The Impact of a Ligand Binding on Strand Migration in the SAM-I Riboswitch

Wei Huang1,2, Joohyun Kim2, Shantenu Jha2,3, Fareed Aboul-ela1

1 Department of Biological Science, Louisiana State University, Baton Rouge, Louisiana, United States of America, 2 Center for Computation & Technology, Louisiana State University, Baton Rouge, Louisiana, United States of America, 3 Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey, United States of America

Abstract

Riboswitches sense cellular concentrations of small molecules and use this information to adjust synthesis rates of related metabolites. Riboswitches include an aptamer domain to detect the ligand and an expression platform to control gene expression. Previous structural studies of riboswitches largely focused on aptamers, truncating the expression domain to suppress conformational switching. To link ligand/aptamer binding to conformational switching, we constructed models of an S-adenosyl methionine (SAM)-I riboswitch RNA segment incorporating elements of the expression platform, allowing formation of an antiterminator (AT) helix. Using Anton, a computer specially developed for long timescale Molecular Dynamics (MD), we simulated an extended (three microseconds) MD trajectory with SAM bound to a modeled riboswitch RNA segment. Remarkably, we observed a strand migration, converting three base pairs from an antiterminator (AT) helix, characteristic of the transcription ON state, to a P1 helix, characteristic of the OFF state. This conformational switching towards the OFF state is observed only in the presence of SAM. Among seven extended trajectories with three starting structures, the presence of SAM enhances the trend towards the OFF state for two out of three starting structures tested. Our simulation provides a visual demonstration of how a small molecule (<500 MW) binding to a limited surface can trigger a large scale conformational rearrangement in a 40 kDa RNA by perturbing the Free Energy Landscape. Such a mechanism can explain minimal requirements for SAM binding and transcription termination for SAM-I riboswitches previously reported experimentally.

Introduction

Riboswitches reveal the versatility of Ribonucleic Acid (RNA) folding, and its remarkable biological impact. They are folded mRNAs that sense cellular metabolite levels and control expression of downstream genes [1–4]. Design of altered or novel riboswitches has been suggested for biomedical engineering applications [5–9]. Riboswitches also represent an important target for the design of novel antibacterials [10–12].

Riboswitches contain an aptamer, which recognizes and binds the metabolite. This binding triggers conformational rearrangement of the expression platform, which controls gene expression. Like other transcriptional riboswitches, the SAM-I riboswitch secondary structure is rearranged upon ligand binding [2,13,14]. The P1 and terminator (T) helices form in the ligand-bound state (Figure 1). This bound state is called the transcription OFF state since the terminator stops transcription. Without ligand the antiterminator (AT) helix forms, preventing formation of P1 and T helices, and allowing transcription (the transcription ON state).

SAM-I and other riboswitches raise the question–how can a small molecule binding to a limited contact surface cause a major folding rearrangement of a much larger RNA? Addressing this question requires consideration of conformational dynamics. X-ray studies of riboswitches have largely focused on the ligand-bound aptamer, truncating the expression domain to suppress conformational dynamics [15–18]. Such dynamic behavior is problematic for high resolution structure determination.

All-atom MD simulations are a major workhorse to tackle conformational dynamics [19–25]. Such methods have been applied to riboswitches, working largely with aptamer X-ray coordinates. These studies have revealed further insights into ligand recognition [26–30], the role of ions [25,31], and contrasted dynamic properties in the liganded and unliganded states [27,29,30,32,33]. Larger-scale, slower dynamic processes have required coarse-grained modeling or directed simulations using unbiased force fields [34,35].

Until recently, however, all-atom MD simulations using unbiased force fields have been generally limited to time scales...
Author Summary

Folding dynamics is crucial for RNA function. Riboswitches are a classic example. A typical riboswitch senses the cellular concentration of a small molecule. By refolding itself into a new structure, the riboswitch converts that information into changes in rates for synthesis of related metabolites. Understanding how the small molecule physically changes RNA structure can help us to target riboswitches, which occur mainly in bacteria, for drug design, or to engineer new riboswitches. This understanding has been blocked because 1) we cannot view intermediate stages experimentally and 2) simulations cannot reach the timescale for the structural conversion. Recent advances in RNA structure modeling enable us to model intermediate states. A new computer specialized for long timescale molecular dynamics (MD) simulations, called Anton, helps us to extend the simulation timescale. We modeled intermediate riboswitch structures, focusing on a reduced segment of the structure-switching region, in order to reduce the time required for a transition. We simulated an MD trajectory in which a small molecule converted the structure of this reduced switching region. Some steps in riboswitch structural transitions are therefore accessible to the newly extended MD timescales. Wider availability of resources like Anton can aid the advancement of riboswitch engineering and novel antibiotic design.

less than microseconds. The birth of a specialized machine designed for MD simulation-Anton [36,37] has increased the timescale limitation up to 200 times compared to simulations with conventional High Performance Computing (HPC) machines. Recent advances in software development and RNA structure modeling have improved the building of RNA models [38–48]. Together with enhanced sampling techniques [49–52] these modeling tools extend the accessible conformational space beyond that available within even the extended MD timescales.

Here we employ all-atom MD simulations to observe the direct effects of ligand binding on the equilibrium between alternative SAM-I riboswitch base pairing configurations. A large gap remains between the timescale required for strand migration and the flavors of the P1 helix propagation, or strand migration from an AT to a P1 helix (also termed the “switching” event [29] or “conformational collapse” [55]). We observed a strand migration event in the presence of SAM converting 3 AT helix base pairs (characteristic of the unbound riboswitch ON state) to competing P1 helix base pairs (characteristic of the OFF state). Overall, our simulations predict that SAM perturbs the reduced Free Energy Landscape (FEL) in a manner that favors conformations with expanded P1 helix base pairing and reduced AT pairing within the competing region, for certain starting geometries. Based on this simulation, we propose a mechanism for ligand-induced conformational switching which is consistent with reported requirements for SAM-I riboswitch function.

Results

Generation and structural analysis of models from MC-Sym

For SAM binding to fully convert an AT helix to a P1 helix may require at least milliseconds, judging from NMR measurements on an analogous strand-switching RNA [53], or longer based on a strand displacement assay [54]. We reasoned that the most rapid effect of SAM binding on the riboswitch would take place if the ligand bound to an intermediate conformation, hybridizing elements of the ON and the OFF state. Figure 2 shows a schematic of the strategy that we used to generate a starting configuration for our MD simulation. In the ON state, one strand of the P1 helix pairs with a downstream segment of the expression domain (removed in the crystallized RNAs) to form the AT helix. We initiated our simulation with a truncated segment fixing a partial P1 helix (two base pairs), and a partial AT helix (seven base pairs). In between a 4 nucleotide competition region can form either a P1 or AT helix. We call this “hybrid” construct 6P1_11AT, since it has the potential to form up to 6 P1 base pairs (characteristic of the ON state) to competing P1 helix base pairs (characteristic of the OFF state). Overall, our simulations predict that SAM perturbs the reduced Free Energy Landscape (FEL) in a manner that favors conformations with expanded P1 helix base pairing and reduced AT pairing within the competing region, for certain starting geometries. Based on this simulation, we propose a mechanism for ligand-induced conformational switching which is consistent with reported requirements for SAM-I riboswitch function.

Figure 1. Schematic illustration of regulation of transcription by the SAM-I riboswitch in response to SAM. (A) At low levels of SAM (left), the riboswitch forms an ensemble of secondary structures, most of which include the formation of the antiterminator (AT) helix. As SAM levels increase, ligand binding induces the formation of the structure shown at the right. SAM is bound to the four helix junction in which the P1 helix prevents formation of the AT by sequestering the “switching strand” (highlighted in red). The formation of the rho-independent terminator hairpin (T) then terminates transcription. The region incorporating the AT/T formation is termed the “expression platform” as it controls gene expression, while the four helix junction is termed the “aptamer”. (B) Closeup of the competing P1, AT, and T helices showing the explicit base sequences.

doi:10.1371/journal.pcbi.1003069.g001
pairs and up to 11 AT base pairs. Our simulations start with three of the four switching base pairs as AT helix, and a boundary nucleotide residue (U110) is positioned equally close to its putative AT or P1 binding partners.

This choice of starting structure allowed us to 1) Work with a segment that was shown experimentally to bind SAM, 2) Include the minimal nucleated P1 helix known to bind SAM, and 3) Maximize the potential number of AT base pairs with the potential to switch to P1 pairing (see “Details of MD simulations” in Supplementary Information (text S1)).

We used MC-Sym [40,56] to sample the placement of the AT helix in the 3D structures and the geometry of the boundary region with the nucleated P1. Previously we showed experimentally that SAM binds to hybrid constructs [57]. Though reduced in affinity, the SAM binding to the hybrids has similar dependence that SAM binds to hybrid constructs [57]. Though reduced in affinity, the SAM binding to the hybrids has similar sensitivity to mutations as with the aptamer. Therefore we assumed that the folding of the portion of the SAM/hybrid riboswitch complex outside of the strand switching region, henceforth referred to as the “aptamer core”, is similar to that in the X-ray structure of the aptamer.

Since the AT helix approximates a canonical A form geometry, the critical local region to be sampled is the three nucleotide segment A109, U110 and A111. These three nucleotides act as a hinge to bridge the partial P1 helix and the nearly complete AT helix. Additionally, an explicit triplet constraint was applied on the three nucleotides highlighted in purple in Figure 2A and 2B (A4, U110 and A136). Two adenosines compete for base pairing with a U (Figure 2B). The scripts used to generate the models can be found in the SI.

An overview of the outcome from MC-Sym sampling is shown in Figure 2 using the pseudo-dihedral angle [58]. Monitoring of the pseudo-dihedral angle (Figure 2C) indicates that MC-Sym has focused on the populated geometries according to the known structures, but also has sampled exhaustively the full range of geometries (Figure 2D). There is a region (between 80 and 170 degrees) that is rarely sampled due to steric clash with the P3 helix coordinates (Figure 2D). Therefore, the results demonstrate that MC-Sym can sample a wide range of the conformational space, while placing the AT helix without steric clashes.

Selection of starting models for MD simulation

Three criteria were used for selecting MC-Sym generated models for MD simulations: 1) Calculated potential energy should be favorable, 2) The SAM binding pocket must be accessible and 3) Coaxial stacking should be present between the P1 and AT helices. The latter constraint was based on experimental observations that SAM binding at pM affinity was detected for RNA constructs which allow the potential for such stacking (“3P1_10AT”), but not for those which do not (“3P1_9AT”) [57]. For reasons explained in supplementary information, we used the Amber99bsc0 force field with the generalized Born (GB) implicit solvent model to calculate free energy. Two (model 51 and model 55) out of the top five ranked in terms of free energy satisfied all the three criteria. Figure S1A shows calculated free energies, while Figure S1B highlights the coaxial stacking for these two models as measured by internucleotide vdW energies. The local geometry of the switching region is displayed schematically for these two structures in Figure 3A and global folds are shown in Figure 3B. The main difference between these two models is that the unpaired 5’ strand of the P1 helix is placed in the two different grooves of the AT helix—in the minor groove of the AT helix for model 51, and in the major groove for model 55 (Figure 3A).

The geometry sampled in model 51 and 55 resembles an RNA triple helix composed of poly(U)-poly(A)-poly(U) from a crystal structure [59]. With limited experimental data, these two models are rationalized as potential models for the intermediate or “transition state” between ON and OFF state.

Strand migration is observed in model 51 with SAM present

Table 1 lists MD trajectories included in this study, using model 51 and 55 and the X-ray coordinates (3NPB) [60] as starting models. Different trajectory evolutions are observed for
model 51 with or without SAM. Strikingly, formation of a complete P1 helix (all 6 Watson-Crick base pairs) is observed at ~1.3 μs for the simulation in the presence of SAM (see Movie S1). Figure 4A displays the time evolution of RMSD for individual base pairs with reference to that in the P1 helix of the X-ray structure. A small RMSD value (deep blue) indicates that the geometry of the nucleobases in a single base pair is close to that observed for the Watson-Crick base pair in the crystal structure. Monitors of classical Watson-Crick hydrogen bonding presence for the base pairs in the P1 helix and the AT helix are presented in Figure S2.

As is apparent from Figure 4A and Figure S2, the time at which the P1 helix was completely formed can be located as indicated with the red arrow in Figure 4A. The lifetime of this conformation spans from frame 6544 to frame 6635 (18.2 ns). The top 4 base pairs in the P1 helix (base pair 1 to 4) maintain the P1-like conformation corresponding to the crystal structure during the remaining simulation in the presence of SAM. Additionally, the electrostatic interactions between the sulfur atom of SAM and the carbonyl oxygen atoms of two U nucleotide residues persist throughout the simulation (Figure 4B) as observed in the repeated simulation on the aptamer domain of the yitJ SAM-I riboswitch (3NPB in the presence of SAM in Table 1).

The short life span of the fully formed P1 helix is linked to fraying of the closing base pair (base pair 6). The two participating nucleotides flip to a cross-strand stacking conformation. The adjacent base pair (base pair 5) is disrupted shortly after the loss of the closing base pair, but reappears at 1.9 μs for 300 ns (altogether 2250 out of 9097 snapshots after the strand migration event display this base pair). The two bases remain proximal (Figure 4C), however, and flip between states involving alternative hydrogen bonding patterns (Movie S1, Figure S3).

A similar plot for the AT helix pairs shows that the destabilization of AT base pairs 1 to 3 precedes complete P1 formation (Figure 4A, Figure S2). The disruption of this AT region is not due to the deficiency in modeling the AT helix since the simulation on the same model without SAM maintains the geometry close to a standard A-form helix for 2 (base pair 2 and 3) out of these three base pairs. Moreover, the 2 base-pair starting partial P1 helix is unstable in the absence of SAM in model 51 (Figure 4A, Figure S2).

During the interval leading up to the strand migration event, P1 helix base pairs 4–6 show a slowly rising trend in RMSD relative to the X-ray coordinates (Figure 4A, Figure S3). Thus the strand migration event is preceded by a fluctuation in which the corresponding nucleotide residues explore a “transition state”. This fluctuation coincides with the loss of AT helix base pairs 2 and 3, which otherwise block P1 helix propagation through P1 base pairs 5 and 6. The RMSD for P1 helix base pairs 3–6 goes down, in some cases dramatically, at the time of the strand migration. The two terminal base pairs drift towards configurations which show only a slightly smaller RMSD relative to X-ray coordinates. Monitors of classical Watson-Crick hydrogen bonding presence for the base pairs in the P1 helix and the AT helix are presented in Figure S2.

As is apparent from Figure 4A and Figure S2, the time at which the P1 helix was completely formed can be located as indicated with the red arrow in Figure 4A. The lifetime of this conformation spans from frame 6544 to frame 6635 (18.2 ns). The top 4 base pairs in the P1 helix (base pair 1 to 4) maintain the P1-like conformation corresponding to the crystal structure during the remaining simulation in the presence of SAM. Additionally, the electrostatic interactions between the sulfur atom of SAM and the carbonyl oxygen atoms of two U nucleotide residues persist throughout the simulation (Figure 4B) as observed in the repeated simulation on the aptamer domain of the yitJ SAM-I riboswitch (3NPB in the presence of SAM in Table 1).

The short life span of the fully formed P1 helix is linked to fraying of the closing base pair (base pair 6). The two participating nucleotides flip to a cross-strand stacking conformation. The adjacent base pair (base pair 5) is disrupted shortly after the loss of the closing base pair, but reappears at 1.9 μs for 300 ns (altogether 2250 out of 9097 snapshots after the strand migration event display this base pair). The two bases remain proximal (Figure 4C), however, and flip between states involving alternative hydrogen bonding patterns (Movie S1, Figure S3).

A similar plot for the AT helix pairs shows that the destabilization of AT base pairs 1 to 3 precedes complete P1 formation (Figure 4A, Figure S2). The disruption of this AT region is not due to the deficiency in modeling the AT helix since the simulation on the same model without SAM maintains the geometry close to a standard A-form helix for 2 (base pair 2 and 3) out of these three base pairs. Moreover, the 2 base-pair starting partial P1 helix is unstable in the absence of SAM in model 51 (Figure 4A, Figure S2).

During the interval leading up to the strand migration event, P1 helix base pairs 4–6 show a slowly rising trend in RMSD relative to the X-ray coordinates (Figure 4A, Figure S3). Thus the strand migration event is preceded by a fluctuation in which the corresponding nucleotide residues explore a “transition state”. This fluctuation coincides with the loss of AT helix base pairs 2 and 3, which otherwise block P1 helix propagation through P1 base pairs 5 and 6. The RMSD for P1 helix base pairs 3–6 goes down, in some cases dramatically, at the time of the strand migration. The two terminal base pairs drift towards configurations which show only a slightly smaller RMSD relative to X-ray coordinates. Monitors of classical Watson-Crick hydrogen bonding presence for the base pairs in the P1 helix and the AT helix are presented in Figure S2.
coordinates than at the start. The RMSD of backbone atoms relative to the X-ray coordinates, however, decreases and remains low after the strand migration event (Figure S3).

Strand migration is dependent on starting geometry

Complete P1 formation does not take place in model 55 within the time scale (1.467 μs) accessible so far for this simulation when SAM is present. Base pair 3 in the P1 helix gets trapped in a state with base pair geometry close to an AU Hoogsteen base pair (U•A cis W.C./Hoogsteen and class XXIII according to reference [61]) (Figure S4). In addition, this state is stabilized by a new hydrogen bond interaction between A4 and SAM, which is not sampled during the simulation of the aptamer domain (the construct for the X-ray study) in the presence of SAM (Figure S4B, C).

Analysis of trajectories in the presence and absence of SAM

In Figures 5 and 6 we visualize the conformational pathways observed for the various trajectories for model 51. The overall fraction of hydrogen bonds in Watson-Crick base pairs from the P1 helix and from the AT helix are used as generalized coordinates. Figure 5 displays the conformational trajectories for a simulation started from model 51 only, and a second simulation from model 51 in complex with SAM.

The results suggest that starting model 51 locates at a branch point in the FEL. The formation of a stable AT helix (high probability for AT Helix Hydrogen Bonding-vertical axis) is favored in the absence of SAM, while the presence of SAM allows model 51 to navigate to other transient states and eventually leads to sampling of the conformation with a complete P1 (high
Figure 5. Visualization for simulations on model 51 over reduced conformational space representation. Two simulations include trajectory starting from model 51 in complex with SAM and trajectory starting from model 51 only. The red star indicates the position of the starting structure on the reduced conformational space. The black and orange arrows show some major transitions in temporal order for model 51 without SAM and with SAM respectively. The fractions of hydrogen bonds, represented by averaged hydrogen bond probability (see definition in Materials and Methods), in the P1 helix and in the AT helix are chosen as generalized reaction coordinates. The color scale is proportional to the logarithm of the population of snapshots on the corresponding reaction coordinates. The local switching regions from some representative snapshots are shown in cartoon representation with P1 in blue and AT in red. SAM is displayed in sphere model.

doi:10.1371/journal.pcbi.1003069.g005

Figure 6. Visualization for simulations over reduced conformational space representation for individual model 51 and model 55 trajectories. The original trajectories (Figures 3 and 5) are shown here on the reduced conformational space separately (top) with (left) and without (right) SAM (A–B). The trajectories with or without SAM for model 55 are shown in a similar representation immediately below (C, D). Trajectories were initiated from frame 6615 (E, F) of the original model 51 trajectory with SAM at which a full 6 base pair P1 helix was observed, or frame 9974 (G, H), which contained 5 of 6 possible P1 helix pairs. The full 6 base pair P1 helix as formed during the original simulation following the strand migration event at 1.3 microseconds proved stable over the course of simulations with and without SAM (E, F, respectively). Simulations starting with frame 9974 displayed evidence of conversion of at least one base pair from P1 to AT helix pairing without SAM (H), whereas at least 4 P1 helix base pairs persisted throughout the simulation in the presence of SAM (G). Lengths of the trajectories are summarized in Table 1.

doi:10.1371/journal.pcbi.1003069.g006
The role of SAM in pre-positioning J1/2 for P1 helix stabilization

In SAM-I aptamer X-ray structures a Mg$^{2+}$ ion observed near the SAM binding site and phosphate moieties in J1/2 and J3/4 [60, 62–64]. We monitored the contact distances between this Mg$^{2+}$ and phosphates in J1/2 in the various trajectories of the SAM-I riboswitch aptamer and hybrid starting models with and without SAM. Our previous simulation on another aptamer of SAM-I riboswitch—metF from T. tengcongenesis [31] indicated cooperativity between this Mg$^{2+}$-phosphate coordination complex and SAM-leading to stabilization of tertiary interactions. Similarly, this effect was also observed in simulations of model 51 and 55 in the presence of SAM (Figure 7). The presence of SAM is correlated with the maintenance of short magnesium contacts with J1/2, while these contact distances increase during the simulations without SAM. Contact distances between Mg$^{2+}$ and phosphates in J3/4 are almost constant in the presence and absence of SAM. For the metF aptamer the correlation between the presence of SAM and short Mg$^{2+}$ contact distances with J1/2 is still maintained (Figure S5, but contact distances with phosphates on J3/4 begin to increase in the absence of SAM. Additionally, for restarted simulations of frame 6615 and 9974, the contacts of this Mg$^{2+}$ ion with J1/2 are still maintained even in the absence of SAM (Figure S6). Overall, these results confirm the stable coordination between the Mg$^{2+}$ ion and J3/4 in the absence of SAM, and the tendency of SAM contact to stabilize an additional coordination with J1/2.

Movies S2, S3, S4, S5, S6, S7 also highlight base moieties attached to nucleotide A7/9 in J1/2 and A80/82 in J3/4. Our earlier study also observed transient formation of a non-adjacent dinucleotide stack between nucleotide bases in J1/2 and J3/4 in simulations with and without SAM [31]. Of the ~12 X-ray SAM-I riboswitch coordinate sets [60, 62–64] all except one (pdb id 3GX3, with SAH bound) show the two nucleotide bases pointing to the same region outside the helix, with the respective bases within 3–7 angstroms proximity. In this study we again observed transient formation of dinucleotide stacking with and without SAM for model 51 and the aptamer, but with alternating stacking geometries (Movies S2, S3, S4, S5, S6, S7). Predominantly the two nucleotides were positioned with favorable stacking energies (Figure S7) but little effect was observed from SAM binding.

Discussion

A mechanism for ligand-induced conformational switching

Overall, we can summarize the results with model 51 MD trajectories as the following: 1) In the absence of SAM, 2–3 starting P1 helix base pairs appear to be unstable, whereas a long AT helix remains stable up to the terminal base pair; 2) In the

Figure 7. Monitor of Mg$^{2+}$ bridging interaction with SAM, and nucleotides on J1/2 and J3/4. A) Geometry of Mg$^{2+}$ binding site in SAM-binding region, involving coordination with electronegative groups on SAM, J1/2 and J3/4. B) Mg$^{2+}$ contact distances with the electronegative groups listed in the legend four trajectories with and without SAM using model 51 and model 55 as starting coordinates. Contacts with phosphates on nucleotides 81–83 in J3/4 are persistent in the presence or absence of SAM. Contacts with two of the electronegative groups on J1/2 (A8 P and G9 O6) become short and persistent when SAM is present only.

doi:10.1371/journal.pcbi.1003069.g007
presence of SAM, a strand migration event is observed after \( \sim 1.3 \mu s \) leading to transient formation of a full 6 base pair P1 helix, at the expense of competing AT helix base pairs; 3) Terminal base pairs within the fully formed P1 helix form transiently in the original simulation, but appear relatively stable in a new trajectory using the snapshot with fully-formed P1 helix as the starting point; 4) The fully formed P1 helix is also relatively stable in a trajectory which starts with the same snapshot even in the absence of SAM. 5) In a trajectory starting with 5 P1 base pairs with SAM, hydrogen bond contacts corresponding to the five base pairs appear slightly more stable than they do in one starting from the same snapshot without SAM.

By contrast, trajectories starting with model 55 result in a conformation in which the competing base pair at the boundary between P1 and AT base pairs forms a non-Watson-Crick pair, while other P1 and AT base pairs are stable. Taken together, these findings indicate that SAM binding promotes P1 helix base pairing at the expense of AT helix pairing, but with qualifications. Certain starting geometries, such as that in which the 5' nucleotides reside near the major groove of the AT helix (as in model 55), may be slow to convert to the P1 helix-forming conformation. Our original simulation of model 51 in the presence of SAM resulted in 4 stable P1 helix base pairs. Thus, SAM binding may have its strongest direct stabilization of P1 helix base pairs near the SAM binding site.

All of these results are consistent with experimental evidence. A minimum length of P1 helix is necessary for SAM binding [13,65], though the presence of a partial AT helix can restore \( \mu M \) SAM binding with a P1 helix as short as 2 base pairs [57]. The latter study indicated that SAM binding affinity increases in model systems as the P1 helix is extended and the AT helix shortened. There are also indications that P1 helix dynamics are reduced by SAM binding [17,54,65] for truncated aptamers.

**Mechanism for stabilization of P1 helix base pairing by SAM**

Earlier we proposed that SAM contacts with J1/2 and indirect stabilization of Mg\(^{2+}\) contacts with J1/2 enhance P1 helix formation [31], and that the contacts with J1/2 block formation of competing conformers [66]. Our simulation suggests that additional enhancement of P1 helix formation arises through direct contact with SAM. The importance of these electrostatic SAM-P1 helix contacts for mediating the ligand binding specificity has been established experimentally [63].

As observed in our earlier simulations [31], direct contacts between SAM and the key G11 nucleotide within the P1 helix are persistent throughout these extended timescale simulations. In addition, we observed shorter contact distances between a bound Mg\(^{2+}\) and at least two electronegative functional groups on J1/2 in the presence of SAM during the simulations starting with model 51 and with the aptamer, than in the absence of SAM (Figures S4, S5, S6). The Mg\(^{2+}\) ion site which we have monitored here, observed in the original X-ray structures, is suspected to form an inner sphere coordination complex [31,67].

Movies shown in supplementary materials (Movies S2, S3, S4, S5, S6, S7) vividly illustrate the interplay between SAM, Mg\(^{2+}\), and the backbones of the J1/2 and J3/4 junctions. Movies without SAM show the Mg\(^{2+}\) surrounded by phosphates from J3/4, with particularly stable coordination with phosphates 81 and 83 (83 and 85 in the aptamer). In the presence of SAM, G9/11 O6 is anchored in a bridging position between the Mg\(^{2+}\) and phosphate groups in J1/2. A recent study identified a cooperative effect between Mg\(^{2+}\) and SAM in SAM-I riboswitch folding, and proposed a role for the same core Mg\(^{2+}\) in pre-organizing folding intermediates for SAM binding [55]. Movies S2, S3, S4, S5, S6, S7 provide a striking illustration of a potential mechanism to explain this cooperativity. This coordination complex could induce a reorientation of the P1 helix, as reported [65], by fixing the position of J1/2.

Favorable non-adjacent dinucleotide stacking between nucleotide bases in J1/2 and J3/4 is observed in model 51 with SAM, but with altered geometry in the absence of SAM (Movies S2 and S3). Altogether, these observations leave an open question as to the role that non-adjacent dinucleotide stacking may play in pre-positioning J1/2 and J3/4 in a manner that is favorable to aptamer formation and P1 helix formation specifically.

**Implications of restarted MD simulations**

The simulation of model 51 in this study shows that the stabilization of the partial P1 helix by SAM anchors the 5' strand of the P1 helix in an orientation that enables this single strand region to compete over the AT helix. This model is reminiscent of an NMR study on a small RNA system showing that the stabilization of a pre-formed helical region by a tetra loop increases the rate of conversion between two different hairpin folds [53]. In the riboswitch, SAM stabilization of the nucleated P1 helix may play a similar role.

Simulations on the same starting coordinates in the absence of SAM displayed the loss of all P1 helix base pairing. Conformations with three or fewer base pairs in the P1 helix may not be stable enough to prevent the formation of the AT helix in the absence of SAM. When a snapshot with fully formed P1 helix (frame 6615) was used as the starting structure an MD trajectory displayed a relatively stable P1 helix even in the absence of SAM. When frame 9974 with 5 P1 helix base pairs was used as the starting structure, all 5 P1 helix pairs remained stable in the presence of SAM. With these starting coordinates, however, as the simulation time approached 1 \( \mu s \), the P1 helix began to show some instability in the absence of SAM. Therefore, differing degrees of shift of the conformational equilibrium amongst a series of conformational intermediates toward the OFF state with SAM facilitate the SAM-I riboswitch function as a dimmer switch. The most dramatic SAM binding effect is on a hybrid conformer with minimal P1 helix base pairing.

**Implications for SAM-I riboswitch folding in active transcription complexes**

In the biological context, it is proposed that SAM binding takes place soon after the transcription of the aptamer-forming segment [68]. The conformation is then locked before full transcription of the antiterminator, a mechanism similar to that indicated for other transcriptional riboswitches [69,70]. Such a mechanism is highly sensitive to the concentrations of reaction components—a recent report indicated that nucleotide levels dramatically alter the degree of kinetic control of a lysine riboswitch within active transcription complexes [71]. For the \( \text{yitJ} \) SAM-I riboswitch, moreover, partial AT helix formation can take place in the non-overlapping region, even in the presence of a full P1 helix. Our simulations indicate that the presence of SAM would prevent strand invasion by this partial AT and dissociation of the P1 helix in this scenario.

High resolution structures of riboswitch aptamers with and without ligand have led to the proposal that many fold according to the “conformational capture” mechanism [72,73]. Typically this mechanism is described as selection by the ligand of a single bound conformer amongst a range of conformations being sampled by the unliganded substrate (Figure 8A). Inclusion of a portion of the expression domain, however, leads to more
dramatic effects of ligand on RNA folding. Secondary structure calculations predict that an equilibrium Boltzmann ensemble for the yitJ SAM-I riboswitch includes some hybrid conformations with partial P1 and partial AT helix [57,66]. A “capture” of these intermediates, according to simulations here, would facilitate rapid propagation of a longer P1 helix. This event would free a sufficient segment of the 3’ strand of the AT to nucleate the downstream Terminator sequence. By contrast, our simulations indicate that in the absence of SAM the AT helix could displace the nucleated P1 helix within the intermediates.

A more precise description of conformational capture in this scenario would be selection of a region of conformational space by the ligand, which then chaperones the RNA towards the aptamer configuration (Figure 8B). Conformations that can bind SAM have reduced free energy relative to those for which RNA and SAM are not in contact, and the reduction in free energy for each conformer is proportional to favorable free energy of binding. We hypothesize that the aptamer folding rate would be accelerated by this mechanism because during the Levinthal sampling process the FEL region that can initiate aptamer formation is widened.

Simulations of RNA folding kinetics [74] concluded that a strand migration pathway would lead to the fastest transition rate for an inter-conversion between two hairpins. Kinetic folding studies for a number of riboswitches [75–79] indicate that P1 helix formation takes place during later stages of the folding pathway. Our simulations and the experimental findings in our previous study [57] therefore raise the possibility of a role for SAM in accelerating P1 helix formation, by facilitating strand migration as the aptamer folding pathway. In this scenario, SAM binding could still facilitate aptamer formation after a portion of the expression domain has been transcribed. In vitro kinetics of SAM-I riboswitch

Figure 8. Schematic of conformational capture and FEL perturbation mechanisms for ligand-induced riboswitch conformational change. Each cartoon representation of the RNA molecule should be viewed as representing an ensemble of secondary structures with closely related topology in the P1/AT helix competition region. Conventional descriptions of “conformational capture” (A) assume sampling of numerous conformations, one of which is “selected” and fixed by ligand binding. If a subset of conformations is susceptible to reduced affinity ligand binding (B), SAM binding can lower the free energy of the partial P1 helix-forming families of conformers (third and fourth from the left in each panel). Note that in this situation, the ligand/aptamer complex can form rapidly through a pathway bypassing the non-P1 helix forming families of conformations (the two on the far left). Such a mechanism may be hypothesized for SAM-induced P1 helix formation in vitro or even during active transcription.

doi:10.1371/journal.pcbi.1003069.g008
folding and transcription termination would then be highly sensitive to mutations in the expression domain, as has been reported for a transcriptional lysine riboswitch [71]. SAM effects on P1 vs. AT length could be tested through NMR measurements on partially labeled SAM-I riboswitch hybrid constructs, or through NMR methods designed to detect minor conformers [80,81].

### Approaching the switching event using MD simulations

In previous work we showed that altered base pairing in the unliganded yifJ SAM-I riboswitch as compared to the bound state extends beyond the P1/AT helix switch [57,66]. The question of the impact of the ligand on the riboswitch conformation is therefore related to large-scale alterations in RNA folding. Although great advancement has been achieved to speedup MD simulations, a complete simulation of folding/unfolding for RNA of this size (~40 kDa) is still not possible. The P1/AT helix switching event alone may take ms or longer, according to data on analogous model systems [53]. This is because the nucleation of the transient state takes up most of the folding time. The propagation step can be faster than the overall folding rate by four orders of magnitude [82]. Therefore, MC-Sym was used to sample a discrete conformational space aiming to identify candidate transient state models with atomic details to bypass the most time-consuming part of the simulation.

In the SI, we discuss measurements from the literature for folding and conformational switching for a range of RNAs. Overall, considering the literature data and estimating a rate constant based upon snapshots observed in the model 51 trajectory as transition states (Tables S1 and S2), the microsecond regime appears plausible for the strand migration within the three base pair stretch simulated in this study. Extension of the MD timescale to the microsecond regime by using Anton now appears to make some intermediate steps of strand migration accessible. The adequacy of force fields and parameters for long timescale simulations for RNA is relatively untested as compared to protein MD [83]. Nonetheless, it seems improbable that a strand exchange observed only in the presence of ligand, and leading to decreased RMSD relative to X-ray coordinates (Figure 4, Figure S3), is solely a result of instabilities or imperfections of the force fields [34]. The two terminal base pairs of the P1 helix are transient in the original model 51 simulation with SAM, although the RMSD relative to the X-ray structure for each base pair remains lower than before the strand invasion. This observation may reflect instabilities in the force field, a genuine tendency towards “fraying” [85,86], or a longer simulation may be required to reach a thermodynamically stable state. The success of this study in observing a strand migration event should motivate efforts to optimize and validate parameters and protocols for long timescale MD simulations for RNA.

### Methods

#### RNA modeling using MC-Sym

The atomic models for the RNA construct described in Figures 2 and 3 were generated using MC-Sym [40] installed locally. The aptamer core (highlighted in blue in the figure) was modeled using its counterpart in the known structure of the yifJ SAM-I riboswitch (PDB ID: 3NPB) [60]. The other parts of the construct were built from the library of small fragment RNA structures, known as Nucleotide Cyclic Motifs (NCMs) [56]. An explicit triplet constraint was applied on the three nucleotides highlighted in purple (A4, U110 and A136). In this way we sampled the 3D space in which these three nucleotides are proximal to each other. In these three nucleotides, the two As are competing for base pairing with a U. The scripts used to generate the models can be found in Appendix S2. Different RMSD threshold values were tested to ensure exhaustive sampling in the local region bridging the partial P1 and the AT helix (A109, U110 and A111). Models with small differences (low pairwise RMSD) in pseudo-diHedral angle of the A109-A111 region were filtered out. Energy minimizations (max step is 2000 or gradient tolerance <1.0) were performed on the atomic structures of the models generated from MC-Sym runs using Nucleic Acid Builder [NAB] [87]. AMBER99sc0 force field [88] and Generalized Born model [89] with an inverse Debye-Huckel length of 0.19 Å⁻¹ [90] were used in the energy minimization procedure. 149 models were generated in this step. This energy minimization is mainly to rebuild the chain connectivity for models generated from MC-Sym without introducing the sampling effect of the force field. Thus, we used MC-Sym to sample the possible placement of the AT helix in the 3D structures and the geometry of the potential nucleation site of the P1 helix close to the SAM binding pocket. The modeling assumed that the folding of the aptamer core is similar to that in the crystal structure of the aptamer domain. After the energy minimization step, models with high van der Waals energy were filtered out. There are two reasons for high van der Waals energy: 1) steric clashes that cannot be released by energy minimization, 2) broken chain connectivity that cannot be bridged during energy minimization.

#### All-atom MD simulation

Models were chosen following the three criteria listed in the results section under “Selection of starting models for MD simulation”. For the models in the presence of SAM, the ligand was placed in the binding pocket while maintaining most of the interactions (except the contacts with the end base pair AU in the partial P1 helix) observed in the crystal structure of the aptamer domain complex (PDB: 3NPB). The simulations are run on Anton [36]. The equilibrated structures for Anton were prepared using local HPC clusters following the MD protocol as described in our previous study [31] (also see “Details of MD simulations” in Text S1). The trajectory was recorded for every 200 ps.

#### Hydrogen bond probability

The definition of hydrogen bond probability (HBP) of the hydrogen bond i at time t is similar to that in reference [91]:

$$HBP_i(t) = \frac{1}{1 + e^{\Delta E_i(t) / T}}$$  \hspace{1cm} \text{(1)}$$

where \(\Delta E_i(t) = E[r(t), \theta(t)] - E[r_0, \theta_0]\), and \(E[r, \theta]\) is defined as

$$E[r, \theta] = - \cos \theta / r$$  \hspace{1cm} \text{\text{(2)}}$$

Here \(r\) is the distance between hydrogen and hydrogen bond acceptor, \(\theta\) is the angle of hydrogen bond donor, hydrogen and hydrogen bond acceptor and scaling constant \(\epsilon = 100\ \text{Å}^2\). In the reference state \(r_0 = 3.2\ \text{Å}\) and \(\theta_0 = \pi/3\ \text{rad}\). The list of hydrogen bonds monitored is listed in Table S3.

### Supporting Information

#### Appendix S1

Predicting the time scale for the propagation step in strand switching.

(DOCX)
Figure S1  (A) Plot of the free energy rank with amber99 bsc0 and GB solvent model. (B) Scatter plot of vdW_A109-U110 versus vdW_U110-A111 clustered into 4 classes. Model 51 and Model 55 both fall within Cluster 2 (blue) indicating favourable VDW interactions through the A109-A11 junction region.

Figure S2  Hydrogen bond analysis of model 51 simulations. Monitor of hydrogen bonds in Watson-Crick base pairs. Here hydrogen bonds are defined with H-bond distance cutoff (<3.5 A) and H-bond angle cutoff (>145°). The order of hydrogen bonds in the P1 helix is as in Table S3.

Figure S3  Clustering analysis for base pair 3–6 in the P1 helix for the model 51 trajectory with SAM. Each base pair was grouped into five states using k-means (k=5) clustering based on the RMSD of nucleobases. (A) State assignments as a function of time for each base pair. The state assigned for each time point is indicated as an open circle. (B) Representative structure for each state. The centroid structure of each state was chosen to represent the state. The population of each state was labeled at the top of individual state. The RMSD for the base pair relative to X-ray coordinates is plotted against the vertical axis on the right. (C) The RMSD for the backbone for P1 helix base pairs 3–6 is plotted along with the same parameter for the trajectory of model 51 without SAM.

Figure S4  Monitor of some local geometries in model 55 simulations. (A) Time evolution of RMSD for individual base pairs in the P1 and the AT helix from simulations on model 55 in the presence (left) and the absence (right) of SAM, monitored as for model 51 in Figure 3. In contrast to the model 51 trajectories, no migration of P1 or AT helix base pairing is observed beyond the initially unpaired hinge position. (B) Local view of the switching region in model 55. The RNA is shown in cartoon representation except A4 and U100 in stick representation. The ligand SAM is also displayed in stick representation with carbon atoms in yellow. (C) Distance monitor during the simulations for model 55 and 3NPB with SAM for a AU Hoogsteen base pair (Top) and a new interaction between SAM and A4 sampled in model 55 (Middle). Monitor of the base stacking between A4 and A111 via vdW interaction energy (Bottom). The Hoogsteen base pairing scheme is shown on the right.

Figure S5  Plots of distance monitor for simulations of the μ/P1 aptamer (PDB ID: 3NPB) in the presence and the absence of SAM.

Figure S6  Plots of distance monitor for restarted simulations of model 51 for frame 6615 and frame 9974 in the presence and the absence of SAM.

Figure S7  (A) Monitor of vdW stacking energy calculated for non-adjacent dinucleotide stacking between nucleotides 7 and 82 for simulations of model 51 with (red) and without (black) SAM. (B) A histogram of calculated vdW values for the two nucleotides for each simulation.

Movie S1  Visualization of trajectory for model 51, with SAM, highlighting strand migration event converting 3 base pairs from AT to P1 helix configuration.

Movie S2  Visualization of SAM, core Mg²⁺, electronegative moieties in J1/2 and J3/4, and the dinucleotide stack between nucleotides A7/9 in J1/2 and A82/84 in J3/4. The bases in the non-adjacent dinucleotide stack are highlighted as sticks. G9/11 O6 is highlighted as a purple or red sphere. The Mg²⁺ ion is depicted as a yellow sphere. Backbone phosphates are also highlighted.

Movie S3  Similar visualization as for S2 for model 51 without SAM

Movie S4  Similar visualization for trajectory involving model 55 with SAM.

Movie S5  Similar visualization for trajectory involving model 55 without SAM.

Movie S6  Similar visualization for trajectory involving the aptamer, starting from x-ray coordinates, with SAM.

Movie S7  Similar visualization for trajectory involving the aptamer, starting from x-ray coordinates, without SAM.

Text S1  Supplementary Text includes descriptions of: Analysis of MC-sym sampling using the pseudo-dihedral angle, Ranking and selection of MC-sym modeled structures for MD simulations, Supplementary Methods (Details of MD simulations), Captions for supplementary figures and movies.

Table S1  Calculated Free Energies (kcal/mol) using RNAeval (30) for formation of variant secondary structures of the riboswitch conformation shown in Figure 1

Table S2  Calculated Free Energies using RNAeval for putative secondary structures and transition states from the study by Wenter et al. 2006.

Table S3  List of hydrogen bonds included in calculations shown in Figures 5&6 and Figure S2.

Acknowledgments
We are most grateful for the opportunity to use Anton generously donated by David E. Shaw. We thank Dr. Markus Dittrich and other staff at PSC for their support in using Anton. W.H. thanks Dr. Le Yan for his help with using the computer clusters at LSU HPC. We thank Dr. Marc Parisien for critical reading of the manuscript. We are grateful for the support of this work with computing resources provided by the LSU Center for Computation and Technology (CCT).

Author Contributions
Conceived and designed the experiments: JK SJ FA. Performed the experiments: WH. Analyzed the data: WH JK. Contributed reagents/materials/analysis tools: WH JK SJ. Wrote the paper: WH FA.
