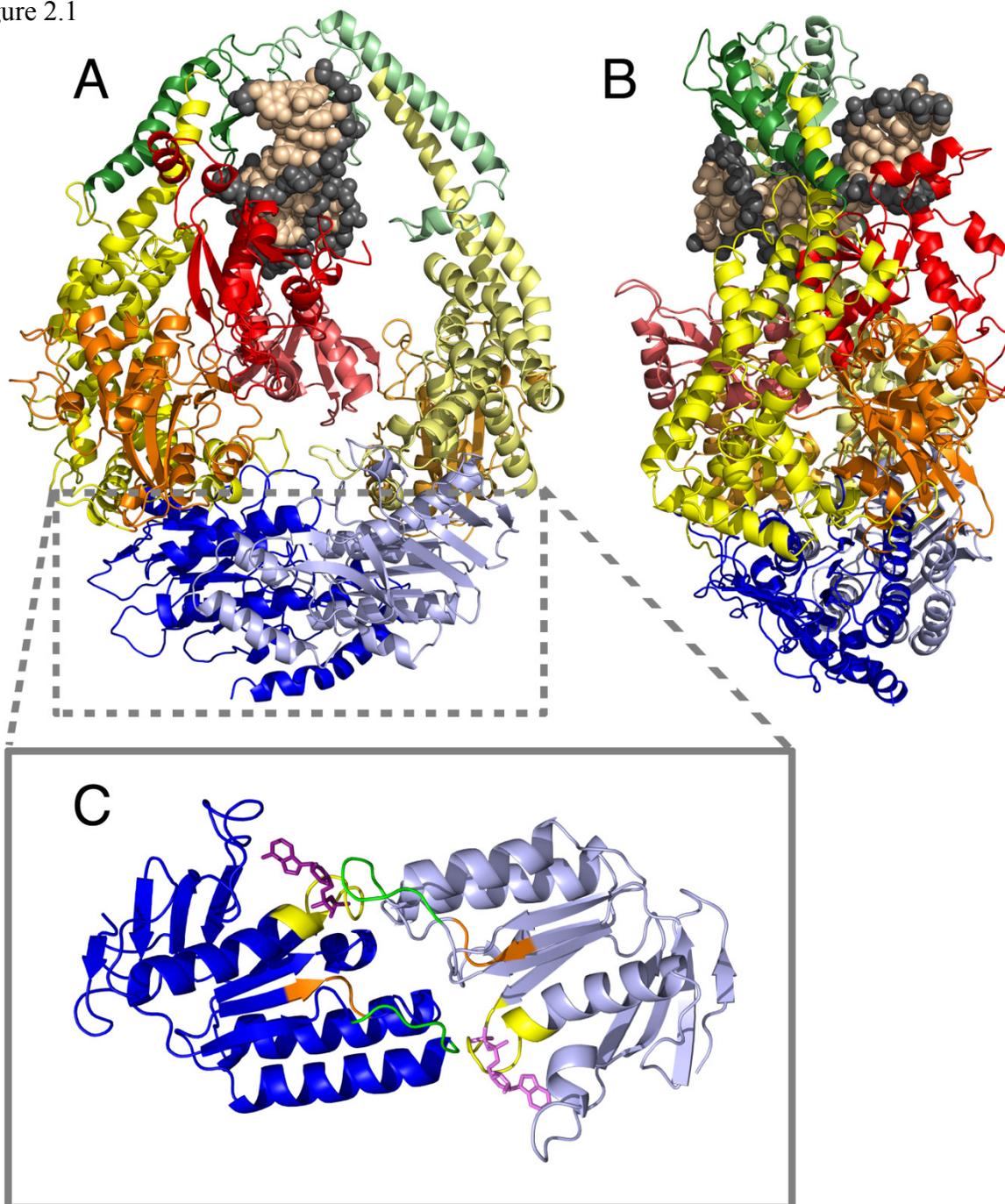


Figure 2.2 The dynamic behavior of the ATPase site in both chains are represented along an arbitrary horizontal time axis. Alterations among three possible nucleotide binding states (ADP/ATP/free) of the nucleotide binding domain are shown along the vertical axis with the help of curves that represent different hydrolysis pattern during functionally important phases of the protein. Functionally distinct states are colored as blue for scanning, pink for mismatch recognition and green for sliding phases.

Figure 2.2

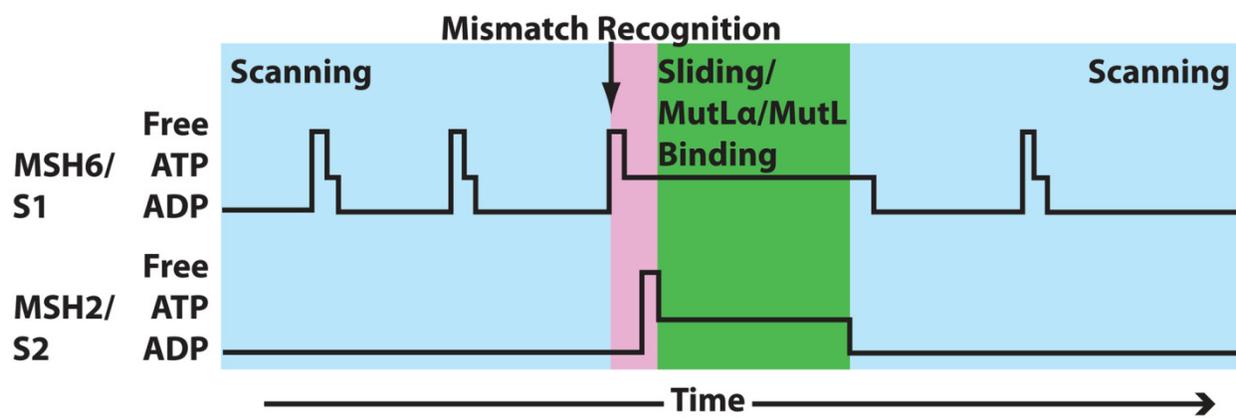


Figure 2.3 Root mean square fluctuation (RMSF) of C α atoms as a function of residue number calculated from the first 10 normal modes (red: without DNA, black: with DNA) and from crystallographic B-factors according to $RMSF_{Xray} = \sqrt{3B/8\pi^2}$ (blue) for MSH6 (A), MSH2 (B), MutS S1 (C), and MutS S2 (D). Discontinuities along the blue curve are due to missing residues in the crystal structures. Protein domains are indicated by colored bars with red, orange, yellow, green, and blue for domains I, II, III, IV, and V. Chain MSH6/MutS S1 are colored with dark shades while MSH2/MutS S2 are indicated by light shades.

Figure 2.3

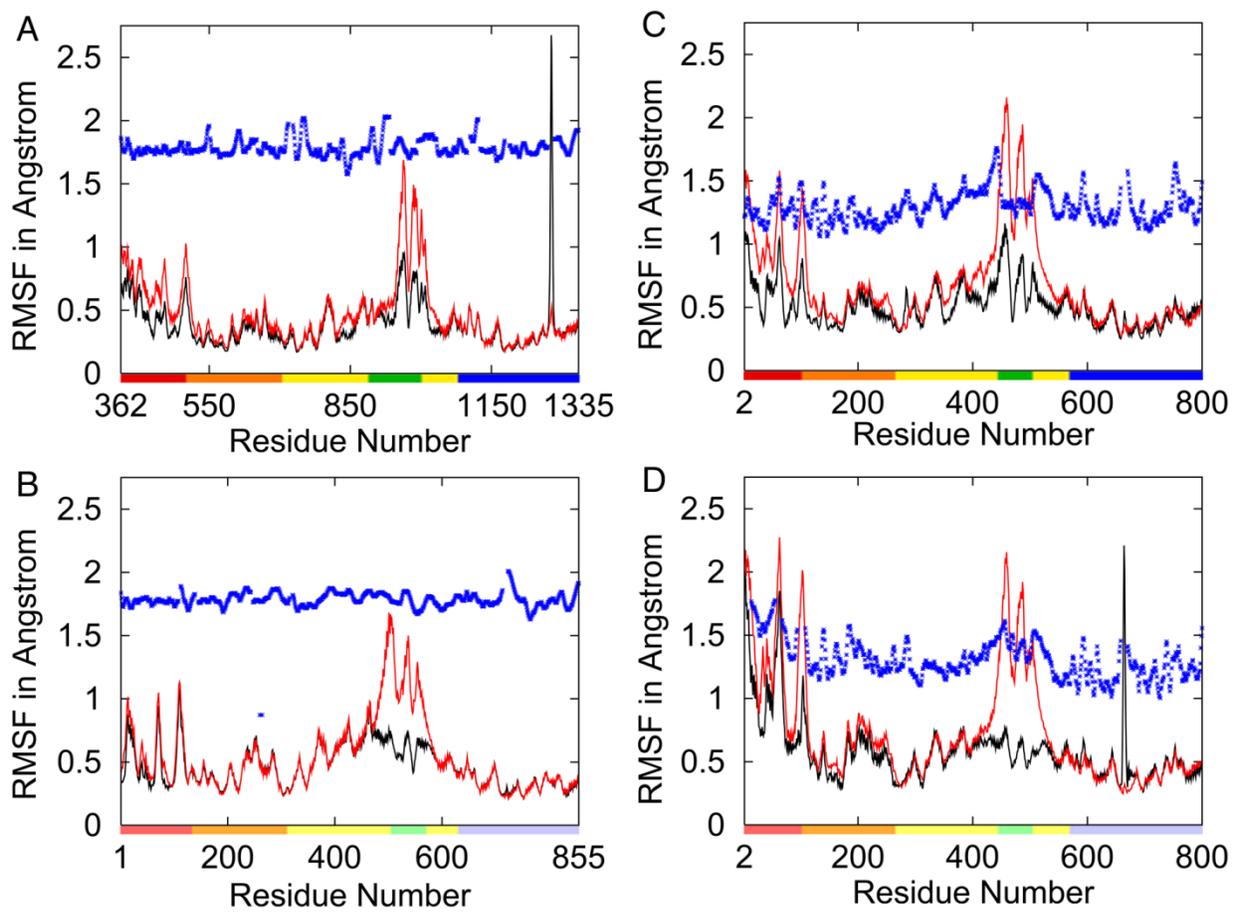


Figure 2.4 Protein backbone showing thermal fluctuations color coded by B-factors calculated from RMSF of the first 10 modes for MSH-free (A), MSH-DNA (B), MutS-free (C), and MutS-DNA (D). The color scale for B-factors is provided at the end of the figure.

Figure 2.4

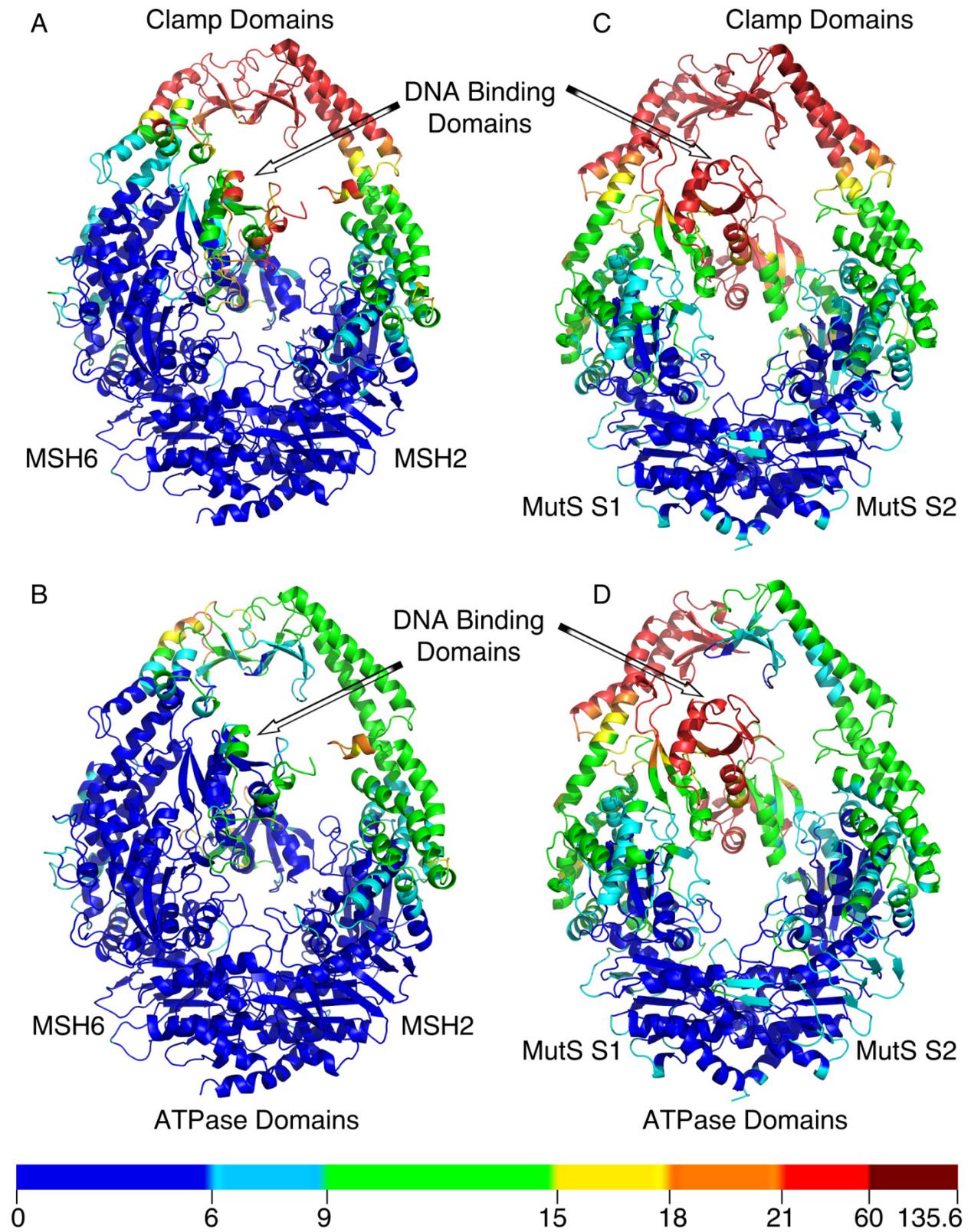


Figure 2.5 Average covariance from the first 10 modes in MSH2-MSH6 (A; upper triangle: MSH-free, lower triangle: MSH-DNA), and MutS (B; upper triangle: MutS-free, lower triangle: MutS-DNA). Protein domains in both chains are indicated by colored bars following the same color scheme as in Figure 2.3.

Figure 2.5

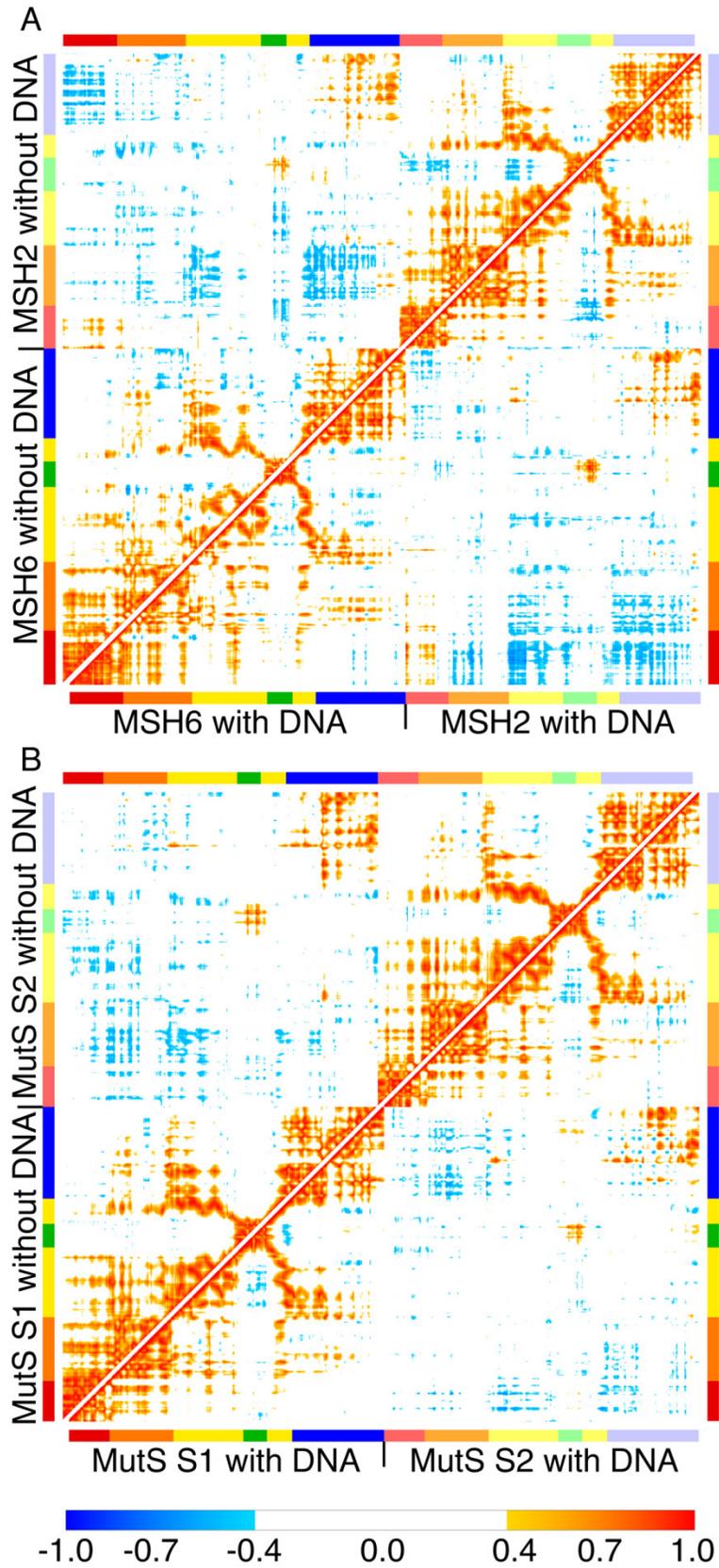


Figure 2.6 Mode motions of MSH-free projected on to the minimized crystal structure for modes 1 (A), 2 (B), 3 (C), 4 (D), and 5 (E). Motions are indicated by colored arrows in the direction of the mode vectors for every 6th residue. Motions involving the clamp, DNA binding, and ATPase domains are shown in green, red, and blue, respectively.

Figure 2.6

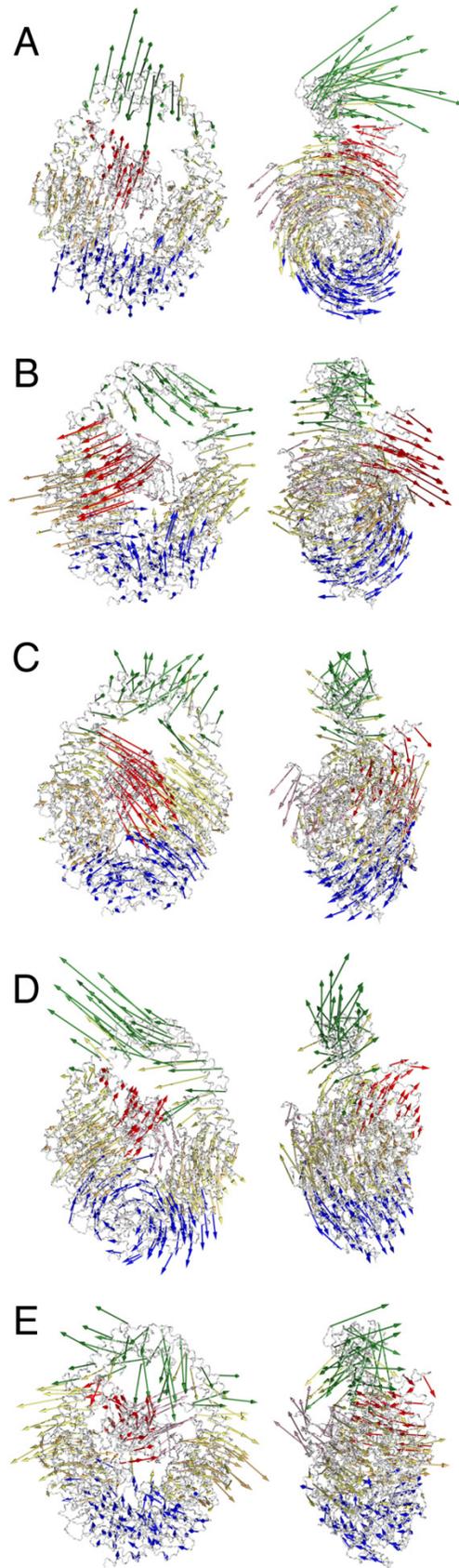
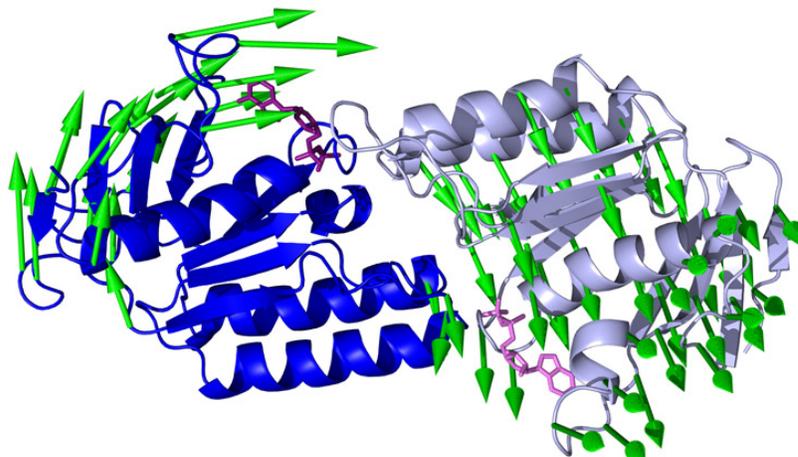


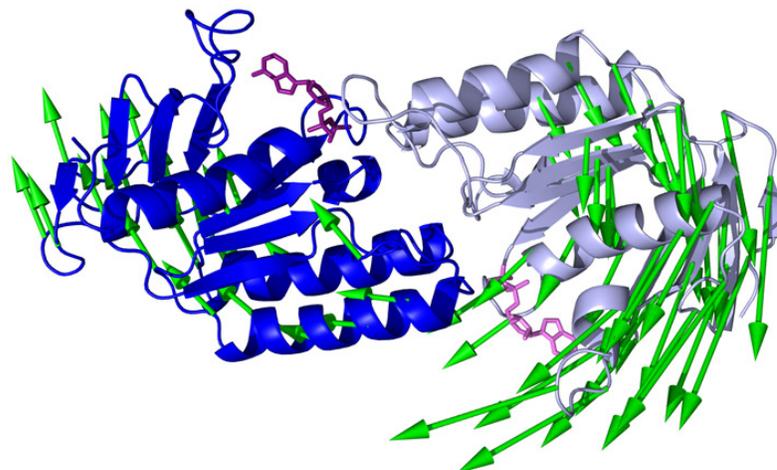
Figure 2.7 Close-up views of the motions in the nucleotide-binding domain of MSH-free during modes 2 (A), 3 (B), and 5 (C). Arrows are placed on each C α atom in every 3 consecutive residues and only displacements of more than 1 Å are shown. The two chains and bound nucleotides follow the same color scheme as Figure 2.1C.

Figure 2.7

A



B



C

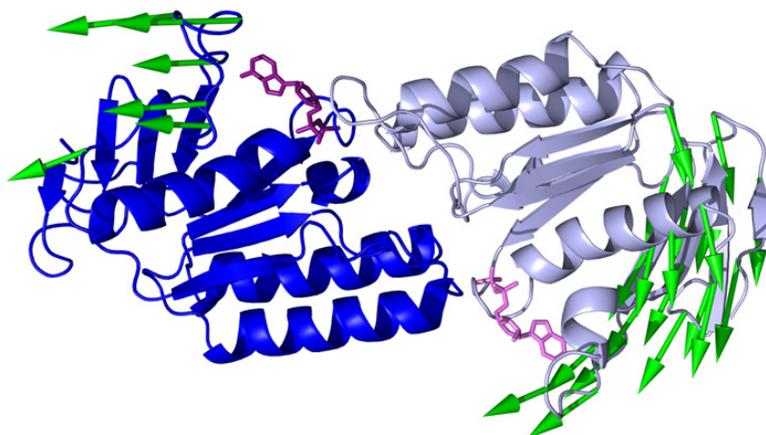
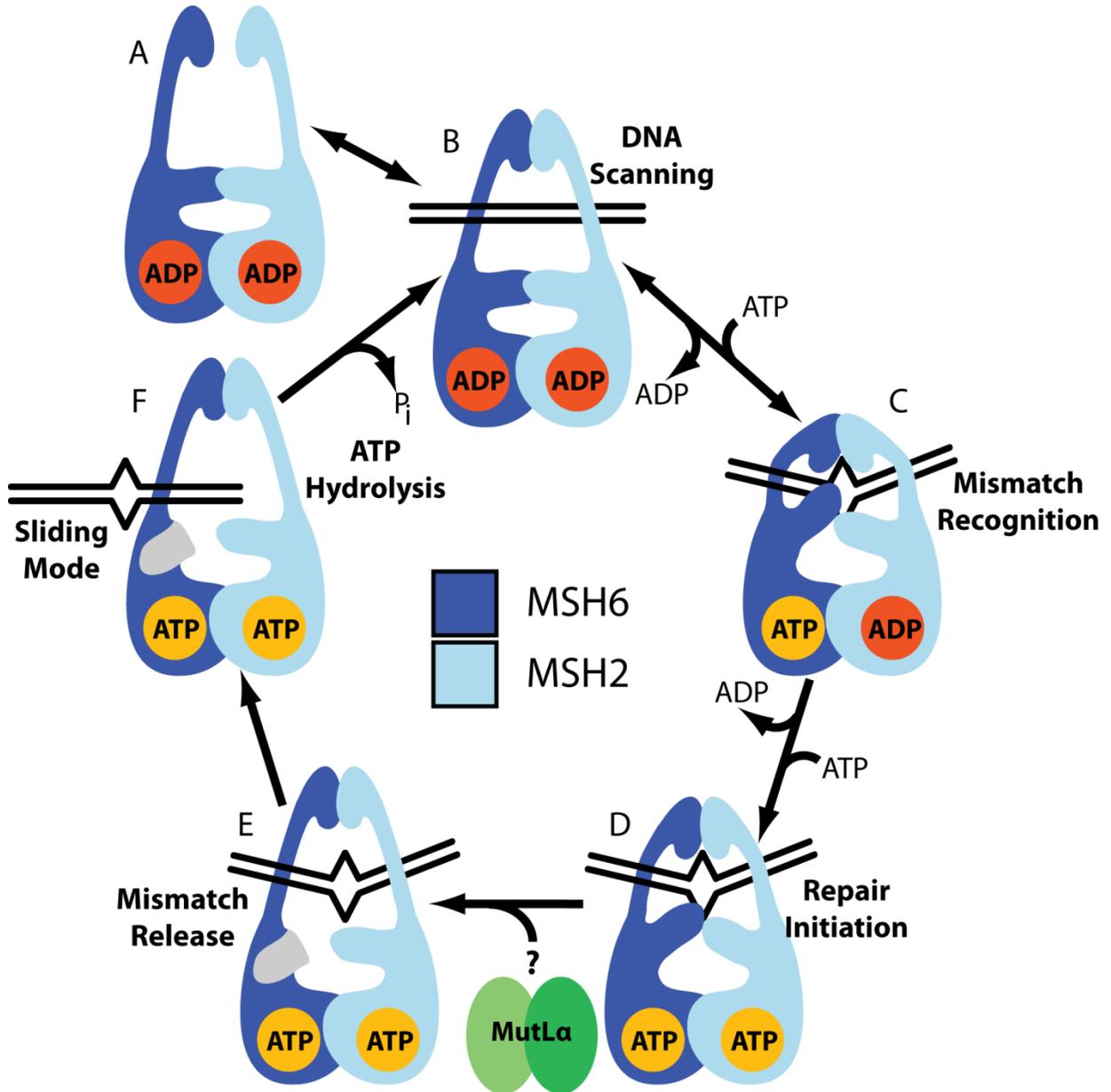


Figure 2.8 Schematic diagram representing distinct conformational states during the functional cycle of MSH2-MSH6 or MutS.

Figure 2.8



Chapter 3

An Analysis of *E. coli* MutS Dynamics from Molecular Dynamics Simulations

3.1 Abstract

Following up on the work presented in Chapter 2, this chapter serves to complement the normal mode analyses by examining the larger scale structural dynamics of the *Escherichia coli* MutS from nine independent 200 ns molecular dynamics simulations. Standard techniques were employed to measure the flexibility of each protein residue and to monitor changes in the protein secondary structure. Finally, using principal component analysis, two distinct principal modes that are likely to be linked to important protein function were identified. The first mode, found in eight of the nine simulations, describes the movement of the S2 DNA binding along the DNA helical axis towards the S1 DNA binding domain. This mode is hypothesized to be involved in ATP-hydrolysis independent movement of MutS along DNA. In the second mode, unique to the simulation with ATP in S1 and ADP in S2, both DNA binding domains moved upwards towards the DNA in a concerted fashion and resembled a DNA bending mode that bends the DNA upon mismatch recognition. The conformational changes observed in both modes also demonstrated coupled motions between the S2 DNA binding domain and the distant ATPase domains. Overall, the results are consistent with our previously proposed functional cycle for mismatch recognition and builds upon the individual states characterized in Chapter 2.

3.2 Introduction

In this chapter, new observations from nine independent simulations of the *Escherichia coli* MutS-DNA structure bound with different ADP and ATP nucleotides (Figure 3.1) are reported. Each system was derived from the 2.27 Å MutS-DNA crystal structure (PDBID: 1W7A) (29) which is bound to two ATP nucleotides. Unresolved coordinates within the structure (residues 660-667 in the S1 monomer) were generated using the loop modeling facility in MODELLER (99) and since MutS is a homodimer, the S1 subunit served as a template for completing missing residues in the S2 subunit. After being fully solvated, each of the nine systems contained over 165,000 atoms and was simulated for at least 200 ns using NAMD (68) along with the CHARMM27/CMAP all-atom force field (71-73) for a collective simulation time of 1.8 μs in the NPT ensemble. Each model is referenced by its nucleotide configuration using the “X:X” notation which corresponds to nucleotides that have been modeled into the S1 and S2 ATPases, respectively (Figure 3.1). For example, ADP:ATP has ADP in S1 and ATP in S2 while NONE:NONE contains no nucleotides in either ATPases.

The positional fluctuation measured from each simulation showed that the protein was generally more flexible than in the crowded crystal structure environment, but this extra mobility was not derived from changes in the major protein secondary structure elements (namely, α -helices and β -sheets). Due to the length of each simulation and complexity of the protein conformational dynamics, direct visualization of each trajectory only provided limited insight. Thus, the method of principal component analysis (PCA) was used to uncover the so-called “essential dynamics” within the protein (100). Similar to NMA, PCA filters out the dominant, functionally relevant (largest amplitude) modes from the local fluctuations in an MD simulation.

Applying PCA to our nine simulations, two new conformational transitions (or modes) were identified and found to be in good agreement with the previously proposed functional cycle for mismatch recognition (101). Additionally, both modes demonstrated coupled motions between the S2 DNA binding domain and the ATPase domains which support an allosteric signaling mechanism in the protein.

3.3 Methods

Analyses

The root mean square fluctuation (RMSF) and other basic distance measurements were calculated using CHARMM (v. c36a1) (64) and interfaced with the MMTSB Tool Set (102).

Crystallographic B-factors were converted to RMSF values by applying the $RMSF = \sqrt{\frac{3B}{8\pi^2}}$

relation. The protein secondary structure for each simulation snapshot was assigned for each monomer (S1 and S2) using the Dictionary of Secondary Structures in Proteins (DSSP) program (103). All molecular images were generated using PyMOL (104).

Principal Component Analysis

The essential dynamics for each simulation were ascertained via PCA (100) by first constructing and then diagonalizing the variance-covariance matrix of the C α positional fluctuations. This produces a set of $3N$ eigenvalues (λ_i) and corresponding eigenvectors (\vec{v}_i) that, collectively, describe the many different motions within a system. Typically, the eigenvalues are arranged in decreasing order such that the first vector, \vec{v}_1 , has λ_1 with the largest amplitude. This first eigenvalue-eigenvector pair is often referred to as the first principal component (PC1) and represents the motions within a given simulation that has the largest average amplitude. The level of similarity between \vec{v}_1 from different simulations was determined by calculating the inner dot product which should equal 1 when the proteins are moving in exactly the same way, 0 when the motions are not correlated, and -1 when protein motions are anti-correlated. All nine

trajectories (with frames extracted from every 10 ps of simulation time) were first superimposed onto the same starting structure and the principal components were calculated using the program Wordom (v. 0.21) (105-106). One method for identifying an eigenvector \vec{v}_1 that better represents multiple simulations is to combine the multiple trajectories into one, analyze the combined trajectory using PCA, and then compare this new PC1 with those obtained from the individual simulations (107-108). Individual trajectories with related motions have been shown to have eigenvalues and eigenvectors that are similar to the combined trajectory (107-108). Thus, the nine trajectories were concatenated into one, analyzed as described above, and compared with \vec{v}_1 and λ_1 from the individual simulation models.

3.4 Results

Protein Secondary Structure Propensities

The evolution of the protein secondary structure for each simulation was monitored using the DSSP program (103). While differences existed between the simulations (mostly attributed to localized changes in highly flexible loop, turn, and bend regions), the two major secondary structure elements, α -helices and β -sheets, remained nearly unchanged. Figure 3.2 shows the propensity of β -sheets and α -helices determined from each simulation and for each residue of each monomer. With the exception of a short β -strand located near residue 268 which lies unimportantly on the periphery of a larger β -sheet, all of the β -sheets that existed in the crystal structure remained present in all nine trajectories for nearly 100% of the simulation time (Figure 3.2A). The minor break located between residues 750 and 760 corresponds to a flexible β -hairpin connecting two flanking β -strands. Differences in the β -sheet propensity between S1 and S2 are essentially indistinguishable. Similarly, Figure 3.2B shows that all of the α -helices that were present in the crystal structure remained intact in both monomers for the all of the simulations with the exception of two short helices near residue 480 (located at the tip of the clamp domains) and residue 734 (located near the base of the ATPases) both of which are unstable due to being solvent exposed. Visual examination of the crystal structure showed that the two helices positioned near residue 188 and residue 589 were originally present as an intermediate between a 3/10-helix and an α -helix and therefore was not classified as being novel. Overall, both the S1 and S2 monomers shared nearly identical helical propensities.

Principal Component Analysis Reveals Unique Motions in the S2 DNA Binding Domain

Figure 3.3 shows the percent contribution of the top 20 eigenvalues, λ_i , (arranged in decreasing order from each of the nine simulations) plotted against the eigenvector index, i (for $i \leq 20$). In all of the simulations, the first eigenvector, \vec{v}_1 , accounted for at least 73% or more of the fluctuations while \vec{v}_2 contributed less than 15%. Table 3.1 shows the overlap between \vec{v}_1 for the different simulations calculated from the inner dot product between any given pair of eigenvectors. The \vec{v}_1 from ATP:ATP (the same nucleotide configuration as the 1W7A crystal structure) had the highest average overlap of 0.74 while the \vec{v}_1 from ATP:ADP had the lowest overlap of 0.45 with the other eight simulations. \vec{v}_1 from the combined trajectory had the maximum average overlap of 0.82 meaning that, with the exception of \vec{v}_1 from ATP:ADP which only had an overlap of 0.58, the motions captured by this eigenvector also accurately and best describe the essential motions involved in each of the individual simulations. Although, \vec{v}_1 from ADP:NONE only had an intermediate level of overlap (0.70) with PC1 from the combined trajectory, visual comparison of ADP:NONE showed that \vec{v}_1 and PC1 had essentially the same motions. Since \vec{v}_1 from ATP:ATP showed the highest overlap (0.93) with PC1 from the concatenated trajectory, the $C\alpha$ atomic displacement for ATP:ATP that corresponds to the largest amplitude movement along PC1 was mapped onto the starting simulation structure in order to visualize the extent of the motions identified in the projection (Figure 3.4). This “porcupine plot” (which only displays $C\alpha$ displacements greater than 3 Å) serves as a good representative for the conserved movement observed in eight out of the nine simulations (excluding ATP:ADP). Parts of the S1 DNA binding domain appear to move slightly upwards towards the DNA in the direction of the mismatch site while the entire S2 DNA binding domain, in a concerted motion with the S2 connector domain, slides under and along the DNA helical axis towards the

relatively stationary S1 DNA binding domain. There is also some minor movement of parts of the clamp domains that interact with the DNA. As well, the S1 core domain moved by a few Angstroms towards the center of the protein while residues in the ATPase dimer interface, which includes a conserved helix-turn-helix motif (see Figure 3.5), moved downwards and away from the base of the protein.

For the ATP:ADP simulation, which showed limited overlap with PC1 from the concatenated trajectory (Table 3.1), a similar “porcupine plot” showing the largest amplitude movement along its own first principal component, \vec{v}_1 , was also mapped onto the starting simulation structure (Figure 3.6). The most unique characteristic about this mode is the significant upward movement of the S2 DNA binding domain towards the DNA. Additionally, this eigenmode showed conformational changes in the S2 connector, S1 DNA binding domain, clamp domains, and ATPases domains (see Figure 3.6) that were similar to the PC1 captured by the concatenated trajectory (which likely accounts for the 0.58 overlap in Table 3.1) but lacked the previously observed motions in the S1 core. \vec{v}_1 from ATP:ADP was also compared with \vec{v}_2 and \vec{v}_3 from the other eight simulations and did not demonstrate any significant overlap (with an average overlap of about 0.2 in both cases). Thus, this mode appears to be exclusive to ATP:ADP.

3.5 Discussion

Protein Flexibility

The C α root mean square deviation (RMSD) for all nine trajectories was previously assessed and the protein was found to be quite stable for a system of this size and simulation length (109) (see also Figure 4.3 in Chapter 4). However, the overall RMSD is not always a good indicator of the local protein flexibility. Instead, the C α RMSF calculated from a simulation is often used to better understand the extent of the local protein dynamics and offers a direct method for comparison with crystallographic B-factors (see Materials and Methods). The RMSFs derived from the simulations were found in general to be higher than the crystal structures (Figure 3.7). Normally, RMSF differences between the simulations and X-ray structures can be attributed to incomplete sampling due to overly short simulation lengths and is exemplified by simulated RMSF values that are lower than those derived from the crystal structure (46). However, considering that our simulations are each over 200 ns long and that they exhibit higher mobility on average, it is more likely that the X-ray structures are restricted from sampling alternate conformations due to crystal packing forces (46).

α -Helix and β -Sheet Propensity

MD simulations can often be used to successfully monitor the evolution of secondary structure elements (110-112). Surprisingly, the β -sheets and α -helices that were present in the crystal structure (Figure 3.1) were essentially maintained throughout the trajectories and the propensities of these dominant secondary elements were extremely high (Figure 3.2). When interpreting these

results, it is also important not to underestimate the CHARMM27 force field effects which has been shown to over-stabilize α -helices (113). Nonetheless, the conservation of the two dominant secondary structure elements on sub- μ s time scales suggests that the mobility in the $C\alpha$ atoms may result from domain-level conformational changes.

Principal Component Analysis

The predominant large scale motion sampled in each trajectory and in the combined trajectory was measured using PCA. With the exception of ATP:ADP, the overlap (dot product) measured between \vec{v}_1 from the combined trajectory and \vec{v}_1 from the other eight simulations was remarkably high (see Table 3.1) which implies that PC1 from the combined trajectory is a good representative of the dominant motion observed in eight of the nine simulations. Figure 3.4 demonstrates the extent of movement from the representative ATP:ATP simulation (which showed the highest overlap) projected along PC1 from the combined trajectory. The largest and most interesting conformational change in the protein comes from the S2 DNA binding domain which moves under and along the local DNA helical axis towards the S1 DNA binding domain in a manner that resembles a mode for sliding along DNA. However, any substantial movement of the DNA would first require the S1 DNA binding domain to relinquish its contacts with the mismatched DNA. In fact, one of the modes from NMA demonstrated the transition of the MutS protein from a repair initiation mode to a sliding clamp conformation which involved opening of the DNA binding cavity and movement of the S1 DNA binding domain away from the mismatched DNA (101). The combination of this NMA mode followed by the movement of the S2 DNA binding domain as described by PC1 could very well account for the conformational

changes necessary for the experimentally observed ATP-hydrolysis independent movement of MutS along DNA after mismatch recognition (39-40). Finally, PC1 was also compared with the 10 lowest frequency modes from our NMA findings (101) and was found to overlap most with the sliding clamp mode, although, the overlap was only about 0.3 (too small to suggest any significant relationship between the modes).

The similarities in the conformational sampling amongst the different simulations were also assessed by clustering the structures based on its C α RMSD ((109) and see also Figure 4.4 in Chapter 4). All nine simulations were successfully grouped into six overlapping clusters, where one of the six clusters was visited only by ATP:ADP which suggested that the ATP:ADP simulation samples slightly different conformations from the other eight simulations. Analysis of the ATP:ADP trajectory using PCA confirmed that the first principal component from this simulation was unique and showed little to no overlap with the \vec{v}_1 , \vec{v}_2 , and \vec{v}_3 from the other eight simulations. Figure 3.6 shows the extent of movement of the protein along \vec{v}_1 from ATP:ADP. The most obvious difference is the large movement of the S2 DNA binding domain which, instead of moving along the DNA helical axis, moves upwards towards the DNA in what resembles a DNA bending action. Interestingly, the ATP:ADP nucleotide configuration was previously identified from NMA as the mismatch binding mode (101). However, no DNA bending mode was identified from NMA and comparison of the DNA bending mode from ATP:ADP with the 10 lowest frequency modes from NMA only showed a maximum overlap of 0.1. DNA bending in MutS has been studied experimentally using atomic force microscopy (48), but the exact mechanism by which the protein bends the DNA is largely unknown. Thus, it is proposed that upon mismatch binding by the S1 DNA binding domain, the S2 DNA binding domain pushes upwards on the DNA and bends it as depicted in Figure 3.6.

In both the DNA sliding and DNA bending modes identified from PCA, the movement of the S2 DNA binding domain was coupled to an elongation of both ATPases (see Figure 3.4 and Figure 3.6). The two nucleotide binding sites and both conserved C-terminal helix-turn-helix motifs (HTH, residues 766-800) (20, 35) were seen moving downwards and away from the base of protein, thereby slightly elongating the MutS structure. We note that this elongation is largely dependent on the movement of the HTH motif (Figure 3.6) which has been previously demonstrated to attenuate ATPase activity and affect dimerization upon being disrupted (35). Therefore, it can be speculated that the elongation of the HTH motifs could play a role in modifying the distance between the ATPase dimer interface which would directly affect ATPase activity since ATP hydrolysis requires that the opposing dimer move closer in order to complete the active site. Thus, an increase in distance between the subunits (governed by the movements of the HTH motifs) would likely inhibit ATP hydrolysis.

3.6 Conclusions

The complex nature of the of the DNA mismatch recognition cycle involves numerous conformational changes in the MutS protein that are dependent upon the bound nucleotides as well as the presence or absence of a DNA mismatch. We have investigated the dynamics of the MutS protein by applying the PCA method to several long simulations of the MutS-DNA complex. Overall, the essential dynamics of MutS derived from our nine simulations complements the NMA work presented in Chapter 2. The observation of coupled motions between the DNA binding domain and the distant ATPases strengthens the case for an allosteric signaling pathway within the protein. More importantly, the identification of two novel conformational modes, one for DNA sliding and one for DNA bending, provides new insight into the DNA mismatch recognition process.

Table 3.1 Pair-wise Overlap of the First Eigenvector From the Nine Simulations and the Combined Trajectory

	Simulations								Combined Trajectory
	NONE:ATP	ATP:ATP	ADP:NONE	NONE:ADP	ADP:ADP	ADP:ATP	ATP:ADP	NONE:NONE	
ATP:NONE	0.78	0.78	0.53	0.67	0.70	0.76	0.28	0.91	0.88
NONE:ATP		0.83	0.63	0.71	0.71	0.73	0.41	0.77	0.89
ATP:ATP			0.68	0.83	0.77	0.70	0.52	0.79	0.93
ADP:NONE				0.62	0.45	0.45	0.52	0.50	0.70
NONE:ADP					0.68	0.59	0.63	0.70	0.86
ADP:ADP						0.67	0.50	0.75	0.86
ADP:ATP							0.46	0.75	0.82
ATP:ADP								0.31	0.58
NONE:NONE									0.90

Figure 3.1 The MutS-DNA structure (center) with the different nucleotide-bound conformations (surrounding).

Figure 3.1

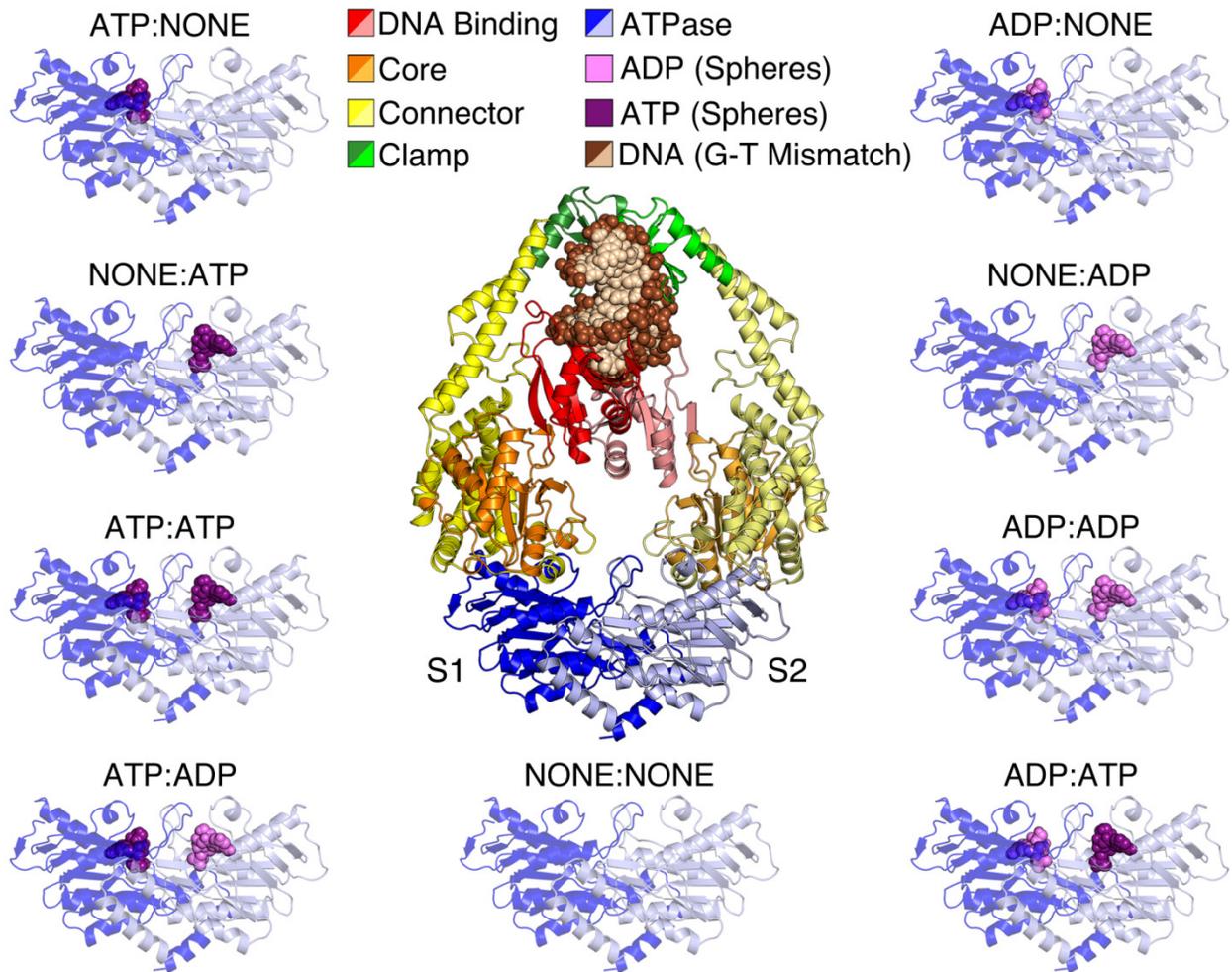


Figure 3.2 Secondary structure propensity for the S1 (red) and S2 (blue) monomers from each simulation. A) β -sheet propensity. B) α -helix propensity. The simulation order shown in A) is the same in B). The colored bar located below each plot corresponds to the five different structural domains and is colored according to the legend in Figure 3.1 and the broken black bars above each plot corresponds to the same type of secondary structure elements that were present in the crystal structure.

Figure 3.2

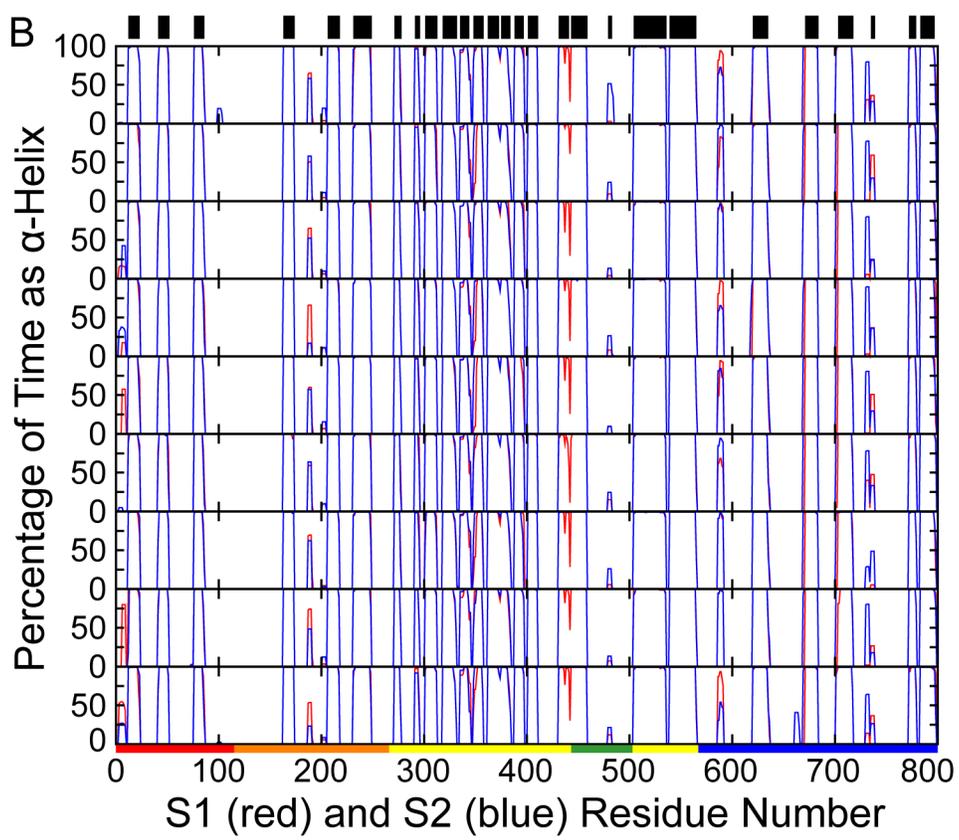
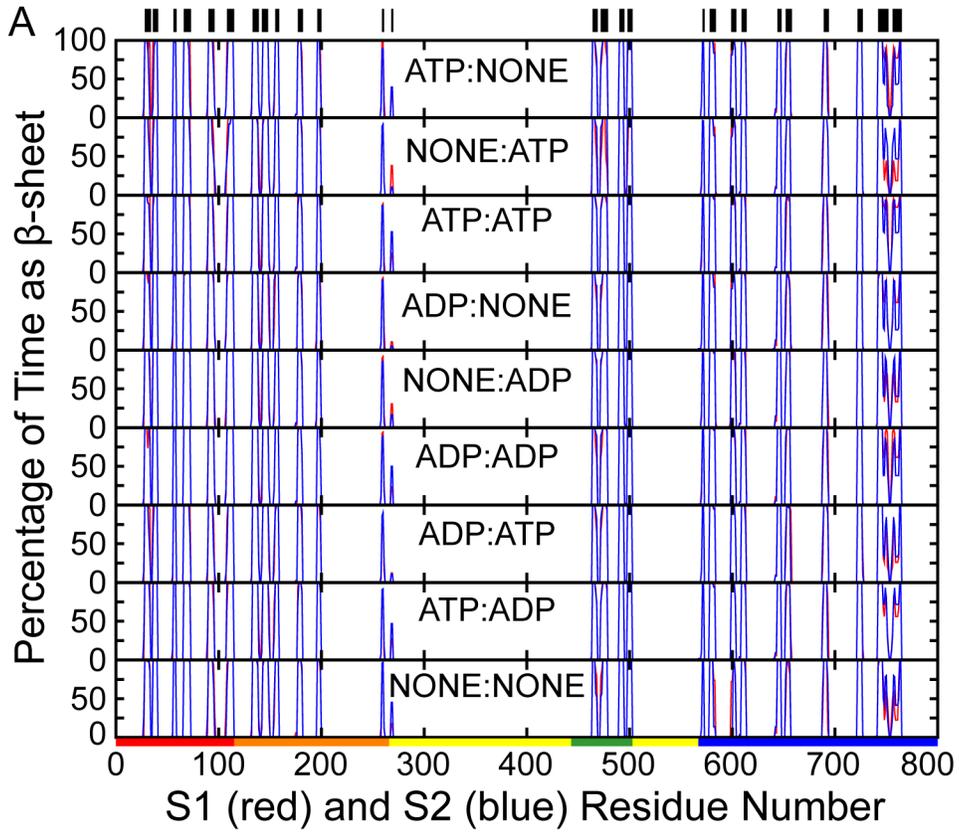


Figure 3.3 The percent contribution of the first 20 eigenvectors for each of the nine simulations.

Figure 3.3

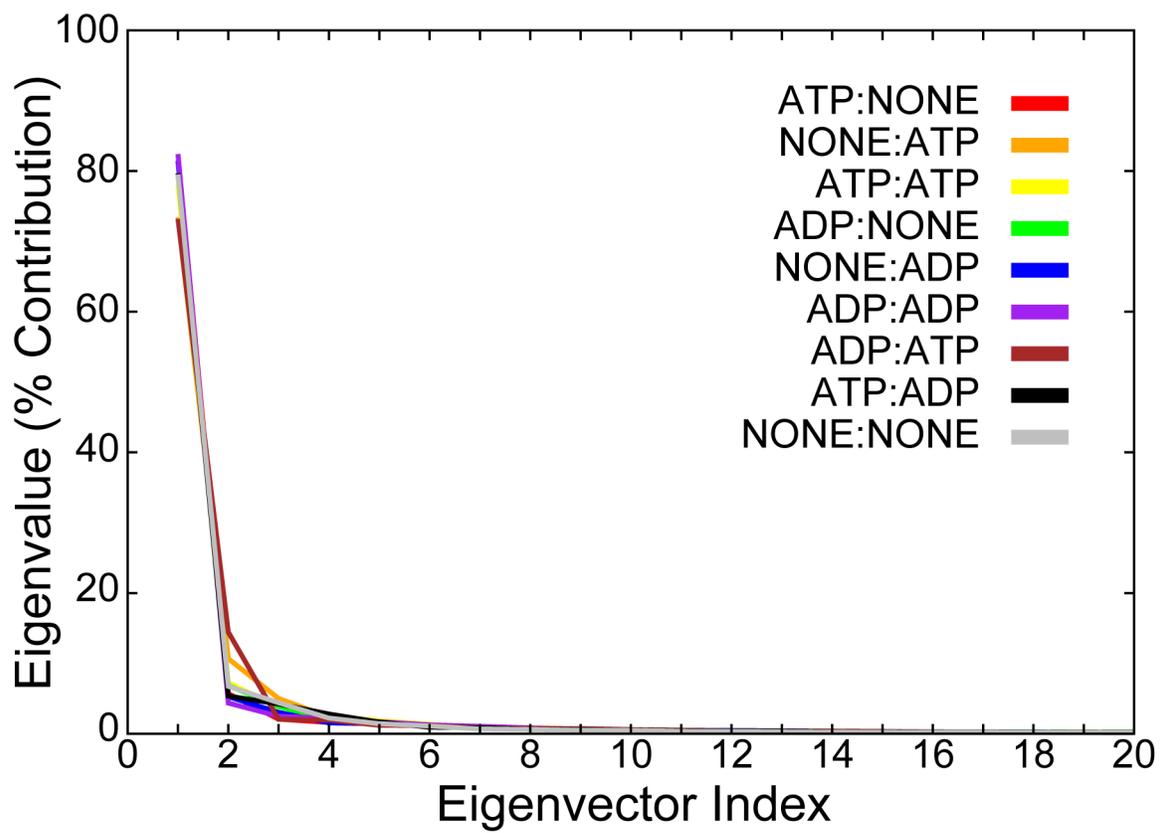


Figure 3.4 Schematic “Porcupine plot” for the ATP:ATP simulation projected onto the first principal component (PC1) of the combined trajectory and mapped onto the starting simulation structure. Each cone points in the direction of motion and the length of the cone represents the amplitude of the fluctuation for each $C\alpha$ atom. For clarity, the cones are colored according to the different structural domains and only motions larger than 3 Å are displayed.

Figure 3.4

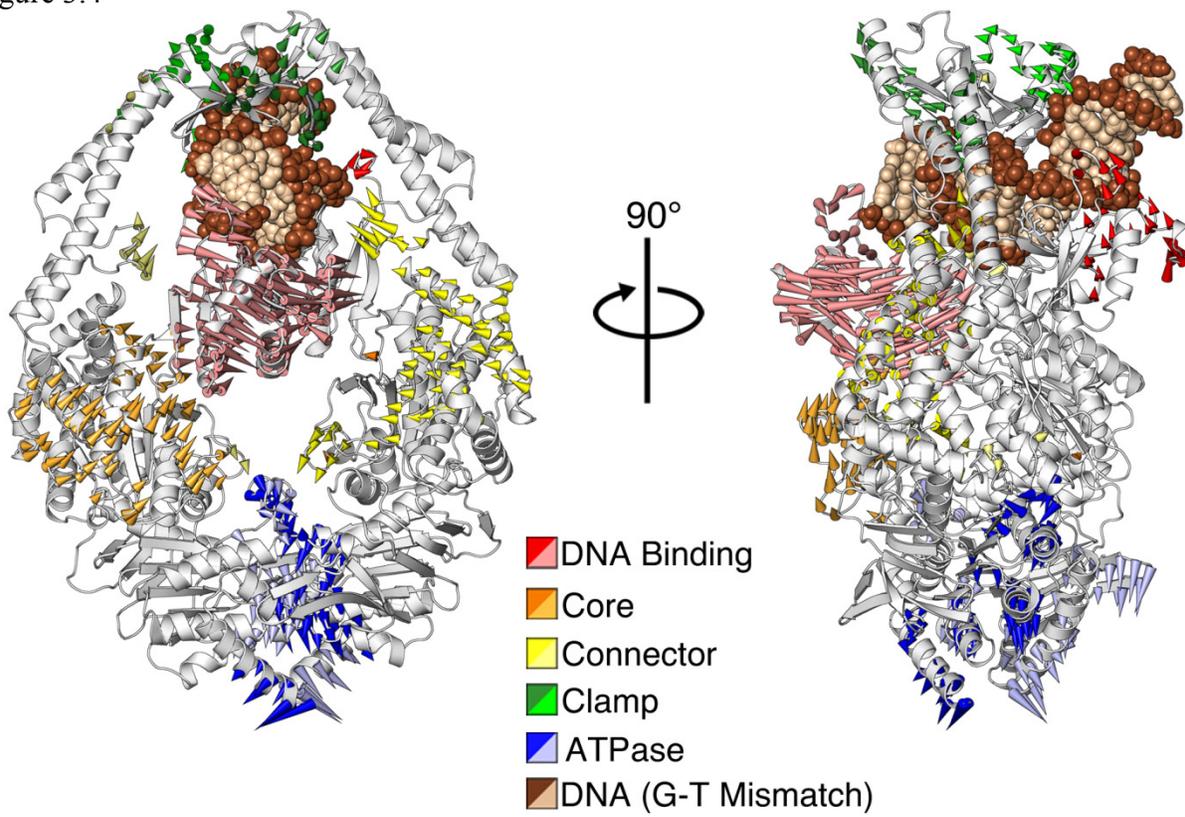


Figure 3.5 MutS ATPase domain with the S1 and S2 subunits colored in dark and light blue, respectively. The conserved helix-turn-helix is colored red and pink for the S1 and S2 subunits, respectively.

Figure 3.5

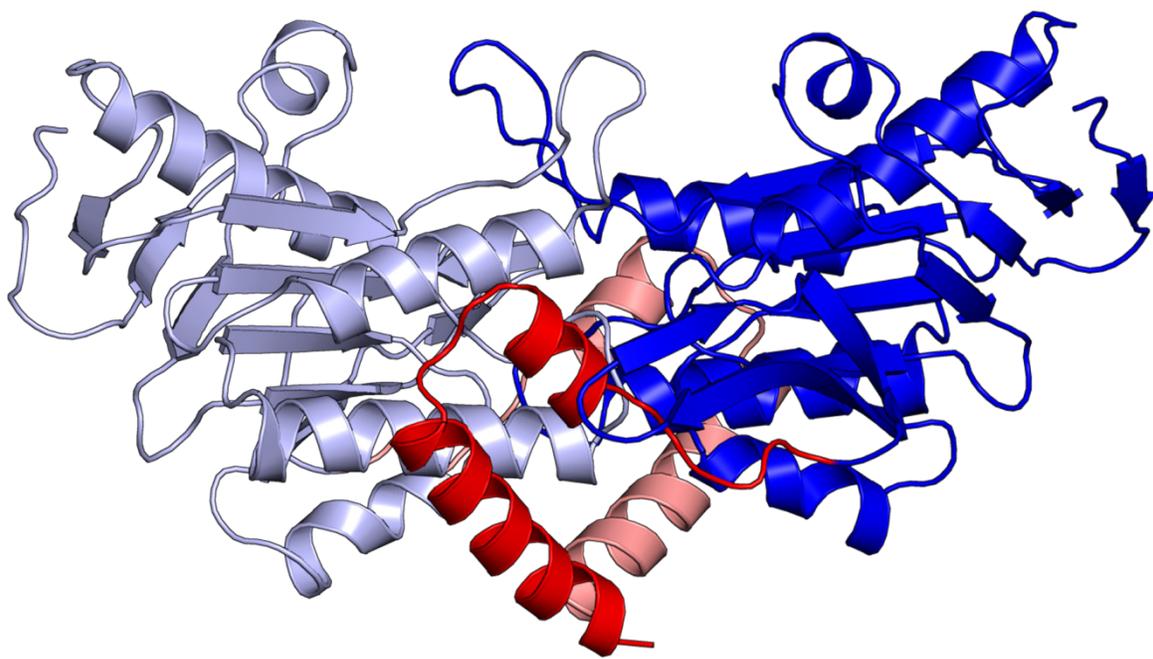


Figure 3.6 Schematic “Porcupine plot” for the ATP:ADP simulation showing its own first principal component mapped onto the starting simulation structure. Each cone points in the direction of motion and the length of the cone represents the amplitude of the fluctuation for each C α atom. For clarity, the cones are colored according to the different structural domains and only motions larger than 3 Å are displayed.

Figure 3.6

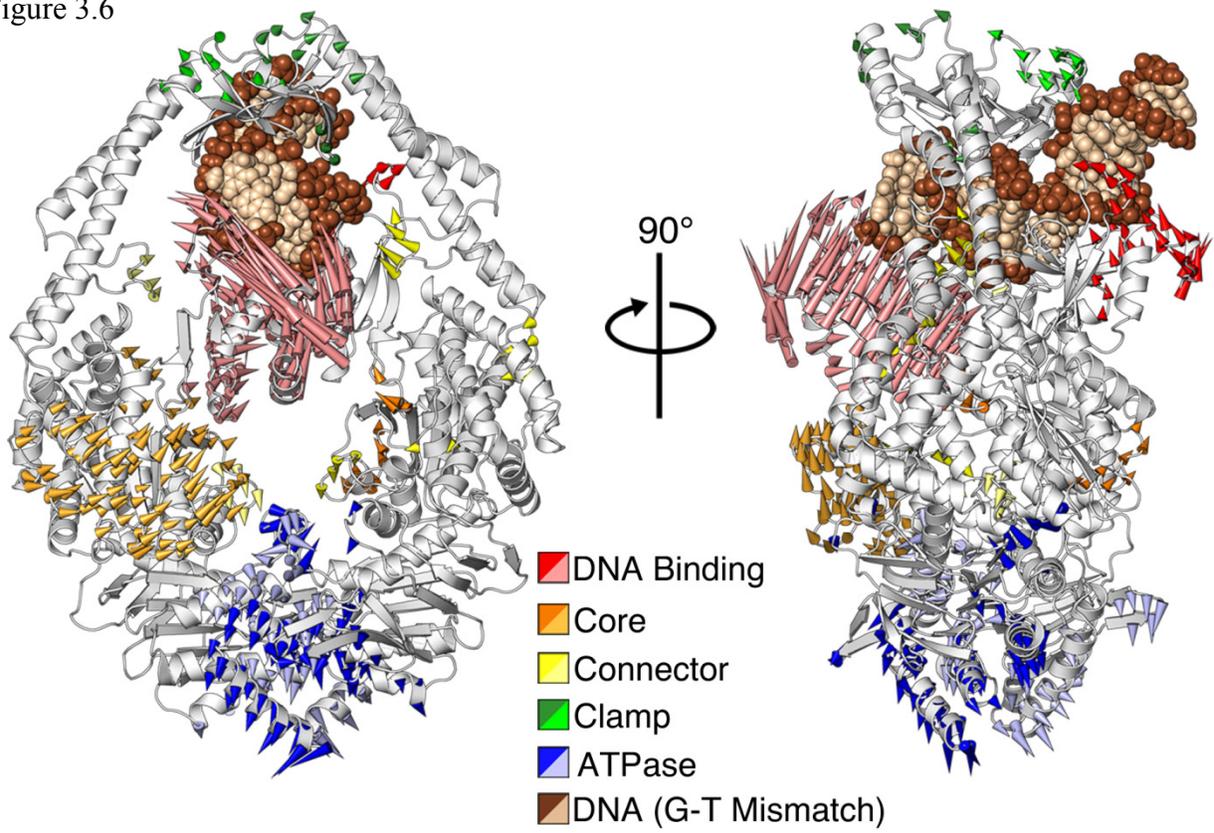
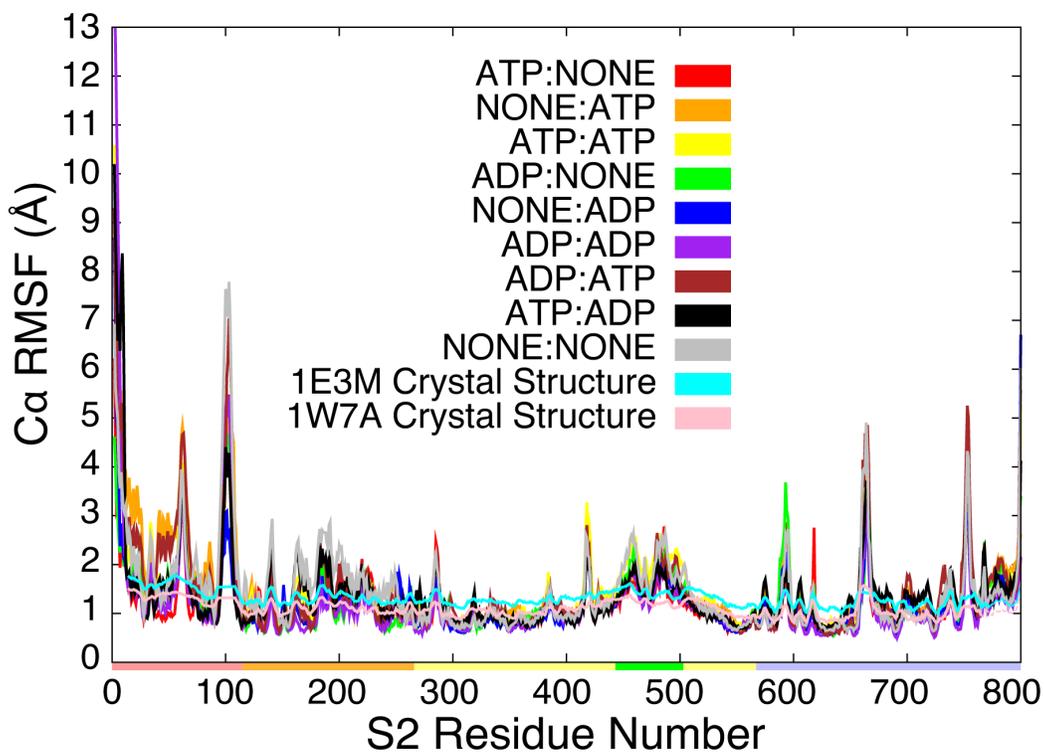
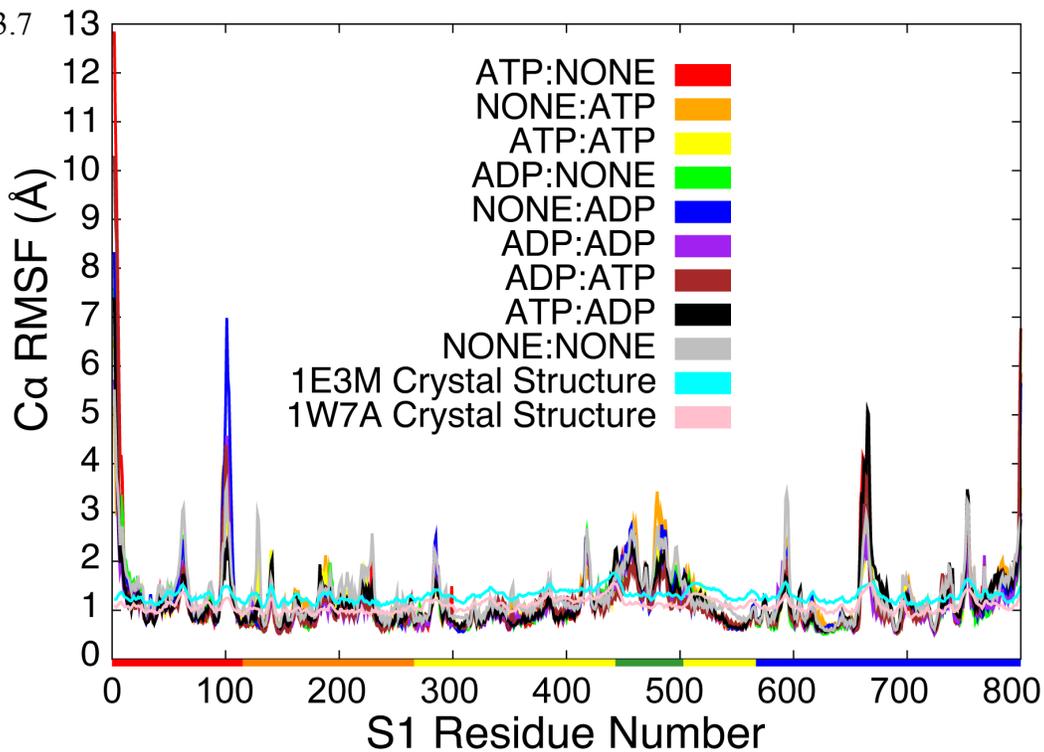


Figure 3.7 The RMSF for the S1 and S2 monomers calculated from the nine MD simulations and from the 1W7A and 1E3M crystal structures. The colored bar located below each panel corresponds to the five different structural domains and is colored according to the legend in Figure 3.1.

Figure 3.7



Chapter 4

Base-Flipping Mechanism in Post-Mismatch Recognition by MutS

Sean M. Law and Michael Feig

Submitted to *Biophys. J.*

4.1 Abstract

DNA mismatch recognition and repair is vital for preserving the fidelity of the genome. Conserved across prokaryotes and eukaryotes, MutS is the primary protein that is responsible for recognizing a variety of DNA mismatches. From molecular dynamics simulations of the *Escherichia coli* MutS-DNA complex, we describe significant conformational dynamics in the DNA surrounding a G·T mismatch that involves weakening of the base pair hydrogen bonding in the base pair adjacent to the mismatch and, in one simulation, complete base opening via the major groove. The energetics of base flipping was further examined with Hamiltonian replica exchange free energy calculations revealing a stable flipped-out state with an initial barrier on the order of about 2 kcal/mol. Furthermore, we observe changes in the local DNA structure as well as in the MutS structure that appear to be correlated with base flipping. Our results suggest a role of base flipping as part of the repair initiation mechanism, most likely leading to sliding clamp formation.

4.2 Introduction

The integrity of the genome is safeguarded from replication errors by an evolutionarily conserved DNA mismatch repair (MMR) pathway. MMR in *E.coli* begins with the mismatch recognition protein, MutS, scanning the DNA for base-base mismatches and small insertion/deletion loops (6). Upon mismatch recognition, MutL binds to MutS followed by further downstream repair events to ultimately restore the parental genotype (7, 10-12, 39-40, 114-115). Defects in the MMR pathway lead to replication and recombination errors and have been linked to hereditary non-polyposis colorectal cancer in humans (116) and are likely to play a role in other types of cancer as well (117).

Crystal structures of prokaryotic MutS and one of its human homologs, MSH2-MSH6, bound to various DNA mismatches, have provided mechanistic insight into the mismatch recognition process (20-21, 24, 29-30, 44-45, 118-120). Heteroduplex DNA bound to MutS (Figure 4.1A) is bent by about 45°-60° towards the major groove at the site of the mismatch. Mismatch specific contacts are made by a conserved F36-X-E38 motif (Figure 4.1B). The F36, first identified in cross-linking studies (26), forms an aromatic ring stack on the 3' side of the mismatched base. Mutation of F36 abolishes mismatch binding *in vitro* and is associated with defective MMR *in vivo* (22, 121-122).

The intrusion of a Phe residue into the duplex stack resembles intercalating residues commonly found in other DNA repair systems such as DNA glycosylases, T4 endonuclease V, and DNA demethylases, all of which involve a base-flipping mechanism (123-124). A similar base-flipping mechanism has also been proposed for MutS (16, 25, 30, 48, 51) but direct evidence has been lacking to date.

A recent FRET study has indicated that the MutS-DNA complex may involve transient intermediate states and exhibit more dynamics than suggested by the crystal structures (50). More detailed insight into the dynamics of the MutS-DNA complex during mismatch recognition is difficult to obtain with biochemical experiments but can be gained from computer simulations. Previous computational studies of MutS and homologs include normal mode analysis (101) and limited molecular dynamics (MD) simulations (32, 125-126). Here, we present results from sub-microsecond MD simulations of the MutS-DNA complex to focus on the details of the post-mismatch recognition process. In particular, we describe the observation of spontaneous base-flipping of the base adjacent to the mismatch site when bound to MutS. Quantitative aspects of the base opening transition were additionally analyzed with the Hamiltonian replica exchange method (HREM) (75). Our results suggest that flipping of the base adjacent to the mismatch is energetically likely in the MutS-DNA complex. Furthermore, it appears that base-flipping may be coupled to conformational changes in the protein suggesting a mechanistic role during repair initiation by MutS.

4.3 Materials and Methods

Simulated Systems and Molecular Dynamics Protocol

MD simulations of *E. coli* MutS in complex with DNA containing a G·T mismatch were carried out with explicit solvent. The starting conformation of the MutS-DNA complex was taken from the crystal structure 1W7A (29). Missing residues 660-667 in the S1 (mismatch binding) monomer were completed using the loop modeling (127) function in MODELLER, version 9 (99). Visual comparison of the model with a recent crystal structure of MutS where the disordered loop was resolved (120) showed no appreciable differences. Missing residues in the homodimeric S2 subunit were modeled after the S1 chain. Histidine ionization states were predicted using PROPKA3.1 (128) and confirmed visually based on the local protein environment. Nine simulations were carried out with all possible combinations of bound ATP, bound ADP, or no nucleotide at either the S1 or S2 ATPase domain. Positioning of the nucleotides was based on resolved nucleotides in the 1E3M (20) and 1W7A (29) crystal structures. The “X:X” notation is used here to denote which nucleotides are bound to the S1 and S2 subunits, respectively (*e.g.* ATP:ADP means that ATP is bound to S1 and ADP is bound to S2 while NONE:NONE is free of nucleotides). In addition to the wild-type system, simulations of an S1-F36A mutant with four different nucleotide combinations (ADP:NONE, ADP:ADP, ADP:ATP, NONE:NONE) were also carried out (see below).

Each structure was solvated using the TIP3P water model (129) and electrically neutralized with sodium ions. The total dimension of each system was approximately 155 Å x 117 Å x 94 Å and contained more than 165,000 atoms. The particle-mesh Ewald method (130) was employed to account for electrostatic interactions. The direct electrostatic sum and Lennard-

Jones (LJ) interactions were truncated at 10 Å with a switching function becoming effective at 8.5 Å and a non-bonded list cutoff at 12.5 Å. The all-atom CHARMM27/CMAP force field was used for all calculations (71-73) and chosen because it has been extensively validated in many other simulations of protein-nucleic acid simulations (54-55, 131), including simulations describing base flipping (132).

Minimization, Equilibration, and MD Protocol

Initial minimization involved 50 steps with the steepest descent method (SD) followed by 10,000 steps of adopted basis Newton-Raphson minimization. During the minimization, a 10 kcal/mol/Å² harmonic restraint was applied to all solute heavy atoms. Following minimization, each structure was gradually heated to 300 K and equilibrated in the NVT ensemble through three consecutive stages: (1) 1.4 ns of MD were carried out during which the solute was restrained as described above but water and ions were allowed to move freely; (2) solute restraints were released gradually over a period of 100 ps; (3) unrestrained MD over 6.4 ns was carried out to further equilibrate the system. All minimization and equilibration steps were carried out using the CHARMM program (64), version c35a1 in conjunction with the MMTSB Tool Set (102). After the initial equilibration phase, each of the nine simulations was then continued for another 200 ns using the NAMD simulation program, version 2.6 (68). The unrestrained NAMD simulations were carried out in the NPT ensemble that was maintained using a Langevin thermostat and barostat with a friction coefficient of 5 ps⁻¹ and a 2 fs integration time step was used in conjunction with SETTLE (133) to holonomically constrain bonds involving hydrogen atoms.

F36A Mutant Set Up and Simulations

Fully equilibrated structures from four wild type simulations (ADP:NONE, ADP:ADP, ADP:ATP, and NONE:NONE) were taken and the S1-Phe36 residue was mutated to an alanine residue using the MMTSB Tool Set (102). These mutated structures were subjected to the solute restraint protocol described above followed by a gradual release of the restraints and then equilibrated for an additional 10 ns (completely free of restraints). Each of the F36A mutant simulations was simulated for a total of 60 ns by using an identical production simulation protocol as described above.

Water Residence Time

The residence time of water molecules located within 4 Å of the G10 base was calculated by using a coordinate correlation function which has been previously used to assess solvent and ion residence times (134-135). Briefly, the water correlation function, $C_{\alpha}(t)$, is written as:

$$C_{\alpha}(t) = \frac{1}{N_{water}} \sum_{i=1}^{N_{water}} \frac{1}{N_{\alpha,i}(0, t_{total} - t)} \sum_{t'=0}^{t_{total}-t} p_{\alpha,i}(t', t'+t; t^*) \quad (4.1)$$

where $p_{\alpha,i}(t', t'+t; t^*)$ is a binary function that is set to 0 unless water molecule i is found within the predefined area α between time t' and $t'+t$. To ignore waters that escape and quickly rebind, the rebinding time t^* was set to 1 ps. The binary function is then accumulated across the total simulation time t_{total} and divided by the number of times $N_{\alpha,i}(0, t_{total} - t)$ water

molecule i is found within the confines of α . Finally, N_{water} corresponds to the total number of water molecules that participate in the residence time calculation. Depending on the overall sampling, water residence times may be fit to a bi-exponential decay function or, in some cases, a tri-exponential function. Alternatively, by taking the natural logarithm of the water correlation function the residence times can be easily obtained by calculating the inverse slope of portions of $\ln(C_\alpha(t))$ that can be fit to a linear curve.

Solvent-Accessible Surface Area

Analysis of the solvent-accessible surface area for the H1' atom in the DNA minor groove of bases near the mismatch site was obtained from the COOR SURF command in CHARMM using a 1.4 Å probe radius (which represents a single water molecule). Solvent-accessible surface areas were then calculated with and without MutS. The reported change in accessibility upon MutS binding, referred to as Δ SASA, is the difference between these two values.

Hamiltonian Replica Exchange Simulations and Free Energy Calculations

To investigate the energetics of base-flipping, umbrella sampling simulations were carried out. A harmonic biasing potential was applied to enhance base-flipping and to obtain sufficient statistical sampling for estimating the free energy profile associated with base-flipping. The reaction coordinate used for the biasing potential is a pseudodihedral angle introduced earlier (136). The pseudodihedral is based on the following four heavy atom sites: 1) Center of mass (COM) of the G9, T22, C11, and G20 bases (flanking the base of interest, C21); 2) the T22

phosphate; 3) the C21 phosphate; and 4) the COM of the C21 base (Figure 4.2A). While other reaction coordinates have been utilized in the past to study base flipping (132, 137-139), this pseudodihedral angle definition provides an improvement over previous methods (136) and has been shown to produce results that are in good agreement with experiment (140). The biasing potential was applied using the miscellaneous mean field potential (MMFP) module (141) of CHARMM and has the following form:

$$w_i(\theta) = \frac{k_i}{2}(\theta - \theta_i)^2 \quad (4.2)$$

where k_i is the force constant set to 100 kcal/mol/rad², θ is the pseudodihedral angle, and θ_i is the target value for the i^{th} window. A total range of 0 to 162.5° was covered in 2.5° increments to result in 66 windows. Instead of conventional umbrella sampling, we used HHREM (75) with 66 replicas corresponding to the umbrella windows to enhance sampling efficiency further. These simulations involved the entire *E. coli* MutS-DNA complex in explicit solvent. They were carried out by using CHARMM (64) in conjunction with the MMTSB Tool Set (102). Starting structures for different replicas were taken from one of the unbiased simulations where base-flipping was observed spontaneously. Each starting structure was initially subjected to 200 ps of equilibration with the biasing potential of a given replica. Each replica was then simulated for 10.5 ns (for a total simulation time of 693 ns for all 66 replicas). Exchanges between neighboring replicas were attempted every 1 ps. 23-37% of the exchanges were successful.

Analysis

Most of the analysis was carried out with the MMTSB Tool Set and CHARMM, version c35a1 based on the 200 ns production time for the unbiased simulations. Protein RMSD values were calculated using $C\alpha$ atoms. The DNA RMSD was calculated by using all heavy atoms, omitting the ultimate and penultimate bases. 1-D potentials of mean force (PMFs) were generated from the replica exchange simulation using WHAM (76) after discarding the first 5 ns as equilibration. 2-D PMFs along additional degrees of freedom were estimated from the HREM simulations (also with the first 5 ns removed as equilibration) using standard WHAM under the assumption that all other degrees of freedom orthogonal to the pseudodihedral angle are thoroughly sampled (142). All structural figures were generated using PyMOL (104).

4.4 Results and Discussion

A series of nine 200 ns MD simulations of MutS in complex with a G·T mismatch containing DNA were analyzed with a primary focus on MutS-DNA interactions and the dynamics of mismatch DNA when bound to MutS. The simulations differed in the nucleotide(s) bound in the ATPase sites since the simulations were initially set to study the effect of different nucleotides on the MutS structure. During the course of the simulations reported here we did not see significant structural perturbations that could be correlated with the type of nucleotide bound to the ATPase domain. In fact, we found that the MutS-DNA complex sampled similar conformations in all nine simulations. The $C\alpha$ RMSDs were all within 3-4 Å relative to the X-ray structure (Figure 4.3). Furthermore, clustering analysis shows that structures from all simulations fall into closely related conformations, with overlapping sampling of conformations belonging to the four largest clusters (Figure 4.4). This suggests that different nucleotides bound in the ATPase domain do not dramatically affect the overall MutS structure on the sub- μ s time scales covered here. Consequently, the simulations discussed here are treated as nine independent simulations of essentially the same system providing a total of 1.8 μ s of sampling of the MutS-DNA complex.

Dynamics of DNA and Base-flipping in the MutS-DNA complex

Overall, the DNA bound to MutS maintained its bent structure in all simulations as indicated by a heavy-atom RMSD of 1-4 Å (Figure 4.3). However, a more detailed analysis of base pair hydrogen bonding revealed significant base dynamics in the vicinity of the mismatch. More specifically, the G/C(-1) base pair adjacent to the mismatch site on the 5'-side of the thymine of

the G·T base pair lost Watson-Crick hydrogen bonding in most of the simulations (Figure 4.5A). The X/Y(\pm N) notation is used here to denote the X/Y base pair relative to the thymine of the G·T mismatch (see Table 4.1). The G·T mismatch remained stable in all but one of the simulations. In that simulation (NONE:ADP) a new N3-O6 hydrogen bond was formed within the same base pair due to shearing of the G·T base pair. The next-neighbor A/T(+1) and C/G(-2) base pairs stably maintained standard hydrogen bonding in all simulations (Figure 4.5A).

The instability of the G/C(-1) base pair was unexpected and involved the loss of N1–N3 hydrogen bonding and at least partial opening of the C21 base into the major groove. In one of the simulations (NONE:NONE) all of the G/C(-1) hydrogen bonds were lost within the first 10 ns and the base subsequently flipped out into the major groove where it remained for the rest of the simulation. This observation appears to be in conflict with previous NMR and MD studies where significant instability of G·T pairs over canonical base pairs has been established (143-144). However, these studies were not conducted in the presence of MutS and therefore do not account for the severe bend in the DNA caused by interactions with the protein (20-21, 48). The bending leads to significant distortions of the grooves near the mismatch site. In particular, the major groove width is reduced to only 13 Å at the G·T mismatch but increased to 18 Å at the G/C(-1) base pair (see Figure 4.6) compared to the major groove width of canonical B-DNA at around 17 Å (145). The narrow major groove at the G·T pair effectively prevents base opening while the wider major groove at the G/C(-1) base pair is more favorable for base opening.

To test a possible role of F36 in stabilizing mismatch base pairing and promoting G/C(-1) base flipping we ran four additional 60 ns simulations of a S1-F36A mutant. We find that mismatch base pairing is stably maintained without F36 (see Figure 4.5B) although the T22 base reorients with different glycosyl rotation angles (see Figure 4.7A-B). Interestingly, we again

observed spontaneous base opening of the G/C(-1) base pair in one of the simulations (ADP:NONE:F36A) in a very similar manner as in the NONE:NONE simulation (see Figure 4.5B). These results suggest that F36 does not play a significant role in either stabilizing mismatch base pairing or promoting base opening of the C/G(-1) base pair.

Progression of the base-flipping process was quantified with the help of a pseudodihedral angle, θ , (see *Methods* section) with negative values as the base opens into the major groove (see Figure 4.8A). Figure 4.9 shows snapshots of key time points during the base opening process. Initially, G/C(-1) was perfectly base paired ($\theta \approx 0$). The base then rapidly lost base pair hydrogen bonds and stacking interactions to reach a semi-open state ($\theta = -40^\circ$) that was stable for a few ns. Further opening led to another intermediate state that was stabilized by hydrogen bonding interactions to the DNA backbone ($\theta = -81^\circ$). This state also persisted for a few ns. Eventually, the C21 base opened entirely at about 10 ns from the beginning of the production phase of the simulation. The base was briefly fully exposed to the solvent environment ($\theta = -130^\circ$) but then began to interact with the DNA backbone of the *opposing* strand ($\theta = -120^\circ$). This conformation persisted from $t \approx 20$ ns to $t \approx 120$ ns. During the remainder of the simulation, C21 moved back towards various semi-open states but without re-forming a fully stacked configuration. C21 base flipping was associated with a change in the C21 backbone ζ torsion angle from around -150° to 150° (see Figure 4.8B, Figure 4.10B, and Figure 4.11A) as generally expected for DNA base flipping (146). Otherwise, the DNA structure remained largely unaffected by the opening of the C21 base on the time scale of our simulations.

Our observation of spontaneous base-flipping in DNA complexed to MutS provides new molecular-level evidence for the previously proposed idea that base-flipping may play a role in mismatch recognition (16, 25, 30, 48, 51). In order to gain more quantitative insight we also

carried out an HREM simulation of the NONE:NONE MutS-DNA complex where sampling along the base-flipping reaction coordinate was enhanced with a total 10.5 ns of simulation time for each replica. The main result is a PMF free energy profile along the base-flipping reaction coordinate (see Figure 4.10A). The PMF has a prominent minimum near 0° for the fully base paired state and a second minimum at around -105° corresponding to the flipped-out state. The two states are estimated to be separated by a 2 kcal/mol energy barrier. To examine the convergence of the PMF we compared it to PMFs with a shorter simulation lengths (7.5 ns/replica and 9 ns/replica) and found negligible change between the 9.5 ns/replica and 10.5 ns/replica PMFs (Figure 4.12). Based on the variation of the PMF over time we roughly estimate the uncertainty to be between 0.1-0.5 kcal/mol. Thus, the HREM simulation confirms the existence of a favorable, flipped-out state. Based on the PMF we calculate that the G/C(-1) base pair is intact ($\theta \geq -20^\circ$) for 69% of the time but the C21 is flipped-out to varying degrees during the remaining 31% with an estimated uncertainty of 5-10% based on the uncertainty of the PMF.

The observed 2 kcal/mol barrier suggests conformational transitions on ns time scales. This is in apparent contradiction with the rarity of full base opening/closing events in the unbiased simulations. In the replica exchange simulations, complete base opening/closing was also never observed for any individual replica although significant sampling overlap from many replicas at each pseudodihedral value (see Figure 4.13) suggests that the PMF presented in Figure 4.10A is realistic. This suggests the presence of significantly higher kinetic barriers in orthogonal degrees of freedom not captured by the projection onto the C21 pseudodihedral angle. One source for such barriers is likely the torsional dynamics of the ζ backbone dihedral with a barrier height estimated to be larger than 5 kcal/mol in previous simulations of base opening (147). Another source for slow base opening/closing kinetics appears to be the presence of long-

lived water molecules at the constrained protein-DNA interface (Figure 4.14A-B). An analysis of residence times of water molecules located within 4 Å of the G10 base from the NONE:NONE simulation found that waters located on the major groove side and within the cavity left by the flipped out C21 base had residence times up to 500 ps (compared to about 50 ps of surface-bound waters, see Figure 4.14C-D) while waters that managed to enter the cramped minor groove side essentially become trapped near the N2, N3, and N9 atoms of the G10 base with even longer residence times in the ns range (Figure 4.14E-F). The presence of these long-lived water molecules likely hinders base closing which cannot be accomplished unless these waters are displaced. This would explain why C21 never fully restacked in the NONE:NONE simulation despite the ζ torsion reverting back to the -150° range near the end of the simulation.

Our results suggest that base opening may occur on sub- μ s time scales since it was observed spontaneously in two of our simulations. Most likely, base opening kinetics are dominated by the kinetic barrier for ζ backbone dihedral transitions. Base closing, on the other hand, appears to involve much longer time scales due to obstruction by long-lived water molecules. This would imply that the flipped-out state may be kinetically stabilized for a long time despite being thermodynamically slightly less favorable than the fully stacked state according to our analysis.

Opening of the 5' Adjacent Base Next to the Mismatch is in Agreement with Experiment

Direct structural evidence for DNA base-flipping in the MutS-DNA complex is lacking, but there is indirect experimental evidence for at least partial opening of the 5' adjacent base next to the mismatch: Prior to the discovery of the MutS structure, chemical footprinting was used to

uncover the interactions between *Thermus aquaticus* MutS and the DNA minor groove (148). MutS-bound DNA with a G·T mismatch was found to be protected on the 3' side of the lesion, but not on the 5' side of the mismatched thymine where the -4, -2, and -1 positions were hyperreactive to oxidative attack. This was attributed to widening of the minor groove when the crystal structure became available (21). To further understand these data, we analyzed the effect of MutS on solvent-accessibility of H1' (the hydrogen attached to the C1' attack site located in the minor groove) from our simulations. We found that without base flipping (in ATP:NONE), access to the -4 base is fully maintained, access to the -2 base is partially hindered, but the -1 base is largely occluded (Figure 4.15). Base flipping (in NONE:NONE), on the other hand, fully exposes the -1 base so that all three bases become vulnerable to oxidative attack as indicated by experiment.

In a more recent study, 2-AP, a fluorescent adenine analogue often used to probe DNA base-flipping, was incorporated into various positions next to a G·T mismatch (52). It was found that the mean fluorescence lifetime increased when the mismatch was bound by MutS. Furthermore, the level of increase in the observed mean fluorescence lifetime was significantly higher when the probe was placed on the 5' side of the mismatch as compared to other positions. This increase was attributed to an increased amplitude of the longest lifetime component, which could be explained by an increased fraction of extrahelical states. Interestingly, the relative population with the increased fluorescence lifetime, calculated by summing up the fractional amplitude of the two longest lifetimes, was estimated to be ~31% (52), the same percentage as the fraction of flipped-out conformations in our HREM simulation. While we believe that a qualitative comparison with the experiment is meaningful, the surprisingly good quantitative agreement is likely fortuitous because of uncertainties in both the experimental and

computational results as well as differences between experiment and simulation. In the experimental results each reported lifetime results from numerous conformers with comparable quenching rates. Furthermore, the experimental study describes opening of an A/T (or 2-AP/T) base pair that is known to have different base opening rates than G/C base pairs (144) as in the MutS-DNA complex studied here.

Changes in MutS as a Result of Base-flipping

Experiments suggest that initial mismatch recognition is followed by several changes in MutS. First, biochemical data indicate altered activity of the ATPase domain as a result of mismatch recognition where ADP is exchanged for ATP and hydrolysis is stalled (8, 24, 84, 149-150). Second, the affinity for MutS-MutL complex formation is increased. Third, a transition to a sliding clamp formation has been suggested to allow MutS to leave the mismatch site after initial recognition so that DNA repair can take place (39-40, 84, 98). Thus, if base-flipping plays a role in post-mismatch recognition, sliding clamp formation, and/or initiation of repair, one would expect that there are correlated changes in the MutS structure either in the DNA binding domains, the core and connector domains where MutL is proposed to bind (31), or in the ATPase domains. In our simulations we identified changes in the DNA binding domains of both chains, local rearrangements in the ATPase domains, but no significant changes in the core and connector domains of MutS correlated with base flipping.

DNA binding domain motions were characterized by a local coordinate system: X corresponds to motion along the DNA helical axis, Y to motion perpendicular to the helical axis towards the tips of the clamp domains, and Z to motion across the S1-S2 dimer interface (see

Figure 4.16A). The S1 DNA binding domain (S1-D1), which interacts specifically with the DNA mismatch, showed a significant shift by, on average, 1-1.5 Å along X in the simulation where the base is flipped out (Figure 4.8C) compared to all of the other simulations where the base did not flip out (Figure 4.16C-E). Motion of S1 along Y and Z was not correlated with base flipping (Figure 4.17A-B). Due to the bent shape of the DNA, this motion effectively moves the domain away from the DNA (see Figure 4.16B). The S2 DNA binding domain shows a significant shift along Z, laterally away from the DNA towards the S1 core and a moderate shift along X and Y (Figure 4.16F-H). As a result of the motion of the S2 domain, MutS-DNA interactions are also reduced (Figure 4.16B) and these motions appear to be closely coupled to base flipping (Figure 4.8D-F). Additional analysis based on data from the HREM simulations confirms a strong correlation between base flipping and the motion of S2 along X, Y, and Z (see Figure 4.10D-F) and to a lesser extent for S1 along X (see Figure 4.10C) but not along Y or Z (Figure 4.17C-D). An apparent correlation between motion of the DNA binding domain and base flipping points at a possible connection with sliding clamp formation which is assumed to involve reduced MutS-DNA interactions.

Functional coupling between mismatch recognition and ATPase activity requires allosteric signaling over 90 Å from the DNA binding domains to the ATPase domains (see Figure 4.1A). Based on the MutS structure it appears that such communication would involve the core and connector domains which provide the structural connection. In fact, there is a string of highly conserved residues from the DNA-binding to the ATPase domains (Figure 4.18). In our simulations we did not observe motions along this pathway that could be uniquely attributed to base flipping but we did identify changes in the ATPase domain itself in the vicinity of the S1 nucleotide binding pocket that appear to be correlated to base-flipping.

Ser668, a conserved residue in MutS homologs, was previously implicated in ATP hydrolysis (29, 120, 151). Crystal structures suggest that Ser668 in the S2 monomer may move closer to the opposing S1 nucleotide binding site by ~ 5 Å to take part in ATP hydrolysis (29, 120) (Figure 4.19A). While the exact mechanism is unclear, it has been postulated that S2 Ser668, located at the end of an α -helix, could convey the positive charge generated from the helix dipole to the γ phosphate to assist with catalysis of the hydrolysis reaction (29, 120). For this structural rearrangement to occur, Asn616, situated in the P-loop (Figure 4.19A), has to move away from the dimer interface to allow Ser668 to complete the active site (29). Significant reduction in ATPase activity following mutation of either Asn616 or Ser668 supports such a mechanism (29, 151).

In our simulations, we found that the S2 Ser668 to S1 Asn616 distance, the backbone conformation of Asn616 in the S1 monomer, and the ability to form a salt bridge between Glu594 of the S1 monomer and Arg667 of the S2 monomer are all correlated with base flipping. Changes in the backbone of Asn616 were measured by the Ψ (N-C α -C-N) torsion angle. Figure 4.8G shows a significant decrease in the Ser668 to Asn616 C α -C α distance from about 9 Å to 5 Å upon base flipping which may promote nucleotide hydrolysis as indicated by the biochemical data. As shown in Figure 4.8H, the Asn616 Ψ angle changes from 125° to -30° at the same time as the base flips out and, as a result, allows Ser668 to approach the S1 active site. A correlation with base flipping is confirmed from the HREM data where base opening appears to limit the C α -C α distance between Ser668 and Asn616 to 6-8 Å instead of 6-12 Å when the base is fully stacked (Figure 4.10G). Similarly, base flipping seems to broaden the conformational sampling of the Asn616 Ψ value to the full range from -50 to 180 degrees while only extended conformations are observed for fully base-stacked DNA (see Figure 4.10H).

Repositioning of S1 Asn616 also allows the S2 Arg667 side chain to relocate and form a salt bridge with S1 Glu594 (Figure 4.8I and Figure 4.19B). The interaction between these two residues further stabilizes the S2 signature loop in which Ser668 resides (see Figure 4.8J) and Arg667 is also positioned to hydrogen bond with the ribose of adenosine (Figure 4.19B). Based on these results, we speculate that base flipping may promote (or restore) the ability to hydrolyze ATP in the S1 binding site. However, there remains uncertainty about the exact mechanism of how variations in MutS-DNA interactions are communicated to the distant ATPase domain.

Known crystal structures of MutS are very similar with either ATP (29) or ADP bound in the ATPase domains (20). This suggests that they more likely represent the post-mismatch recognition state where ATP hydrolysis is stalled and MutS is poised for sliding clamp formation. The simulation results suggest that this structure may promote base flipping in DNA which in turn seems to initiate sliding clamp formation. The correlated changes in the ATPase domain suggest a connection to ATPase activity. The coincidence of apparent changes in the ATPase domain and DNA binding domain as a result of base flipping would be consistent with a previously suggested role of ATP hydrolysis during sliding clamp formation (39, 41). However, this idea is inconsistent with a hydrolysis-independent model for sliding clamp formation that is supported by other studies (40, 84).

4.5 Conclusions

Sub- μ s computer simulations were used to report direct structural evidence for DNA base-flipping in the large MutS-DNA system. The instability in DNA base pairing was found to be specific for the 5' adjacent base pair instead of the mismatch. This is in contrast to previous hypotheses, but appears to be in good agreement with experimental data. Energetic analysis of the base-flipping process confirmed the existence of a stable flipped-out state in the presence of MutS and revealed an estimated 2-2.5 kcal/mol activation energy barrier for base flipping. Kinetic rates for base flipping were estimated to be in the μ s range due to slow DNA backbone and water rearrangements. Further analysis of changes in the MutS structure suggests that base flipping leads to motions of the DNA-binding domains away from the DNA and more subtle changes in the ATPase domain.

Taken together, our simulations suggest that base flipping may be the key step that allows MutS to transition from the post-mismatch recognition complex to the sliding clamp formation. We hope that our results will motivate further computational and experimental studies to better understand the mechanistic details of DNA mismatch repair initiation by MutS.

4.6 Acknowledgements

We thank Dr. Shayantani Mukherjee, Dr. Katarzyna Maksimiak, Dr. Yi-Ming Cheng, Dr. Richard Venable, and Hugh Crosmun for discussions and Dr. Chresten Søndergaard for help with the determination of protein ionization states in the presence of DNA. We also acknowledge access to computational resources from the High Performance Computing Center at Michigan State University. This work was supported by the National Science Foundation [MCB-0447799]; the National Institutes of Health [GM 092949]; TeraGrid [TG-MCB090003]; and the Center for Biological Modeling at Michigan State University [S. M. L.].

Table 4.1 DNA Sequence Used in All MutS Simulations*

	<u>G/-</u>	<u>T/A</u>	<u>G/C</u>	<u>A/T</u>	<u>C/G</u>	<u>C/G</u>	<u>A/T</u>	<u>C/G</u>	<u>G/C</u>		<u>A/T</u>	<u>C/G</u>	<u>C/G</u>	<u>G/C</u>	<u>T/A</u>	<u>C/G</u>	<u>G/C</u>	<u>A/T</u>	
	<u>(-9)</u>	<u>(-8)</u>	<u>(-7)</u>	<u>(-6)</u>	<u>(-5)</u>	<u>(-4)</u>	<u>(-3)</u>	<u>(-2)</u>	<u>(-1)</u>		<u>(+1)</u>	<u>(+2)</u>	<u>(+3)</u>	<u>(+4)</u>	<u>(+5)</u>	<u>(+6)</u>	<u>(+7)</u>	<u>(+8)</u>	
3'...	G18	T17	G16	A15	C14	C13	A12	C11	G10	G9	A8	C7	C6	G5	T4	C3	G2	A1	...5'
5'...	---	A14	C15	T16	G17	G18	T19	G20	C21	T22	T23	G24	G25	C26	A27	G28	C29	T30	...3'

* The DNA sequence is identical to the crystal structure found in reference (29) and the G·T mismatch is shown in bold.

Figure 4.1 X-ray crystal structure of *E. coli* MutS (20). (A) MutS is colored with respect to its DNA binding domains (red/pink), connector domains (orange or pale orange), core domains (yellow or pale yellow), clamp domains (green or pale green), and ATPase domains (blue or pale blue). The DNA is colored in beige (bases) and brown (backbone). Bound nucleotides are omitted for clarity. (B) A conserved F36-Xaa-E38 motif interacts with the G·T mismatch through the DNA minor groove. The protein is colored in green, the mismatch in pink, and the G/C(-1) base pair with the 5' adjacent base C21 in yellow. The bifurcated base pair hydrogen bond in the G·T mismatch and hydrogen bonding between E38 and T22 are shown as black dotted lines.

Figure 4.1

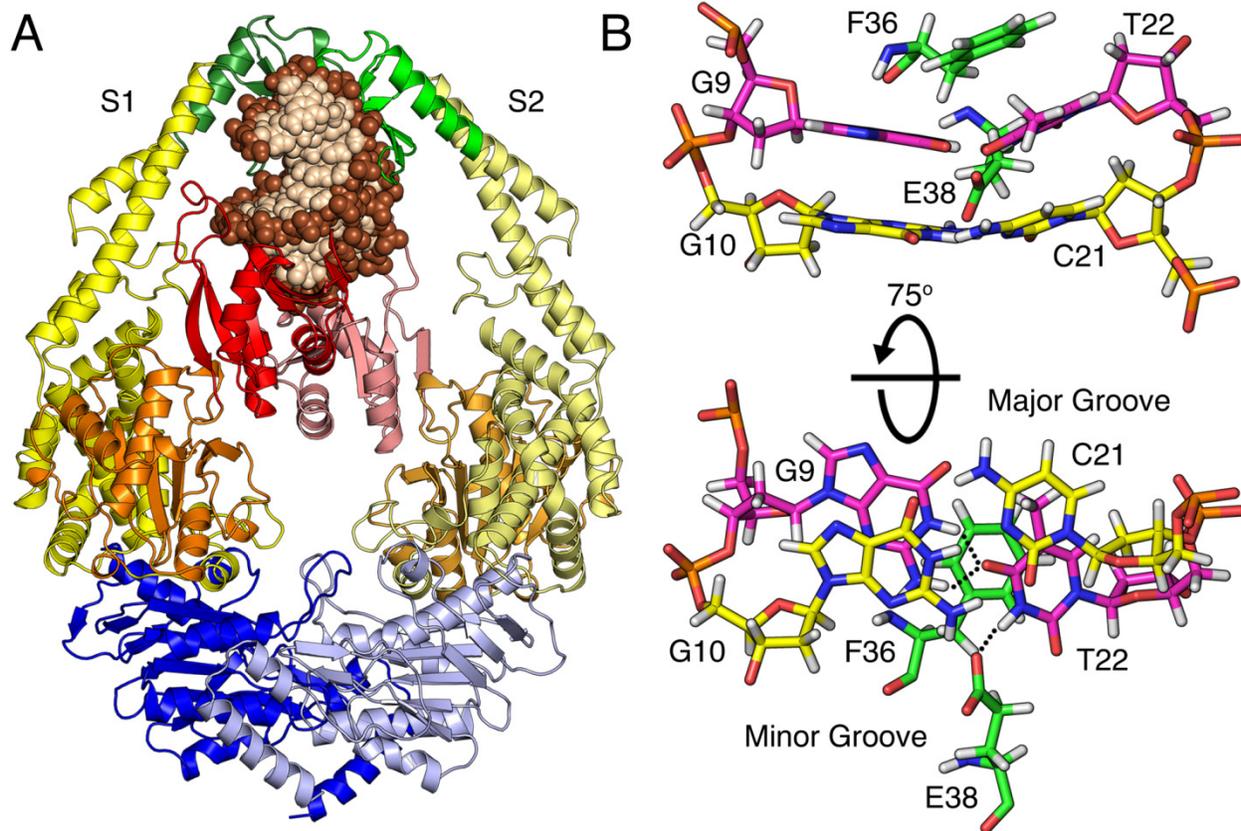


Figure 4.2 Pseudodihedral angle definition (see *Methods*).

Figure 4.2

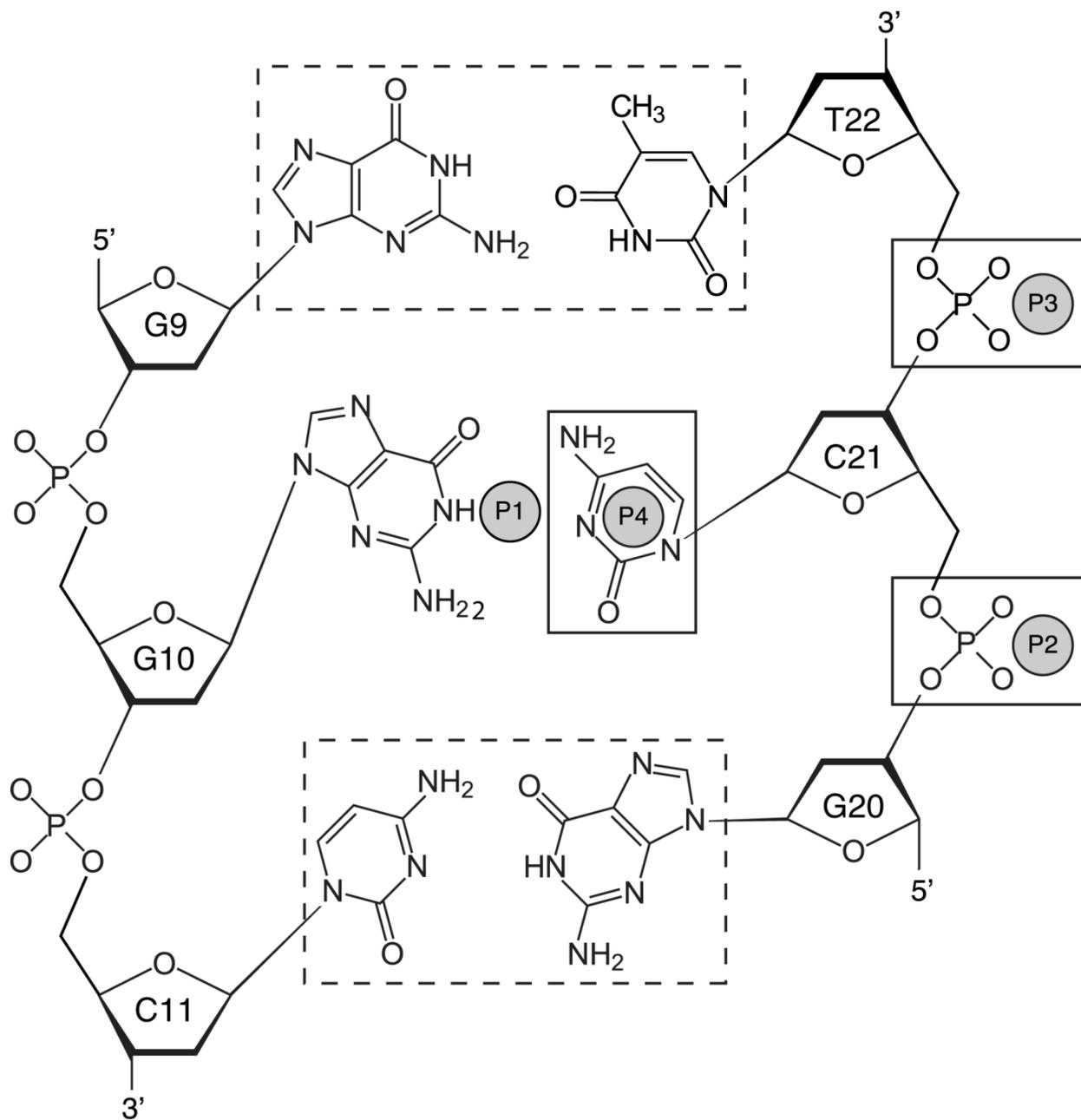


Figure 4.3 $C\alpha$ protein RMSD (red) and heavy atom DNA RMSD (black) for each of the nine simulation models calculated with respect to a common starting structure. Refer to *Methods* for the notation used to describe each simulation model.

Figure 4.3

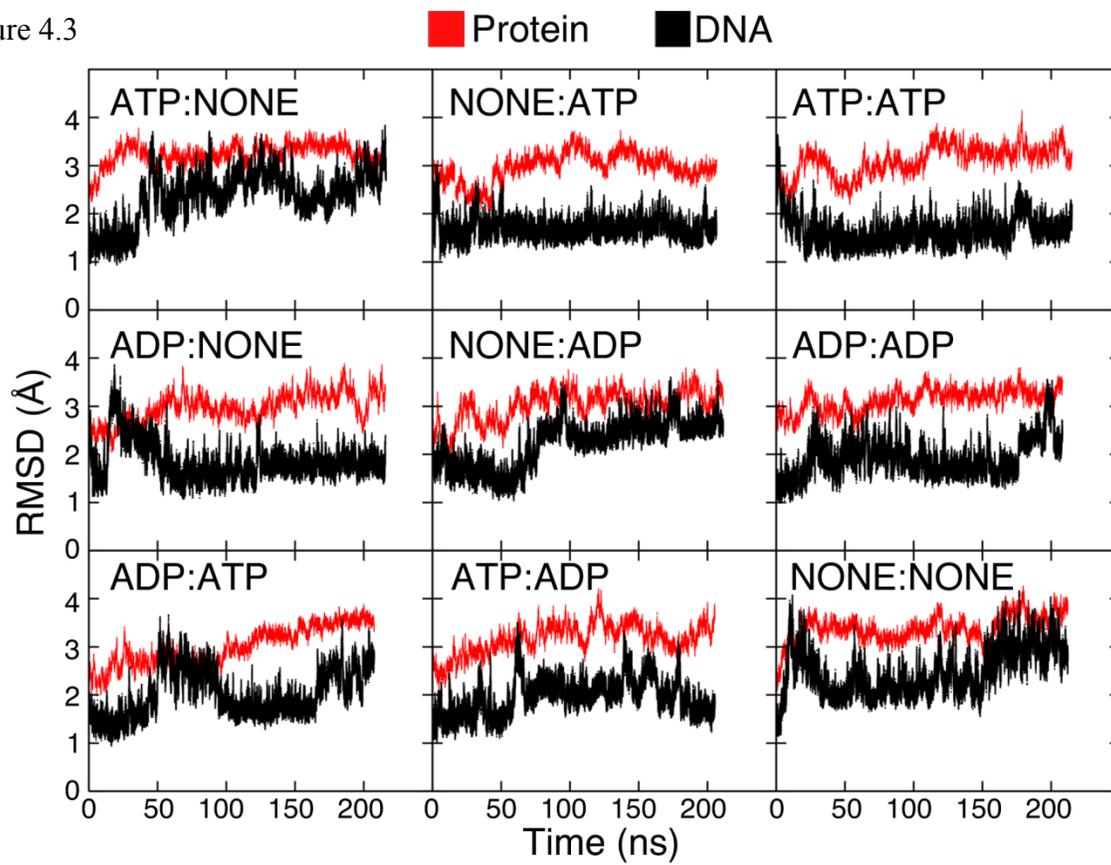


Figure 4.4 K-means clustering (implemented in the MMTSB Tool Set (102)) of the nine simulations using a 2.5 Å radius (large gray overlapping circles). Structures were extracted at every 1 ns of production simulation and clustered based on the C α RMSD. The area of each cluster (six colored circles) is proportional to the number of structures in that cluster and the individual colored slices show the contribution of structures from the nine different simulations. The colored edges correspond to the sampling of each simulation and the length of the edge is proportional to the C α RMSD between any two connected cluster centers. The largest C α RMSD of 2.3 Å was between cluster 2 and cluster 4.

Figure 4.4

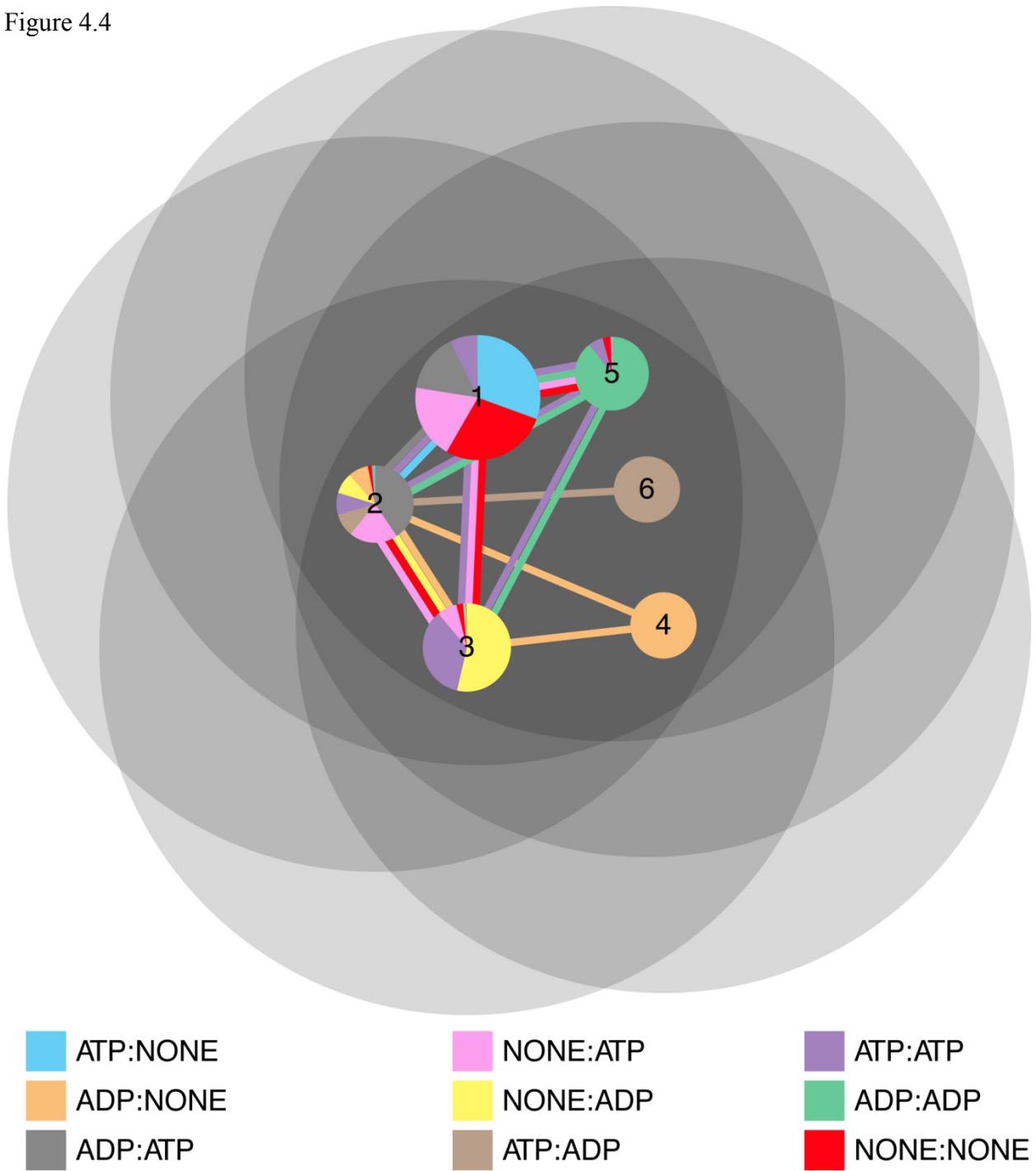


Figure 4.5 DNA base pair hydrogen bonding for C/G(-2), G/C(-1), G·T mismatch, and A/T(+1) base pairs from N3-N1 (C/G base pairs), N1-N3 (A/T base pairs), and N1-O4 (G·T mismatch) distance time series in each simulation are described here. Typical hydrogen bond distances of 3 Å are shown as blue dotted lines. (A) Wild type simulations with different nucleotide combinations. (B) F36A mutant simulations.

Figure 4.5

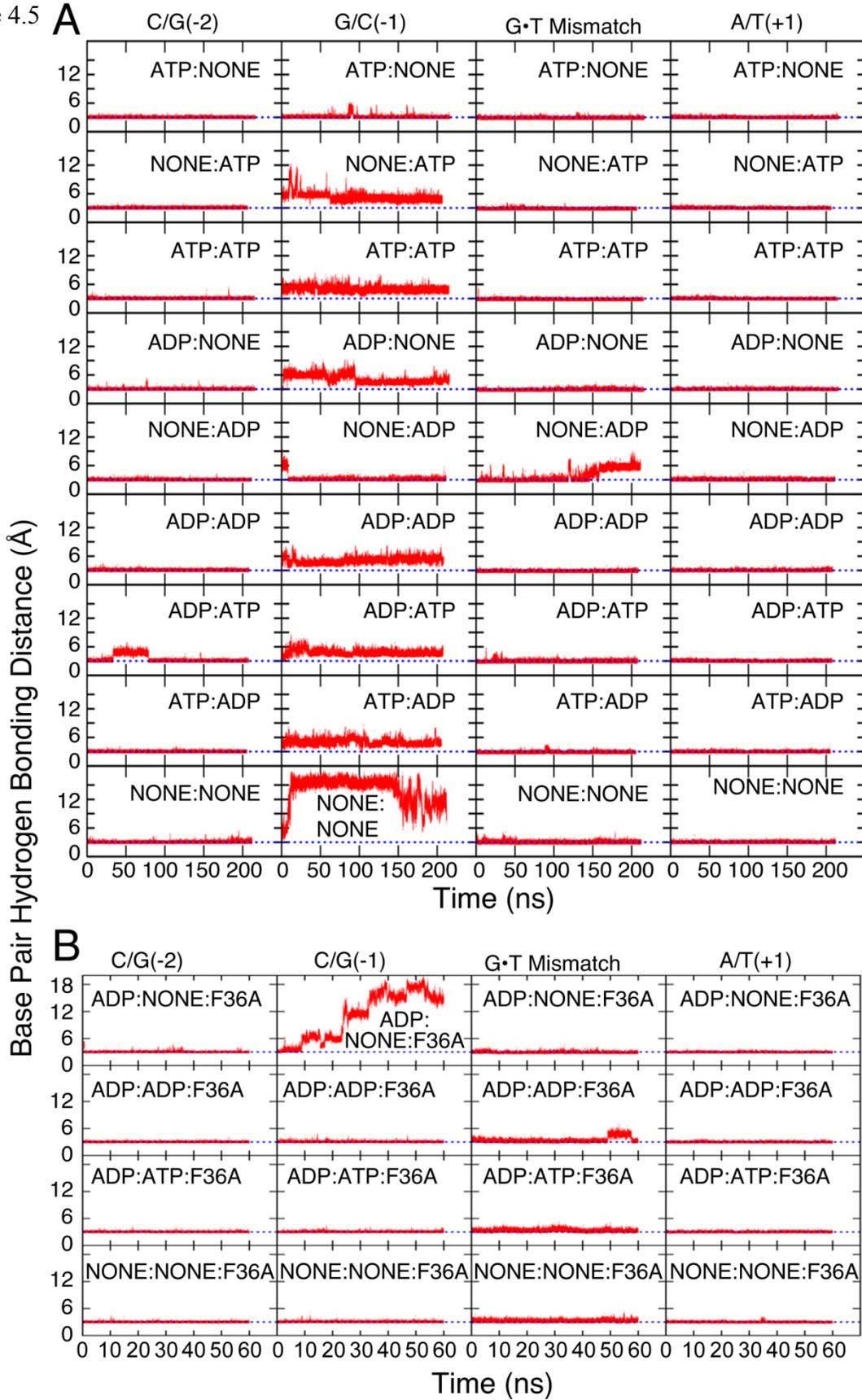


Figure 4.6 Comparison of the G·T and G/C(-1) major groove widths from the unbiased NONE:NONE simulation. The solid blue line corresponds to a canonical major groove width of 17 Å estimated from the B-DNA crystal structure 1BNA (145). Major groove widths were calculated using the 3DNA program v2.0 (152).

Figure 4.6

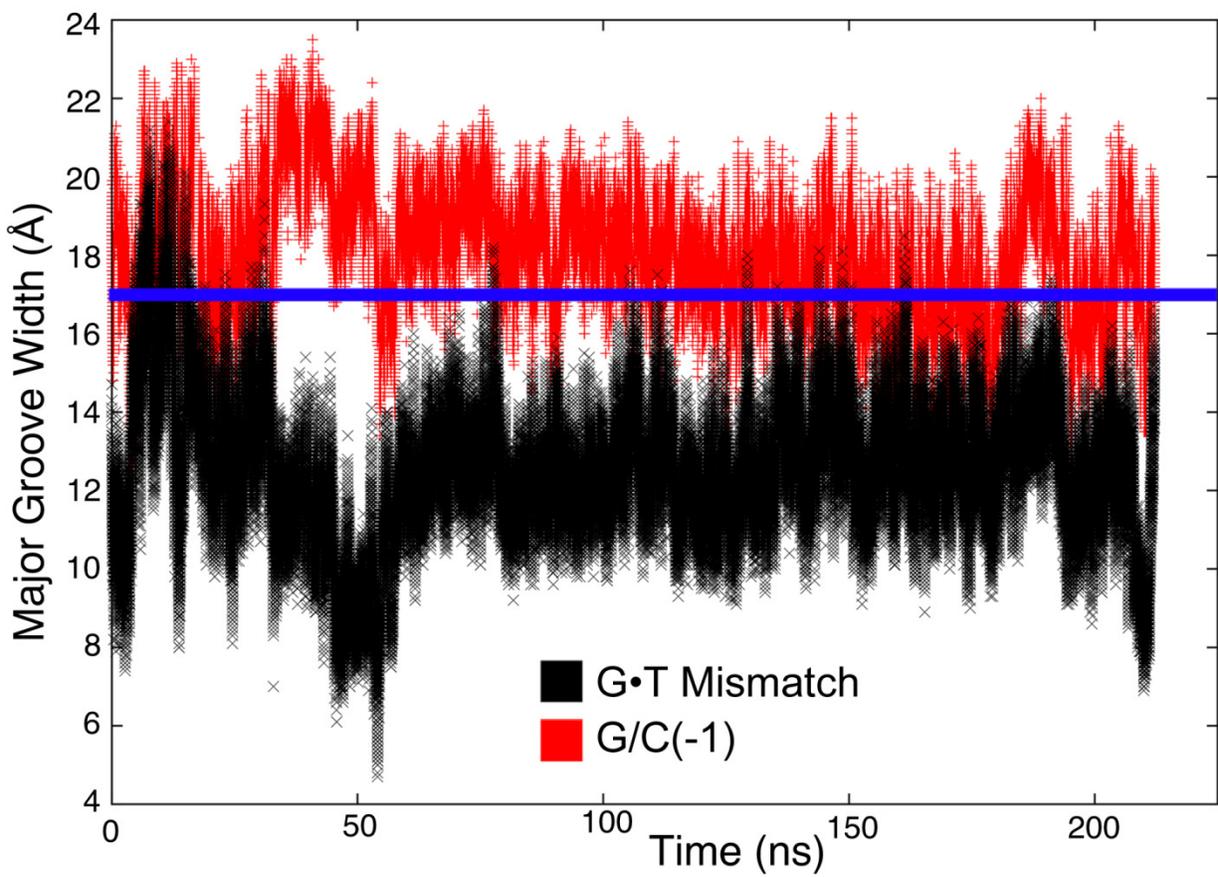


Figure 4.7 Comparison of T22 glycosyl rotation angle, χ , from the unbiased wild type and F36A mutant simulations.

Figure 4.7

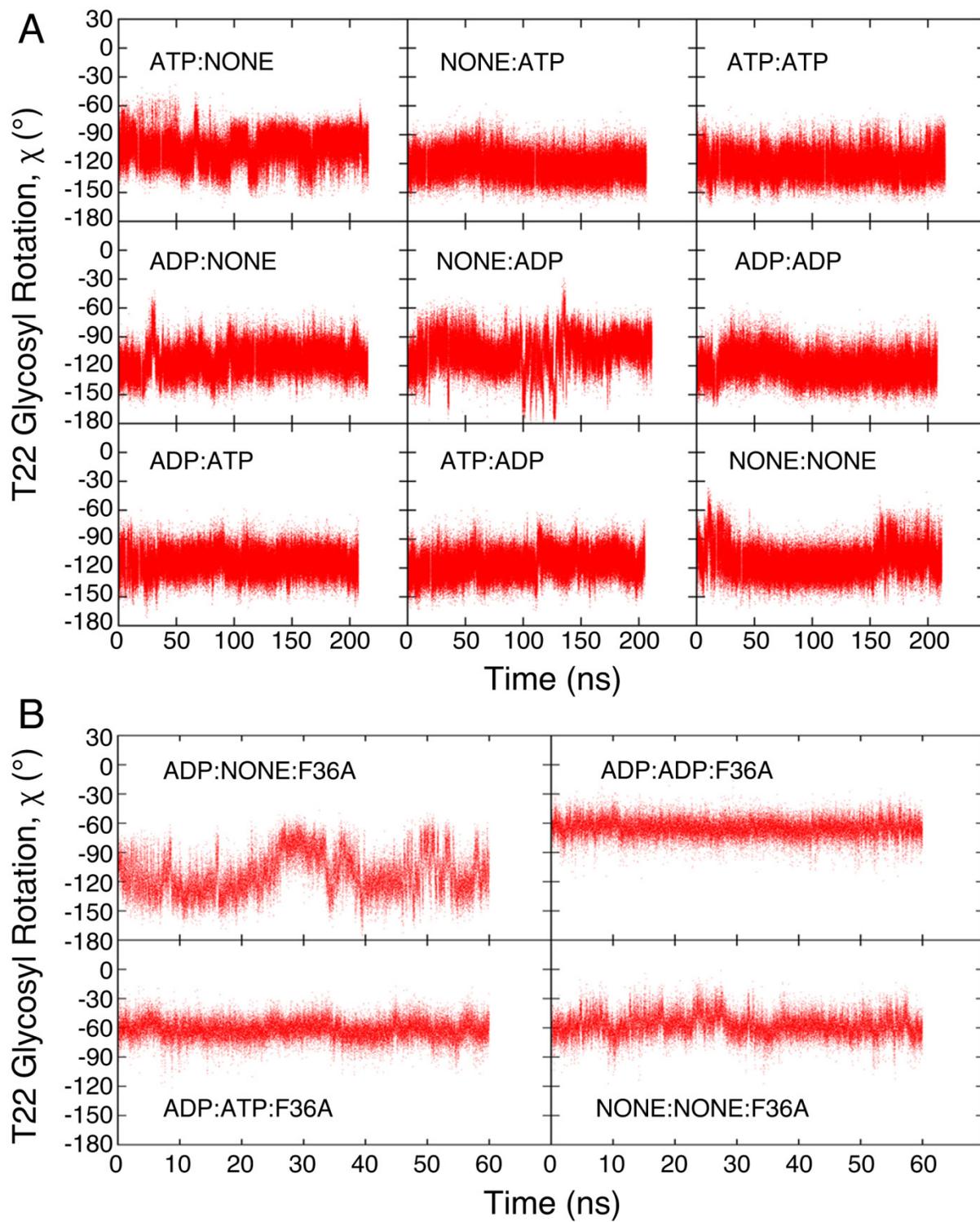


Figure 4.8 Correlation of C21 base-flipping in NONE:NONE simulation with various structural quantities (see *Methods* for definitions): (A) Pseudodihedral angle. (B) C21 backbone ζ torsion angle. (C) Movement of the S1 DNA binding domain (S1-D1) along X. (D) Movement of S2-D1 along X. (E) Movement of S2-D1 along Y. (F) Movement of S2-D1 along Z. (G) S2 Ser668 to S1 Asn616 C α -C α distance. (H) S1 Asn616 Ψ backbone torsion angle. (I) Salt bridge distance between S2 Arg667 and S1 Glu594 measured between heavy atoms. A distance of 3 Å corresponding to hydrogen bonding is shown as a blue line. (J) C α -RMSD of the S2 signature loop.

Figure 4.8

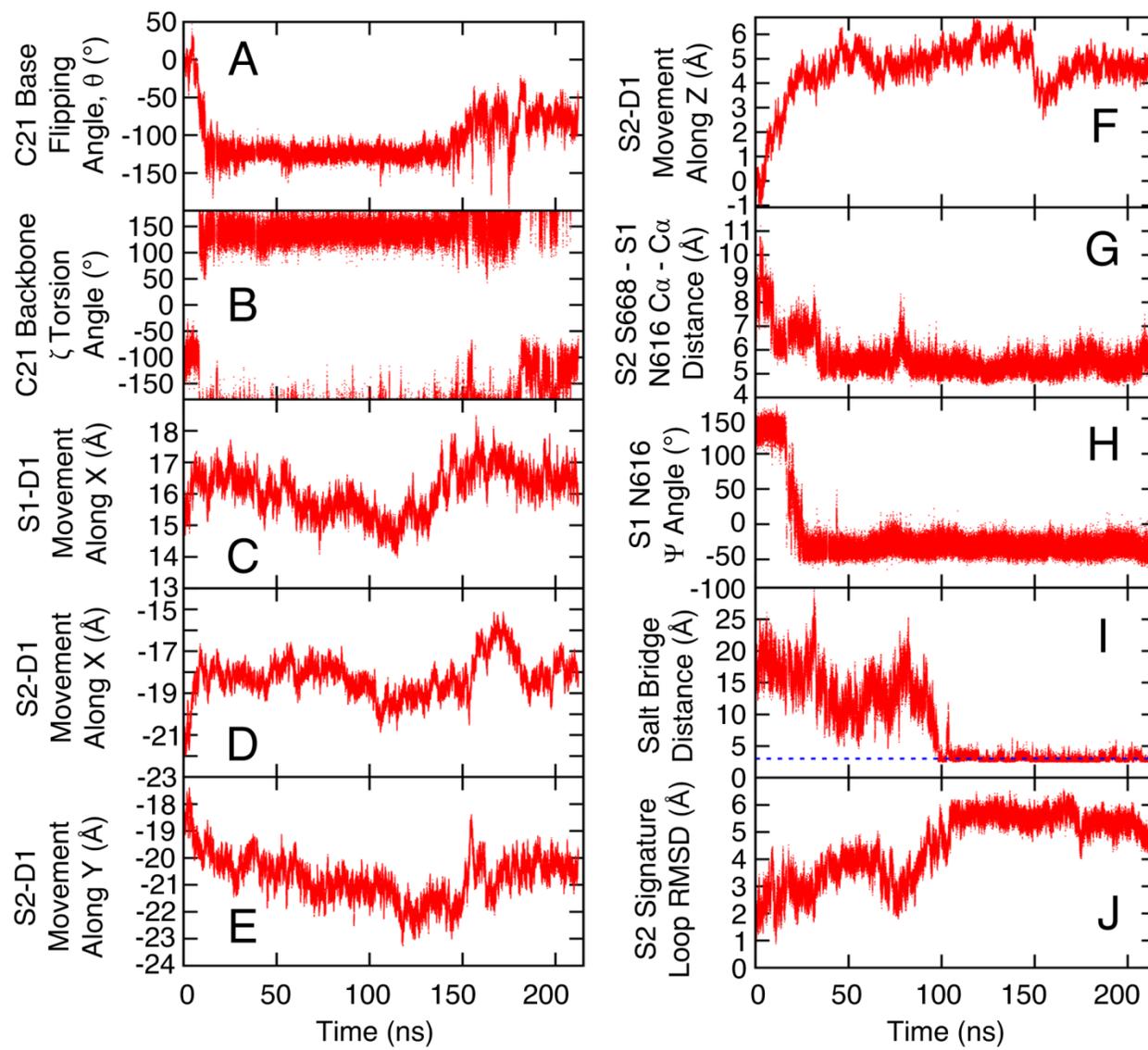


Figure 4.9 Snapshots from the NONE:NONE simulation of base flipping progress viewed from the major groove. Protein, water, and additional DNA are omitted for clarity. The G-T is colored in pink, G/C(-1) in yellow, and C/G(-2) in grey. The red arrow indicates C21.

Figure 4.9

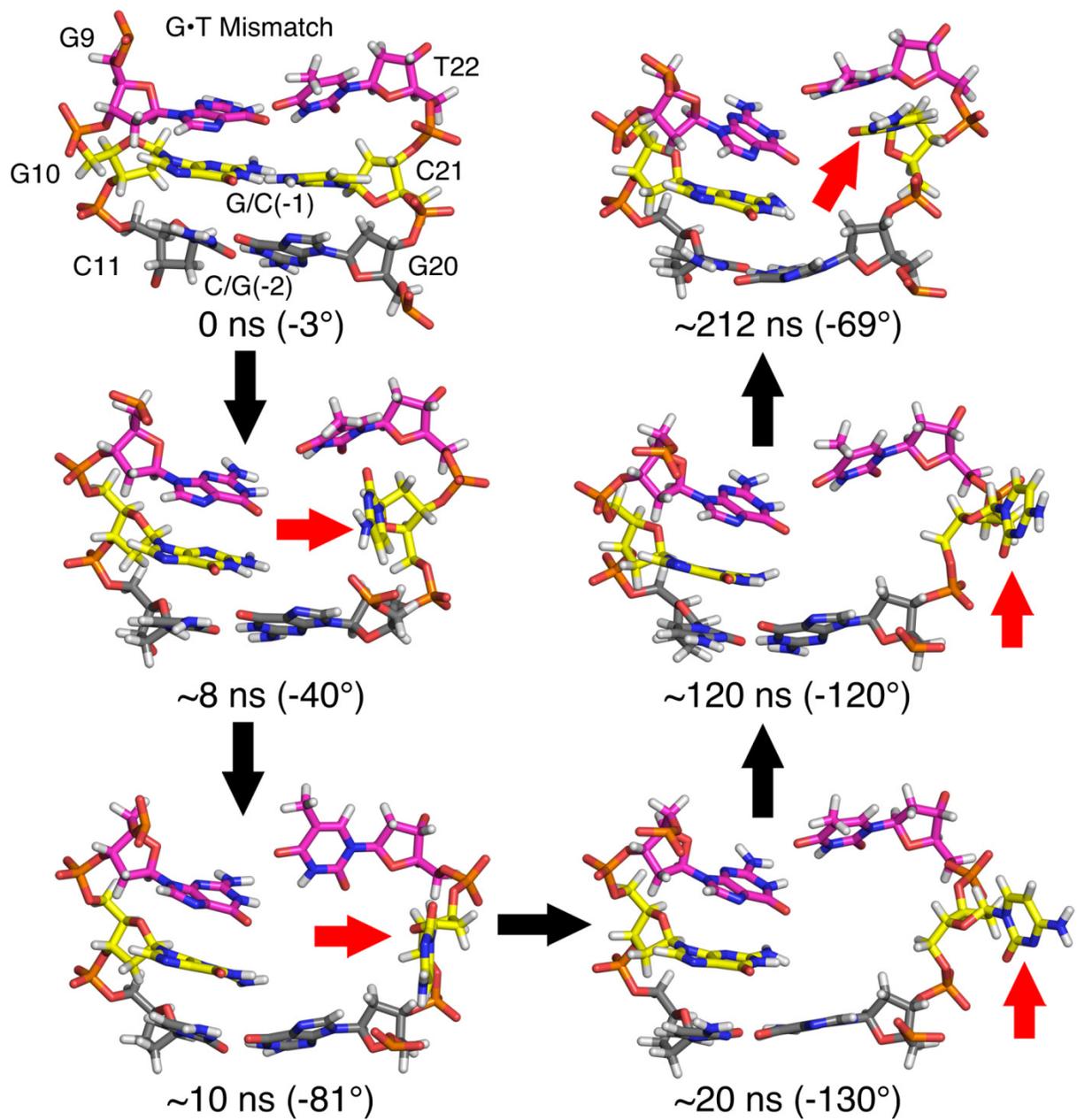


Figure 4.10 Free energy profiles from the HREM simulation: (A) Free energy of base flipping (10.5 ns/replica). (B) C21 backbone ζ torsion angle vs. base flipping. (C) Movement of S1-D1 along X vs. base flipping. (D) Movement of S2-D1 along X vs. base flipping. (E) Movement of S2-D1 along Y vs. base flipping. (F) Movement of S2-D1 along Z vs. base flipping. (G) S2 Ser668 to S1 Arg616 C α -C α distance vs. base flipping. (H) S1 Arg616 Ψ backbone torsion angle vs. base flipping.

Figure 4.10

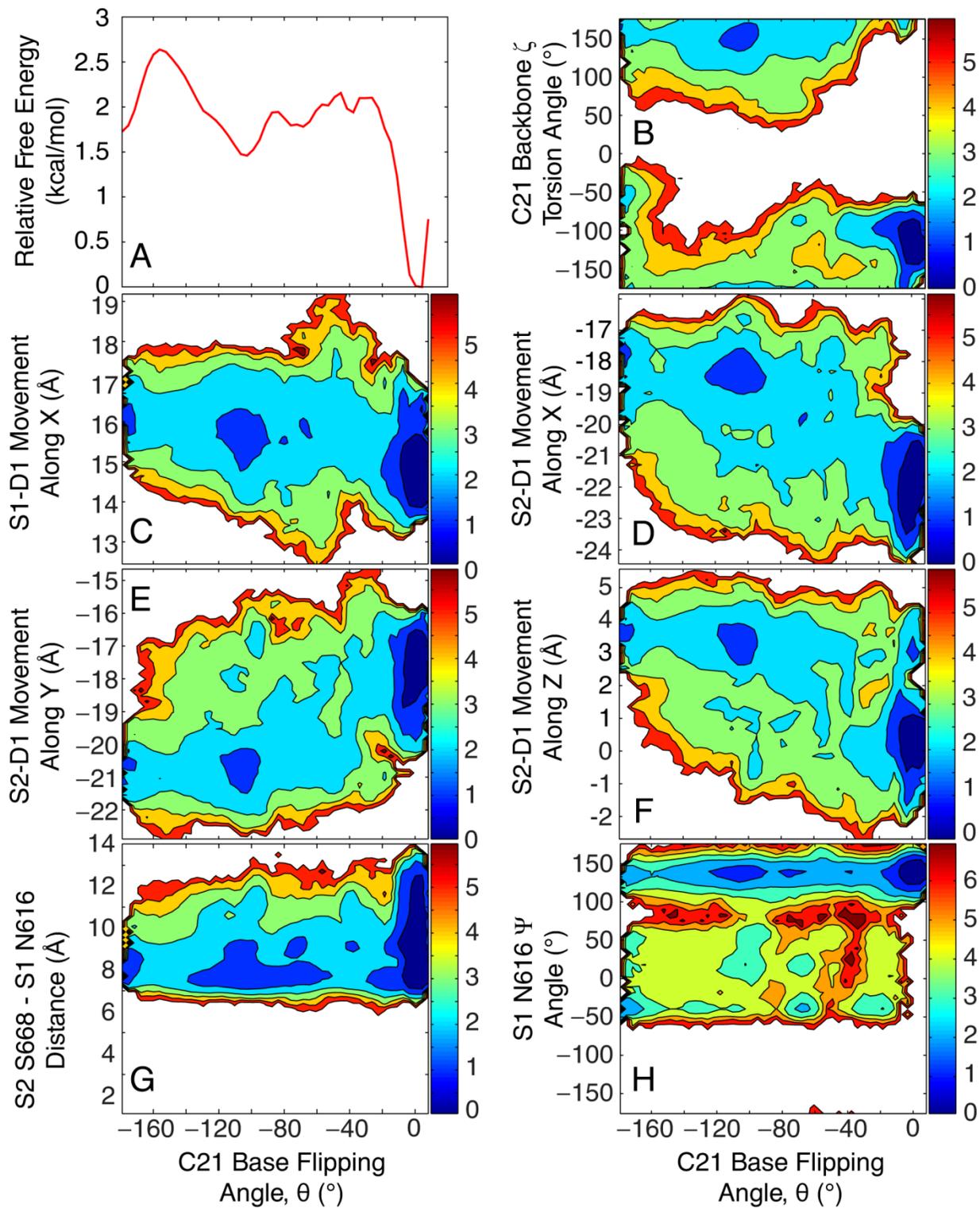


Figure 4.11 Comparison of the C21 backbone ζ torsion angle from the (A) unbiased wild type and (B) F36A mutant simulations.

Figure 4.11

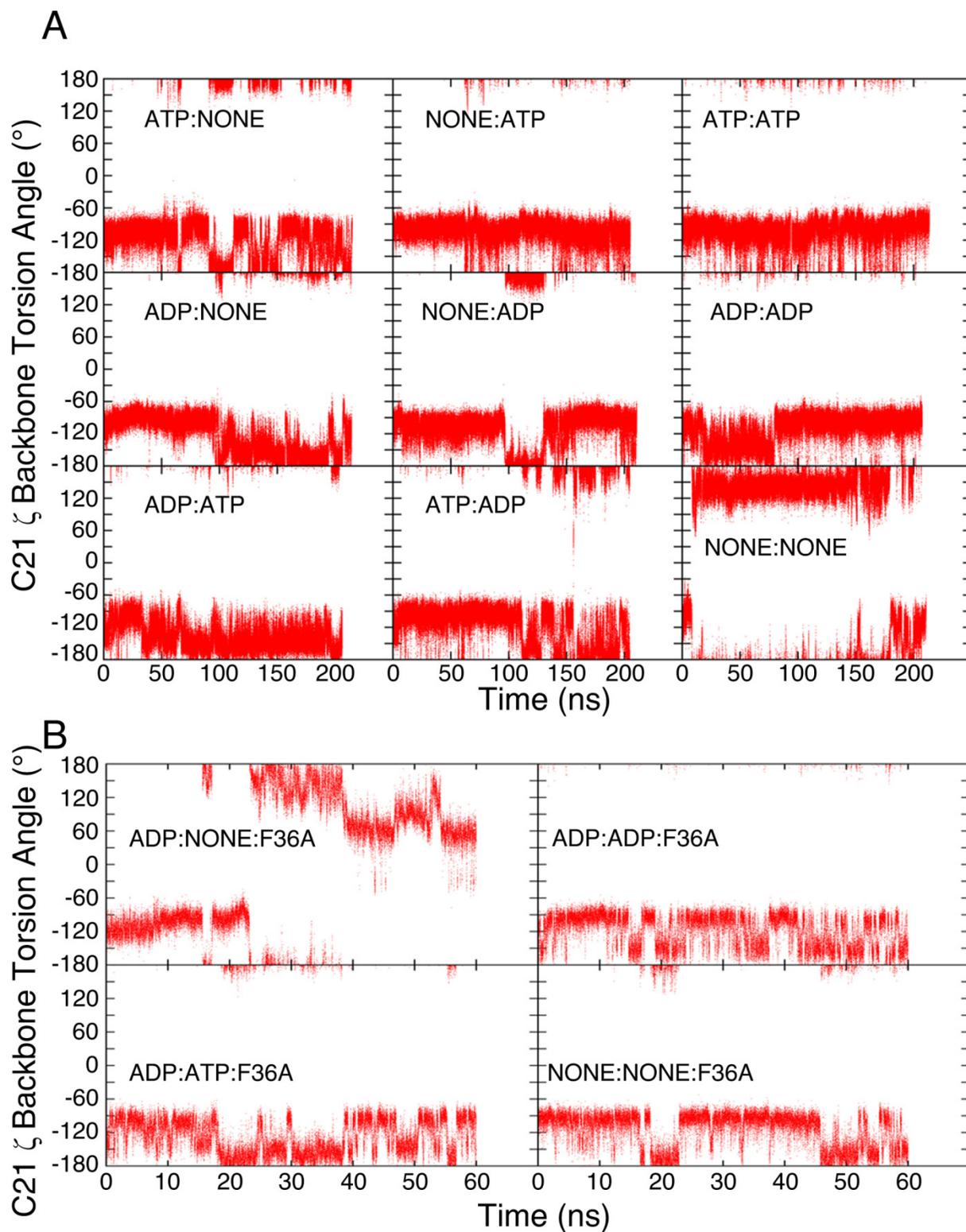


Figure 4.12 Free energy profiles from the same HREM simulation but generated from different length simulations. The PMF from Figure 4.10A is included here for comparison (red).

Figure 4.12

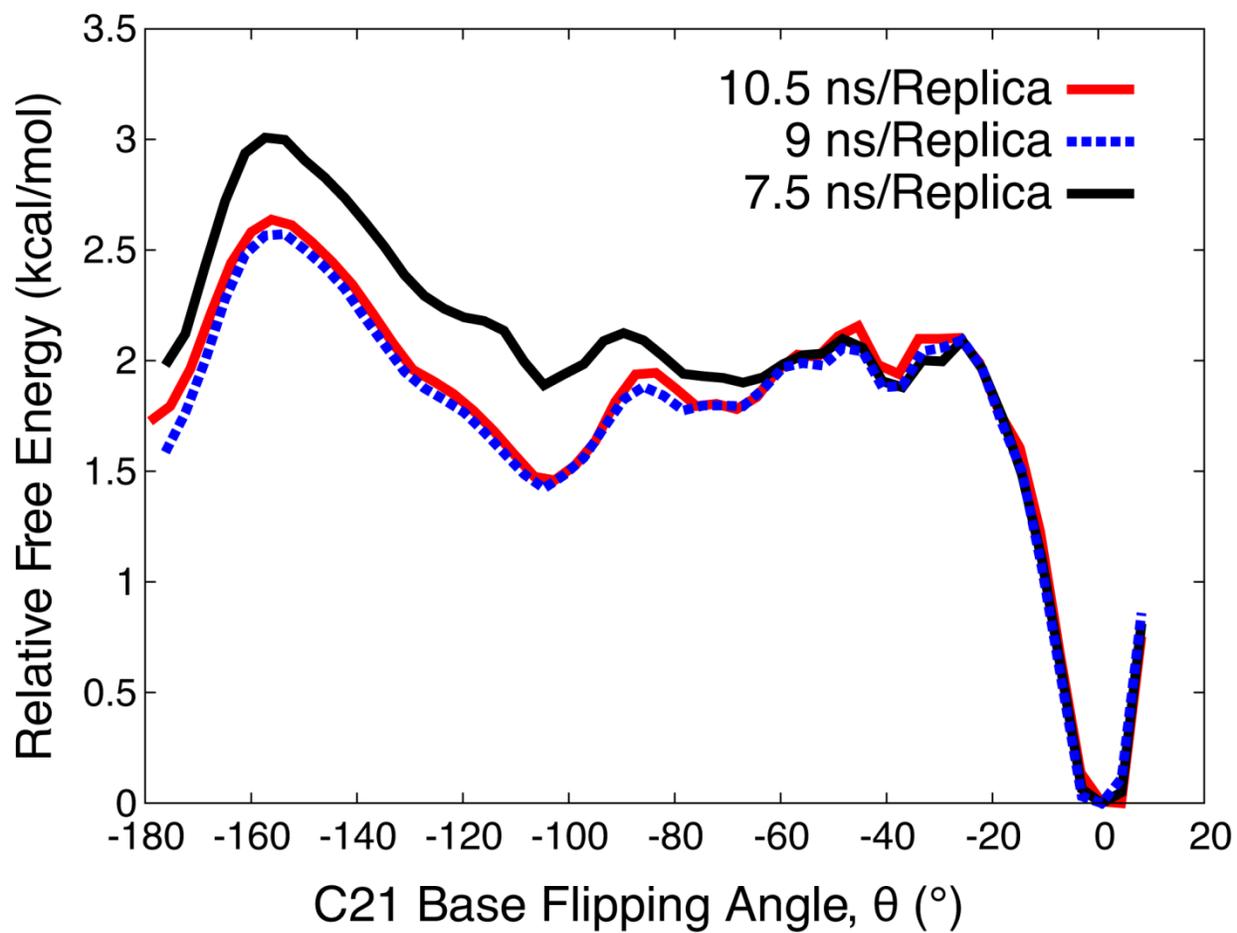


Figure 4.13 HREM sampling overlap. Thick red bars correspond to the range of equilibrium pseudodihedral angles, θ_i , prescribed by a given harmonic potential, while thin black lines correspond to the actual range of pseudodihedral angles, θ , that is sampled by each replica (see *Methods*).

Figure 4.13

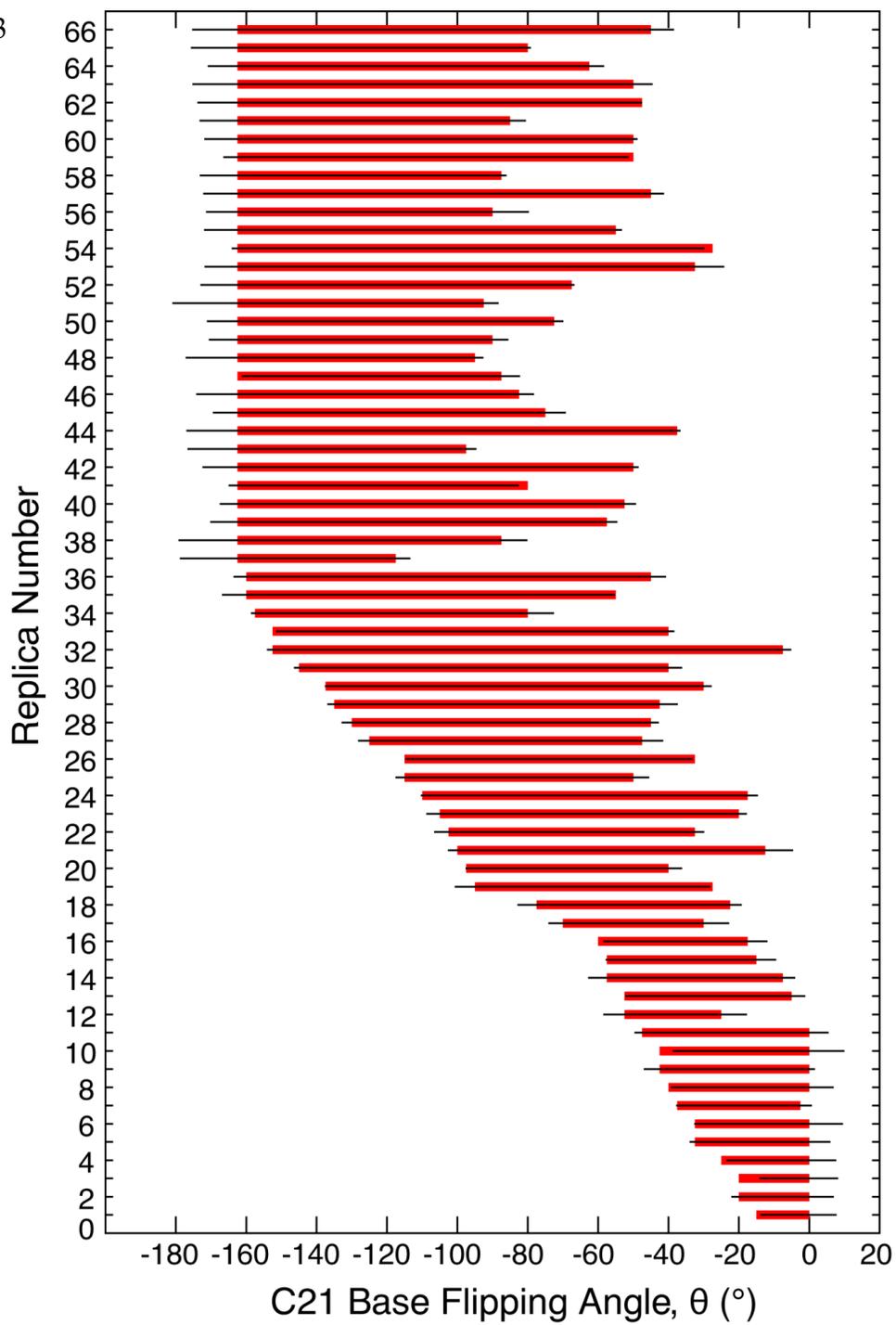


Figure 4.14 Water residence time calculations from the NONE:NONE simulation (see *Methods*). (A)-(B) Diagram depicting the water molecules that have entered into the minor groove side as a result of C21 base flipping. The S1 DNA binding domain is shown as a gray surface, the G/C(-1) base pair is colored yellow, the G·T mismatch is colored in magenta, the rest of the DNA is colored in brown (in (B) only), and water molecules located within 4 Å of the G10 base are colored as orange, blue, and green spheres (only waters within a 4 Å radius of the G10 base are used for the residence time calculation). Additional waters within 6 Å of the G10 base are colored in red and were included to illustrate the crowded environment. For clarity, only protein atoms within a 10 Å radius of the G10 base are shown. The white arrow points to the narrow 6 Å wide channel that is created when C21 is flipped out. Orange spheres correspond to fast moving waters with residence times less than 500 ps while green and blue spheres correspond to trapped waters with long residence times in the 1-10 ns range. (C)-(D) Water residence times for water molecules on the surface of the protein (away from the DNA) (C) before and (D) after C21 base flipping (note that the time is in picoseconds). (E)-(F) Water residence times for water molecules located within 4 Å of the G10 base (E) before and (F) after C21 base flipping. The black lines in (C)-(F) correspond to the inverse slope used to calculate the accompanying residence time.

Figure 4.14

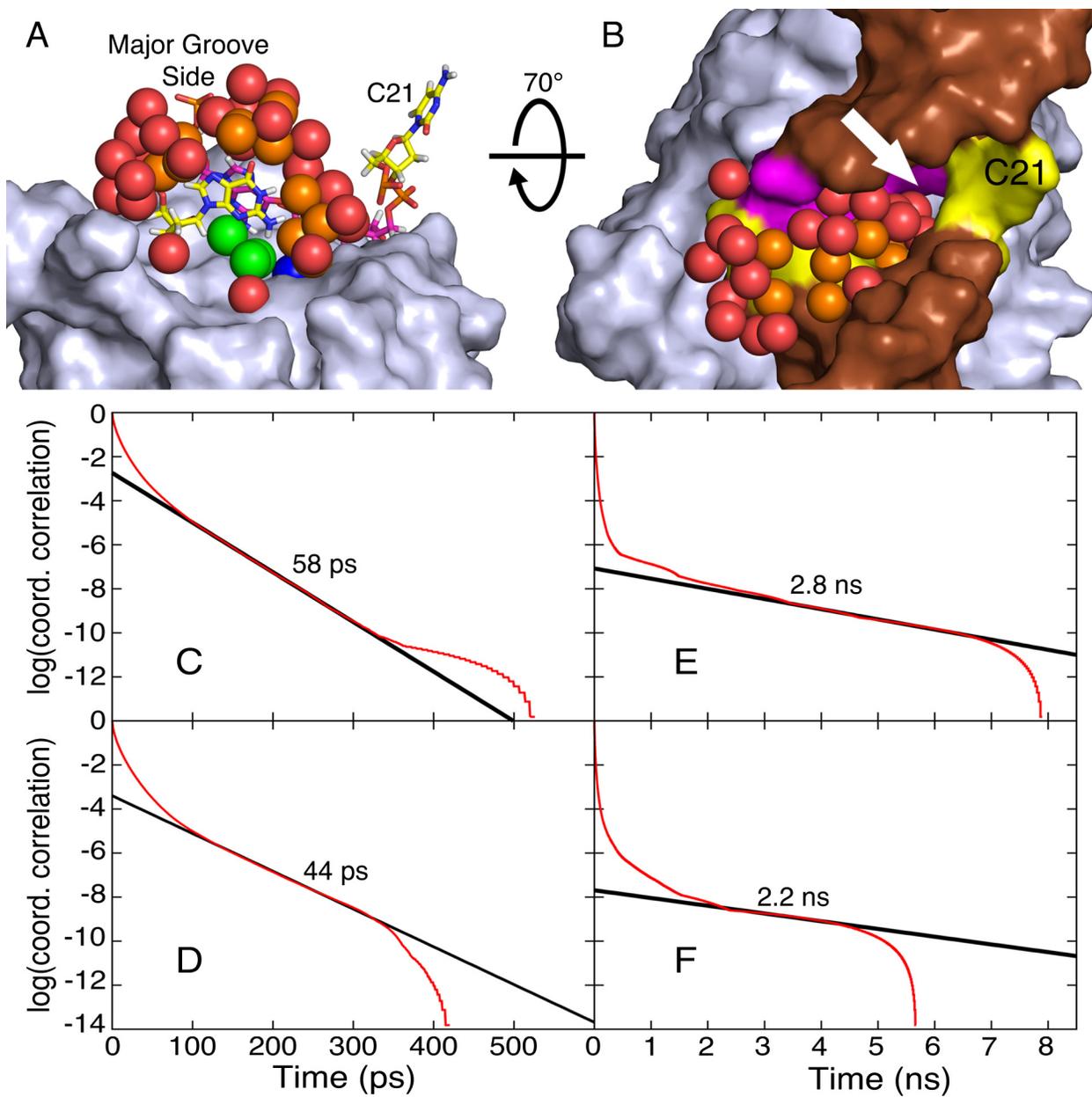


Figure 4.15 Solvent-accessible surface area (SASA) calculations from NONE:NONE and ATP:NONE simulations. Each plot show the time series for the change in solvent-accessible surface area (Δ SASA) of the H1' atom (bound to C1' on the minor groove side) upon binding to MutS. Results from the NONE:NONE simulation (top), where the base is flipped out after 10 ns, are compared with the ATP:NONE simulation (bottom), where the base remains stacked.

Figure 4.15

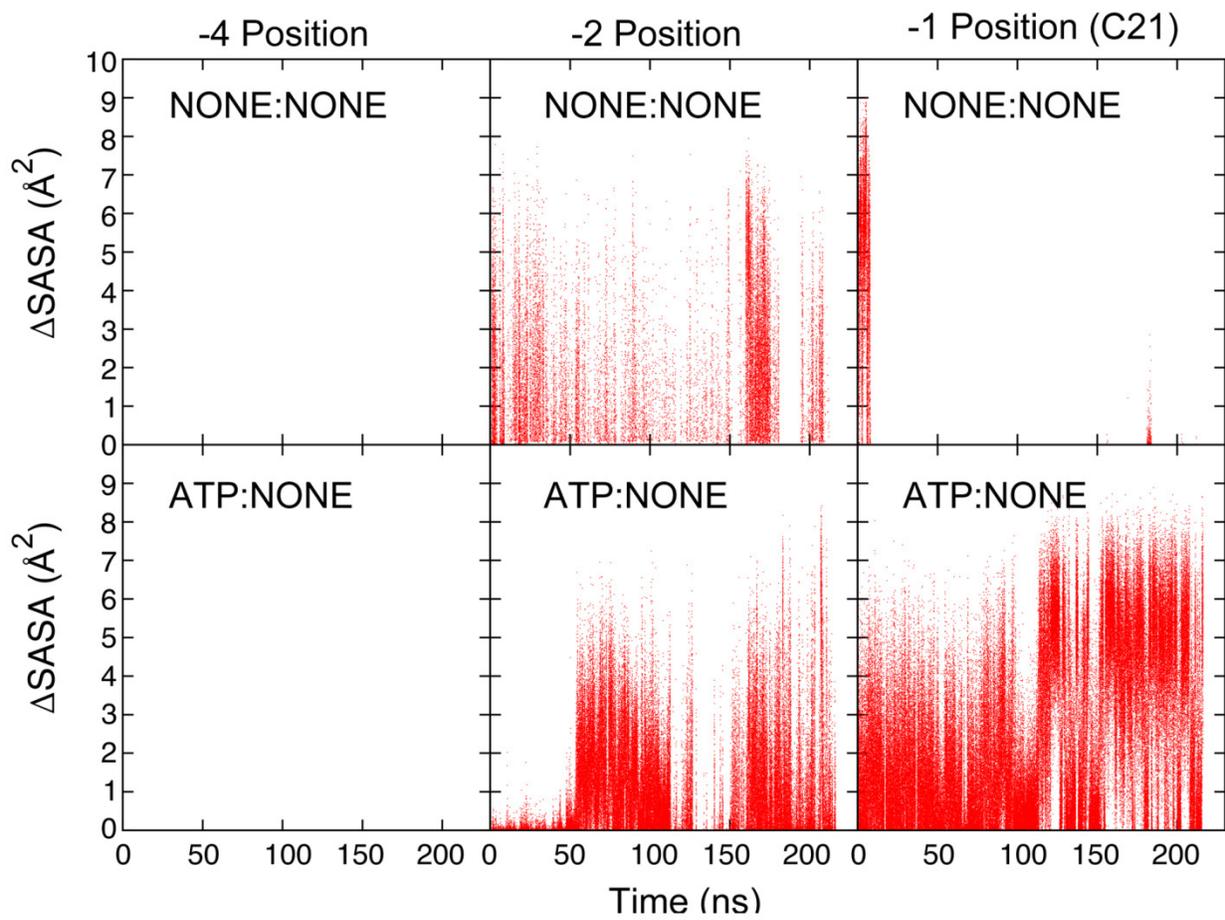


Figure 4.16 Protein domain motions. (A) Side and back view of the starting MutS-DNA complex along with the three orthogonal vectors, X, Y, and Z, used to describe the protein domain motions. The S1 DNA binding domain is red, the S2 DNA binding domain is pink, the DNA is brown, and the rest of the protein is colored white. Vector X is identical to the DNA helical axis, vector Y points upwards towards the phosphorous of cytosine 7, and vector Z is perpendicular to X and Y and points towards the S1 core domain. (B) Side and back view of the S1 and S2 DNA binding domains bound to DNA after ~200 ns of simulation time. The final simulation structure is colored in green and the orientation of the starting structure is identical to (A). (C)-(H) Comparison of the range of motion along the three orthogonal axes X, Y, and Z between trajectories with and without base flipping. The trajectory where base flipping is observed is colored in black and the remaining eight trajectories where no base flipping is observed is collectively colored in red. Movement of the S1 DNA binding domain is shown in (C) – (E) while movement of the S2 DNA binding domain is shown in (F) – (H).

Figure 4.16

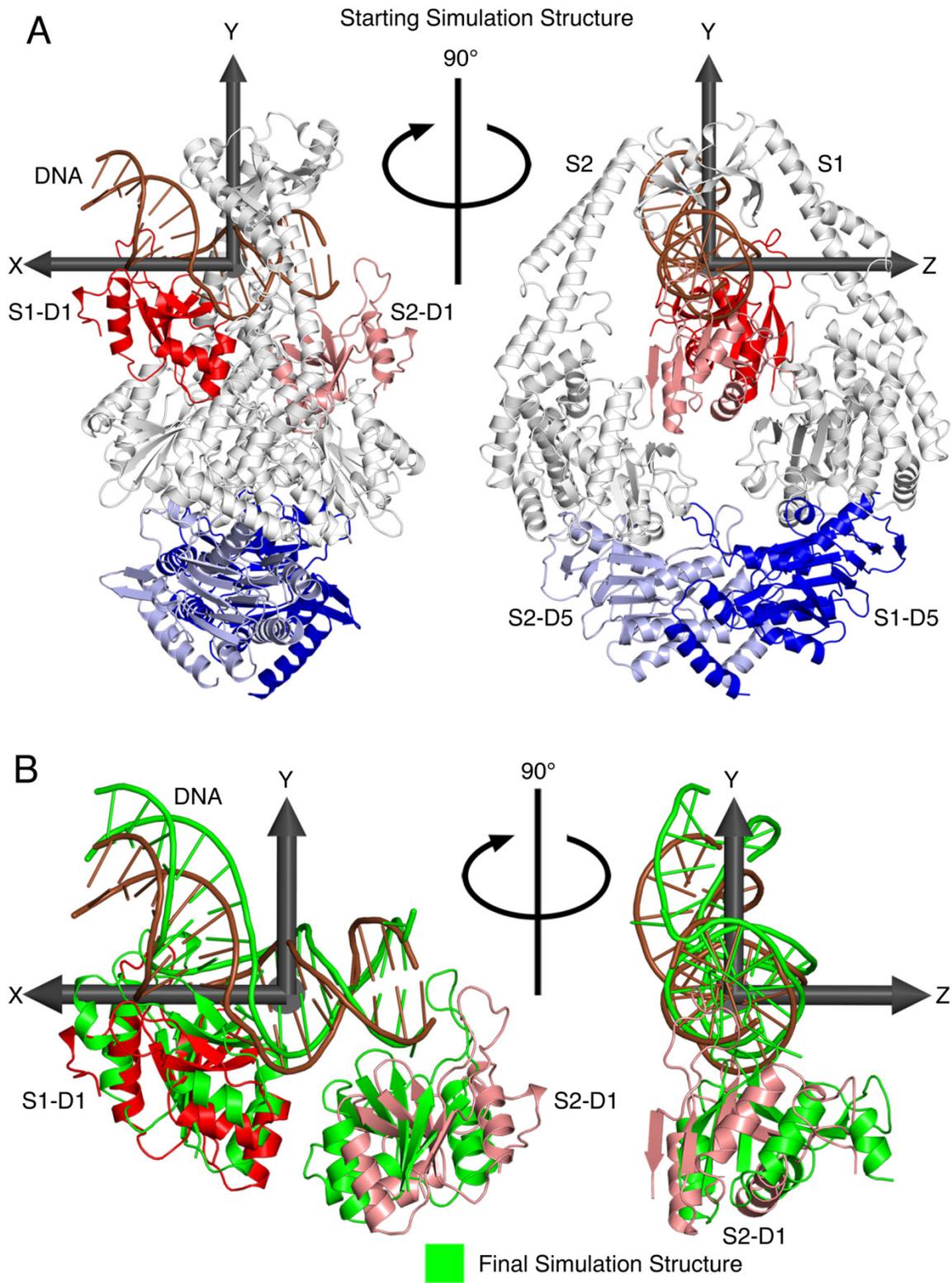


Figure 4.16 (Cont'd)

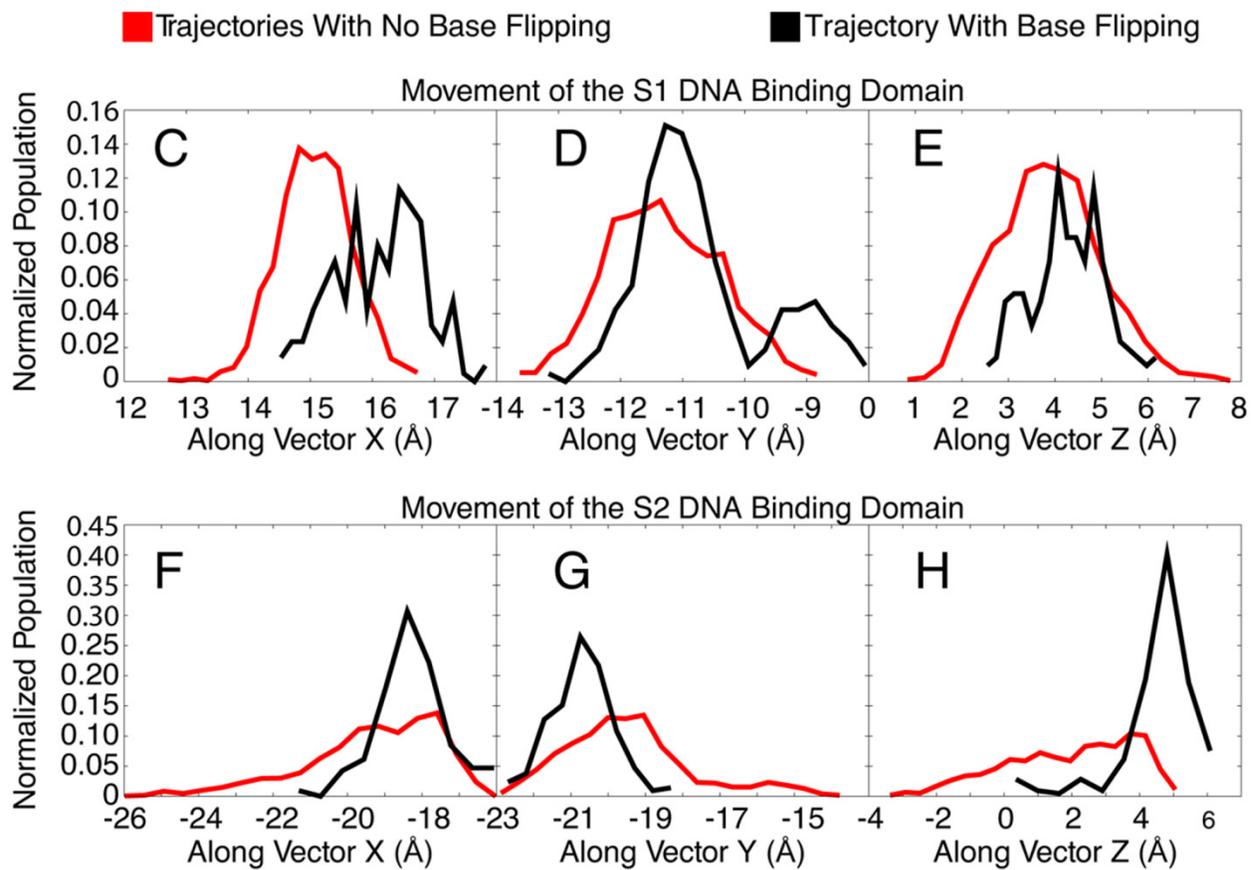


Figure 4.17 S1 DNA binding domain movement from unbiased and HREM (NONE:NONE) simulations along Y and Z directions. (A) Movement of the S1 DNA binding domain (S1-D1) along Y. (B) Movement of S1-D1 along Z. (C) Free energy profile of the movement of S1-D1 along Y with respect to the base opening angle. (D) Free energy profile of the movement of S1-D1 along Z with respect to the base opening angle.

Figure 4.17

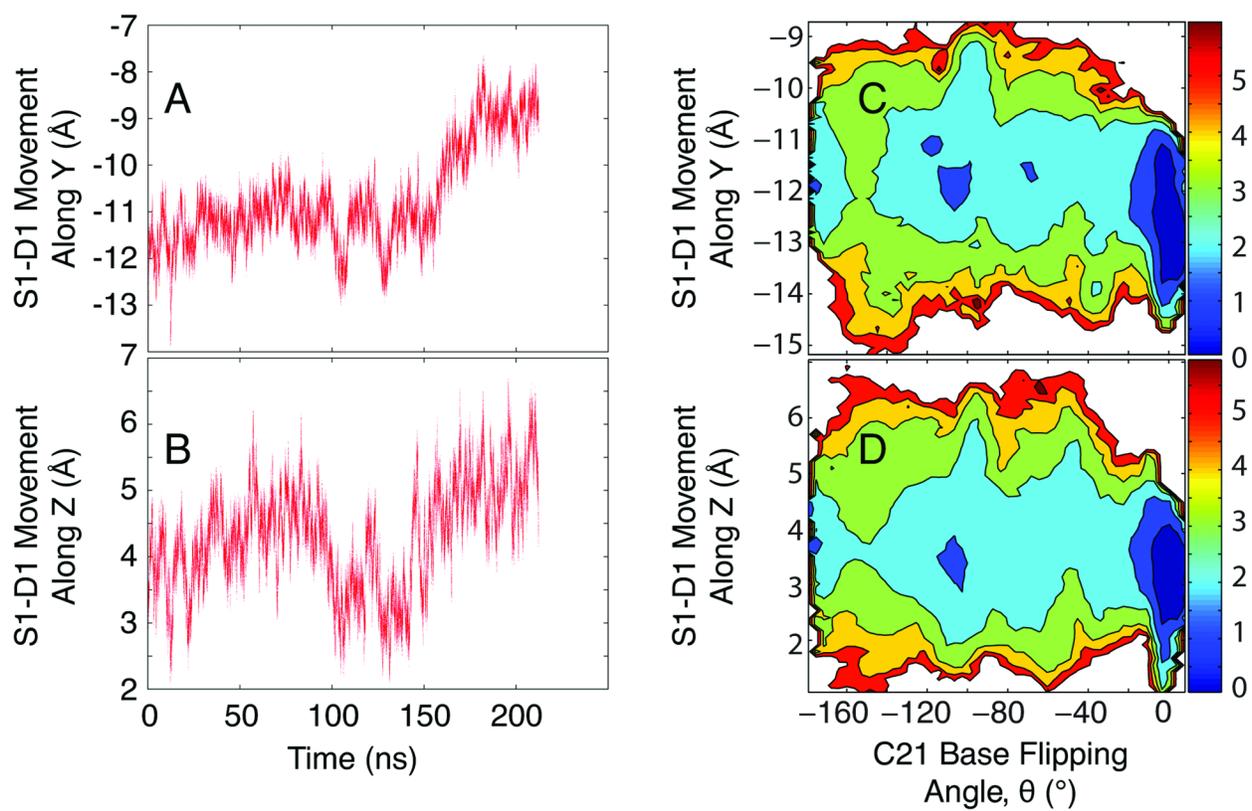


Figure 4.18 Allosteric signaling from the DNA binding domain to the ATPase domains. A structure-based sequence alignment was used (30) to map highly conserved residues onto the NONE:NONE model. The protein is shown as white ribbons and is in the same orientation as the full length structure found in Figure 4.1A. The conserved residues are highlighted as red spheres for the S1 monomer only. The same residues are conserved in the S2 monomer as well. DNA, water, and nucleotides have been omitted for clarity.

Figure 4.18

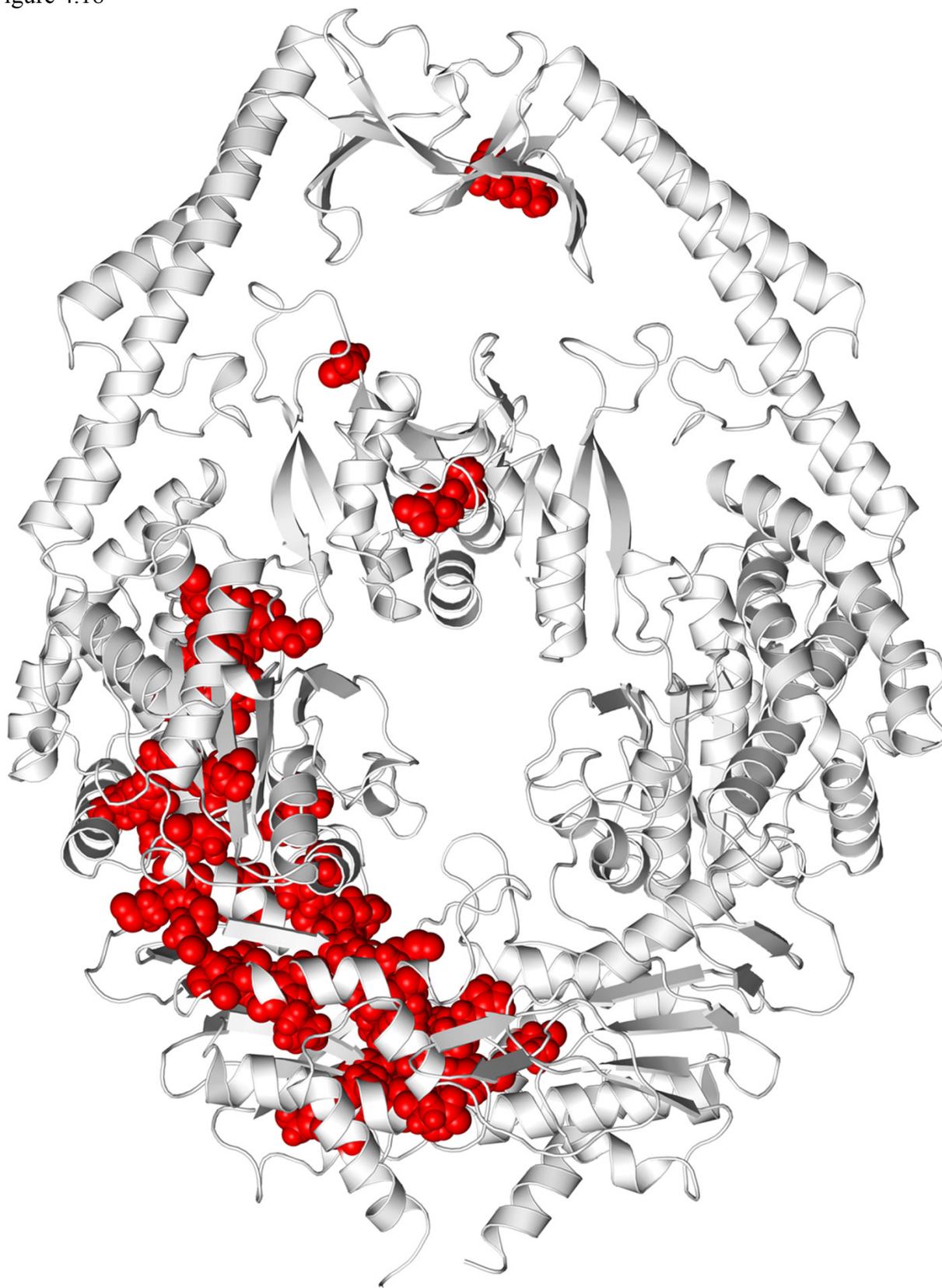
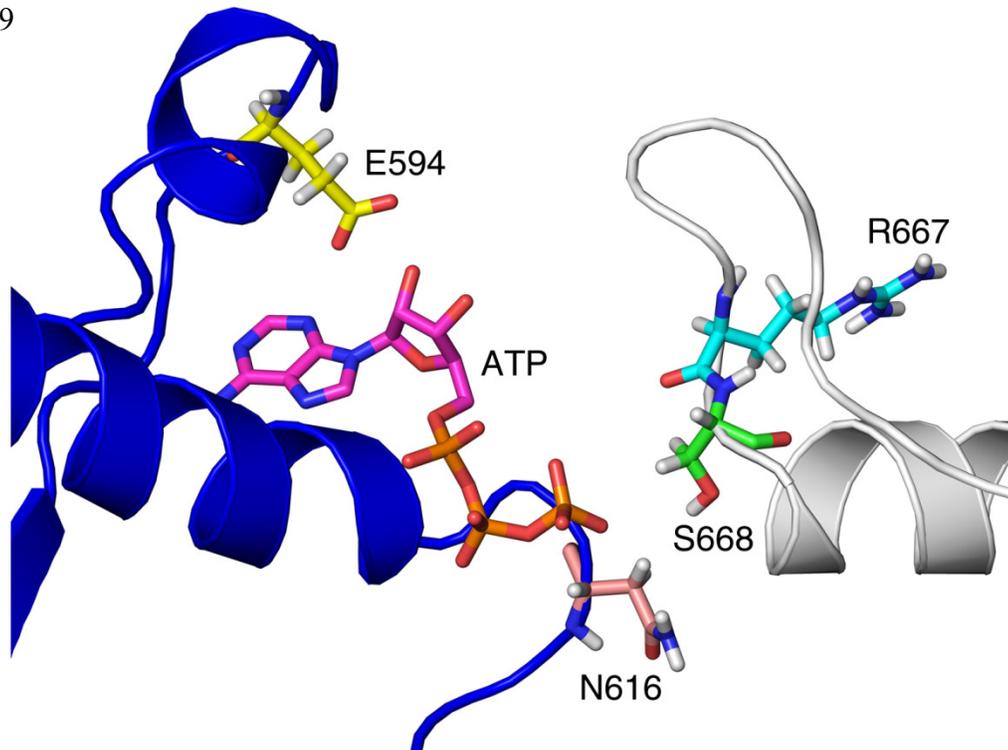


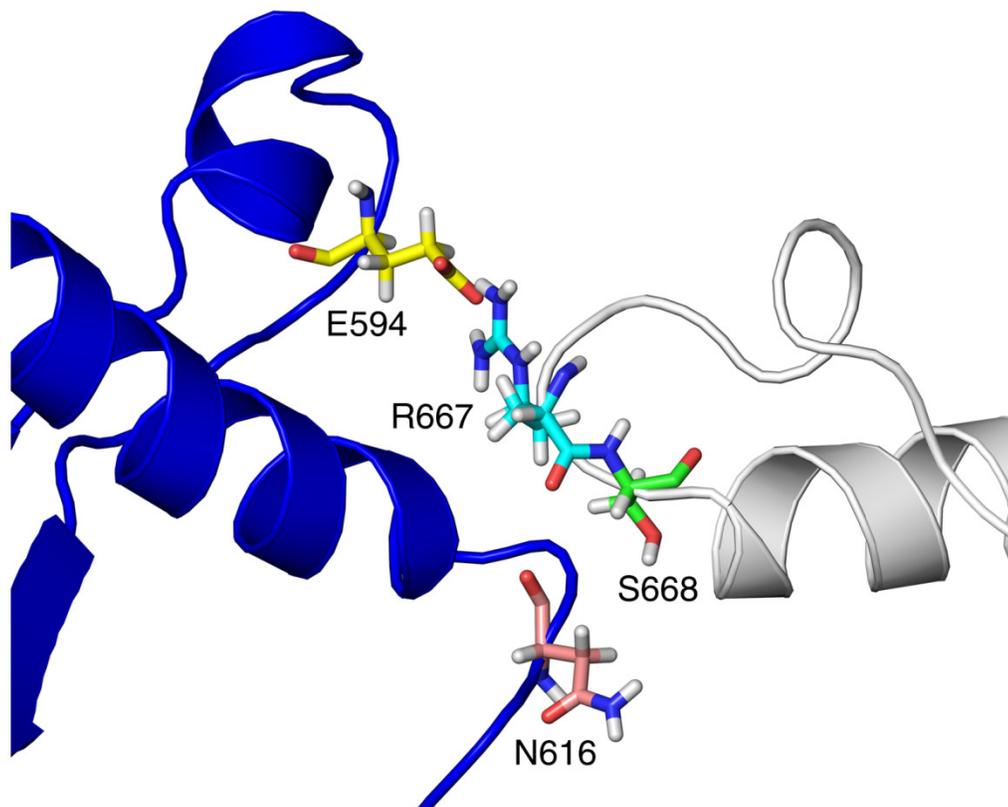
Figure 4.19 Visualization of the effects of base flipping on the ATPase domains. The S1 and S2 monomers are colored in blue and white, respectively. Ser668 is colored green, Asn616 is colored pink, Arg667 is colored cyan, Glu594 is colored yellow, and ATP is colored magenta. (A) Starting simulation structure modeled from the 1W7A crystal structure. ATP is modeled in for reference but is absent in the base-flipping simulation. (B) A post-base-flipping conformation from the base-flipped trajectory showing the Asn616 Ψ angle reorientation, salt bridge formation between Arg667 and Glu594, and stabilization of Ser668 and the S2 signature loop.

Figure 4.19

A



B



Chapter 5

A Path-Based Reaction Coordinate for Biased Sampling of Nucleic Acid Translocation

Sean M. Law, Afra Panahi, and Michael Feig

Afra Panahi contributed to the work on the Weighted Histogram Analysis Method.

5.1 Introduction

Protein-nucleic acid interactions are involved in many biological processes, including DNA replication, transcription, translation, DNA repair, DNA degradation, DNA packaging and homologous recombination. In many of these processes, translocation of proteins along double-stranded DNA plays a central role. Experimental investigations of protein-DNA translocation over the past decade include studies of helicases (153-158), FtsK DNA translocase (159-161), and the Zif268 zinc-finger protein (162). Computational studies have provided additional insight into the mechanisms of protein-DNA translocation and serve as a platform for developing new hypotheses (163-164). One such example is the observation of spontaneous forward and backward translocation of nucleic acid in the RNA polymerase II system during sub-microsecond MD simulations (53). While these results have extended the understanding of the eukaryotic transcription process, such findings are rare because translocation more commonly occurs on millisecond-second time scales.

Millisecond time scales and beyond remain largely inaccessible with conventional constant-temperature MD simulations but can be studied with enhanced sampling methods such as umbrella sampling. Such methods aid in overcoming kinetic barriers by biasing sampling along a suitable reaction coordinate. The choice of the reaction coordinate is often straightforward and may consist of an intramolecular distance, a torsional angle, or root mean square deviations (RMSD) from a given target structure. Protein-nucleic translocation, however, is more difficult to describe with a simple reaction coordinate since the underlying dynamics is often complex with both rotation and translation. Further complications arise when the nucleic acid is deformed during interactions with a protein.

In an earlier investigation using a so-called “path-search algorithm”, Ishida examined the free energy for branch migration in a RuvA-Holliday junction DNA complex by employing the umbrella sampling method (165). In their study, 68 harmonic biasing potentials were applied to 64 individual phosphorus atoms in addition to four other center-of-mass restraints (applied to four central bases) located in the heart of the Holliday junction. In each umbrella window, the 68 independent reaction coordinates were biased along a specific direction vector towards a target position, ultimately resulting in branch migration by a single base position. However, this procedure assumes that each phosphorus atom must migrate towards the exact position of the next target phosphorus atom and this method is limited to branch migration by only one base. More recently, Golosov *et al.* used targeted molecular dynamics (TMD) along with a more sophisticated reaction coordinate in order to study the translocation step in DNA replication by DNA polymerase I (166). However, the complex coordinate chosen to characterize the DNA motion is specific for DNA polymerase I and cannot be easily generalized to other nucleic acid translocation systems.

In the current study, we introduce a generalized path-based restraint that can be used in umbrella sampling simulations to enhance the sampling of a system (or a subset of the system) along any specific 3-dimensional path. For nucleic acid translocation, individual paths for each strand are pre-defined and the movement of each nucleotide is biased via its center-of-mass projection onto its corresponding path towards a target location relative to the path. The use of a center-of-mass projection allows the entire nucleotide to freely sample different conformations while maintaining an overall motion that runs parallel to the translocation path. We have tested our path-based restraint on the Hin recombinase protein-DNA complex (167) by translocating the DNA both in the forward and backward directions using umbrella sampling. The resulting

motion clearly follows a screw-like movement of the DNA that would be difficult to capture with general umbrella sampling potentials.

In the following, the theoretical basis of umbrella sampling is reviewed briefly, the proposed new biasing function for translocation is introduced, and a sample application is described.

5.2 Methods

Umbrella sampling

The time scale associated with transitions between two states depends on the kinetic barrier separating the two states. The rate of conformational transitions can be accelerated effectively by flattening the barrier. This idea is exploited in umbrella sampling where a biasing potential $U_{umbrella}(\xi)$ along a reaction coordinate ξ is applied (168). Simulations are then carried out with a biased energy function

$$E_{biased}(r^N) = E_{unbiased}(r^N) + U_{umbrella}(\xi) \quad (5.1)$$

The resulting sampling can be reweighted to obtain the potential of mean force (PMF) as a function of ξ according to:

$$w(\xi) = -U_{umbrella}(\xi) + w_{biased}(\xi) + f \quad (5.2)$$

where the biased free energy, $w_{biased}(\xi)$, is obtained from the sampling probability function $p_{biased}(\xi)$ in the biased simulation as

$$w_{biased}(\xi) = -kT \ln p_{biased}(\xi) \quad (5.3)$$

and f is a constant (with respect to ξ) and defined according to

$$e^{-\beta f} = \frac{\int e^{-\beta E_{biased}(r^N)} dr^N}{\int e^{-\beta E_{unbiased}(r^N)} dr^N} \quad (5.4)$$

The choice of reaction coordinate is key to successful application of the umbrella sampling method. Ideally, progress along the reaction coordinate would coincide closely with the minimum energy transition path between the two states with minimal sampling of low-energy states in orthogonal directions. Then, the umbrella sampling method can effectively guide the transition between two states over significant kinetic barriers.

In practice, the use of a single umbrella is often not sufficient to study transitions in more complex systems (*i.e.* DNA translocation). Instead, a series of i simulations is carried out with biasing functions $U_{umbrella,i}(\xi)$ that progress along the reaction coordinate in a stepwise fashion with sampling limited to overlapping ranges of ξ . As a result, piecewise PMFs $w_i(\xi)$ are obtained from each simulation which subsequently need to be combined into a total PMF $w(\xi)$ along the entire range of ξ . The combination of multiple PMFs into a single PMF along the entire transition path is possible with the weighted histogram analysis method (WHAM) (169) or the more recently introduced multi-state Bennett acceptance ratio method (MBAR) (170).

Path-Based Restraint Potential

An arbitrary 3-dimensional path is represented by a set of P points that are connected via $P-1$ piece-wise smooth cubic spline equations, $S_q(t_q)$ ($q=1, \dots, P-1$). $t_q \in [0,1]$ describes interpolated positions on the q^{th} spline piece between spline points $S_q(0)$ and $S_q(1) = S_{q+1}(0)$. Based on the piece-wise t_q values, a continuous reaction coordinate ξ is introduced as

$\xi = q + t_q$ with positions on the entire path given as $S(\xi)$ so that, for example,

$$S(\xi = 1.5) = S_1(t_q = 0.5) \text{ and } S(\xi = 14.92) = S_{14}(t_q = 0.92).$$

A path-based restraint potential can then be defined by comparing the projection of the center of mass (COM) of a set of atoms onto such a path with a given reference value according to:

$$U_{umbrella}(\xi) = K(\xi - \xi_0)^2 \quad (5.5)$$

where K is the force constant in units of kcal/mol (since ξ has arbitrary units), ξ is the instantaneous COM projection in the moving system, and ξ_0 is the target value at which the umbrella potential becomes zero. Because only the projection onto the path is restrained the system is free to explore conformational space orthogonal to the path thereby allowing, for example, a protein-nucleic acid complex to respond structurally when the protein translocates along the nucleic acid.

If multiple COMs are projected onto different parts of the path and the biasing potential is used to advance the entire structure along the path (as in protein-nucleic acid translocation) it is more convenient to use ξ and ξ_0 values relative to initial values from a reference structure rather than absolute values. It then becomes possible to advance multiple points along different sections of a given path with a single target value. In that case the biasing potential becomes:

$$U_{umbrella}(\xi_1, \dots, \xi_M) = \sum_{\alpha=1}^M K_{\alpha} (\xi_{\alpha} - \xi_{\alpha,initial} - \xi_0)^2 \quad (5.6)$$

where ξ_{α} and $\xi_{\alpha,initial}$ correspond to the moving and initial COM projections for the α^{th} set of atoms, respectively, and ξ_0 is the relative change of all COM projections from their initial values.

We will now discuss how to obtain ξ , the projection of an atom or center of mass (COM) of a set of atoms onto a given path. A projection onto a line is defined as the closest point on the path from a given reference point. In the most general case this involves an analytically intractable quintic equation with potentially multiple solutions. In order to simplify the problem, we approximate the path at ξ by a secant vector that passes through the points $S(\xi_0 + \Delta\xi)$ and $S(\xi_0 - \Delta\xi)$. Because of the umbrella potential, sampling of ξ is assumed to be sufficiently close to ξ_0 for the approximation of a locally linear path to be reasonable. In order to obtain a good approximation of the part of the path sampled with a given umbrella potential, we choose $\Delta\xi = 1 / \sqrt{K}$. The projection of a given COM onto the linear path can then be solved analytically with a unique solution. If ξ is not close to ξ_0 or if it is unknown, such as at the beginning of a

simulation, tangent vectors are first defined from the P-1 spline points and subsequently used to find the closest spline segment and to obtain an initial estimate of the projection, ξ_{guess} . Then a refined estimate is obtained as above by using the secant vector that passes through the points at $(\xi_{guess} + \Delta\xi)$ and $(\xi_{guess} - \Delta\xi)$. To ensure that a closer point does not exist in the two neighboring spline pieces, we also project the COM onto the neighboring splines using the same protocol. From our experience, this two-step approach for determining an initial value of ξ is sufficiently robust.

The path-based restraint potential was implemented in CHARMM (v. c36a4).

Generalized Weighted Histogram Analysis Method

Calculation of unbiased PMFs from umbrella sampling simulations (74) is commonly addressed using WHAM (169) (171-173). WHAM is often applied to obtain a PMF for the (single) reaction coordinate that is used during umbrella sampling but an extension of WHAM to multi-dimensional umbrella potentials or to the calculation of PMFs for other reaction coordinates than the one(s) used in the umbrella potential is not so straightforward. In the case of a multi-dimensional biasing potential with M coordinates $\xi_1 \dots \xi_M$ one obtains:

$$p_{unbiased}(\xi_1, \dots, \xi_M) = e^{\beta U_{umbrella}(\xi_1, \dots, \xi_M)} p_{biased}(\xi_1, \dots, \xi_M) e^{-\beta f} \quad (5.7)$$

The PMF for a different reaction coordinate η is formally given according to:

$$\begin{aligned}
P_{unbiased}(\eta) &= \int \delta(\eta - \eta'(\xi_1, \dots, \xi_M)) P_{unbiased}(\xi_1, \dots, \xi_M) d\xi_1 \dots d\xi_M \\
&= e^{-\beta f} \times \\
&\quad \int \delta(\eta - \eta'(\xi_1, \dots, \xi_M)) e^{\beta U_{umbrella}(\xi_1, \dots, \xi_M)} P_{biased}(\xi_1, \dots, \xi_M) d\xi_1 \dots d\xi_M \\
&\equiv e^{-\beta f} \Xi(\eta)
\end{aligned}
\tag{5.8}$$

$P_{unbiased}(\eta)$ can be obtained numerically by first constructing the M -dimensional histogram for $P_{unbiased}(\xi_1, \dots, \xi_M)$ according to Eq. (5.7) and then accumulating all of the elements where $\eta(\xi_1, \dots, \xi_M)$ falls into a given bin in a one-dimensional histogram from N umbrella windows:

$$P_{unbiased}(\eta_j) = \sum_{i_1=1}^{N_1} \dots \sum_{i_M=1}^{N_M} P_{unbiased}(\xi_{1,i_1}, \dots, \xi_{M,i_M}) \Big| \eta(\xi_{1,i_1}, \dots, \xi_{M,i_M}) \in [\eta_j, \eta_j + \Delta\eta]
\tag{5.9}$$

where $\eta_j = \eta_{\min} + j\Delta\eta$ and $\xi_{i,j} = \xi_{i,\min} + j\Delta\xi_i$ with histogram bin widths $\Delta\eta$ and $\Delta\xi_i$. In practice, the explicit construction of the M -dimensional histogram is not desirable for large numbers of M because of computer memory limitations and because the limited length of typical simulations does not generate sufficient sampling for conventional multidimensional WHAM to converge. Instead, the histogram for $P_{unbiased}(\eta)$ can be accumulated on the fly without explicitly constructing the M -dimensional histogram. The combination of Eq. (5.7) and (5.9) gives:

$$P_{unbiased}(\eta_j) = \left(e^{-\beta f} \times \sum_{i_1=1}^{N_1} \dots \sum_{i_M=1}^{N_M} e^{\beta U_{umbrella}(\xi_{1,i_1}, \dots, \xi_{M,i_M})} p_{biased}(\xi_{1,i_1}, \dots, \xi_{M,i_M}) \Big| \eta \in [\eta_j, \eta_j + \Delta\eta] \right) \quad (5.10)$$

Since $p_{biased}(\xi_{1,i_1}, \dots, \xi_{M,i_M})$ is obtained from the simulation as the fraction of samples where all of the variables $\xi_1 \dots \xi_M$ fall into a particular bin in the M -dimensional histogram, $P_{unbiased}(\eta)$ can be calculated by directly accumulating $e^{\beta U_{umbrella}(\xi_1, \dots, \xi_M)}$ from each simulation frame into the bin where $\eta(\xi_1, \dots, \xi_M) \in [\eta_j, \eta_j + \Delta\eta]$:

$$P_{unbiased}(\eta_j) = \frac{e^{-\beta f}}{n} \sum_{i=1}^n e^{\beta U_{umbrella}(\xi_1, \dots, \xi_M)} \Big| \eta(\xi_1, \dots, \xi_M) \in [\eta_j, \eta_j + \Delta\eta] \quad (5.11)$$

with the summation now running over the n samples from the simulation. Eq. (5.11) applies to any number of variables ξ_j and is therefore valid in general for any biasing term

$U_{umbrella}(r^N)$ and any reaction coordinate $\eta(r^N)$.

In order to obtain the relative free energy shifts, f_k , between k overlapping umbrella windows, an iterative approach is used in the WHAM formalism. This step is independent of the reaction coordinate used to calculate the PMF from $p_{unbiased}(\eta)$ and therefore the standard WHAM formalism can be applied:

$$e^{-\beta f_k} = \int d\xi_1 \dots d\xi_M \sum_{i=1}^N \frac{n_i e^{-\beta U_{umbrella,k}(\xi_1, \dots, \xi_M)}}{\sum_{j=1}^N n_j e^{-\beta [U_{umbrella,j}(\xi_1, \dots, \xi_M) - f_j]}} P_{biased,i}(\xi_1, \dots, \xi_M) \quad (5.12)$$

where the umbrella potential depends on the variables $\xi_1 \dots \xi_M$. The integral is solved numerically through discrete summation. For a small number of variables ξ_i ($M < 4$) Eq. (5.12) could be used directly, but for the more general case it is again desirable to avoid the explicit construction of the multi-dimensional histogram and instead accumulate the right side of Eq. (5.12) on the fly. This is accomplished by expressing Eq. (5.12) as:

$$e^{-\beta f_k} = \sum_{i=1}^N \sum_{l=1}^{n_i} \frac{e^{-\beta U_{umbrella,k}(\xi_{1,i,l}, \dots, \xi_{M,i,l})}}{\sum_{j=1}^N n_j e^{-\beta [U_{umbrella,j}(\xi_{1,i,l}, \dots, \xi_{M,i,l}) - f_j]}} \Delta \xi_1 \dots \Delta \xi_M \quad (5.13)$$

where for each sample l of simulation i with a total of n_i samples, the umbrella potential variables $\xi_{1,i,l} \dots \xi_{M,i,l}$ are determined and then used to calculate

$U_{umbrella,k}(\xi_{1,i,l}, \dots, \xi_{M,i,l})$ and $\sum_{j=1}^N n_j e^{-\beta [U_{umbrella,j}(\xi_{1,i,l}, \dots, \xi_{M,i,l}) - f_j]}$ (for a

more detailed derivation of Eq. (5.13) see references (170, 174)). Once the f_k values are determined, the PMF along η can be calculated according to Eq. (5.11).

Simulation of Hin-recombinase translocation

As a test system we chose DNA translocation in the Hin-recombinase complex due to its small size and high-resolution (2.3 Å) crystallographic data for the structure of the Hin 52-mer peptide bound to a 14 base pair DNA oligomer (16). All crystallographic waters and unpaired bases located at the DNA ends were removed from the crystal structure (PDBID: 1HCR). This initial structure was used as the reference for defining two paths for the two strands of the DNA backbone along which DNA is translocated. More specifically, the phosphorous atoms from residues 4-15 on strand 1 and phosphorous atoms from residues 18-29 on strand 2 are used as the spline points for the first and second paths, respectively. The restraint potential was then applied to the projection onto the respective path of the center of mass of the heavy atoms of residues 5-13 (strand 1) and residues 19-27 (strand 2) with a total of 18 independent reaction coordinates (for 9 bases on each strand). We define forward translocation as the movement of the biased DNA nucleotides from their initial reference position towards the A15/T17 base pair (or towards the protein C-terminus) and backward translocation as the movement of the biased DNA nucleotides towards the G3/C29 base pair (or towards the protein N-terminus) (see Table 5.1 and Figure 5.1C).

Umbrella sampling of the forward and backward translocation process was achieved by increasing/decreasing ξ_0 from +/-0.05 to +/-1.3 in 0.025 increments (resulting in a total of 52 windows in each direction) with final structures from each umbrella used as the input for the next window. Initially, 300 ps simulations with a force constant of $K=200$ kcal/mol were carried out at each value of ξ_0 . The final structures at each value of ξ_0 were then simulated with a reduced force constant of $K=100$ kcal/mol for 1 ns per window (after 80 ps of additional equilibration). A high force constant was chosen initially to obtain starting structures close to the target ξ_0 values

but production simulations were run with a lower force constant to improve sampling overlap between adjacent windows.

All simulations were conducted in implicit solvent using the GBMV method in CHARMM (175-176) and the default parameters as specified in the MMTSB Tool Set along with the following GBMV parameters: $\beta = -12$, $S_0 = 0.65$ (177). In addition, atomic radii developed for nucleic acids by Banavali and Roux were used instead of the standard van der Waals radii (178). All simulations were performed using the program CHARMM (version c36a4c) (64) along with the CHARMM27/CMAP force field (71-72) for proteins and nucleic acids and in combination with the MMTSB Tool Set (102). A force switching function, made effective at 15 Å and truncated at 17 Å, was used in the calculation of non-bonded interactions with the cutoff for list generation set at 20 Å. Langevin dynamics was applied to all heavy atoms with a 10 ps^{-1} friction coefficient (179). SHAKE (180) was used to constrain bond lengths involving hydrogen atoms along with a 1.5 fs integration time step. The starting structure was subjected to energy minimization followed by a gradual heating from 100-300 K over the course of 60 ps. The final structure was then utilized as the initial input for umbrella sampling as described above.

The energetics of DNA translocation is described by a single reaction coordinate, $\eta(\xi_1, \dots, \xi_{18})$ that is derived from the individual 18 projection reaction coordinates as follows:

$$\eta(\xi_1, \dots, \xi_{18}) = \frac{\sum_{i=1}^{18} \xi_i}{18} \quad (5.14)$$

Therefore, η can be viewed as the average translocation of the individual bases. While η depends on ξ_i in this case, any other reaction coordinate η could be chosen as well. The free energy profile for forward and backward translocation along η was calculated by using the generalized WHAM method described above for the data from the 1 ns production sampling for each of the 2x52 umbrellas.

5.3 Results

DNA Translocation

Umbrella sampling with the new path-based biasing potential was applied in simulations of DNA translocation in the Hin-recombinase system. During both the forward ($t_o > 0$) and backward translocation ($t_o < 0$) processes, both the protein and the DNA remained stable. Final snapshots from selected windows are shown in Figure 5.1. During the umbrella sampling simulations the DNA followed the pre-defined path dictated by its DNA backbone and moved in a screw-like fashion in either direction while largely retaining its internal structure. However, some degree of local distortion was observed, especially during forward translocation. Moreover, fraying of the A15/T17 base pair (in which no biasing potential was applied) was observed during forward translocation.

Free Energy Profile for DNA Translocation

Figure 5.2 shows the PMF for DNA translocation as a function of the average translocation reaction coordinate, $\eta(\xi_1, \dots, \xi_{18})$ (see Eq. (5.14)). A minimum is located near $\eta = 0$ which corresponds to the native binding state. Translocation away from this state by $\eta \approx \pm 0.5$ (approximately equivalent to translocation by one-half base pairs) causes a free energy increase by about 10 kcal/mol and 12.5 kcal/mol in the backward and forward directions, respectively. Movement of the DNA by a full base pair resulted in further increases in the relative free energy to about 30 kcal/mol for $\eta \approx -1.0$ and about 60 kcal/mol for $\eta \approx +1.0$. Overall, the PMF shows

that backward translocation is favored over forward translocation, but DNA movement in general is highly unfavorable in this system.

5.4 Discussion

This work describes the development of a new path-based restraint that can be used in umbrella sampling or Hamiltonian replica exchange simulations of more complex conformational transitions that cannot be easily captured with simple geometric reaction coordinates. In particular, the path-based potential is useful for the simulation of nucleic acid translocation. A first application involves a small protein-DNA system, Hin-recombinase bound to its target sequence, which has been previously studied in a different context using short MD simulations (181-182). As a result of the biasing potential, the DNA follows a screw-like movement in both directions by one full base pair rather than just direct translational motion. The umbrella sampling results show that the protein can stay bound to the DNA and is capable of forming new interactions during the translocation process.

Based on the free energy profile obtained from our simulations, backward movement appears to be favored vs. forward movement but in both directions DNA translocation appears to be highly unfavorable in this system. This suggests that Hin-recombinase does not scan DNA to find its target sequence. This finding is consistent with the experimental observation that Hin-recombinase binds specifically only to its target DNA base sequence (167).

Our approach for enhancing sampling along a given path provides greater flexibility in biased simulations of conformational dynamics than established biased sampling methods since it supports more complex motions than what can be generally described with simple geometric coordinates and provides better control of the entire transition path than RMSD-based restraints. The path-based biasing potential is also less restrictive than RMSD-based restraints because only the projection of selected atoms onto a given path is restrained and the system is free to move in

orthogonal degrees of freedom. While the application of the path-based restraint potential is illustrated here for the case of DNA translocation it is applicable in a wide variety of contexts. Other cases where such a potential could be applied may involve protein folding, DNA base flipping, or the passage of molecules through channels or nanopores.

5.5 Conclusion

DNA translocation is a complex event that plays a fundamental role in cell division, gene transcription, and mismatch recognition. Computational methods offer alternative efforts for studying this essential process. The work presented here overcomes some of the disadvantages of previous methods and provides a more intuitive approach for examining DNA translocation. As well, a free energy profile was calculated using a generalized WHAM and showed significantly high barriers associated with DNA translocation. Finally, unlike unrestrained systems, we cautiously remind the reader that adding additional restraints to a simulation can be helpful answering complex biological questions but extra care is needed when interpreting the validity of the results.

Table 5.1 DNA Sequence Used in All Hin Recombinase Simulations*

5'...	G3	T4	T5	T6	T7	T8	G9	A10	T11	A12	A13	G14	A15	...3'
3'...	C29	A28	A27	A26	A25	A24	C23	T22	A21	T20	T19	C18	T17	...5'

*The DNA sequence is identical to the crystal structure found in reference (167) with the exception that unpaired terminal bases are omitted.

Figure 5.1 Snapshots from forward and backward DNA translocation in Hin-recombinase. The reference protein structure is white and the protein structure resulting from DNA translocation is colored in dark grey in all snapshots. Each base pair is colored independently of the others in order to track the translocation process with respect to the reference structure. A) Forward translocation by one full base step. B) Forward translocation by half of a base step. C) The reference protein-DNA structure. D) Backward translocation by half of a base step. E) Backward translocation by a full base step.

Figure 5.1

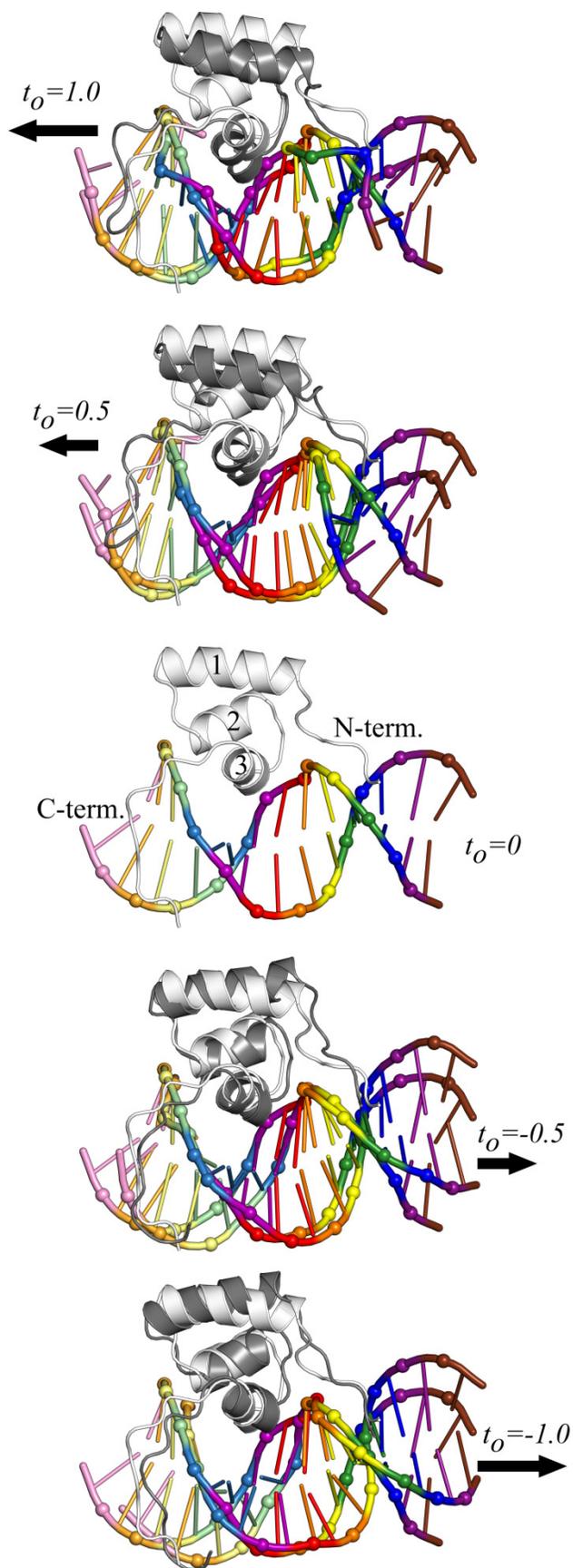
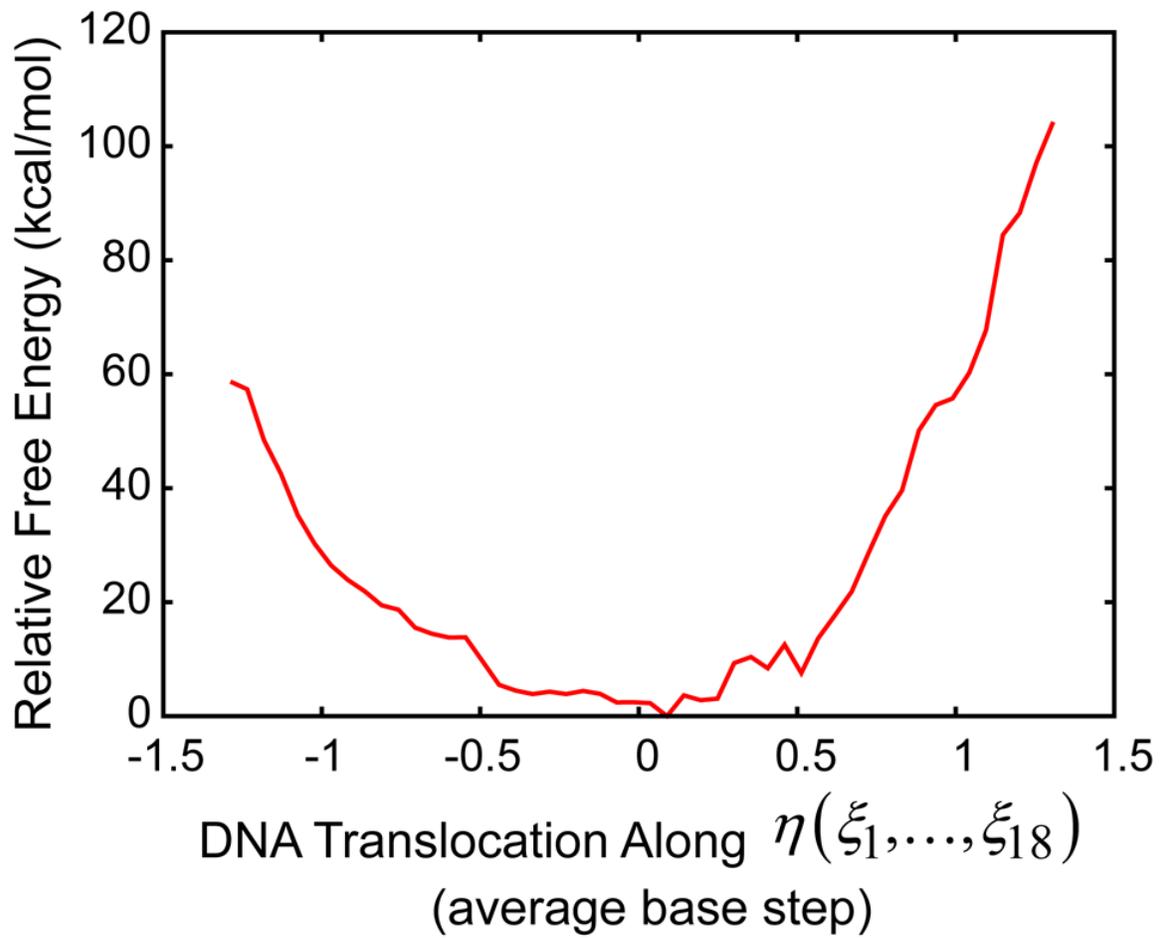


Figure 5.2 Free energy profile for DNA translocation.

Figure 5.2



Chapter 6

Conclusions and Perspectives

6.1 Conclusions and Perspectives

MutS and MSH2-MSH6 Conformational Dynamics

Following the crystallization of the first two prokaryotic MutS structures in 2000 (20-21), understanding the protein structure-function relationship became a major focus in the field of DNA mismatch repair. From a number of biochemical studies, it was suggested that MutS takes on various conformational states depending on the type of DNA (homoduplex vs. heteroduplex) that it is bound to and the presence or absence of different ATP/ADP nucleotides. However, almost all of the high resolution X-ray structures of MutS to date are bound to a mismatch and appear virtually identical even when bound by different nucleotides (20, 29) or different mismatches (45). Thus, the current picture of the MutS conformational dynamics has been largely influenced by the limited number of available crystal structures. Over the past decade, a number of new methods such as AFM (183), FRET (50), and deuterium exchange mass spectrometry (31, 184) have been used to investigate the DNA mismatch repair process but none of these methods offer sufficient resolution to accurately describe the various MutS conformational states. In contrast, computational approaches have proven in the past to be able to reveal alternative conformational states beyond the crystal structure and they also provide important molecular-level details that can serve as an excellent complement to experiments.

The work described in Chapters 2 and 3 were aimed at improving the relatively static picture of MutS conformational dynamics and to help create a detailed mechanism for DNA mismatch recognition that incorporates both experimental as well as structural data. In Chapter 2, four distinct conformational states (characterized as either a DNA binding mode, DNA scanning and mismatch mode, a repair initiation mode, or a sliding clamp formation mode) were identified

from NMA and found to be conserved across the different MutS and MSH2-MSH6 systems. These results suggest that the prokaryotic and eukaryotic proteins are not only structurally alike but share considerable similarities in their overall dynamics. Additionally, correlated movements between the DNA binding domains and the ATPase domains were found in several of the biologically relevant modes which support the idea of a long-range allosteric signaling pathway between the distant domains. From these observations, a detailed functional cycle for mismatch recognition was established based on the conserved low-frequency modes and available experimental data. Extending the work from Chapter 2, Chapter 3 examined the protein essential dynamics from nine sub- μ s long MD simulations of the *E. coli* MutS-DNA system using principal component analysis. The PCA results contributed two additional conformational states (a DNA sliding mode that follows sliding clamp formation and a DNA bending mode) to the functional mechanism. These modes also demonstrated similar motional coupling between the DNA binding cavity and the distant ATPases as well as between adjacent ATPases.

Altogether the proposed mechanism has improved our understanding of the MutS conformational dynamics and provides a framework for developing a more comprehensive picture of the mismatch recognition cycle. Future work should focus on validating each step of the functional cycle, possibly by introducing new mutations in the protein that would prevent the transition between the different conformational states. Also, the role of MutL within the mismatch recognition cycle is still rather unclear and warrants further attention.

Does MutS Employ a DNA Base-flipping Mechanism?

Since 2003, it has been repeatedly speculated that the MutS protein binds to the mismatch, bends the DNA, and uses a base-flipping mechanism to flip out one of the mismatched bases (16, 25, 30, 48, 51). This proposition is not necessarily surprising when considering the fact that MutS is a DNA repair protein that inserts a conserved residue (phenylalanine) into the DNA minor groove much like to several other DNA repair systems (DNA demethylases, DNA glycosylases, and T4 endonuclease V) which employ the same technique in order to promote base opening (123-124). Thus, it is reasonable to believe that MutS may also utilize a similar base-flipping mechanism during post-mismatch recognition. However, direct evidence for a base-flipping mechanism has been lacking to date. In Chapter 4, the DNA dynamics surrounding a G·T mismatch were analyzed. The mismatch base pair, which forms a bifurcated hydrogen bond in the crystal structure, was found to be more stable than expected but the 5' adjacent base next to the mismatched thymine was observed to be much more dynamic and, in one case, the 5' adjacent base flipped out of the helical stack via the major groove. To the best of our knowledge, this is the first reported case of spontaneous DNA base-flipping from an unbiased protein-DNA MD simulation. The direct visualization of the base-flipping process is significant not only for the MutS system but also in general for the study of protein-DNA interactions. In other DNA repair systems, the damaged base is often flipped and then removed. However, it is not clear why the 5' adjacent base is flipped out rather than the mismatch itself. This may have to do with the fact that the mismatch is stabilized through its interactions with the conserved phenylalanine. A flipped out base could serve as an amplification of the mismatch recognition signal and possibly play a role in DNA unbending (48).

To further validate these observations, it may be possible to use a disulphide cross-linking strategy (124), where the 5' adjacent base is replaced by a disulphide-modified cytosine

and an appropriate residue in the DNA binding domain is mutated to cysteine, to capture the base in its open state. However special care is required to ensure that the disulphide-modified cytosine is not recognized as a mismatch by MutS and that the engineered cysteine residue does not affect the general protein function. It may also be interesting to see from future MD simulations whether or not the 5' adjacent base still experiences enhanced dynamics when the mismatch is replaced by a normal Watson-Crick base pair.

Homoduplex vs. Heteroduplex DNA

The affinity of MutS for heteroduplex DNA is well understood but exactly how the protein differentiates between damaged and undamaged DNA is still largely unclear (16). A hypothetical model for DNA scanning describes the free energy landscape as a rugged terrain with deep minima that correspond to locations along the DNA that are highly flexible due to the presence of a mismatch (36). However, brute force simulations are not useful for studying DNA translocation because MutS has been shown to move along DNA at a rate (2 ms/base) that is well beyond the time scales currently available by straight MD. Instead, alternative enhanced sampling techniques such as umbrella sampling (74) are required. Chapter 5 details the development of a new path-based restraint, its successful application to drive the forward and backward translocation of DNA in the Hin recombinase test system, and the determination of the corresponding free energy profile along an intuitive translocation reaction coordinate from the biased simulations. Future application of these same techniques to study DNA translocation in the MutS-DNA system would greatly improve the current understanding of how the protein distinguishes between mismatch DNA and homoduplex DNA. For example, instead of only

looking at DNA translocation starting from a mismatch-bound state, it may be interesting to replace the mismatch with a regular base pair and then translocate this DNA through MutS. Since MutS is known to bind tightly to a mismatch then the free energy barrier for moving the mismatch away from this low energy state would be high. However, one would expect the free energy barrier to be relatively flat when translocating along homoduplex DNA. Furthermore, since MutS also bends the DNA upon mismatch binding, it may also be interesting to measure DNA bending as a function of DNA translocation.

Future Directions

The body of work presented in this dissertation has provided new insight into the MutS DNA mismatch repair protein. One of the main challenges of studying MutS (and its homologs) using computer simulation methods is the large size of the protein which has limited the conformational sampling to sub- μ s time scales. Recent experimental studies have demonstrated the importance of understanding the interactions between MutS and the downstream repair protein, MutL, (9, 31) as well as MutS tetramerization effects on longer length mismatch-containing DNA (185). However, examining these larger complicated systems using all-atom MD simulations would be prohibitive from a computational standpoint. One possible solution is to decrease the overall resolution of the system by using a coarse-graining method whereby each protein or nucleic acid residue is represented by one or more coarse-grained particles (depending on the level of accuracy required). This kind of approach has been explored in our group through the development of an accurate coarse-graining model for protein-nucleic acids systems,

PRIMO/PRIMONA (186), and would be most suitable for the future investigation of multi-protein MutS-DNA complexes.

References

References

1. Watson, J. D., and F. H. Crick. 1953. Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* 171:737-738.
2. Loeb, L. A., and T. A. Kunkel. 1982. Fidelity of DNA Synthesis. *Annu Rev Biochem* 51:429-457.
3. Friedberg, E. C., G. C. Walker, and W. Siede. 1995. DNA Repair and Mutagenesis. ASM Press, Washington, D.C.
4. Hsieh, P., and K. Yamane. 2008. DNA Mismatch Repair: Molecular Mechanism, Cancer, and Ageing. *Mech. Ageing Dev.* 129:391-407.
5. Modrich, P., and R. Lahue. 1996. Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology. *Annu Rev Biochem* 65:101-133.
6. Su, S. S., and P. Modrich. 1986. Escherichia Coli Muts-Encoded Protein Binds to Mismatched DNA Base Pairs. *Proc. Natl. Acad. Sci. U. S. A.* 83:5057-5061.
7. Grilley, M., K. M. Welsh, S. S. Su, and P. Modrich. 1989. Isolation and Characterization of the Escherichia Coli MutL Gene Product. *J. Biol. Chem.* 264:1000-1004.
8. Acharya, S., P. L. Foster, P. Brooks, and R. Fishel. 2003. The Coordinated Functions of the E-Coli MutS and MutL Proteins in Mismatch Repair. *Mol. Cell* 12:233-246.
9. Winkler, I., A. D. Marx, D. Lariviere, R. J. Heinze, M. Cristovao, A. Reumer, U. Curth, T. K. Sixma, and P. Friedhoff. 2011. Chemical Trapping of the Dynamic MutS-MutL Complex Formed in DNA Mismatch Repair in Escherichia Coli. *J Biol Chem* 286:17326-17337.
10. Au, K. G., K. Welsh, and P. Modrich. 1992. Initiation of Methyl-Directed Mismatch Repair. *J. Biol. Chem.* 267:12142-12148.
11. Cooper, D. L., R. S. Lahue, and P. Modrich. 1993. Methyl-Directed Mismatch Repair Is Bidirectional. *J. Biol. Chem.* 268:11823-11829.

12. Grilley, M., J. Griffith, and P. Modrich. 1993. Bidirectional Excision in Methyl-Directed Mismatch Repair. *J. Biol. Chem.* 268:11830-11837.
13. Viswanathan, M., and S. T. Lovett. 1998. Single-Strand DNA-Specific Exonucleases in Escherichia Coli. Roles in Repair and Mutation Avoidance. *Genetics* 149:7-16.
14. Viswanathan, M., and S. T. Lovett. 1999. Exonuclease X of Escherichia Coli. A Novel 3'-5' Dnase and Dnaq Superfamily Member Involved in DNA Repair. *J Biol Chem* 274:30094-30100.
15. Burdett, V., C. Baitinger, M. Viswanathan, S. T. Lovett, and P. Modrich. 2001. In Vivo Requirement for Recj, Exovii, Exoi, and Exox in Methyl-Directed Mismatch Repair. *Proc. Natl. Acad. Sci. U. S. A.* 98:6765-6770.
16. Kunkel, T. A., and D. A. Erie. 2005. DNA Mismatch Repair. *Annu Rev Biochem* 74:681-710.
17. Schofield, M. J., and P. Hsieh. 2003. DNA Mismatch Repair: Molecular Mechanisms and Biological Function. *Annu Rev Microbiol* 57:579-608.
18. Joseph, N., V. Duppatla, and D. N. Rao. 2006. Prokaryotic DNA Mismatch Repair. *Prog Nucleic Acid Res Mol Biol* 81:1-49.
19. de Las Alas, M. M., R. A. M. de Bruin, L. Ten Eyck, G. Los, and S. B. Howell. 1998. Prediction-Based Threading of the hMSH2 DNA Mismatch Repair Protein. *FASEB J* 12:653-663.
20. Lamers, M. H., A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. de Wind, and T. K. Sixma. 2000. The Crystal Structure of DNA Mismatch Repair Protein MutS Binding to a G•T Mismatch. *Nature* 407:711-717.
21. Obmolova, G., C. Ban, P. Hsieh, and W. Yang. 2000. Crystal Structures of Mismatch Repair Protein MutS and Its Complex with a Substrate DNA. *Nature* 407:703-710.
22. Yamamoto, A., M. J. Schofield, I. Biswas, and P. Hsieh. 2000. Requirement for Phe36 for DNA Binding and Mismatch Repair by Escherichia Coli MutS Protein. *Nucleic Acids Res.* 28:3564-3569.

23. Schofield, M. J., F. E. Brownell, S. Nayak, C. W. Du, E. T. Kool, and P. Hsieh. 2001. The Phe-X-Glu DNA Binding Motif of MutS - the Role of Hydrogen Bonding in Mismatch Recognition. *J. Biol. Chem.* 276:45505-45508.
24. Lebbink, J. H. G., D. Georgijevic, G. Natrajan, A. Fish, H. H. K. Winterwerp, T. K. Sixma, and N. de Wind. 2006. Dual Role of MutS Glutamate 38 in DNA Mismatch Discrimination and in the Authorization of Repair. *EMBO J.* 25:409-419.
25. Holmes, S. F., K. D. Scarpinato, S. D. McCulloch, R. M. Schaaper, and T. A. Kunkel. 2007. Specialized Mismatch Repair Function of Glu339 in the Phe-X-Glu Motif of Yeast MSH6. *DNA Repair* 6:293-303.
26. Malkov, V. A., I. Biswas, R. D. Camerini-Otero, and P. Hsieh. 1997. Photocross-Linking of the NH2-Terminal Region of Taq MutS Protein to the Major Groove of a Heteroduplex DNA. *J. Biol. Chem.* 272:23811-23817.
27. Tama, F., M. Valle, J. Frank, and C. L. Brooks. 2003. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 100:9319-9323.
28. Van Wynsberghe, A., G. H. Li, and Q. Cui. 2004. Normal-Mode Analysis Suggests Protein Flexibility Modulation Throughout RNA Polymerase's Functional Cycle. *Biochemistry* 43:13083-13096.
29. Lamers, M. H., D. Georgijevic, J. H. Lebbink, H. H. Winterwerp, B. Agianian, N. de Wind, and T. K. Sixma. 2004. ATP Increases the Affinity between MutS ATPase Domains. Implications for ATP Hydrolysis and Conformational Changes. *J Biol Chem* 279:43879-43885.
30. Warren, J. J., T. J. Pohlhaus, A. Changela, R. R. Iyer, P. L. Modrich, and L. S. Beese. 2007. Structure of the Human MutS Alpha DNA Lesion Recognition Complex. *Mol. Cell* 26:579-592.
31. Mendillo, M. L., V. V. Hargreaves, J. W. Jamison, A. O. Mo, S. Li, C. D. Putnam, V. L. Woods, and R. D. Kolodner. 2009. A Conserved MutS Homolog Connector Domain Interface Interacts with MutL Homologs. *Proc. Natl. Acad. Sci. U. S. A.* 106:22223-22228.
32. Mukherjee, S., and M. Feig. 2009. Conformational Change in MSH2-MSH6 Upon Binding DNA Coupled to ATPase Activity. *Biophys. J.* 96:L63-65.

33. Bjornson, K. P., and P. Modrich. 2003. Differential and Simultaneous Adenosine Di- and Triphosphate Binding by MutS. *J. Biol. Chem.* 278:18557-18562.
34. Antony, E., and M. M. Hingorani. 2004. Asymmetric ATP Binding and Hydrolysis Activity of the *Thermus Aquaticus* MutS Dimer Is Key to Modulation of Its Interactions with Mismatched DNA. *Biochemistry* 43:13115-13128.
35. Biswas, I., G. Obmolova, M. Takahashi, A. Herr, M. A. Newman, W. Yang, and P. Hsieh. 2001. Disruption of the Helix-U-Turn-Helix Motif of MutS Protein: Loss of Subunit Dimerization, Mismatch Binding and ATP Hydrolysis. *J. Mol. Biol.* 305:805-816.
36. Gorman, J., A. Chowdhury, J. A. Surtees, J. Shimada, D. R. Reichman, E. Alani, and E. C. Greene. 2007. Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex MSH2-MSH6. *Mol. Cell* 28:359-370.
37. Gradia, S., S. Acharya, and R. Fishel. 2000. The Role of Mismatched Nucleotides in Activating the hMSH2-hMSH6 Molecular Switch. *J. Biol. Chem.* 275:3922-3930.
38. Gradia, S., S. Acharya, and R. Fishel. 1997. The Human Mismatch Recognition Complex hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell* 91:995-1005.
39. Allen, D. J., A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich, and J. D. Griffith. 1997. MutS Mediates Heteroduplex Loop Formation by a Translocation Mechanism. *EMBO J.* 16:4467-4476.
40. Gradia, S., D. Subramanian, T. Wilson, S. Acharya, A. Makhov, J. Griffith, and R. Fishel. 1999. hMSH2-hMSH6 Forms a Hydrolysis-Independent Sliding Clamp on Mismatched DNA. *Mol. Cell* 3:255-261.
41. Blackwell, L. J., K. P. Bjornson, D. J. Allen, and P. Modrich. 2001. Distinct MutS DNA-Binding Modes That Are Differentially Modulated by ATP Binding and Hydrolysis. *J. Biol. Chem.* 276:34339-34347.
42. Blackwell, L. J., D. Martik, K. P. Bjornson, E. S. Bjornson, and P. Modrich. 1998. Nucleotide-Promoted Release of HmutS Alpha from Heteroduplex DNA Is Consistent with an ATP-Dependent Translocation Mechanism. *J. Biol. Chem.* 273:32055-32062.

43. Sixma, T. K. 2001. DNA Mismatch Repair: MutS Structures Bound to Mismatches. *Curr. Opin. Struct. Biol.* 11:47-52.
44. Lamers, M. H., H. H. Winterwerp, and T. K. Sixma. 2003. The Alternating ATPase Domains of MutS Control DNA Mismatch Repair. *EMBO J* 22:746-756.
45. Natrajan, G., M. H. Lamers, J. H. Enzlin, H. H. Winterwerp, A. Perrakis, and T. K. Sixma. 2003. Structures of Escherichia Coli DNA Mismatch Repair Enzyme MutS in Complex with Different Mismatches: A Common Recognition Mode for Diverse Substrates. *Nucleic Acids Res.* 31:4814-4821.
46. Hunenberger, P. H., A. E. Mark, and W. F. van Gunsteren. 1995. Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J Mol Biol* 252:492-503.
47. Kato, R., M. Kataoka, H. Kamikubo, and S. Kuramitsu. 2001. Direct Observation of Three Conformations of MutS Protein Regulated by Adenine Nucleotides. *J. Mol. Biol.* 309:227-238.
48. Wang, H., Y. Yang, M. J. Schofield, C. W. Du, Y. Fridman, S. D. Lee, E. D. Larson, J. T. Drummond, E. Alani, P. Hsieh, and D. A. Erie. 2003. DNA Bending and Unbending by MutS Govern Mismatch Recognition and Specificity. *Proc. Natl. Acad. Sci. U. S. A.* 100:14822-14827.
49. Engel, A., and D. J. Muller. 2000. Observing Single Biomolecules at Work with the Atomic Force Microscope. *Nat. Struct. Biol.* 7:715-718.
50. Sass, L. E., C. Lanyi, K. Weninger, and D. A. Erie. 2010. Single-Molecule FRET TACKLE Reveals Highly Dynamic Mismatched DNA-MutS Complexes. *Biochemistry* 49:3174-3190.
51. Tessmer, I., Y. Yang, J. Zhai, C. W. Du, P. Hsieh, M. M. Hingorani, and D. A. Erie. 2008. Mechanism of MutS Searching for DNA Mismatches and Signaling Repair. *J. Biol. Chem.* 283:36646-36654.
52. Nag, N., B. J. Rao, and G. Krishnamoorthy. 2007. Altered Dynamics of DNA Bases Adjacent to a Mismatch: A Cue for Mismatch Recognition by MutS. *J Mol Biol* 374:39-53.

53. Feig, M., and Z. F. Burton. 2010. RNA Polymerase Ii with Open and Closed Trigger Loops: Active Site Dynamics and Nucleic Acid Translocation. *Biophys. J.* 99:2577-2586.
54. Villa, E., A. Balaeff, and K. Schulten. 2005. Structural Dynamics of the Lac Repressor-DNA Complex Revealed by a Multiscale Simulation. *Proc. Natl. Acad. Sci. U. S. A.* 102:6783-6788.
55. Woo, H. J., Y. Liu, and R. Sousa. 2008. Molecular Dynamics Studies of the Energetics of Translocation in Model T7 RNA Polymerase Elongation Complexes. *Proteins* 73:1021-1036.
56. Eargle, J., A. A. Black, A. Sethi, L. G. Trabuco, and Z. Luthey-Schulten. 2008. Dynamics of Recognition between Trna and Elongation Factor Tu. *J. Mol. Biol.* 377:1382-1405.
57. Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 1958. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* 181:662-666.
58. Brooks, C. L., M. Karplus, and B. M. Pettitt. 1988. *Proteins : A Theoretical Perspective of Dynamics, Structure, and Thermodynamics.* J. Wiley, New York.
59. Karplus, M., and J. A. McCammon. 2002. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* 9:646-652.
60. McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of Folded Proteins. *Nature* 267:585-590.
61. Karplus, M., and G. A. Petsko. 1990. Molecular Dynamics Simulations in Biology. *Nature* 347:631-639.
62. Mackerell, A. D. 2004. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* 25:1584-1604.
63. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* 4:187-217.

64. Brooks, B. R., C. L. Brooks, III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. 2009. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 30:1545-1614.
65. Becker, O. M., A. D. M. Jr, B. Roux, and M. Watanabe. 2001. Computational Biochemistry and Biophysics. CRC Press, New York, NY.
66. Weiner, P. K., and P. A. Kollman. 1981. Amber - Assisted Model-Building with Energy Refinement - a General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* 2:287-303.
67. Scott, W. R. P., P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren. 1999. The Gromos Biomolecular Simulation Program Package. *J Phys Chem A* 103:3596-3607.
68. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 26:1781-1802.
69. Shaw, D. E. 2009. Anton: A Specialized Machine for Millisecond-Scale Molecular Dynamics Simulations of Proteins. *Abstracts of Papers of the American Chemical Society* 238:-.
70. Shaw, D. E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. 2010. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* 330:341-346.
71. MacKerell, A. D., M. Feig, and C. L. Brooks, III. 2004. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J Am Chem Soc* 126:698-699.
72. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 102:3586-3616.

73. Foloppe, N., and A. D. MacKerell. 2000. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* 21:86-104.
74. Torrie, G. M., and J. P. Valleau. 1977. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* 23:187-199.
75. Fukunishi, H., O. Watanabe, and S. Takada. 2002. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* 116:9058-9067.
76. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The Weighted Histogram Analysis Method for Free-Energy Calculations of Biomolecules. I. The Method. *J. Comput. Chem.* 13:1011-1021.
77. Hayward, S., and B. L. de Groot. 2008. Normal Modes and Essential Dynamics. *Methods in Molecular Biology (Clifton, N.J.)* 443:89-106.
78. Cui, Q., and I. Bahar. 2006. Normal Mode Analysis : Theory and Applications to Biological and Chemical Systems. Chapman & Hall/CRC, Boca Raton.
79. Ma, J. 2005. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* 13:373-380.
80. Modrich, P. 2006. Mechanisms in Eukaryotic Mismatch Repair. *J. Biol. Chem.* 281:30305-30309.
81. Peltomaki, P. 2003. Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. *J Clin Oncol* 21:1174-1179.
82. Gu, L. Y., Y. Hong, S. McCulloch, H. Watanabe, and G. M. Li. 1998. ATP-Dependent Interaction of Human Mismatch Repair Proteins and Dual Role of PcnA in Mismatch Repair. *Nucleic Acids Res.* 26:1173-1178.
83. Jacobs-Palmer, E., and M. M. Hingorani. 2007. The Effects of Nucleotides on MutS-DNA Binding Kinetics Clarify the Role of MutS ATPase Activity in Mismatch Repair. *J. Mol. Biol.* 366:1087-1098.

84. Mazur, D. J., M. L. Mendillo, and R. D. Kolonder. 2006. Inhibition of Msh6 ATPase Activity by Mismatched DNA Induces a MSH2(ATP)-MSH6(ATP) State Capable of Hydrolysis-Independent Movement Along DNA. *Mol. Cell* 22:39-49.
85. Gorbalenya, A. E., and E. V. Koonin. 1990. Superfamily of UvrA-Related Ntp-Binding Proteins - Implications for Rational Classification of Recombination Repair Systems. *J. Mol. Biol.* 213:583-591.
86. Fiser, A., R. K. G. Do, and A. Sali. 2000. Modeling of Loops in Protein Structures. *Protein Sci.* 9:1753-1773.
87. Mackerell, A. D., M. Feig, and C. L. Brooks. 2004. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* 25:1400-1415.
88. Li, G. H., and Q. Cui. 2002. A Coarse-Grained Normal Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca²⁺-ATPase. *Biophys. J.* 83:2457-2474.
89. Tama, F., F. X. Gadea, O. Marques, and Y. H. Sanejouand. 2000. Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins-Structure Function and Genetics* 41:1-7.
90. DeLano, W. L. 2002. The PyMOL Molecular Graphics System.
91. Van Wynsberghe, A. W., and Q. Cui. 2005. Comparison of Mode Analyses at Different Resolutions Applied to Nucleic Acid Systems. *Biophys. J.* 89:2939-2949.
92. Lu, M., B. Poon, and J. Ma. 2006. A New Method for Coarse-Grained Elastic Normal-Mode Analysis. *Journal of Chemical Theory and Computation* 2:464-471.
93. Hays, J. B., P. D. Hoffman, and H. X. Wang. 2005. Discrimination and Versatility in Mismatch Repair. *DNA Repair* 4:1463-1474.
94. Mitra, R., B. M. Pettitt, G. L. Rame, and R. D. Blake. 1993. The Relationship between Mutation-Rates for the (C-Center-Dot-G) → (T-Center-Dot-a) Transition and Features of T-Center-Dot-G Mismatch Structures in Different Neighbor Environments, Determined by Free-Energy Molecular Mechanics. *Nucleic Acids Res.* 21:6028-6037.

95. Hiratsuka, T. 1994. Nucleotide-Induced Closure of the ATP-Binding Pocket in Myosin Subfragment-1. *J. Biol. Chem.* 269.
96. Bilwes, A. M., C. M. Quezada, L. R. Croal, B. R. Crane, and M. I. Simon. 2001. Nucleotide Binding by the Histidine Kinase Chea. *Nat. Struct. Biol.* 8:353-360.
97. Janas, E., M. Hofacker, M. Chen, S. Gompf, C. van der Does, and R. Tampe. 2003. The ATP Hydrolysis Cycle of the Nucleotide-Binding Domain of the Mitochondrial ATP-Binding Cassette Transporter Mdl1p. *J. Biol. Chem.* 278:26862-26869.
98. Pluciennik, A., and P. Modrich. 2007. Protein Roadblocks and Helix Discontinuities Are Barriers to the Initiation of Mismatch Repair. *Proc. Natl. Acad. Sci. U. S. A.* 104:12709-12713.
99. Sali, A., and T. L. Blundell. 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* 234:779-815.
100. Amadei, A., A. B. M. Linszen, and H. J. C. Berendsen. 1993. Essential Dynamics of Proteins. *Proteins-Structure Function and Genetics* 17:412-425.
101. Mukherjee, S., S. M. Law, and M. Feig. 2009. Deciphering the Mismatch Recognition Cycle in MutS and MSH2-MSH6 Using Normal-Mode Analysis. *Biophys. J.* 96:1707-1720.
102. Feig, M., J. Karanicolas, and C. L. Brooks, III. 2004. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* 22:377-395.
103. Kabsch, W., and C. Sander. 1983. Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22:2577-2637.
104. Schrödinger, L. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.3.
105. Seeber, M., M. Cecchini, F. Rao, G. Settanni, and A. Caflisch. 2007. Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* 23:2625-2627.

106. Seeber, M., A. Felling, F. Raimondi, S. Muff, R. Friedman, F. Rao, A. Caflisch, and F. Fanelli. 2011. Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J. Comput. Chem.* 32:1183-1194.
107. Brigo, A., K. W. Lee, G. I. Mustata, and J. M. Briggs. 2005. Comparison of Multiple Molecular Dynamics Trajectories Calculated for the Drug-Resistant Hiv-1 Integrase T66i/M154i Catalytic Domain. *Biophys. J.* 88:3072-3082.
108. van Aalten, D. M. F., A. Amadei, A. B. M. Linssen, V. G. H. Eijssink, G. Vriend, and H. J. C. Berendsen. 1995. The Essential Dynamics of Thermolysin - Confirmation of the Hinge-Bending Motion and Comparison of Simulations in Vacuum and Water. *Proteins-Structure Function and Genetics* 22:45-54.
109. Law, S. M., and M. Feig. 2011. Base-Flipping Mechanism in Post-Mismatch Recognition in MutS. *Submitted for publication.*
110. Zheng, X., K. Diraviyam, and D. Sept. 2007. Nucleotide Effects on the Structure and Dynamics of Actin. *Biophys. J.* 93:1277-1283.
111. van Gunsteren, W. F., P. H. Hunenberger, A. E. Mark, P. E. Smith, and I. G. Tironi. 1995. Computer Simulation of Protein Motion. *Comput. Phys. Commun.* 91:305-319.
112. Sekijima, M., C. Motono, S. Yamasaki, K. Kaneko, and Y. Akiyama. 2003. Molecular Dynamics Simulation of Dimeric and Monomeric Forms of Human Prion Protein: Insight into Dynamics and Properties. *Biophys. J.* 85:1176-1185.
113. Mittal, J., and R. B. Best. 2010. Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin Hp35 and Pin Ww Domains in Explicit Water. *Biophys. J.* 99:L26-28.
114. Lahue, R. S., K. G. Au, and P. Modrich. 1989. DNA Mismatch Correction in a Defined System. *Science* 245:160-164.
115. Hargreaves, V. V., S. S. Shell, D. J. Mazur, M. T. Hess, and R. D. Kolodner. 2010. Interaction between the Msh2 and Msh6 Nucleotide-Binding Sites in the Saccharomyces Cerevisiae Msh2-Msh6 Complex. *J. Biol. Chem.* 285:9301-9310.
116. Kolodner, R. D., N. R. Hall, J. Lipford, M. F. Kane, M. R. Rao, P. Morrison, L. Wirth, P. J. Finan, J. Burn, P. Chapman, and et al. 1994. Human Mismatch Repair Genes and Their

Association with Hereditary Non-Polyposis Colon Cancer. *Cold Spring Harb Symp Quant Biol* 59:331-338.

117. Li, G. M. 2008. Mechanisms and Functions of DNA Mismatch Repair. *Cell Res.* 18:85-98.
118. Junop, M. S., G. Obmolova, K. Rausch, P. Hsieh, and W. Yang. 2001. Composite Active Site of an ABC ATPase: MutS Uses ATP to Verify Mismatch Recognition and Authorize DNA Repair. *Mol. Cell* 7:1-12.
119. Alani, E., J. Y. Lee, M. J. Schofield, A. W. Kijas, P. Hsieh, and W. Yang. 2003. Crystal Structure and Biochemical Analysis of the MutS•ADP•Beryllium Fluoride Complex Suggests a Conserved Mechanism for ATP Interactions in Mismatch Repair. *J. Biol. Chem.* 278:16088-16094.
120. Lebbink, J. H. G., A. Fish, A. Reumer, G. Natrajan, H. H. K. Winterwerp, and T. K. Sixma. 2010. Magnesium Coordination Controls the Molecular Switch Function of DNA Mismatch Repair Protein MutS. *J. Biol. Chem.* 285:13131-13141.
121. Bowers, J., T. Sokolsky, T. Quach, and E. Alani. 1999. A Mutation in the MSH6 Subunit of the *Saccharomyces Cerevisiae* MSH2-MSH6 Complex Disrupts Mismatch Recognition. *J. Biol. Chem.* 274:16115-16125.
122. Drotschmann, K., W. Yang, F. E. Brownell, E. T. Kool, and T. A. Kunkel. 2001. Asymmetric Recognition of DNA Local Distortion - Structure-Based Functional Studies of Eukaryotic MSH2-MSH6. *J. Biol. Chem.* 276:46225-46229.
123. Yang, W. 2008. Structure and Mechanism for DNA Lesion Recognition. *Cell Res.* 18:184-197.
124. Yang, C. G., C. Yi, E. M. Duguid, C. T. Sullivan, X. Jian, P. A. Rice, and C. He. 2008. Crystal Structures of DNA/RNA Repair Enzymes AlkB and ABH2 Bound to dsDNA. *Nature* 452:961-965.
125. Salsbury, F. R. 2010. Effects of Cisplatin Binding to DNA on the Dynamics of the *E. Coli* MutS Dimer. *Protein Peptide Lett* 17:744-750.

126. Salsbury, F. R., J. E. Clodfelter, M. B. Gentry, T. Hollis, and K. D. Scarpinato. 2006. The Molecular Mechanism of DNA Damage Recognition by MutS Homologs and Its Consequences for Cell Death Response. *Nucleic Acids Res.* 34:2173-2185.
127. Fiser, A., M. Feig, C. L. Brooks, III, and A. Sali. 2002. Evolution and Physics in Comparative Protein Structure Modeling. *Acc Chem Res* 35:413-421.
128. Bas, D. C., D. M. Rogers, and J. H. Jensen. 2008. Very Fast Prediction and Rationalization of Pka Values for Protein-Ligand Complexes. *Proteins* 73:765-783.
129. Jorgensen, W. L. 1981. Quantum and Statistical Mechanical Studies of Liquids .10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers - Application to Liquid Water. *J Am Chem Soc* 103:335-340.
130. Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald - an Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98:10089-10092.
131. Golosov, A. A., and M. Karplus. 2007. Analysis of the Translocation Step in DNA Replication by DNA Polymerase I with Computer Simulations. *Biophys. J.*:225a-225a.
132. Huang, N., N. K. Banavali, and A. D. MacKerell, Jr. 2003. Protein-Facilitated Base Flipping in DNA by Cytosine-5-Methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* 100:68-73.
133. Miyamoto, S., and P. A. Kollman. 1992. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* 13:952-962.
134. Feig, M., and B. M. Pettitt. 1999. Sodium and Chlorine Ions as Part of the DNA Solvation Shell. *Biophys. J.* 77:1769-1781.
135. Garcia, A. E., and G. Hummer. 2000. Water Penetration and Escape in Proteins. *Proteins* 38:261-272.
136. Song, K., A. J. Cambell, C. Bergonzo, C. del los Santos, A. P. Grollman, and C. Simmerling. 2009. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *J. Chem. Theory Comput.* 5:3105-3113.

137. Banavali, N. K., and A. D. MacKerell, Jr. 2002. Free Energy and Structural Pathways of Base Flipping in a DNA GCGC Containing Sequence. *J Mol Biol* 319:141-160.
138. Varnai, P., and R. Lavery. 2002. Base Flipping in DNA: Pathways and Energetics Studied with Molecular Dynamic Simulations. *J Am Chem Soc* 124:7272-7273.
139. Hagan, M. F., A. R. Dinner, D. Chandler, and A. K. Chakraborty. 2003. Atomistic Understanding of Kinetic Pathways for Single Base-Pair Binding and Unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.* 100:13922-13927.
140. Nikolova, E. N., E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei, and H. M. Al-Hashimi. 2011. Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* 470:498-U484.
141. Beglov, D., and B. Roux. 1997. An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* 101:7821-7826.
142. Banavali, N. K., and B. Roux. 2005. Free Energy Landscape of A-DNA to B-DNA Conversion in Aqueous Solution. *J Am Chem Soc* 127:6866-6876.
143. Varnai, P., M. Canalia, and J. L. Leroy. 2004. Opening Mechanism of G Center Dot T/U Pairs in DNA and RNA Duplexes: A Combined Study of Imino Proton Exchange and Molecular Dynamics Simulation. *J Am Chem Soc* 126:14659-14667.
144. Moe, J. G., and I. M. Russu. 1992. Kinetics and Energetics of Base-Pair Opening in 5'-D(Cgcgaattcgcg)-3' and a Substituted Dodecamer Containing G.T Mismatches. *Biochemistry* 31:8421-8428.
145. Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1981. Structure of a B-DNA Dodecamer - Conformation and Dynamics .1. *P Natl Acad Sci-Biol* 78:2179-2183.
146. Chen, Y. Z., V. Mohan, and R. H. Griffey. 1998. Effect of Backbone Zeta Torsion Angle on Low Energy Single Base Opening in B-DNA Crystal Structures. *Chem. Phys. Lett.* 287:570-574.
147. Feig, M., R. Zacharias, and B. M. Pettitt. 2001. Conformations of an Adenine Bulge in a DNA Octamer and Its Influence on DNA Structure from Molecular Dynamics Simulations. *Biophys. J.* 81:352-370.

148. Biswas, I., and P. Hsieh. 1997. Interaction of MutS Protein with the Major and Minor Grooves of a Heteroduplex DNA. *J Biol Chem* 272:13355-13364.
149. Bjornson, K. P., D. J. Allen, and P. Modrich. 2000. Modulation of MutS ATP Hydrolysis by DNA Cofactors. *Biochemistry* 39:3176-3183.
150. Antony, E., and M. M. Hingorani. 2003. Mismatch Recognition-Coupled Stabilization of MSH2-MSH6 in an ATP-Bound State at the Initiation of DNA Repair. *Biochemistry* 42:7682-7693.
151. Acharya, S., and K. Patterson. 2010. Mutations in the Conserved Glycine and Serine of the MutS ABC Signature Motif Affect Nucleotide Exchange, Kinetics of Sliding Clamp Release of Mismatch and Mismatch Repair. *Mutat Res-Fund Mol M* 684:56-65.
152. Lu, X. J., and W. K. Olson. 2003. 3dna: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* 31:5108-5121.
153. Dillingham, M. S., D. B. Wigley, and M. R. Webb. 2000. Demonstration of Unidirectional Single-Stranded DNA Translocation by PcrA Helicase: Measurement of Step Size and Translocation Speed. *Biochemistry* 39:205-212.
154. Dillingham, M. S., D. B. Wigley, and M. R. Webb. 2002. Direct Measurement of Single-Stranded DNA Translocation by PcrA Helicase Using the Fluorescent Base Analogue 2-Aminopurine. *Biochemistry* 41:643-651.
155. Soutanas, P., M. S. Dillingham, P. Wiley, M. R. Webb, and D. B. Wigley. 2000. Uncoupling DNA Translocation and Helicase Activity in PcrA: Direct Evidence for an Active Mechanism. *EMBO J.* 19:3799-3810.
156. Caruthers, J. M., and D. B. McKay. 2002. Helicase Structure and Mechanism. *Curr. Opin. Struct. Biol.* 12:123-133.
157. Enemark, E. J., and L. Joshua-Tor. 2006. Mechanism of DNA Translocation in a Replicative Hexameric Helicase. *Nature* 442:270-275.
158. Gyimesi, M., K. Sarlos, and M. Kovacs. 2010. Processive Translocation Mechanism of the Human Bloom's Syndrome Helicase Along Single-Stranded DNA. *Nucleic Acids Res.* 38:4404-4414.

159. Massey, T. H., C. P. Mercogliano, J. Yates, D. J. Sherratt, and J. Lowe. 2006. Double-Stranded DNA Translocation: Structure and Mechanism of Hexameric Ftsk. *Mol. Cell* 23:457-469.
160. Sivanathan, V., M. D. Allen, C. de Bekker, R. Baker, L. K. Arciszewska, S. M. Freund, M. Bycroft, J. Lowe, and D. J. Sherratt. 2006. The Ftsk Gamma Domain Directs Oriented DNA Translocation by Interacting with Kops. *Nat. Struct. Mol. Biol.* 13:965-972.
161. Lowe, J., A. Ellonen, M. D. Allen, C. Atkinson, D. J. Sherratt, and I. Grainge. 2008. Molecular Mechanism of Sequence-Directed DNA Loading and Translocation by Ftsk. *Mol. Cell* 31:498-509.
162. Takayama, Y., D. Sahu, and J. Iwahara. 2010. Nmr Studies of Translocation of the Zif268 Protein between Its Target DNA Sites. *Biochemistry* 49:7998-8005.
163. Yu, J., T. Ha, and K. Schulten. 2007. How Directional Translocation Is Regulated in a DNA Helicase Motor. *Biophys. J.* 93:3783-3797.
164. Yu, J., T. Ha, and K. Schulten. 2006. Structure-Based Model of the Stepping Motor of PcrA Helicase. *Biophys. J.* 91:2097-2114.
165. Ishida, H. 2010. Branch Migration of Holliday Junction in RuvA Tetramer Complex Studied by Umbrella Sampling Simulation Using a Path-Search Algorithm. *J. Comput. Chem.* 31:2317-2329.
166. Golosov, A. A., J. J. Warren, L. S. Beese, and M. Karplus. 2010. The Mechanism of the Translocation Step in DNA Replication by DNA Polymerase I: A Computer Simulation Analysis. *Structure* 18:83-93.
167. Feng, J. A., R. C. Johnson, and R. E. Dickerson. 1994. Hin Recombinase Bound to DNA - the Origin of Specificity in Major and Minor-Groove Interactions. *Science* 263:348-355.
168. Torrie, G. M., and J. P. Valleau. 1974. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* 28:578-581.
169. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* 13:1011-1021.

170. Shirts, M. R., and J. D. Chodera. 2008. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* 129:124105-124114.
171. Ferrenberg, A. M. 1989. Efficient Use of Monte Carlo Simulation Data. In Department of Physics. Carnegie-Mellon University, Pittsburgh, PA. 86.
172. Ferrenberg, A. M., and R. H. Swendsen. 1988. New Monte Carlo Technique for Studying Phase Transitions. *Phys Rev Lett* 61:2635-2638.
173. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo Data Analysis. *Phys Rev Lett* 63:1195-1198.
174. Souaille, M., and B. Roux. 2001. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* 135:40-57.
175. Lee, M. S., M. Feig, F. R. Salsbury, and C. L. Brooks. 2003. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations. *J. Comput. Chem.* 24:1348-1356.
176. Lee, M. S., F. R. Salsbury, and C. L. Brooks. 2002. Novel Generalized Born Methods. *J. Chem. Phys.* 116:10606-10614.
177. Chocholousova, J., and M. Feig. 2006. Balancing an Accurate Representation of the Molecular Surface in Generalized Born Formalisms with Integrator Stability in Molecular Dynamics Simulations. *J. Comput. Chem.* 27:719-729.
178. Banavali, N. K., and B. Roux. 2002. Atomic Radii for Continuum Electrostatics Calculations on Nucleic Acids. *J. Phys. Chem. B* 106:11026-11035.
179. Pastor, R. W., B. R. Brooks, and A. Szabo. 1988. An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms. *Mol Phys* 65:1409-1419.
180. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* 23:327-341.

181. Komeiji, Y., and M. Uebayasi. 1999. Change in Conformation by DNA-Peptide Association: Molecular Dynamics of the Hin-Recombinase-Hixl Complex. *Biophys. J.* 77:123-138.
182. Komeiji, Y., and M. Uebayasi. 1999. Molecular Dynamics Simulation of the Hin-Recombinase - DNA Complex. *Mol Simulat* 21:303-324.
183. Jia, Y., L. Bi, F. Li, Y. Chen, C. Zhang, and X. Zhang. 2008. Alpha-Shaped DNA Loops Induced by MutS. *Biochem Biophys Res Commun* 372:618-622.
184. Mendillo, M. L., C. D. Putnam, A. O. Mo, J. W. Jamison, S. Li, V. L. Woods, and R. D. Kolodner. 2010. Probing DNA- and ATP-Mediated Conformational Changes in the MutS Family of Mismatch Recognition Proteins Using Deuterium Exchange Mass Spectrometry. *J. Biol. Chem.* 285:13170-13182.
185. Jiang, Y., and P. E. Marszalek. 2011. Atomic Force Microscopy Captures MutS Tetramers Initiating DNA Mismatch Repair. *EMBO J* Advance Online Publication.
186. Gopal, S. M., S. Mukherjee, Y. M. Cheng, and M. Feig. 2010. Primo/Primona: A Coarse-Grained Model for Proteins and Nucleic Acids That Preserves near-Atomistic Accuracy. *Proteins* 78:1266-1281.