Catalytically Requisite Conformational Dynamics in the mRNA-Capping Enzyme Probed by Targeted Molecular Dynamics[†]

Robert V. Swift*[,‡,§] and J. Andrew McCammon[‡,§,||,⊥]

Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, Department of Pharmacology, and Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California, 92039-0365

Received January 4, 2008; Revised Manuscript Received February 8, 2008

ABSTRACT: The addition of a N7-methyl guanosine cap to the 5′ end of nascent mRNA is carried out by the mRNA-capping enzyme, a two-domain protein that is a member of the nucleotidyltransferase superfamily. The mRNA-capping enzyme is composed of a catalytic nucleotidyltransferase domain and a noncatalytic oligonucleotide/oligosaccharide binding (OB) domain. Large-scale domain motion triggered by substrate binding mediates catalytically requisite conformational rearrangement of the GTP substrate prior to the chemical step. In this study, we employ targeted molecular dynamics (TMD) on the PBCV-1 capping enzyme to probe the global domain dynamics and internal dynamics of conserved residues during the conformational transformation from the open to the closed state. Analysis of the resulting trajectories along with structural and sequence homology to other members of the superfamily allows us to suggest a conserved mechanism of conformational rearrangements spanning all mRNA-capping enzymes and all ATP-dependent DNA ligases. Our results suggest that the OB domain moves quasi-statically toward the nucleotidyltransferase domain, pivoting about a short linker region. The approach of the OB domain brings a conserved RxDK sequence, an element of conserved motif VI, within proximity of the triphosphate of GTP, destabilizing the unreactive conformation and thereby allowing thermal fluctuations to partition the substrate toward the catalytically competent state.

Conformational flexibility plays an essential role in regulating the functional activity of many proteins and enzymes. The magnitude of protein motion ranges from small local atomic vibrations that occur on a femto- to picosecond time scale, to side-chain rotation that occurs on a nanosecond or longer time scale, to larger global domain motion that typically occurs on the time scale of micro- to milliseconds. It is often these large-scale global dynamics that play an intrinsically important role in regulating protein and enzyme activity. For instance, global domain motions may be induced by an allosteric effector, influencing downstream functional events. Kern and Zuiderweg (*1*) offer an excellent review of these types of motions. Alternatively, in enzymatic catalysis, substrate binding may trigger domain closure over the active site, resulting in the formation of favorable interactions that facilitate catalysis. This mechanism of substrate-induced protein motion is known as induced fit and was first described by Koshland in 1958 (*2*).

Structural and experimental evidence suggests that mRNA-capping enzymes undergo catalytically requisite global domain motion consistent with the induced fit mechanism prior to the chemical step. These enzymes carry out the second in a set of three enzymatic reactions that result in the formation of a *N*-7-methyl-GMP cap on the 5′ end of nascent mRNA in eukaryotes and their viruses (*3*). In the first of these three steps, the 5′-phosphate is cleaved from mRNA by a triphosphatase. The resulting 5′-diphosphate end is guanylated by an mRNA-capping enzyme in the second step. Finally, N-7 of guanosine is methylated by a methyltransferase. The formation of the 5′ cap is the first post-transcriptional processing step and plays an essential role in mRNA splicing, nuclear export, mRNA stability, and translation initiation (*4*–*6*). Additionally, mutational analysis of the encoding gene has demonstrated that the *Saccharomyces cervisiae* capping enzyme is essential in maintaining cell viability (*7*).

mRNA-capping enzymes belong to a larger superfamily of enzymes called the nucleotidyltransferase superfamily, whose members include ATP- and NAD^+-dependent DNA ligases, ATP-dependent RNA ligases, and mRNA-capping enzymes. Members are distinguishable by their common structural elements and chemistry. Common structural features include a catalytic N-terminal domain, called the nucleotidyltransferase domain, and a noncatalytic C-terminal domain (*8*). The nucleotidyltransferase domain is common to all superfamily members and contains five conserved motifs, I, III, IIIa, IV, and V, that form the substrate-binding pocket, while the structural elements of the C-terminal domain differ among members. In mRNA-capping enzymes and ATP-dependent DNA ligases, the C-terminal domain

[†] This work was supported in part by Grants from the National Science Foundation (NSF), MCB-0506593 and MCA-93S013, to J.A.M. and the National Institutes of Health (NIH), GM31749, to J.A.M.

* To whom correspondence should be addressed. Telephone: 858-822-2771. Fax: 858-534-4974. E-mail: rswift@mccammon.ucsd.edu.

[‡] Department of Chemistry and Biochemistry.

[§] Center for Theoretical Biological Physics.

[||] Department of Pharmacology.

[⊥] Howard Hughes Medical Institute.

10.1021/bi8000209 CCC: $40.75
Published on Web 03/11/2008

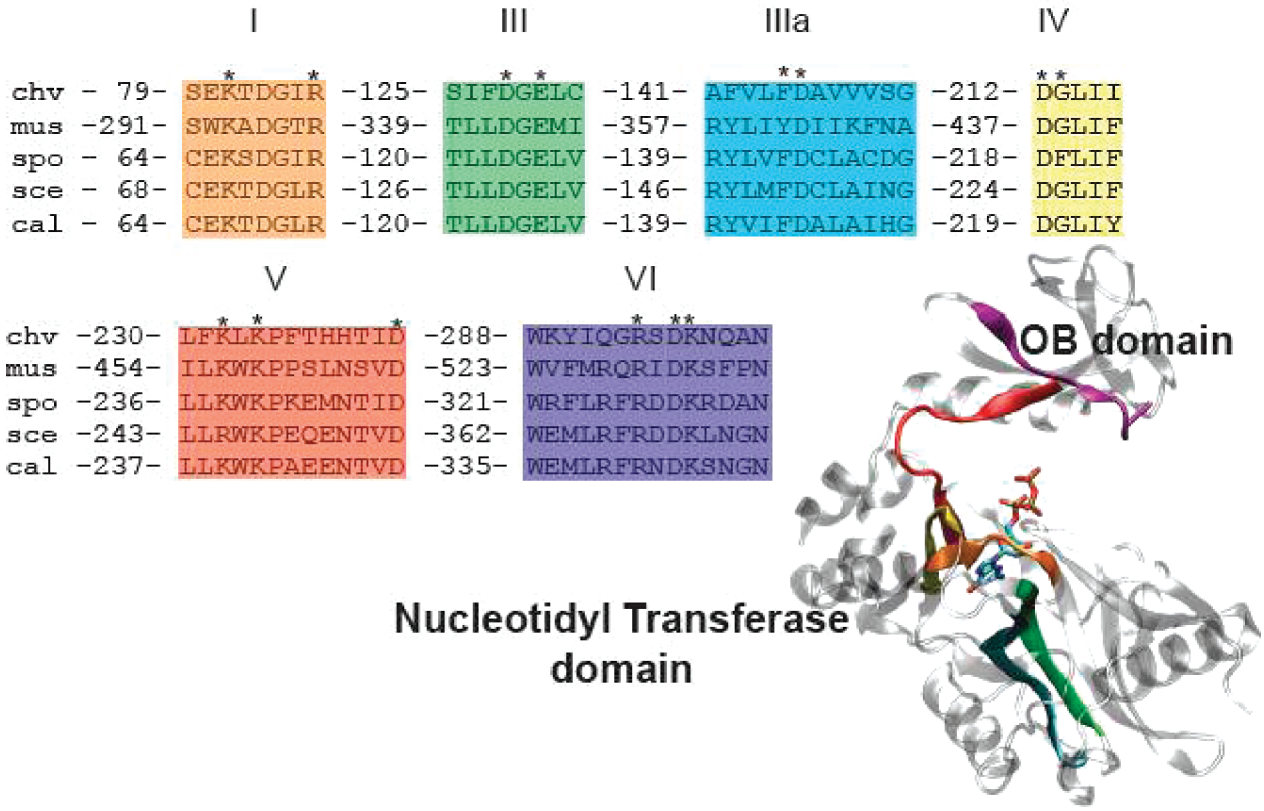

FIGURE 1: Conserved motifs in capping enzymes. Alignments include *Chlorella* virus PBCV1 (chv), mouse (mus), *Schizosaccharomyces pombe* (spo), *S. cerevisiae* (sce), and *Candida albicans* (cal). The amino acid number preceding the motif is given by the number preceding the motif. Essential residues are denoted with an asterisk. The structural locations of the motifs are mapped by color, and the N- and C-terminal domains are labeled nucleotidyltransferase domain and OB domain, respectively.

contains an oligonucleotide/oligosaccharide binding (OB)[1] fold that includes a sixth motif implicated in intermediate formation during catalysis (*9*) (Figure 1). All members share a common chemical mechanism, whereby a nucleotide is transferred to the 5′ end of a polynucleotide via a covalent nucleotide−enzyme intermediate (*8*). In the mRNA-capping enzyme, chemistry is initiated when a conserved lysine in motif I attacks α phosphate of GTP, concertedly displacing pyrophosphate and resulting in the formation of a GMP−Lys intermediate. After mRNA binding, GMP is transferred to the 5′ end of the bound mRNA, completing the reaction.

Large global rearrangements of the OB domain prior to the chemical step in catalysis have been implicated by structural evidence (*10*). The mRNA-capping enzyme from *Paramecium bursaria Chlorella* virus, PBCV-1, has been crystallized with GTP bound in the active site in two distinct conformations: "open" and "closed" (Figure 2). In the open state, the OB domain is distal from the nucleotidyltransferase domain; the centers of mass of the two domains are separated by 30.3 Å. The triphosphate of GTP is in an unreactive conformation, nearly orthogonal to the nucleophilic lysine. In the closed state, the OB domain is proximal to the active site and the triphosphate has undergone conformational isomerization; the centers of mass of the two domains are separated by 25.5 Å, and the triphosphate is in the catalytically requiste in-line conformation. Subsequent circular dichroism (CD) and fluorescence studies carried out on the yeast capping enzyme demonstrated that two distinct conformational states exist in the presence and absence of GTP (*11*). These results bolster the claim that global rearrangements are an intrinsic part of the catalytic repertoire of the capping enzyme. Additionally, mutational analysis of the mouse capping enzyme identified 16 amino acids essential for activity, 14 of which were localized in the six conserved motifs (*12*) (Figure 1). These results validated an earlier mutational study of the *S. cerevisiae* capping enzyme (*13*).

While structural and experimental studies have made great progress in illuminating the requisite conformational rearrangements that occur within the mRNA-capping enzymes during catalysis, our knowledge of these rearrangements is incomplete. A broader understanding of the process requires knowledge of the internal dynamics that occur during the conformational transition from the open to the closed state and how these dynamics mediate the requisite isomerization. Consequently, the goal of this study is to understand these dynamics. In particular, we hope to identify those amino acids essential in mediating the transition from the open, unreactive state to the closed, reactive state and determine if these residues are consistent with essential conserved residues found during mutational analysis of the mouse and *S. cerevisiae* capping enzymes (*12, 13*).

Internal protein dynamics may be probed by numerically integrating the equations of motion arising from a protein parametrized with a classical force field, a technique called molecular dynamics (MD) (*14*). In practice, the fastest atomic vibrations limit the size of the numerical time steps to 1−2

[1] Abbreviations: OB, (oligonucleotide/oligosaccharide) binding; CD, circular dichroism; MD, molecular dynamics; TMD, targeted molecular dynamics; rmsd, root-mean-square deviation; PDB, Protein Data Bank; NMN, nicotinic mononucleotide.

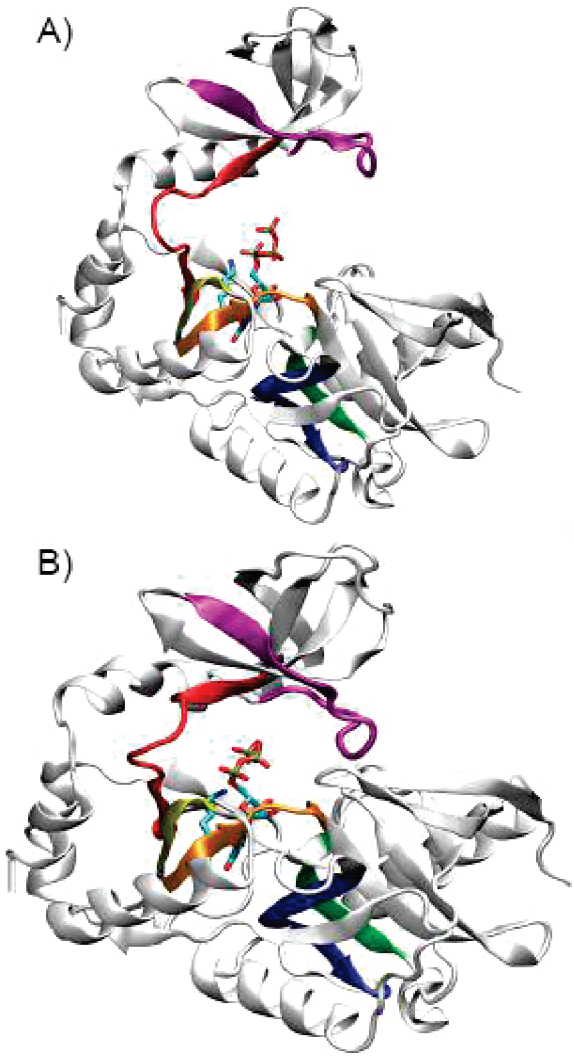

FIGURE 2: (A) Open crystal structure illustrating the unreactive triphosphate conformation. The motifs are colored the same as in Figure 1, and GTP and nucleophilic K82 are rendered in licorice. The centers of mass of the two domains are separated by 30.3 Å. (B) Closed crystal structure illustrating the reactive conformation. The centers of mass of the two domains are separated by 25.5 Å.

fs, which puts a practical restriction on simulation length to the tens to hundreds of nanoseconds, preventing observation of global domain motions that occur on the micro- to millisecond time scale. To access long time scale motions, an energetic bias applied along a reaction coordinate can be added to the classical force field, resulting in forces that drive the enzyme from one conformational state to another. In targeted molecular dynamics (TMD), the energetic bias is applied in a time-dependent fashion, giving rise to a force that quasi-statically decreases the root-mean-square deviation (rmsd) between a moving and target structure. While TMD allows us to observe global domain motion during a computationally accessible time scale, it suffers from two drawbacks.

First, there is no guarantee that the minimum free-energy pathway will be followed. Changing the initial conditions may result in different pathways. Additionally, thermal fluctuations lead to statistical uncertainty for the pathway determined from any given set of initial conditions (*15*). Other, more rigorous chain of state methods, such as the nudged elastic band (*16, 17*) and the harmonic Fourier bead (*18*) methods, optimize object functions connecting multiple images of the system between two end points to yield an optimal path. While these methods are more likely to generate the minimum potential and minimum free-energy path, respectively, they are more complex to perform and are computationally more expensive than TMD. A simpler alternative to the chain of state methods, steered molecular dynamics (SMD) (*19*), applies a time-dependent linear restraint to the center of mass of a group of atoms but, similar to TMD, is not guaranteed to move the system along the lowest free-energy pathway. Furthermore, SMD requires one to define a reasonable reaction coordinate describing the molecular mechanism of interest. This can be difficult in the high dimensional space occupied by proteins and enzymes.

Second, TMD does not immediately yield the free energy of the pathway. If one wishes to calculate the free energy of a pathway, a computationally more intensive technique, such as umbrella sampling, must be used (*20*). The work of Ravindranathan, Gallicchio, and Levy provides an excellent example of the umbrella-sampling technique to calculate the free energy of domain closure in the two-domain ribose-binding protein.

Despite the shortcomings of TMD, its simple implementation and computational efficiency make it an attractive method to explore the qualitative characteristics of conformational isomerization. Since it was pioneered by Schlitter and co-workers (*21*), TMD has been used to model conformational transitions in a variety of systems. A few recent examples include the activation of the CCK2R G-protein-coupled receptor (*22*), the coupling of C-loop closure and channel gating in the nicotinic acetylcholine receptor (*23*), and the activation mechanisms of gelsolin, an actin polymerization regulatory protein (*24*). In this work, TMD was used to study the internal dynamics that occur as the PBCV-1 capping enzyme moves from the open, unreactive state to the closed, reactive state.

MATERIALS AND METHODS

Molecular Model and Computational Procedures. Structural coordinates for the open, unreactive and closed, reactive states were taken from chain A and B, respectively, of the Protein Data Bank (PDB) code 1CKM. Both the open and closed structures consist of the crystallographically resolved residues 11–327. All crystallographic waters were retained in the simulation. Histidine protonation states were specified for the open state using WHATIF (*25*). The results were manually verified, with each protonation state agreeing well with the surrounding microenvironment. All other basic residues were protonated, while all acidic residues were deprotonated. Because TMD requires that the molecular composition be the same for both the moving and target structures, the protonation states of the closed target conformation were assigned to match those of the open state. To test the validity of approximating the protonation states of the closed conformation using the pK_a values calculated from the open crystal structure, WHATIF was used to recalculate pK_a values of all titratable residues in the closed conformation. While acidity changes were observed, none were significant enough to merit protonation-state reassignment. All other hydrogens and protein force-field parameters were assigned for both states in the xleap module of

AMBER8 according to the AMBER94 force field (*26*). The Carlson parameters were assigned to GTP (*27*), and both states were immersed in a 10 Å solvent octahedron of pre-equilibrated TIP3P waters (*28*). Three sodium ions were added to neutralize the system charge. Each electrically neutral, solvated system comprised of an identical number of 40 495 atoms. To remove spurious crystal-packing contacts, a set of 26 000 minimization steps were independently carried out on both the open and closed states in the Sander module of Amber 8 (*29*). The first 6000 steps were performed in three 2000 step cycles. Hydrogen was relaxed during the first 2000 steps, holding all other atoms fixed. Hydrogen, water, and ions were relaxed during the next 2000 steps. In the last cycle, the protein backbone was held fixed, minimizing all other atoms. Finally, a constraint-free minimization was carried out over the last 20 000 steps, of which the first 19 500 were the steepest descent, while the last 500 used the conjugate-gradient algorithm. After minimization, each system was restrained harmonically with a 5 kcal mol^{-1} Å^{-2} force constant and equilibrated at 300 K at a constant volume for 30 ps using the langevin thermostat to maintain temperature (*30*) and the SHAKE algorithm to constrain hydrogens bound to heavy atoms (*31*). The restraints were subsequently released, and both systems were then equilibrated at 1 atm for an additional 50 ps using a Berendsen barostat (*32*). Targeted molecular dynamics was then carried out in the NPT ensemble as implemented in the sander module of Amber 8 by applying the following time-dependent, harmonic energy bias to all α carbons and GTP atoms in the open state:

$$U_{\mathrm{TMD}} = \frac{1}{2}Nk(\mathrm{rmsd} - \rho(t))^2$$

where N gives the number of atoms included in the bias, k is the harmonic force constant, and rmsd is the root-mean-square deviation between a configuration at time point t and the target configuration. The reference rmsd value at time t is given by $\rho(t)$. When $\rho(t)$ is monotonically decreased, the moving structure is gradually driven toward the target structure. A truly quasi-static equilibrium transformation occurs in the limit as the force constant goes to 0 and the time goes to infinity; consequently, smaller force constants and longer simulation times represent more accurate models but come at a greater computational cost. To systematically choose the best balance between efficiency and accuracy, four independent simulations with k vaues of 5, 4, 3, and 1.5 kcal mol^{-1} Å^{-2} were carried out over 1 ns each. We inspected the trajectories in VMD (*33*) and observed that all four simulations approached the target structure. We desired the lowest harmonic force constant that produced the sought after transformation; therefore, the 1.5 kcal mol^{-1} Å^{-2} force constant was selected and the simulation time was extended to 2 ns. SHAKE was used during the TMD simulation to constrain all hydrogen heavy-atom bonds. Simulations were carried out on the DataStar machine at the San Diego Supercomputer Center, saving snapshots every 500 fs for analysis. Postsimulation analysis was performed in VMD and Matlab with a variety of customized scripts.

RESULTS

Quasi-rigid Domain Motion during the Open to Closed Transformation. TMD was used to probe the internal protein dynamics in the PBCV-1 capping enzyme during the transition from the open, unreactive conformation to the closed, reactive conformation. Figure 3A shows a profile of the rmsd as a function of the simulation progress. For a given time, t, we define progress as the fraction of total simulation length contained in the interval (0, t). It is clear that the profile is a decreasing function without any jumps or discontinuities, indicating a smooth transition from the initial to final states. To explore the global domain dynamics during the transition between states, the α-carbon rmsd of the OB domain was calculated following alignment to the first snapshot of the OB domain and again following a rigid-body protein transformation according to the alignment matrix that minimized the rmsd between trajectory snapshots of the nucleotidyltransferase domain and the first snapshot of the nucleotidyltransferase domain. The α-carbon rmsd of the nucleotidyltransferase domain was also calculated following alignment to the first snapshot of the nucleotidyltransferase domain. The results are shown in Figure 3B. Figure 3C shows the distance between the centers of mass of the OB and nucleotidyltransferase domains as a function of the simulation progress, illustrating that the domains approach one another. These results suggest that both the OB and nucleotidyltransferase domains move as quasi-rigid bodies

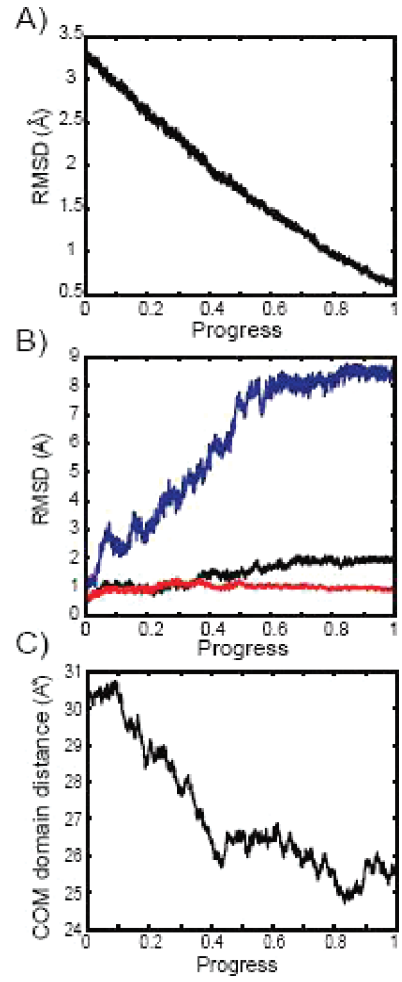

FIGURE 3: (A) Total α-carbon and GTP rmsd. (B) Blue, rmsd of the OB domain following structure alignment to reference the nucleotidyltransferase domain; black, rmsd of the OB domain following alignment of the OB domain to reference the OB domain; and red, rmsd of the nucleotidyltransferase domain following alignment to reference the nucleotidyltransferase domain. (C) Distance between centers of mass of the OB and nucleotidyltransferase domains.

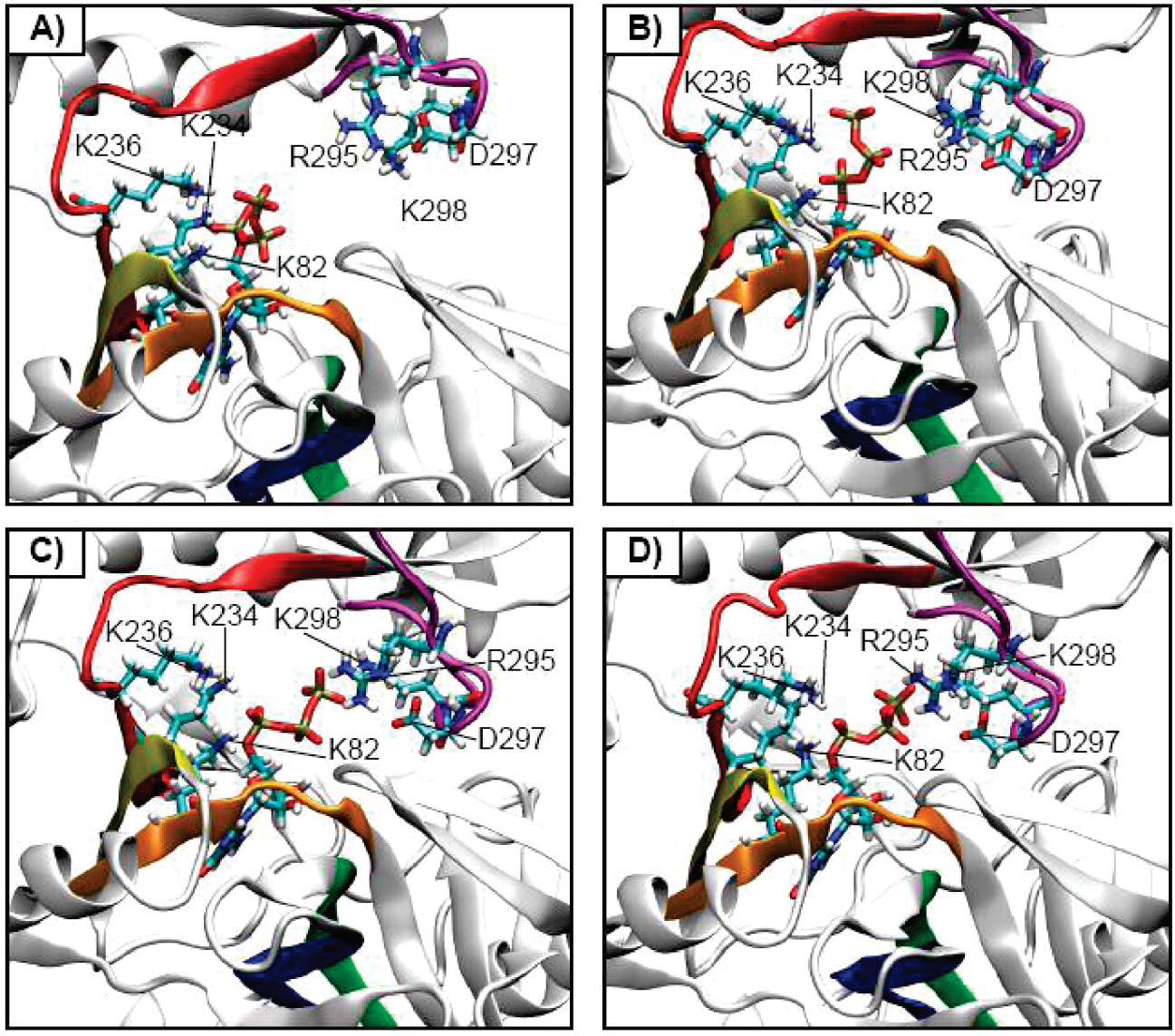

FIGURE 4: Discontinuously spaced snapshots observed during the TMD simulation at (A) 4% simulation progress, (B) 60% simulation progress, (C) 78% simulation progress, and (D) 90% simulation progress.

toward one another about a hinge in the short region linking the two domains. Similar rigid-body rotations have been observed in other two-domain proteins, such as adenylate kinase (*34*), ribose-binding protein (*20*), and transferrin (*35*), and are reviewed by Gerstein, Lesk, and Chothia (*36*). Thus, the observed quasi-rigid body translations are in good agreement with other structurally similar proteins.

Conserved Active-Site Network Mediates Triphosphate Rearrangement. Figure 4 shows four snapshots taken at discontinuously spaced intervals over the trajectory that illustrate key interactions that the enzyme may use to stabilize intermediate conformations en route to the reactive state. In Figure 4A, the simulation has progressed 4% toward completion and the enzyme is largely in the same conformational state observed in the open crystal structure. Stabilized by interactions between the γ-phosphate oxygens and K236, the β-phosphate oxygens and K82, and the α-phoshpate oxygens and K234, the triphosphate is folded back over the K82 nucleophile in an unreactive conformation. The OB domain is distal from the active site in an open conformation.

Figure 4B captures a representative conformation 60% into the simulation. The OB domain has moved into proximity of the active site bringing R295 and K298, members of conserved motif VI, into proximity of the triphosphate tail, which has rotated away from its initial state into a nearly vertical orientation. In this upright position, it interacts with K82, K234, and K236, members of conserved motifs I, IV, and V, respectively, in the nucleotidyltransferase domain, and R295 and K298 in the OB domain.

As the simulation progresses, the triphosphate moves through the vertical orientation observed in Figure 4B and into a new state, shown in Figure 4C. This new state, occurring 78% into the simulation, is characterized by interactions between the γ-phosphate and K298 in the OB domain and the α-phosphate and K234 in the nucleotidyltransferase domain. The final stages of the simulation entail further rearrangement of the triphosphate tail into the catalytically requisite in-line conformation. A representative conformation of this late phase is shown in Figure 4D. The simulation has progressed 90% to completion. The bulk of the triphosphate rearrangement has occurred. However, the final in-line conformation has not yet been attained.

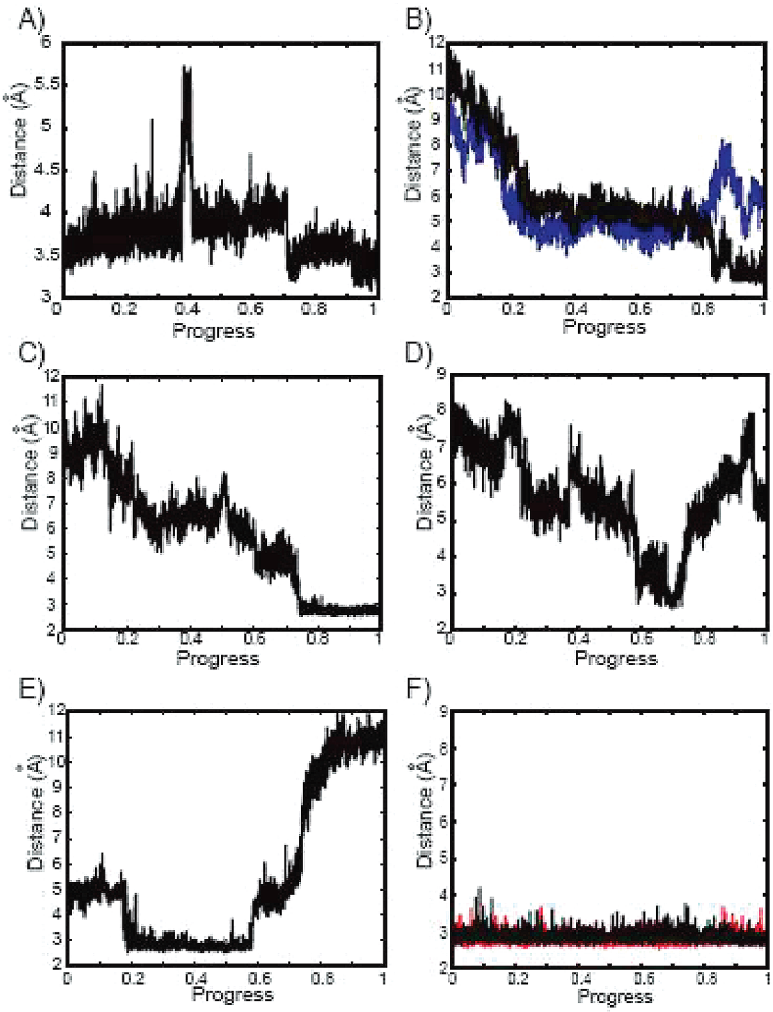

FIGURE 5: (A) Distance between the K82 nucleophile and α-phosphate electrophile. (B) Black, distance between the NH2 of R295 and the O2B of GTP; and blue, distance between the NH1 of R295 and the O1G of GTP. (C) Distance between the nitrogen of the side chain of K298 and the O3G of GTP. (D) Distance between the nitrogen of the side chain of K234 and the O2G of GTP. (E) Distance between the nitrogen of the side chain of K236 and the O2G of GTP. (F) Blue, distance between the nitrogen of the side chain of K298 and the OD2 of D297; black, distance between the NE2 of R295 and the OD2 of D297; and red, distance between the NH2 of R295 and the OD2 of D297.

Although not illustrated in Figure 4, the triphosphate isomerizes into the final in-line conformation over the last 10% of the simulation.

While Figure 4 provides a good pictorial representation of the different stages involved in the conformational transformation, examining GTP−side-chain bond distances of those residues involved as functions of the simulation progress permits a more detailed analysis. These bond distances are illustrated in Figure 5. Figure 5A charts the distance from the K82 nucleophile to the α-phosphate electrophile as a function of the simulation progress. The distance gradually increases over the first 70% of the simulation before discontinuously decreasing as a result of a rotation about the bond connecting the O5′ oxygen of ribosyl to the α-phosphate. The distance decreases over the last 10% of the simulation as the triphosphate group attains its final inline conformation. The anomalous large spike that occurs 40% into the distance time series results from a rotation about the α-phosphate O5′ bond. This rotation briefly swings the α-phosphate away from K82 before returning to its equilibrium position. Because the formation of the enzyme−guanylate intermediate is facilitated by the increasing proximity between the K82 nucleophile and the α-phosphate electrophile, the approach of these two atoms over the last 20% of the trajectory is consistent with movement into the catalytically competent state. To move toward the reactive, closed state, hydrogen bonds stabilizing the enzyme−substrate complex in the unreactive conformation must rupture and transiently be replaced by interactions that allow thermal fluctuations to carry the triphosphate tail of GTP into its final, reactive conformation.

Consideration of parts B−D of Figure 5 gives qualitative insight into how this conformational evolution might progress. Key events that will now be discussed occur 60% into the simulation and are pictorially represented in Figure 4B. Figure 5B shows the displacement between the guanidinium nitrogen of R295 and the nonbridging β- and γ-phosphate oxygens. The distances are initially correlated to the displacements separating the centers of mass of the two domains, rapidly decreasing over the first 30% of the simulation to meta-stable equilibrium distances of roughly 5.5 Å that are maintained until the simulation is 80% complete. Figure 5C shows a similar progression for the distance between the amine of K298 and O3G, a nonbridging γ-phosphate oxygen. Figure 5D shows that the distance between the amine of K234 and O2G decreases over the

first 60% of the simulation. This decrease is attributable to both the triphosphate tail rotating toward the side chain and the attractive electrostatic force that result from the approach of the negatively charged triphosphate oxygen toward the positively charged terminal amine of K234. At 60% progress, the rapid decrease in the distance between the terminal amine of K234 and O2G is anticorrelated with a 2 Å increase in the distance between K236 and O2G (Figure 5E), indicating that O2G exchanges hydrogen-bonding partners, replacing K236 with K234. After hydrogen-bond exchange, there is brief period of stability that extends as the simulation progresses from 60 to 70% completion. After this transient period of stability, both distances increase because R295 and K298 form energetically favorable interactions with triphosphate as it moves into an in-line conformation with respect to the nucleophile.

Figure 5F demonstrates the evolution of the interactions between the carboxyl of D297 and the guanidinium of R295 as well as the amine of K298. The red and black curves show that interactions between both nitrogens of the guanidinium moiety of R295 and the carboxyl group of D297 are conserved over the course of the simulation. As a result of these two strong, short interactions, the configurational space available to the R295 side chain is restrained, limited by the smaller volume of the configurational space accessible to the shorter D297 side chain. As a result, the interacting pair samples a smaller volume of configuration space than either would sample independently. The reduced mobility allows both the guanidinum nitrogens of R295 not involved in contacting the carboxyl group of D297 to readily form two contacts with nonbridging oxygens of the triphosphate tail of GTP (Figure 5B).

Nucleotidyl Transferase Domain Guanosine Stabilization. As illustrated in Figure 1, the GTP-binding pocket is lined by a set of five conserved motifs. Several residues within these motifs make key interactions with the guanosine group of GTP that maintains the relative positions and orientations of the ribofuranose ring and guanine base as the OB domain closes over the active site and the triphosphate rotates into a reactive conformation (Figure 6). Motif IIIa and IV each contribute a hydrophobic side chain that interacts with the guanine forming a stabilizing hydrophobic sandwich. Motiff IIIa contributes F146 that forms π-stacking interactions with the aromatic guanine. On the opposite side of the guanine plane, I216 is contributed by motif IV. Table 1 shows that these interactions undergo small, ~0.2 Å fluctuations around their mean, signifying that they constitute a firm binding pocket. In addition, the butyl amine of K188 protrudes into the deep end of the active site, donating a hydrogen bond to the O6 of GTP, adding additional stabilization. K188 is independent of the six motifs but, as sequence analysis demonstrates, is either conserved or substituted by a potential hydrogen-bonding residue in the surveyed capping enzymes (*9*). Although not as extensively stabilized as the guanine base, ribofuranose of GTP also makes stabilizing interactions with the surrounding active site. Extending from motif I, the NH1 of R87 donates a hydrogen bond to the O2′ of ribofuranose. This interaction is complemented by E131, a motif III residue that accepts a hydrogen bond from ribofuranoses 2′ hydroxyl hydrogen. As Table 1 demonstrates, both the R87 and E131 interactions are relatively stable around their means. Consequently, they form a steady

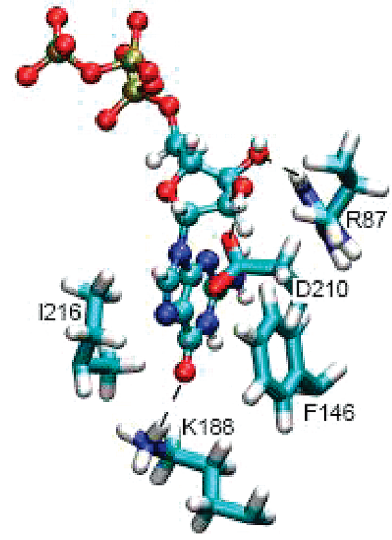

FIGURE 6: Protein−guanosine interactions that persist during the observed global domain motions. The representative conformation occurs 50% into the simulation. For the sake of clarity, only protein side chains involved in conserved interactions with guanonsine are shown.

Table 1: Guanosine−Protein Interactions[a]

motif	protein	GTP	mean (rmsd)
I	R87: NH1	ribose: O2′	3.20 (0.44)
III	E131: OE1	ribose: Ho3′	3.59 (0.53)
IIIa	F146: phenyl COM	guanine: COM	3.83 (0.24)
IV	I216: center of mass	guanine: COM	4.24 (0.218)
	K188: NZ	guanine: O6	2.93 (0.26)

[a] Interactions between conserved active-site residues and the guanosine moiety. COM indicates the center of mass of the residue. The rmsd values were measured with respect to the first frame of the trajectory and are shown in parenthesis. The rmsd value reported is the average over the entire simulation reported in angstroms.

interaction network with the ribofuranose moiety. The net effect of the five residues is to stabilize the guanosine group. This is a critically important role that constrains the electrophilic α-phosphate proximal to the K82 nucleophile, thereby allowing the occurrence of the first chemical step following triphosphate conformational isomerization.

DISCUSSION

Mechanism of Conformational Rearrangement. While the exchange of hydrogen-bonding partners observed in the simulation is a result of an external bias applied to GTP, the exchange does provide a useful context within which a model of molecular events, rationalized in terms of only the electrostatic and thermal driving forces available *in vivo*, may be posited. The OB domain approach toward the active site is triggered by GTP binding and results in the addition of three potential hydrogen bonds with the triphosphate tail of GTP: two from the guanidinium moiety of R295 and one from the terminal amine of K298. The addition of these interactions may destabilize the pre-existing hydrogen bond between the terminal amine of K236 and the nonbridging γ-phosphate oxygen, O2G. The hydrogen-bond destabilization is likely sufficient enough that thermal fluctuations will eventually rupture the bond, thereby allowing thermal fluctuations to rotate the triphosphate toward the reactive state. Figure 4B gives a qualitative, pictorial representation

that may resemble the active site during this period of transition. Once the hydrogen bond with K236 is broken, two hydrogen bonds between the nonbridging oxygens on the β- and γ-phosphate and the guanidinium of R295 may stabilize the rotated state sufficiently to prevent re-equilibration into the unreactive ground state. Additional thermal fluctuations may result in further rotation of the triphosphate tail toward the terminal amines of K234 and K298. As the distance between these amines and the nonbridging γ-phosphate oxygens decreases, an attractive electrostatic force will arise, biasing the triphosphate to fluctuate into the final closed, reactive conformation. This putative model is consistent with independent mutational studies carried out by Sawya and Shuman (*12*) on the mouse capping enzyme and Wang and co-workers on the *S. cerevisiae* capping enzyme (*13*). Each of these studies showed that mutation of the R295 homologue to alanine, as well as the conservative replacement with lysine, eliminated wild-type activity. Both arginine and lysine are positively charged, basic residues; consequently, enzyme activity must require a bidenate hydrogen-bond donor at this position. TMD simulation results suggest an explanation consistent with this requirement. While a positively charged lysine may be sufficient to destabilize the K236−O2G hydrogen bond, single coordination of the triphosphate is insufficient to stabilize the ensemble of intermediate conformers populated during isomerization.

According to our model of conformational isomerization, destabilization of the unreactive state is principally dependent upon interactions between R295 and the triphosphate tail of GTP. Our simulations suggest that these interactions are facilitated by two conformationally restraining hydrogen bonds between the guanidinium moiety of R295 and the carboxyl group of D297. These short, strong interactions restrain the volume of configurational space accessible to both residues and allow R295 to interact more readily with the triphosphate tail. Accordingly, even though D297 does not interact directly with GTP, it is critically important in influencing the course of the catalytically requisite conformational isomerization. This implication is in good agreement with mutational analysis performed by Sawaya and Shuman on the mouse capping enzyme (*12*) and Wang and coworkers on the *S. cerevisiae* capping enzyme (*13*). These studies independently verified that, in both the mouse and *S. cerevisiae* capping enzymes, mutation of the D297 homologue to alanine eliminated wild-type activity. Moreover, each study showed that wild-type activity could not be recovered by replacement of the D297 homologue with asaparagine and that the conservative replacement with glutamate only partially recovered wild-type activity in the mouse capping enzyme. Hydrogen bonds cannot be made between the guanidinium moiety of R295 and the methyl group of alanine, and only a single, weak hydrogen bond can be made with the acetamide group of asparagine. As a result, these studies imply that both hydrogen-bond interactions between the two guanidinium nitrogens of the R295 moiety and the carboxyl group of either aspartate or glutamate are essential for enzyme activity. This is consistent with the suggested restraining role that D297 plays on R295 dynamics. Furthermore, our simulation results are bolstered by the partial recovery of enzyme activity brought about by conservative glutamate substitution. The glutamate mutant has three additional degrees of conformational freedom, so that the glutamate−arginine pair is less restrained than in the wild-type enzyme. The increased fluctuations of the guanidinium decrease its availability to interact with the triphosphate tail of GTP, thereby disfavoring conformational interconversion, resulting in decreased overall activity.

While the triphosphate tail of GTP is undergoing dynamic rearrangements, a critical network of conserved interactions, identified as essential in alanine scans of the mouse and *S. cerevisiae* capping enzymes, stabilize the guanosine moiety of GTP. Essential residues include R87 from motif I, E131 from motif III, F142 from motif IV, and K188. The stabilization that they confer is critical in maintaining the relative proximity of the α-phosphate of GTP to the terminal amine nucleophile of K82 during the conformational isomerization of the triphosphate tail of GTP.

C-Terminal Domain Homology and Mechanistic Conservation. While all members of the nucleotidyltransferase superfamily are structurally characterized by a noncatalytic C-terminal domain and a catalytic N-terminal domain, the structural elements and catalytic role of the C-terminal domain differ among superfamily members. Consequently, to put the results of this study in broader biological context, it is useful to briefly survey the similarities and differences of C-terminal domains among superfamily members while further dividing the superfamily into three subsets.

NAD^+-dependent DNA ligases constitute the first subset of enzymes and contain a structural domain, 1a, fused to the nucleotidyltransferase domain that interacts with the NMN-leaving group in a manner analogous to the interactions between motif VI and GTP in the capping enzymes (*37*). ATP-dependent RNA ligases make up the second subset and can be further broken down into subtypes I and II by the absence of motifs III and IIIa in type-I ligases (*38*). The C-terminal domains in both types of ligase are composed of helical bundles that differ in number and topology (*38, 39*). Additionally, it was shown that the C-terminal domain in type-II ligases is not essential for the formation of the nucleotide enzyme intermediate. The third subset is composed of ATP-dependent DNA ligases and mRNA-capping enzymes. In ATP-dependent DNA ligases and mRNA-capping enzymes, the C-terminal domain is made up of an OB fold and a sixth conserved motif with a constituent RxDK sequence that is essential for the formation of the enzyme−nucleotide intermediate (*9, 12, 40*). This categorization is in accordance with the evolutionary mechanism of biochemical specificity among superfamily members proposed by Shuman and Lima (*8*). In their proposal, contemporary biochemical specificity was attained by fusion of structurally different ancillary domains to an ancient fundamental catalytic unit, which is structurally and functionally homologous to the nucleotidyltransferase domain. This categorization allows us to infer that our proposed mechanism of structural rearrangement that occurs prior to the first step of catalysis is conserved across all eukaryotic capping enzymes and ATP-dependent DNA ligases. Our mechanism provides a dynamic context to rationalize the functional roles of residues deemed essential by mutational analysis and complements previous structural explanations. Additionally, the intimate relationship between structure and function implies that similarly conserved mechanisms likely span members of the other two superfamily subcategories.

The TMD methodology allowed us to generate an ensemble of structures that allowed us to posit a mechanism

of conformational isomerization consistent with mutational analysis of the mouse and *S. cerevisiae* capping enzymes. The structural relationships between superfamily members allowed us to infer the generality of the mechanism to include all mRNA-capping enzymes and ATP-dependent DNA ligases. Despite these extensions to the existing body of knowledge regarding nucleotidyltransferase superfamily members, the study was limited to the first conformational isomerization step prior to the first chemical step. Furthermore, no attempt was made to quantitatively anchor the conformational rearrangements on a thermodynamic footing. Consequently, the limitations of the study naturally prompt several additional questions that merit further investigation. First, what is the relationship between global domain motion and internal dynamics during active-site remodeling prior to and following subsequent catalytic steps? The *syn* to *anti* rotation about the *N*-glycosidic bond of guanosine after the first chemical step is one example (*8*). Second, what are the thermodynamic forces that drive these large global rearrangements, and how are they affected by individual catalytic steps during the guanylation of nascent mRNA? For instance, on the basis of the functional requirements of the capping enzyme, we expect that the presence of GTP in the active site will both lower the energetic barrier of C-terminal domain closure and the free energy of the closed state relative to the apo state. Similarly, it is expected that guanylation of K82 will both destabilize the closed state and lower the energetic barrier of the C-terminal domain opening relative to the apo state. Detailed analysis of the thermodynamics and internal dynamics that occur throughout the catalytic cycle promise to move us toward a more complete understanding of the fundamental nature of oligonucleotide ligation and mRNA capping.

ACKNOWLEDGMENT

R.V.S. thanks members of the J.A.M. group for their considerate input during manuscript development, in particular, Xiaolin Cheng for his insightful discussions regarding TMD, as well as Rommie Amaro for her careful reading of the manuscript.

REFERENCES

1. Kern, D., and Zuiderweg, E. R. (2003) The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol. 13*, 748–757.
2. Koshland, D. E. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A. 44*, 98–104.
3. Shuman, S. (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol. 66*, 1–40.
4. Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell 78*, 657–668.
5. Hamm, J., and Mattaj, I. W. (1990) Monomethylated cap structures facilitate RNA export from the nucleus. *Cell 63*, 109–118.
6. Shatkin, A. J. (1985) mRNA cap binding proteins: Essential factors for initiating translation. *Cell 40*, 223–224.
7. Mao, X., Schwer, B., and Shuman, S. (1996) Mutational analysis of the *Saccharomyces cerevisiae* ABD1 gene: Cap methyltransferase activity is essential for cell growth. *Mol. Cell. Biol. 16*, 475–480.
8. Shuman, S., and Lima, C. D. (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr. Opin. Struct. Biol. 14*, 757–764.
9. Sriskanda, V., and Shuman, S. (1998) Mutational analysis of *Chlorella* virus DNA ligase: Catalytic roles of domain I and motif VI. *Nucleic Acids Res. 26*, 4618–4625.
10. Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell 89*, 545–553.
11. Bougie, I., Parent, A., and Bisaillon, M. (2004) Thermodynamics of ligand binding by the yeast mRNA-capping enzyme reveals different modes of binding. *Biochem. J. 384*, 411–420.
12. Sawaya, R., and Shuman, S. (2003) Mutational analysis of the guanylyltransferase component of Mammalian mRNA capping enzyme. *Biochemistry 42*, 8240–8249.
13. Wang, S. P., Deng, L., Ho, C. K., and Shuman, S. (1997) Phylogeny of mRNA capping enzymes. *Proc. Natl. Acad. Sci. U.S.A. 94*, 9573–9578.
14. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev. 106*, 1589–15615.
15. Apostolakis, J., Ferrara, P., and Caflisch, A. (1999) Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *J. Chem. Phys. 110*, 2099–2108.
16. Mills, G., and Jónsson, H. (1994) Quantum and thermal effects in H_2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys. Rev. Lett. 72*, 1124–1127.
17. Mills, G., Jónsson, H., and Jacobsen, K. W. (1998) Nudged elastic band method for finding minimum energy paths of transitions, in *Classical and Quantum Dynamics in Condensed Phase Simulations* (Berne, B. J., Ciccotti, G., and Coker, D. F., Eds.) pp 385–404, World Scientific, Singapore.
18. Khavrutskii, I. V., Arora, K., and Brooks, C. L., III (2006) Harmonic Fourier beads method for studying rare events on rugged energy surfaces. *J. Chem. Phys. 125*, 174108–174115.
19. Isralewitz, B., Gao, M., and Schulten, K. (2001) Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol. 11*, 224–230.
20. Ravindranathan, K. P., Gallicchio, E., and Levy, R. M. (2005) Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol. 353*, 196–210.
21. Schlitter, J., Engels, M., and Kruger, P. (1994) Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *J. Mol. Graphics 12*, 84–89.
22. Marco, E., Foucaud, M., Langer, I., Escrieut, C., Tikhonova, I. G., and Fourmy, D. (2007) Mechanism of activation of a G protein-coupled receptor, the human cholecystokinin-2 receptor. *J. Biol. Chem. 282*, 28779–28790.
23. Cheng, X., Wang, H., Grant, B., Sine, S. M., and McCammon, J. A. (2006) Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. *PLoS Comput. Biol. 2*, e134-1184.
24. Lee, H. S., Robinson, R. C., Joo, C. H., Lee, H., Kim, Y. K., and Choe, H. (2006) Targeted molecular dynamics simulation studies of calcium binding and conformational change in the C-terminal half of gelsolin. *Biochem. Biophys. Res. Commun. 342*, 702–709.
25. Vriend, G. (1990) WHATIF: A molecular modeling and drug design program. *J. Mol. Graphics 8*, 52–56.
26. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. 117*, 5179–5197.
27. Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem. 24*, 1016–1025.
28. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926–935.
29. Case, D. A., Darden, T. A., Cheatham, I. T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) *Amber 8*, University of California at San Francisco, San Francisco, CA.
30. Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992) Langevin dynamics of peptides—The frictional dependence of

isomerization rates of *N*-acetylalanyl-*N*′-methylamide. *Biopolymers 32*, 523–535.
31. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of Cartesian equations of motion of a system with constraints—Molecular dynamics of *N*-alkanes. *J. Comp. Phys. 23*, 327–341.
32. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.
33. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33–38.
34. Lou, H. F., and Cukier, R. I. (2006) Molecular dynamics of apo-adenylate kinase: A principal component analysis. *J. Phys. Chem. B 110*, 12796–12808.
35. Rinaldo, D., and Field, M. J. (2003) A computational study of the open and closed forms of the N-lobe human serum transferrin apoprotein. *Biophys. J. 85*, 3485–3501.
36. Gerstein, M., Lesk, A. M., and Chothia, C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry 33*, 6739–6749.
37. Gajiwala, K. S., and Pinko, C. (2004) Structural rearrangement accompanying NAD^+ synthesis within a bacterial DNA ligase crystal. *Structure 12*, 1449–1459.
38. El Omari, K., Ren, J., Bird, L. E., Bona, M. K., Klarmann, G., LeGrice, S. F. J., and Stammers, D. K. (2006) Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J. Biol. Chem. 281*, 1573–1579.
39. Nandakumar, J., Shuman, S., and Lima, C. D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell 127*, 71–84.
40. Doherty, A. J., and Suh, S. W. (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res. 28*, 4051–4058.

BI8000209