A Mechanism for S-adenosyl methionine Assisted Formation of a Riboswitch Conformation: A Small Molecule with a Strong Arm

A Mechanism for S-adenosyl methionine Assisted Formation of a Riboswitch Conformation: A Small Molecule with a Strong Arm

Wei Huang[1], Joohyun Kim[2], Shantenu Jha[2]*, and Fareed Aboul-ela[1]*

[1]Department of Biological Sciences and [2]Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803

*address correspondence to Fareed Aboul-ela (faboul@lsu.edu) or Shantenu Jha (sjha@cct.lsu.edu). Mailing Address: Fareed Aboul-ela, Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803 or Shantenu Jha, Center for Computation and Technology,

Abstract

The SAM-I riboswitch mediates expression of proteins involved in sulfur metabolism via formation of alternative conformations in response to binding by SAM. Models for kinetic trapping of the RNA in the bound conformation require annealing of non-adjacent mRNA segments during a transcriptional pause. The entropic cost required to bring non-adjacent segments together should slow the folding process. To address this paradox, we performed molecular dynamics simulations on the SAM-I riboswitch aptamer domain with and without SAM, starting with the X-ray coordinates of the SAM-bound RNA. Individual trajectories are 200 ns, amongst the longest reported for an RNA of this size. We applied principle component analysis (PCA) to explore the global dynamics differences between these two trajectories.. We observed a conformational switch between a stacked and non-stacked state of a non-adjacent dinucleotide in the presence of SAM. In the absence of SAM the coordination between a bound magnesium ion and the phosphate of A9, one of the nucleotides involved in the dinucleotide stack, is destabilized. An electrostatic potential map reveals a "hot spot" at the Mg binding site in the presence of SAM. These results suggest that SAM binding helps to position J1/2 in a manner that is favorable for P1 helix formation.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abbreviations

SAM	**S-adenosyl methionine**
SAH	**S-adenosyl homocystein**
AAT	**anti-anti-terminator**
AT	**anti-terminator**
T	**terminator**
RMSD	**root mean square deviation**
RMSF	**root mean square fluctuation**
ESP	**ElectroStatic Potential**

Introduction

Most folded, functional RNAs require conformational flexibility to perform their biological role(1). Often RNA folding occurs as a response to specific signals, such as the binding of a protein, small molecule ligand or metal ion(2). A riboswitch is an autonomous *cis*-regulatory element that can control gene expression by sensing levels of small molecule metabolites, such as amino acids, nucleotides, or enzyme cofactors(3,4). Riboswitches are widely distributed in genomic sequences(5,6). They have been proposed as RNA targets for novel small molecule drugs(7). Riboswitches contain an aptamer domain that can bind to the effector ligand with high specificity(8-10). The mechanism of riboswitch function involves formation of alternative secondary structures in the presence and absence of the metabolite ligand(8,9). It is not clear how these interactions between small molecules and RNAs are coupled to the RNA secondary structures.

The SAM-I riboswitch appears in the 5' untranslated region (5'-UTR) of the mRNA of a number of bacteria. It modulates expression of genes involved in the biosynthesis of methionine, cysteine, SAM and other sulfur containing metabolites in response to the concentration of SAM *in vivo*(11-14). Genetic and biochemical studies (11-13,15) propose the following model for the regulatory mechanism of SAM-I riboswitches (Figure 1a). In the absence of SAM, the SAM-I riboswitch forms the anti-terminator (AT), which prevents formation of the terminator (T) that is required for ρ-independent transcription attenuation. Binding by SAM triggers formation of the anti-anti-terminator (AAT) that occupies the 5' strand of the AT. Under these conditions, the formation of the T facilitates transcription termination, turning off the expression of a set of downstream genes that are involved in sulfur-related metabolic pathways. An important question raised by this model is how SAM binding facilitates AAT formation.

Studies of other riboswitches (16,17) have pointed towards the importance of kinetic factors in determining ligand binding and riboswitch conformations. It has been proposed that SAM binds to the AAT and fixes the conformation before the polymerase can complete the synthesis of AT and T forming segments of the 5'-UTR mRNA(18,19). However, the formation of the AAT requires base pairing between two distal mRNA segments to form the P1 helix. This process requires the RNA to overcome an entropic barrier, which is expected to slow the folding process(20). The barrier may be overcome if the polymerase pauses long enough to allow folding before completing synthesis of the AT-forming segments. The model of kinetic control is strengthened further if ligand binding can accelerate the folding of the aptamer domain, as has been suggested for purine-binding riboswitches(17,21).

Molecular Dynamics simulations have been used to address similar mechanistic questions in protein dynamics(22,23). These methods have limitations due to the large number of degrees of freedom, computational demands and the difficulties of preparing accurate forcefields(24,25). Distinct characteristics of nucleic acids further complicate the application to RNA: the highly negatively charged backbone, the complicated role of the

ionic environment and distinctive hierarchical folding mechanisms(26-29). Nonetheless, it is widely accepted that simulation protocols for RNA systems are mature enough to provide physical insights regarding atomistic details(26,27,29-31). For example, when applied carefully simulations have produced valuable insights, cross validated by experimental data, regarding ionic and electrostatic interactions(32), ribozyme folding and catalysis (33) and ligand-RNA interactions(34) for RNAs up to the ~25 Kd range.

In this manuscript, we report long timescale (200 ns) molecular dynamics simulations of the SAM-1 riboswitch aptamer based upon the crystal structure in the AAT conformation from *Thermoanaerobacter tengcongensis* in complex with SAM(35). This is one of the longest simulations to date of an RNA of this size (32.2 Kd)(36). We repeated the simulation with SAM absent. Principal components analysis (PCA) of the covariance matrix for the atomic coordinates based upon each trajectory has been performed. In the presence of SAM the MD trajectory preserves critical contact distances between the RNA and SAM, as well as with a bound magnesium ion. Many of these contacts involve tight interactions with the J1/2 junction connecting the 5' end of the RNA with the P2 helix. When the ligand is removed, contacts between the magnesium and the phosphate backbone of J1/2 are destabilized. Intriguingly, we observe a non-adjacent dinucleotide stack, which, though not present in the T. *tencongensis* X-ray coordinates, has appeared in a more recent SAM-I riboswitch from another organism. Altogether, our findings suggest a hypothesis for how SAM binding can accelerate and/or stabilize AAT formation by constraining the position of J12, thus reducing the entropic barrier for the annealing of the P1 helix.

Methods

The aptamer domain of *T.tengcongensis* SAM-I riboswitch and its complex with SAM has been submitted to Molecular Dynamics (MD) simulation for 200 ns using an explicit water solvent model. The 2.9 Å crystal structure of SAM bound to the aptamer domain of SAM-I riboswitch (35) is used as the starting point. The simulations with SAM are designated as SAM_TRAJ, while the simulations without ligand are denoted as WOSAM_TRAJ. A second simulation was carried out with identical protocols and length as SAM_TRAJ. Some analysis plots for the second trajectory with SAM are included in the supplementary material. From this point on, we discuss only SAM_TRAJ and WoSAM_TRAJ, concentrating on those aspects of SAM_TRAJ that were preserved in the trajectory of the second simulation with SAM present (unless otherwise noted).

Molecular Dynamics Simulation

The starting structure is the crystal structure of SAM binding to the aptamer domain of the SAM-I riboswitch in the AAT conformation(35) (PDB ID: 2GIS). In WOSAM_TRAJ, SAM is directly removed from the x-ray crystal structure and replaced with solvent water. Water and two magnesium ions in the crystal structure are kept. All the simulations are performed using NAMD (37) with amber99bsc0 correction force field(25). Parameters for SAM are from the Generalized Amber Force Field (GAFF) (38) and missing

parameters are calculated using ANTECHAMBER (39) using the AM1-bcc model(40). The starting structures are prepared using the LEaP module in AMBER(41). The RNA molecules are solvated in a cubic solvent box of TIP3P waters (42) with a 16 Å padding in all directions. Sodium ions are placed in the most negative position around RNA in order to neutralize charges of the system. Energy minimizations are carried out to remove bad contacts. Starting from 0 K, the temperature is raised 10 K for every 10,000 steps and is held constant after reaching the desired temperature (300 K) using temperature reassignment. Harmonic constraints were applied on the RNA and gradually removed over the equilibration steps using scale from 1.0 to 0 with 0.2 increment per 50,000 steps. MD simulations are performed in the NPT ensemble. The Langevin dynmamics was used for constant temperature control, with the value of Langevin coupling coefficient and the Langevin temperature set to 5 ps and 300 K, respectively. The pressure was maintained at 1 atm using the Langevin piston method with a period of 100 fs and decay times of 50 fs. The time step is 2 fs for both equilibration and production phase. Bond lengths between hydrogens and heavy atoms are constrained using SHAKE(43). The long-range electrostatics is treated with the Particle Mesh Ewald (PME) method (44) with a cutoff distance 12 Å. The trajectory is recorded for every 10 ps.

RMSD and RMSF calculation

We have calculated RMSD values including heavy atoms only. Both the X-ray structure and the average structure have been used as references, as noted in the text and figure captions. We have also calculated the RMSF (45) which provides information about the atomic fluctuations over the trajectory broken down by residue. Both the RMSD and RMSF are calculated using the ptraj module in AMBER(41).

Details regarding the PCA, mapping of the Electrostatic Potential (ESP),and Clustering analysis are in the Supplementary Information (SI).

 Results

We aimed to submit the X-ray coordinates of the SAM-I riboswitch aptamer domain to a relatively long timescale (200 ns) MD simulation. The resulting trajectory samples the energetically accessible conformational space in the vicinity of the X-ray coordinates. We repeated the calculations and subsequent analysis with SAM removed to determine the effect of the ligand on the trajectory. We carefully monitored the overall deviation of coordinates from the starting structure and from an average structure. In this way, we could determine when our simulations had equilibrated, and detect signs of instabilities in coordinates which might indicate intrinsic problems with the structure or simulation parameters.

RMSD and RMSF of the binding pocket

The residues that have atoms within 5 Å of SAM were defined as the binding pocket (Figure 1b, c). The RMSD of all heavy atoms in the binding pocket with reference to the crystallographic structure is presented in Figure 2. Each snapshot has been superimposed on the crystallographic coordinates to remove the overall translational and rotational motion. In SAM_TRAJ, the binding pocket is relatively stable throughout the trajectory, though the full RMSD shows an increase in the final ~10% of the 200 ns simulation (see discussion section). The interactions between the ligand and the residues in the binding pocket are maintained (See below). However, in the absence of SAM, the RMSDs gradually increase up to 100 ns. This result indicates instabilities in the crystallographic structures when the ligand is removed. The RMSFs of residues in the binding pocket increase in the absence of SAM (SI Figure 1). Altogether, these results suggest that the binding pocket configuration requires the presence of SAM to maintain a stable fold. Over the course of our simulations, SAM_TRAJ fluctuates about an equilibrium structure which is close to the X-ray coordinates (SI Figure 2). WoSAM_TRAJ, on the other hand, does not reach a stable, equilibrium structure after 200 ns of simulations (SI Figure 2).

Essential dynamics of the SAM-I riboswitch

The MD simulation samples the 3N-dimensional space corresponding to the positions of the *N* atoms. We used PCA as a guide to identify significant dymamic processes from this very large dataset. The first several eigenmodes indicate the directions of global motions and the eigenvalues indicate the magnitude of the motion along each eigenmode. The eigenmodes associated with the largest eigenvalues have the largest contribution to essential dynamics(46). The approach contrasts with conventional normal mode dynamics, which detects relatively trivial harmonic motions. The projection of the trajectory on an eigenmode, which is called the principle component, provides the time-dependent dynamics of the system along this eigenmode.

Here PCA is applied to both SAM_TRAJ and WOSAM_TRAJ. The eigenvalue drops steeply in the first 10 eigenmodes, and the first 5 eigenmodes account for about 72% of the fluctuation of the trajectories (SI Figure 3). The RMSF plots of both trajectories along their first 5 eigenmodes are displayed in SI Figure 3, which present the fluctuations on an atomic basis. Based on these plots, the regions that are responsible for the most variance in atomic position can be identified. A clustering analysis identified three major structure clusters over the course of the trajectory. These are plotted as a function of the first two principle components in Figure 3a, alongside representative snapshots of each cluster for WOSAM_TRAJ. The plot shows that the PCA and clustering analysis have identified similar dynamic processes in WOSAM_TRAJ..

Snapshots representing each cluster reveal a large scale relative motion of helices P1 and P3 (Figure 3, see also SI Movie 1). The P1 helix moves in the opposite direction relative to the P3 helix along the 1st eigenmode in WOSAM_TRAJ (SI figure 4). The time evolution of the first principle component and the distance monitor shows that the P1 helix and the P3 helix separate in the 1st 100 ns and then move close together in the 2nd 100 ns (Figure 3, SI Figures 3 and 4). Like a pair of chopsticks, the two helices move in

7

phase but in opposite directions. The change from positive to negative correlation for these two helices upon removal of SAM is highlighted in SI Figure 4.

The residues that display spike patterns in both trajectories are highlighted in SI Figure 3. These residues (A14, U34 and A51) are highly exposed to the solvent in the crystallographic structure(35). Residue A9 has a spike pattern along the first 5 eigenmodes in SAM_TRAJ. A9 has the opposite orientation as compared to other nucleotides in single stranded junction J1/2, but it is placed close to residues in J3/4. Aside from the A9 and A14 residues, motions in the presence of SAM are confined to the P4 helix. This is the only large scale secondary structure element which does not contact SAM. Overall, the PCA analysis indicates that SAM binding contributes to the formation of a stable core tertiary fold, in agreement with FRET and gel mobility studies(47).

Transient formation of a non-adjacent dinucleotide stack

In order to understand the continuing dynamic behavior observed for A9 in the presence of SAM according to the PCA analysis we inspected more closely the movement of this residue over the course of the trajectory. We observed that A9 and U63 transiently form a non-adjacent dinucleotide stack resembling a motif originally observed in a ribosomal RNA (48,49) (Figure 4a). The LJ energy between nucleotides A9 and U63 was used to monitor this non-adjacent dinucleotide stack. Strongly favorable (negative) values indicate that A9 and U63 are stacked, while we interpret values near zero as unstacked. Some sample snapshots are presented (Figure 4a) and a movie (SI Movie 2). Multiple instances of the transient stack take place, typically holding for a period of tens of nanoseconds.

The sample snapshots show that only one stacked mode between A9 and U63 is found in SAM_TRAJ. WOSAM_TRAJ adopts the same stacking mode as SAM_TRAJ at the beginning of the simulation. Before 20 ns the glycosidic angle of the nucleotide U63 flips from *anti* to *syn* (Figure 4a, b). Later, a transition between stacked state and unstacked state of A9 and U63 is coupled with a shift in glycosidic angle from *anti* to *syn* on A9 and from *anti* to *high-anti* on U63 in SAM_TRAJ (Figure 4a, b and c).

A histogram showing the occupancy of glycosidic angles in Figure 4c gives an indication of the "essential" nature of the motion involving A9 in SAM_TRAJ. This residue undergoes a biphasic "jump" motion between stacked and unstacked configurations. By contrast, in WOSAM_TRAJ motions involving A9, though present, may be harmonic, as indicated by the near-gaussian distribution of glycosidic angle occupancies. U63 may be undergoing anharmonic motion in WoSAM_TRAJ. The correlation between the non-adjacent dinucleotide stack and the glycosidic angle of A9 in SAM_TRAJ is again apparent, since both monitors show a biphasic distribution of values over the course of the trajectory.

Coordination of the specific binding magnesium ion with J1/2 and J3/4 phosphates

In the crystallographic structures of the SAM-I riboswitch, a specific magnesium binding site is found between the junction J1/2 and the junction J3/4(35). This specific bound magnesium ion is also similarly found in a recent crystallographic structure of the yit-J SAM-I riboswitch (**Ailong Ke, personal communication**). The magnesium ion apparently counters the electrostatic repulsion between the phosphate backbone of J1/2 and that of J3/4. Additionally, this magnesium binding site is close to the SAM binding pocket. We monitored the distances of this magnesium to negatively charged atoms on the RNA during our simulated trajectories (Figure 5b). In WoSAM_TRAJ, the contact distances between the magnesium ion and the phosphate backbone of J1/2 lengthen, while those with the phosphate backbone of J3/4 hold constant. The S-turn motif in J3/4 is preserved during the simulation. In SAM_TRAJ, all of the magnesium distances are constant through the simulation. We defined a pseudo-residue by combining those phosphates which contact the magnesium ion in the X-ray structure, and monitored the radius of gyration of this residue over the course of the two trajectories (SI Figure 5). The effective radius of gyration is reduced significantly in SAM_TRAJ as compared to WoSAM_TRAJ.

These results suggest that binding of magnesium and SAM might have a synergistic effect on the stabilization of the J1/2 junction. In addition, the coordination of this magnesium ion with the 3' oxygen atom on A9 may be correlated with the non-adjacent dinucleotide stack state of A9 and U63 (Figure 3a and Figure 5c). Coupled motions are also apparent for the backbone torsion angles (α, γ, δ, ε, ζ) of A9 (SI Figure 6).

An ESP map confirms the likely presence of an electrostatic "hot spot" at this site in the presence of SAM (Figure 6a). Closeups of the hot spot for representatives of each of the three clusters obtained from SAM_TRAJ and WOSAM_TRAJ are shown in Figure 6b. The maps show the hotspot surrounding the Mg in a pocket tightly bounded by J1/2 and J3/4 in SAM_TRAJ snapshots, but focussed more narrowly near the backbone of J3/4 in WOSAM_TRAJ. As forcefields for divalent ions are seen as problematic for nucleic acid simulations(33), we performed a series of shorter simulations in which Mg ions were removed and only monovalent Na ions were present. A high occupancy of the binding site of the removed Mg coordinating the J1/2 and J3/4 phosphate groups was observed for Na ions, with two specific sodium ions in particular located permanently near the phosphates of G10 and A11. More Na ions seem to occupy the site with SAM present than with SAM absent (SI Movie 3 and SI Movie 4).

Interactions between the ligand SAM and the SAM-I riboswitch

Residues in the binding pocket of the SAM-I riboswitch and SAM form a network of hydrogen bonds(35). To explore the interactions between SAM and the SAM-I riboswitch aptamer, we monitored the distance between donor and acceptor for possible hydrogen bonds as indicated in the crystallographic structure (Figure 7). As shown in the figure, all these contacts are maintained during the simulation. Of these distances, the interactions between the methionine chain moiety of SAM and G11/ G58 persist most rigidly through the simulation. In addition, we monitored hydrogen bonding using the *hbond* command in the *ptraj* module of Amber10(41) for all hydrogen donors and

acceptors identified involving SAM binding (SI Table 1). In this analysis, all possible hydrogen bond pairs were tracked. The majority of the hydrogen bond pairs were observed between the carboxyl group/the amino group on the methionine chain moiety of SAM and G11/G58 in the SAM-I riboswitch. Moreover, 3 of the 4 hydrogen bonds with highest occupancy occur between the carboxyl group of SAM and G11.

Discussion

We have used MD simulations to probe the mechanism for coupling between SAM binding and SAM-I riboswitch conformation and secondary structure. Trajectories obtained from our long timescale simulated trajectories have been analyzed by PCA and clustering analysis and ESP calculations. Our trajectories indicate that SAM interactions with J1/2 are major contributors to binding. Moreover, these interactions help to fix J1/2 into a configuration that is favorable for P1 helix formation, which has been proposed to be the kinetic barrier to AAT formation[20]. Comparison of the SAM_TRAJ and WoSAM_TRAJ leads us to suggest that some details of the bound SAM RNA conformation are stabilized when SAM is present in the binding pocket.

Stabilization of the riboswitch aptamer core in the presence of SAM

PCA analysis clearly suggests that a core region of the aptamer structure, defined as those secondary structure elements that contact SAM with at least one residue (P1, J1/2 and P3,) are stabilized when SAM is bound. P4 and some regions of P2, which do not have direct contact with SAM, undergo more dynamic fluctuations in SAM_TRAJ . Two exceptions to this rule are A9, in J1/2, and A14, the bulged nucleotide within the P2 helix. Interestingly, these two residues have the lowest reported electron density in the submitted X-ray data. The apparent correlation in their motion, as indicated by the PCA plots (SI Figure 3) appears to reflect the fact that adjacent secondary structure elements are undergoing very little motion. The intervening residues (10-13) are stacked and are held rigid by interactions with SAM and magnesium. The unstacked status of A9 is reinforced by the transient formation of the non adjacent dinucleotide stack with U63 (see below).

J1/2 positioned for formation of the P1 helix

The observation of extraordinarily stable contact distances between G11 and SAM is consistent with the high sequence conservation observed at this position. Mutation at this position eliminates the SAM binding[11]. This is consistent with previous findings that the length of the methionine moiety is crucial for tight binding of SAM to the aptamer domain[11]. The interactions between SAM and G11 are in the form of hydrogen bonding. This may help explain the discrimination by the riboswitch between SAM and its analog S-adenosyl homocystein (SAH), which is restricted to adopt certain conformations (35,50).

The monitor of magnesium coordination suggests that J1/2 cannot maintain its compact configuration as seen in the crystallographic structure in the absence of SAM. This conclusion is reinforced by the overall increase in the radius of gyration of a pseudo-residue defined by the group of phosphates involved in coordinating the magnesium (SI figure 5). However, a hot spot of negative elecstrostatic potential seems to preserve the magnesium coordination throughout the simulation in the presence of SAM. This result is consistent with biophysical studies which indicate that the compact global folding of the SAM-I riboswitch is stabilized by SAM binding(47). Altogether, these results suggest that the SAM binding affects the placement of J1/2. SAM binding would then appear to lower the entropic cost of pulling the 5' and 3' of the P1 helix together by pre-positioning J1/2, thus facilitating formation of the P1 helix.

Though there is no experimental report of Mg ions being replaced by Na ions in a stably folded SAM-I riboswitch, we have performed simulations with only the latter ion present. These simulations, together with the ESP maps, indicate that the Mg binding site is located within a pocket of strongly negative electrostatic potential in the SAM-bound form. The appearance of the ESP maps for SAM_TRAJ (Figure 6b) suggests that the contact between SAM and J1/2 positions the latter close to J3/4, setting up a region of electrostatic repulsion. Close contacts (<7 angstroms; ~3.5 angstroms for 2 contacts) observed in the X-ray structure and in our simulations for the Mg ion with phosphate moieties (figure 5) are consistent with inner shell binding(51).

The function of the non-adjacent dinucleotide stack

In SAM_TRAJ, the non-adjacent dinucleotide stack of A9 and U63 is transiently formed during the simulation. Though the non-adjacent stack is not observed in the X-ray structure of *T. tencongensis* SAM-I riboswitch RNA, an analogous non-adjacent dinucleotide stack (A10 and A84) is observed in the recent crystallographic structure of the yitJ S-box riboswitch **(Ailong Ke, personal communication)**. Our simulations suggest that the dinucleotide stack may be a conserved feature of SAM-I riboswitches, though this tertiary interaction is apparently not stably maintained. Coordination of the specific binding magnesium and phosphate atom on A9 is coupled with this dinucleotide stack state (Figure 4a and Figure 5c) in SAM_TRAJ, though not in the second SAM trajectory (see supplementary material). When the stack is lost, the coordination of the magnesium and phosphate atom on A9 is shortened slightly. It is possible that this non-adjacent dinucleotide stack stabilizes the tertiary interactions between J1/2 and J3/4.

Although the simulation including bound SAM shows transient formation of, loss, and reformation, of the A9/U63 stack, the stack is maintained consistently through WoSAM_TRAJ. As mentioned above, the significance of conformational changes observed in WoSAM_TRAJ, including the observation of alternate glycosidic angles, is difficult to judge. Although we have used forcefield potentials which have been explicitly optimized with glycosidic angles in mind(25), with no published apo-structure available we have no means of cross validating the details of WoSAM_TRAJ (see below). What we can say from this trajectory is that in the absence of SAM, A9 will have a tendency to

seek alternative conformations. One correlate will be a weakening of the coordination between the specific bound magnesium and the phosphate of A9 mentioned above.

Robust features of long timescale simulations

Since the free-form structure of the aptamer domains have not yet been available publicly or determined experimentally, the free-form structure was simply obtained by removing SAM from the coordinates of the bound complex. However, the goal of this study is not to probe the conformation of the free-form SAM-I riboswitch. The limited timescales available to MD are not sufficient to provide meaningful information in the absence of an experimental determination of the equilibrium structure. Our aim has been to understand effects that SAM binding has on the tertiary structure and dynamics of the aptamer domain.

The simulations reported here are amongst the longest all-atom trajectories reported for an RNA of this size to date(36). Several hundred nanosecond simulations require significant computing resources. Coarse-grained models or knowledge-based models could be an alternative but protocols are still in development(52). Recent simulations have begun to press the limits of MD simulations well into the tens of nanoseconds range or longer for biologically relevant RNAs(33,34). The time evolutions of the root-mean-square deviation (RMSD) of all heavy atoms (atoms except hydrogen atom) with reference to the crystal structure and average structure during simulations are shown in SI Figure 2. As shown in the figure, the RMSD values with reference to the crystallographic structure are maintained within 6 Å, which are reasonable for RNAs that have complicated tertiary folding. The RMSD values using the average structure as a reference stays around 3 Å over the simulation. The larger RMSDs, and larger scale dynamic behavior identified by PCA analysis in the absence of SAM, imply that the aptamer structure as represented in the X-ray coordinates requires the presence of bound SAM for full stability. The first 3 eigenmodes show a strong correlation between the first and second halves of SAM_TRAJ (SI Figure 7), in contrast to the cross correlation map observed between the two halves of WOSAM_TRAJ. We conclude that the essential dynamic modes are consistent through SAM_TRAJ but not through WOSAM_TRAJ.

An apparent jump in the RMSD for SAM_TRAJ during the last 20 ns of the trajectory indicates that at this stage the simulation is beginning to deviate sharply from the experimental coordinates. Until 160 ns, the RMSD stays within less than 5 angstroms of the X-ray structure. Accumulation of small errors in the initial input are a concern over a long MD trajectory(31). Recent progress in refinement of forcefield parameters and protocols has increased confidence in long timescale MD simulations of protein dynamics(53). Few if any MD trajectories of 100 ns or longer have been reported for RNA(31). For these reasons, we ran a second 200 ns MD simulation with SAM present (SI Figure 8). The second trajectory maintained an RMSD relative to the experimental structure of <~5 angstroms. Unless otherwise noted, we highlighted only features of the SAM_TRAJ which were present also during the second trajectory.

12

Effects of transcriptional pausing

The critical decision step in SAM-I mediated regulation is the competence of the 3'
sequence of the P1 helix to form AAT or AT conformations. A folding simulation (20)
study has proposed that SAM can bind to the aptamer after formation of the P3 helix,
then accelerate overall AAT folding. Several studies (19,54) have proposed that
transcriptional pausing or the kinetics of folding has an effect on this genetic decision
step. However, the details of how transcriptional pausing influences riboswitch regulation
are unclear. The actual number of nucleotides that have been transcribed at the time when
SAM binding occurs is unknown.

Although tertiary structures are more brittle than the secondary structures for RNA
molecules, as evidenced in the absence of SAM of this study, it is believed that it is the
tertiary structures that account for the kinetic arrest of the structure rearrangement of
RNA molecules from mechanical pulling studies(55,56). The comparison of the
trajectories of the riboswitch with and without SAM bound suggests a mechanism
through which SAM binding can stabilize the formation of the AAT conformation
relative to an alternative conformer, such as the AT. Alternatively, by aiding the
positioning of J12 SAM binding may accelerate AAT formation during transcription as
predicted by models for kinetic control of riboswitches. The placement of J1/2 facilitated
by SAM binding increases the probability, and therefore the rate, of the encounter
between the 5' and the 3' ends of the P1 helix. The results here are consistent with a
kinetic effect arising from the binding of SAM, trapping the AAT conformation before
the full AT sequence is synthesized by the polymerase.

Conclusion and future work

The results from MD simulations lead us to propose that the positioning of J1/2 upon
SAM binding helps to overcome the entropic barrier to formation of the P1 helix. SAM
binding appears to facilitate the coordination of a magnesium ion to phosphates within
the J1/2 strand. This proposed mechanism can be tested by single molecule experiments
(56) that can explore the energy landscape of the folding or folding kinetics studies(57,58)
with and without SAM. Interestingly, a non-adjacent dinucleotide stack that is present in
a recent crystallographic structure of the SAM-I riboswitch emerges in our simulations.
This finding leads us to suggest that the interaction may form transiently in the
T.tencongensis SAM-I riboswitch RNA as well.

It is important to note that little structural or dynamic information regarding the AT
conformation is currently available at this point. Future studies of transcriptional pausing,
dynamic equilibria between AAT, AT, and T forms, and, in particular, how these factors
are correlated with SAM binding, will be necessary to reveal the regulatory mechanism
behind this system. Understanding the molecular basis of riboswitch recognition of SAM
and the coupling between ligand binding and RNA secondary structure is crucial for the
design of novel riboswitch binding small molecules. Such designed ligands can provide a
powerful tool to further probe the flexibility of riboswitch RNA(59), and the link between
RNA conformational dynamics and function(1).

References

1. Al-Hashimi, H.M. and Walter, N.G. (2008) RNA dynamics: it is about time. *Current Opinion in Structural Biology*, **18**, 321-329.

2. Cruz, J.A. and Westhof, E. (2009) The dynamic landscapes of RNA architecture. *Cell*, **136**, 604-609.

3. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, **5**, 451 - 463.

4. Grundy, F.J. and Henkin, T.M. (2004) Regulation of gene expression by effectors that bind to RNA. *Current Opinion in Microbiology*, **7**, 126-131.

5. Kazanov, M., Vitreschak, A. and Gelfand, M. (2007) Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics*, **8**, 347.

6. Barrick, J. and Breaker, R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology*, **8**, R239.

7. Blount, K.F. and Breaker, R.R. (2006) Riboswitches as antibacterial drug targets. *Nat Biotech*, **24**, 1558-1564.

8. Blouin, S., Mulhbacher, J., Penedo, J.C. and Lafontaine, D.A. (2009) Riboswitches: Ancient and Promising Genetic Regulators. *ChemBioChem*, **10**, 400-416.

9. Henkin, T.M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes & Development*, **22**, 3383-3390.

10. Lim, J., Winkler, W.C., Nakamura, S., Scott, V. and Breaker, R.R. (2006) Molecular-recognition characteristics of SAM-binding riboswitches. *Angewandte chemie int. ed.*, **45**, 964-968.

11. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. and Breaker, R.R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Struct. Biology*, **10**, 701-707.

12. Mcdaniel, B.A.M., Grundy, F.J., Artsimovitch, I. and Henkin, T.M. (2003) Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci, USA*, **100**, 3083-3088.

13. Epshtein, V., Mironov, A.S. and Nudler, E. (2003) The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc. Natl. Acad. Sci, USA*, **100**, 5052-5056.

14. Wang, J.X. and Breaker, R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. . *Biochemistry and Cell Biology*, **86**, 157-168.

15. Grundy, F.J. and Henkin, T.M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol Microbiol*, **30**, 737 - 749.

16. Wickiser, J.K., Cheah, M.T., Breaker, R.R. and Crothers, D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**, 13404-13414.

17. Gilbert, S.D., Stoddard, C.D., Wise, S.J. and Batey, R.T. (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology*, **359**, 754-768.

18. Montange, R.K. and Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annual Review of Biophysics*, **37**, 117-133.

19. Tomsic, J., Mcdaniel, B.A., Grundy, F.J. and Henkin, T.M. (2008) Natural variability in S-Adenosylmethionine (SAM)-dependent riboswitches: S-Box elements in Bacillus subtilis exhibit differential sensitivity to SAM in vivo and in vitro. *Journal of Bacteriology*, **190**, 823-833.

20.   Whitford, P.C., Schug, A., Saunders, J., Hennelly, S.P., Onuchic, J.N. and Sanbonmatsu, K.Y. (2009) Nonlocal helix formation is key to understanding S-Adenosylmethionine-1 riboswitch function. *Biophysical Journal*, **96**, L7-L9.

21.   Rieder, R., Lang, K., Graber, D. and Micura, R. (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem*, **8**, 896 - 902.

22.   Mccammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature*, **267**, 585-590.

23.   Dodson, G.G., Lane, D.P. and Verma, C.S. (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Reports*, **9**, 144-150.

24.   Showalter, S.A. and Bruschweiler, R. (2007) Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *Journal of Chemical Theory and Computation*, **3**, 961-975.

25.   Perez, A., Marchan, I., Svozil, D., Šponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of {alpha}/{gamma} conformers. *Biophys. J.*, **92**, 3817-3829.

26.   Chu, V.B., Bai, Y., Lipfert, J., Herschlag, D. and Doniach, S. (2008) A repulsive field: advances in the electrostatics of the ion atmosphere. *Current Opinion in Chemical Biology*, **12**, 619-625.

27.   Mackerell, A.D., Jr. and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Current Opinion in Structural Biology*, **18**, 194-199.

28.   Foloppe, N., Nilsson, L. and Mackerell, A.D., Jr. (2001) Ab initio conformational analysis of nucleic acid components: Intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers*, **61**, 61-76.

29.   Mcdowell, S.E., Špačková, N.A., Šponer, J. and Walter, N.G. (2007) Molecular dynamics simulations of RNA: An *in silico* single molecule approach. *Biopolymers*, **85**, 169-184.

30.   Aduri, R., Psciuk, B.T., Saro, P., Taniga, H., Schlegel, H.B. and Santalucia, J. (2007) AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *Journal of Chemical Theory and Computation*, **3**, 1464-1475.

31.   Hashem, Y. and Auffinger, P. (2009) A short guide for molecular dynamics simulations of RNA systems. *Methods*, **47**, 187-197.

32.   Auffinger, P., Bielecki, L. and Westhof, E. (2003) The Mg2+ Binding Sites of the 5S rRNA Loop E Motif as Investigated by Molecular Dynamics Simulations. *Chemistry & Biology*, **10**, 551-561.

33.   Krasovska, M.V., Sefcikova, J., Réblová, K., Schneider, B., Walter, N.G. and Šponer, J. (2006) Cations and hydration in catalytic RNA: Molecular dynamics of the Hepatitis Delta Virus Ribozyme. *Biophysical Journal*, **91**, 626-638.

34.   Villa, A., Wohnert, J. and Stock, G. (2009) Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch. *Nucl. Acids Res.*, **advance access**.

35.   Montange, R.K. and Batey, R.T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172-1175.

36.   Auffinger, P. (2006) In Šponer, J. and Lankaš, F. (eds.), *Computational Studies of RNA and DNA*, pp. 283–300.

37.   Phillips, J.C., Braun, R., Wei, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, **26**, 1781-1802.

38.   Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry*, **25**, 1157-1174.

39. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, **25**, 247-260.

40. Jakalian, A., Bush, B.L., Jack, D.B. and Bayly, C.I. (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry*, **21**, 132-146.

41. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Jr, K.M.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, **26**, 1668-1688.

42. Jorgensen, W.L. and Duffy, E.M. (2000) Prediction of drug solubility from Monte Carlo simulations. *Bioorganic & Medicinal Chemistry Letters*, **10**, 1155-1158.

43. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, **23**, 327-341.

44. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, **103**, 8577-8593.

45. Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, **450**, 964-972.

46. Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, **17**, 412-425.

47. Heppell, B. and Lafontaine, D.A. (2008) Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. *Biochemistry*, **47**, 1490-1499.

48. Conn, G.L., Draper, D.E., Lattman, E.E. and Gittis, A.G. (1999) Crystal structure of a conserved ribosomal protein-RNA complex. *Science*, **284**, 1171-1174.

49. Wimberly, B.T., Guyman, R., Mccutcheon, J.P., White, S.W. and Ramakrishnan, V. (1999) A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell*, **97**, 491-502.

50. Lu, C., Smith, A.M., Fuchs, R.T., Ding, F., Rajashankar, K., Henkin, T.M. and Ke, A. (2008) Crystal structures of the SAM-III/SMK riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat Struct Mol Biol*, **15**, 1076-1083.

51. Draper, D.E., Grilley, D. and Soto, A.M. (2005) Ions and RNA folding. *Annual Review of Biophysics & Biomolecular Structure*, **34**, 221-243.

52. Jonikas, M.A., Radmer, R.J., Laederach, A., Das, R., Pearlman, S., Herschlag, D. and Altman, R.B. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189-199.

53. Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O. and Shaw, D.E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, **19**, 120-127.

54. Wickiser, J.K., Winkler, W.C., Breaker, R.R. and Crothers, D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell*, **18**, 49 - 60.

55. Wong, T.N., Sosnick, T.R. and Pan, T. (2007) Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci*, **104**, 17995-18000.

56. Greenleaf, W.J., Frieda, K.L., Foster, D.A.N., Woodside, M.T. and Block, S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, **319**, 630-633.

57. Lang, K., Rieder, R. and Micura, R. (2007) Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res.*, **35**, 5370–5378.

58.   Shcherbakova, I. and Brenowitz, M. (2008) Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting. *Nat. Protocols*, **3**, 288-302.

59.   Murchie, A.I.H., Davis, B., Isel, C., Afshar, M., Drysdale, M.J., Bower, J., Potter, A.J., Starkey, I.D., Swarbrick, T.M., Mirza, S. *et al.* (2004) Structure-Based Drug Design targeting an inactive RNA Conformation: Exploiting the flexibility of HIV-1 TAR RNA. *Journal of molecular biology*, **336**, 625-638.

60.   Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.

61.   Xin, Y., Laing, C., Leontis, N.B. and Schlick, T. (2008) Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*, **14**, 2465-2477.

Figures

Figure 1 (a) Schematic of the regulatory mechanism of the SAM-I riboswitch. This model has been previously described(11-13,15). In the presence of SAM, pairing of non-adjacent regions (highlighted in blue) forms the P1 helix of the anti-antiterminator (AAT) structure, which prevents formation of the antiterminator (AT). Under these conditions, the ρ independent terminator (T) hairpin can form to achieve transcriptional attenuation. (b) Three-dimensional crystallographic structure of the SAM-I riboswitch from T. *tencongensis*(35). A closeup is shown of the SAM binding pocket, highlighting the positions of two of the residues that are observed to contact the RNA. (c) Secondary structure representation of tertiary interactions observed in the crystallographic (Residues in red are involved in formation of the SAM binding pocket. The direction of the arrow is from 5' to 3'. The base pairs are automatically identified using RNAVIEW(60). The structural motifs are annotated as ref.(61)).

Figure 2 Time evolution of RMSD of the binding pocket and the ligand SAM relative to X-ray coordinates (35). In the presence of SAM (SAM_TRAJ, red), the RMSD is stable, while in the absence of SAM (WoSAM_TRAJ, blue), the RMSD increases until approximately 100 ns.

Figure 3. Clustering and Principal Component Analysis (PCA) point towards a chopstick-like motion involving P1 and P3 helices in the absence of SAM.
(a) Projections of snapshots of the WOSAM-TRAJ are plotted against the first two principal components and color coded according to a k=3 k-means clustering: cluster 0: magenta, cluster 1, green, cluster 2, magenta. Representative snapshots from each cluster are also shown. This plot indicates that snapshots can be broadly clustered into two groups (cluster 0 and cluster 2) with cluster 1 representing a group with characteristics similar to those of cluster 2. The projection along PC1 broadly separates the clusters, while projection along PC2 completes the separation between clusters 0 and 2. Structures of representative snapshots indicate that clusters are distinguished by a dramatic change in relative position of P1 and P3. (b) From top to bottom: The time evolution of the first principle component of WoSAM_TRAJ. (each snapshot is color coded as in Figure 3a). RMSD for each snapshot in WOSAM_TRAJ relative to the representative snapshots for cluster 0 (cyan curve) and for cluster 2 (magenta curve). The distance between the Center of Mass (COM) of P1 and P3 for WOSAM_TRAJ (blue) and for SAM_TRAJ (red). During the first half of WOSAM_TRAJ, P1 and P3 helices move apart (clusters 0 and 2), then they move back together during the second half of the trajectory (cluster 1).

Figure 4. A9/U63 dinucleotide stack state monitor and the correlation with glycosidic angles for the respective residues.
(a) Monitor of Lennard-Jones energy (E_{VDW}) of interaction between the A9 and U63 residues involved in a non-adjacent stack in SAM_TRAJ (red) and WoSAM_TRAJ (blue). Snapshots above the panel of the A9/U63 stacking geometry show that the intervals during which E_{VDW} is lowered correspond to the dinucleotide stacked state.; **(b)** monitor of the glycosidic angle of residues involved in the non-adjacent stack in all the simulations. Histograms showing occupancy of the stacking states and glycosidic angles for A9 and U63 are on the right side of each figure respectively. A biphasic distribution for the A9 glycosidic angle sis indicative of anharmonic motions.

Figure 5. Relaxation of the magnesium coordination with J1/2 in the absence of SAM and its coordination with the backbone of nucleotide A9 is correlated with the A9/U63 dinucleotide stack state (Figure 4). (a) Residues near the specific magnesium binding site; (b) distance monitor between the specific binding magnesium and atoms on the phosphate backbone (Red: SAM_TRAJ, blue: WoSAM_TRAJ). Coordination of the magnesium ion with phosphates in J1/2 is stabilized in SAM_TRAJ; (c) distance monitor between the specific binding magnesium and O3' in A9 in SAM_TRAJ. A slight shift is correlated with the non-adjacent dinucleotide stack (see text and Figure 4).

Figure 6. Electrostatic Potential Maps (ESP) highlighting two orthogonal slices through a Mg binding site. a) ESP map for the X-ray coordinates with SAM bound. SAM is highlighted in yellow. J1/2 is highlighted in orange, J3/4 in magenta and the position of the Mg is highlighted by a green sphere. Note that phosphate groups appear as red and some electropositive functional groups as blue. A large region of negative potential appears on all sides of the Mg ion, bordered tightly by J1/2 and J3/4. B) Closeups of ESP in the region around the Mg site for three representative snapshots of SAM_TRAJ (left) and WOSAM_TRAJ (right). The snapshots are chosen from the centroid of each of the three main clusters. In SAM_TRAJ snapshots, the hotspot of negative potential is tightly surrounded by J1/2 and J3/4, whereas it is concentrated near J3/4 only in the absence of SAM.

Figure 7. The distances between hydrogen bond donor and acceptor of hydrogen bonds between SAM and the SAM-I riboswitch RNA in MD trajectory (Nitrogen atoms are colored in blue, and oxygen atoms are colored in red). The contacts between SAM and G11 and G58 are extremely stable during the simulation. For the hydrogen bonding monitor, see SI Table 1.

(a)

(b)

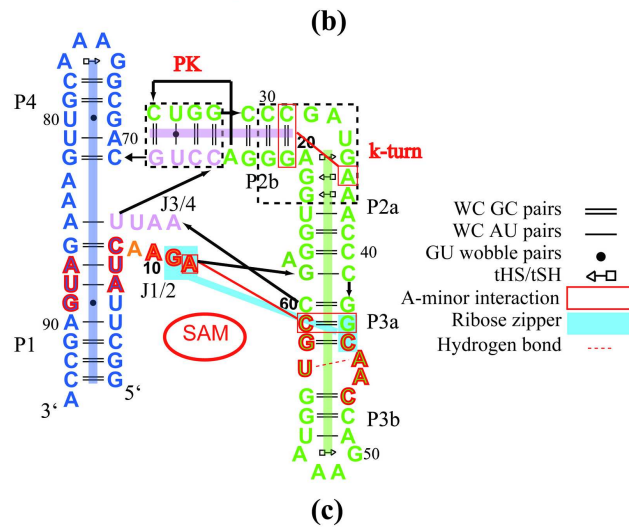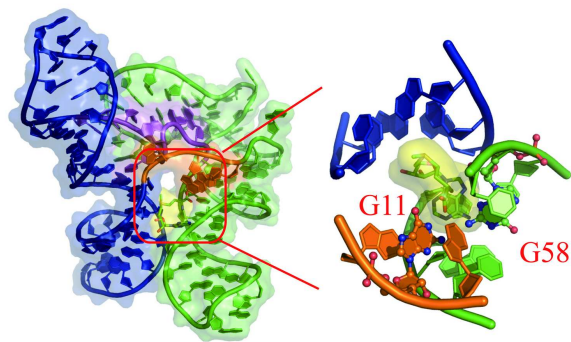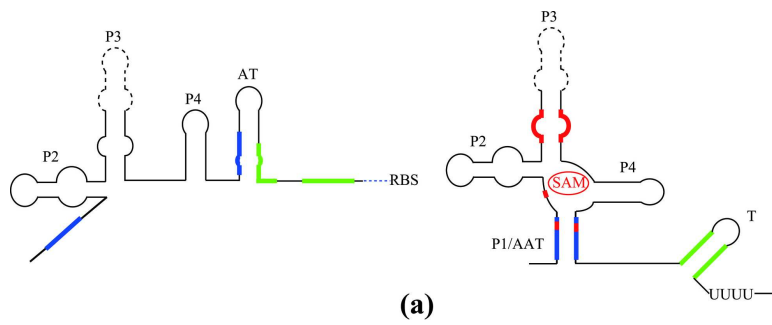(c)

Figure 1
83x130mm (600 x 600 DPI)

Figure 6a
83x76mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
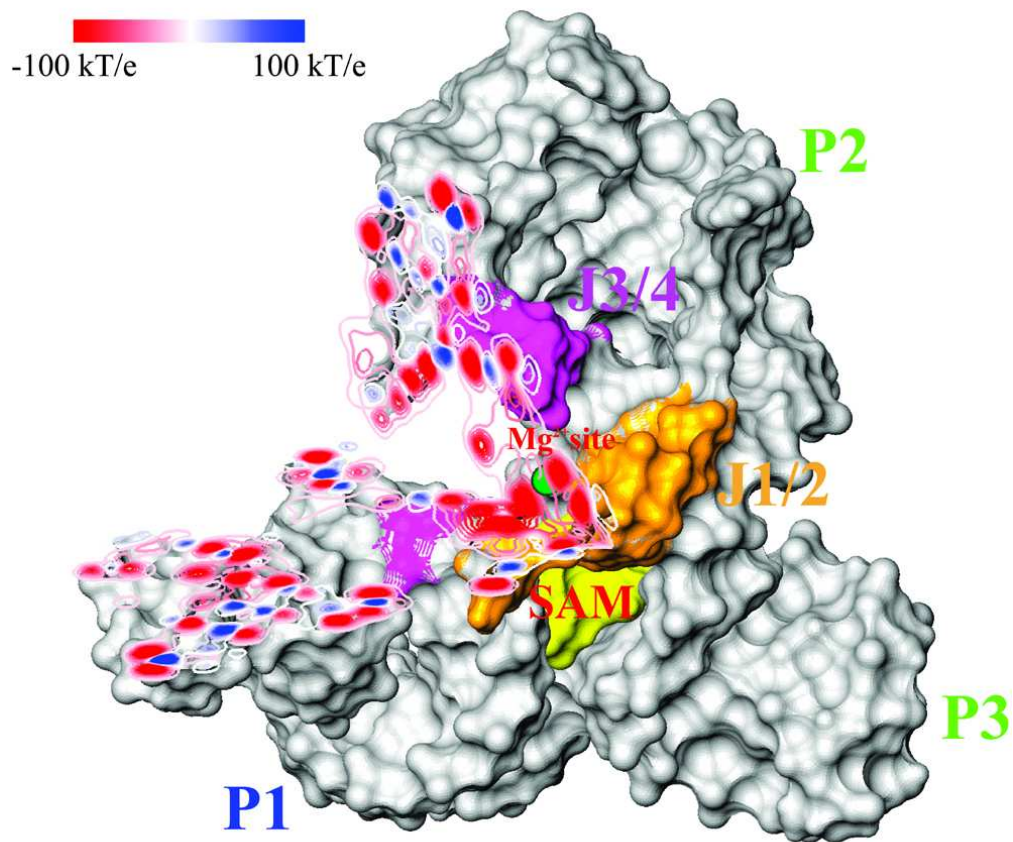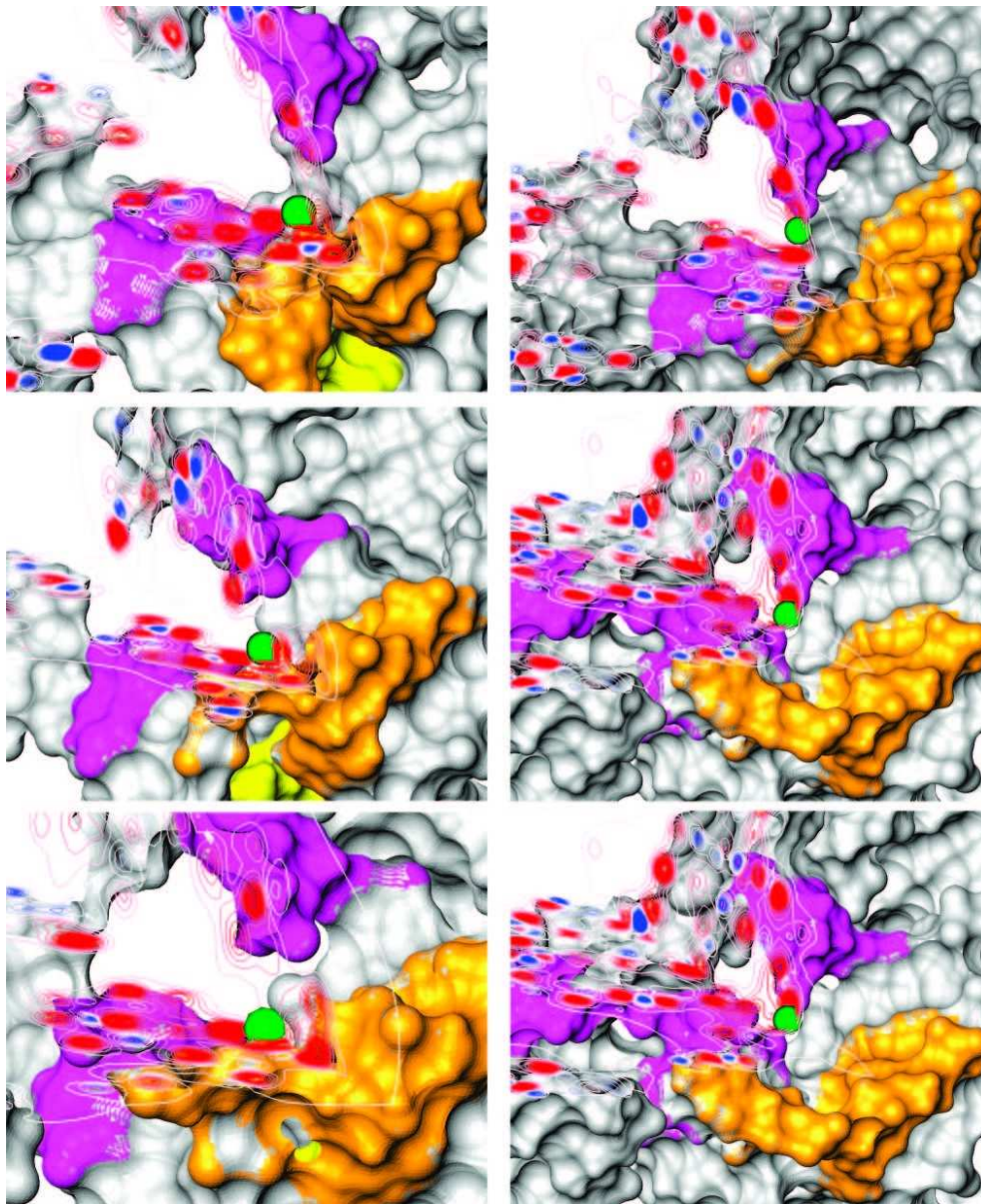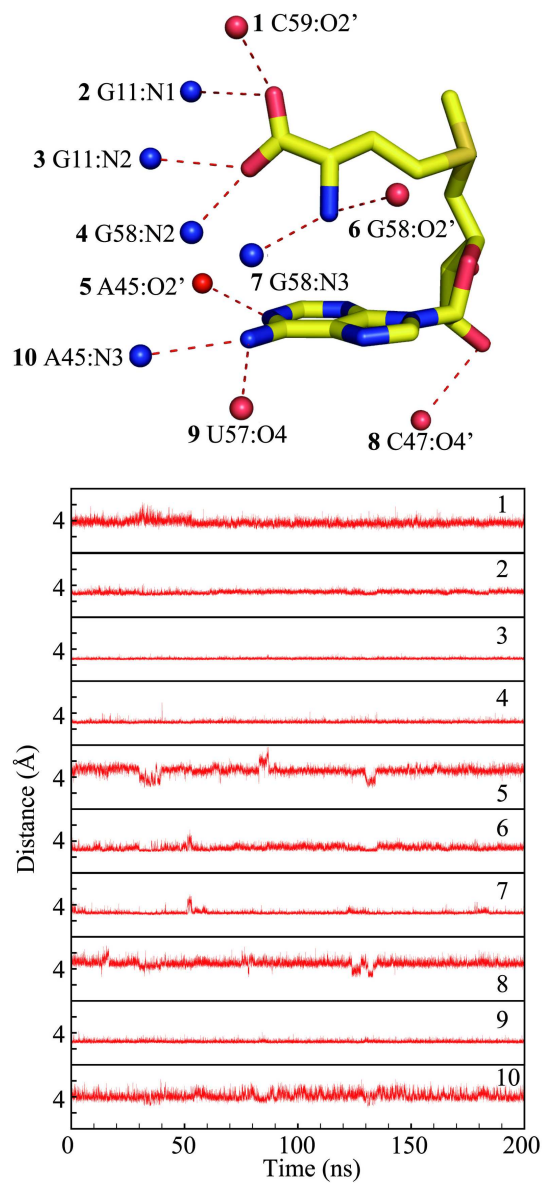49
50
51
52
53
54
55
56
57
58
59
60

Figure 6b
358x434mm (72 x 72 DPI)

Figure 7
70x150mm (600 x 600 DPI)