

Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine

Supplementary Information

Sunny D. Gilbert¹, Robert P. Rambo², Daria Van Tyne¹, and Robert T. Batey^{1*}

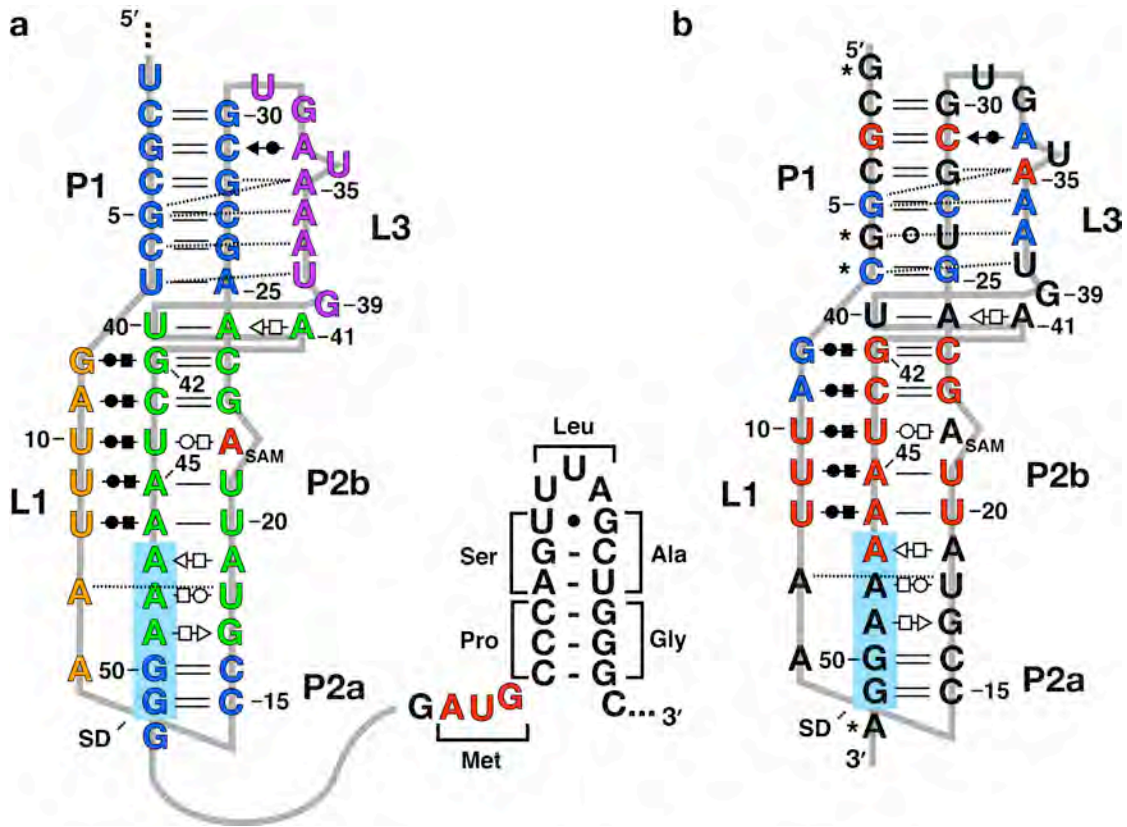
¹Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado, Boulder, CO 80309, USA; ²Life Science Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Bldg 6 Room 2105, Berkeley, CA 94720, USA

*Author to whom correspondence should be addressed

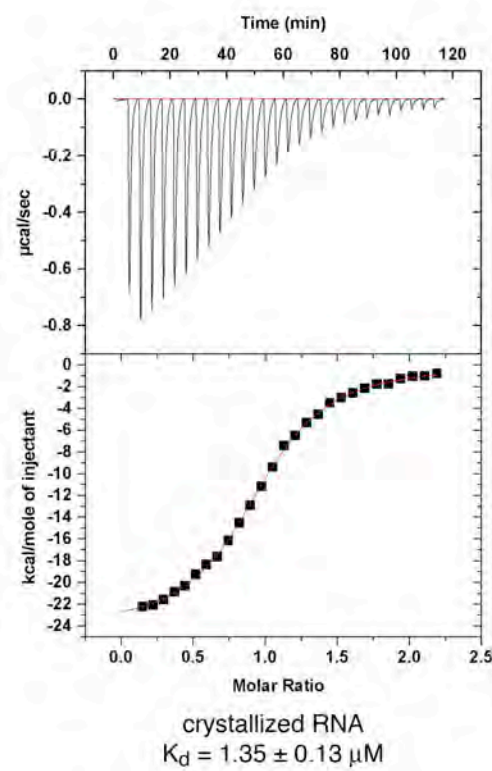
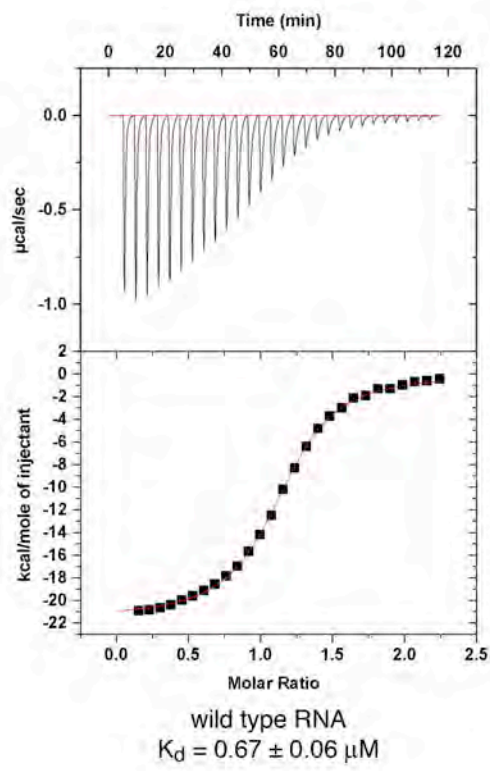
Telephone: (303) 735-2159

Fax: (303) 492-5894

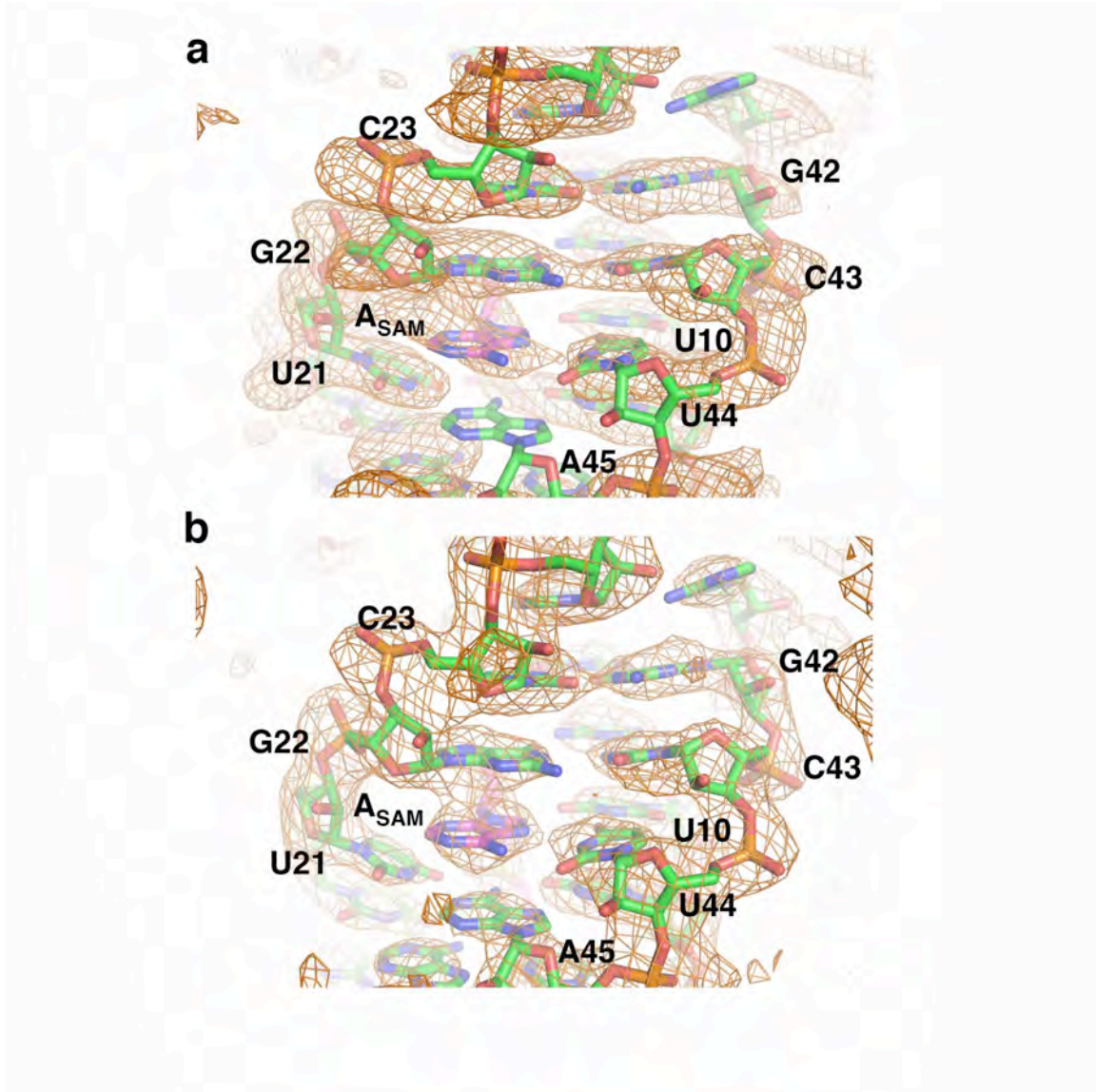
E-mail: robert.batey@colorado.edu



Supplementary Figure 1 Secondary structure representation of wildtype and crystallized RNAs. (a) Sequence of the wild type RNA and also the RNA used in the chemical probing experiments. A 21-nucleotide extension was added to the 3' end in order to ascertain the effect of this sequence on the unliganded RNA. Amino acids correspond to the first six residues of the putative *metX* gene, as identified by the high degree of homology to other identified *metX* genes. (b) Sequence and secondary structure of the crystallized SAM-binding mRNA pseudoknot upstream of the putative *metX* gene from a Sargasso Sea metagenomic environmental sequence¹. The nomenclature for the stems and loops (P1-P2 and L1-L3, respectively) is derived from standard naming of H-type pseudoknots. Red nucleotides are those whose identity is >95% conserved and blue corresponds to >80% conservation of identity. Starred residues indicate positions that were altered from the wild type sequence but retain their pattern of phylogenetic conservation (See **Supplementary Table 1**). Each change was made for a specific purpose: U1G for efficient transcription initiation by T7 polymerase, G52A for efficient 3'-end processing by a 3'-HδV ribozymes, and (C6G, U7C, A25G, G26U) to add a cesium/iridium ion binding site for phasing.



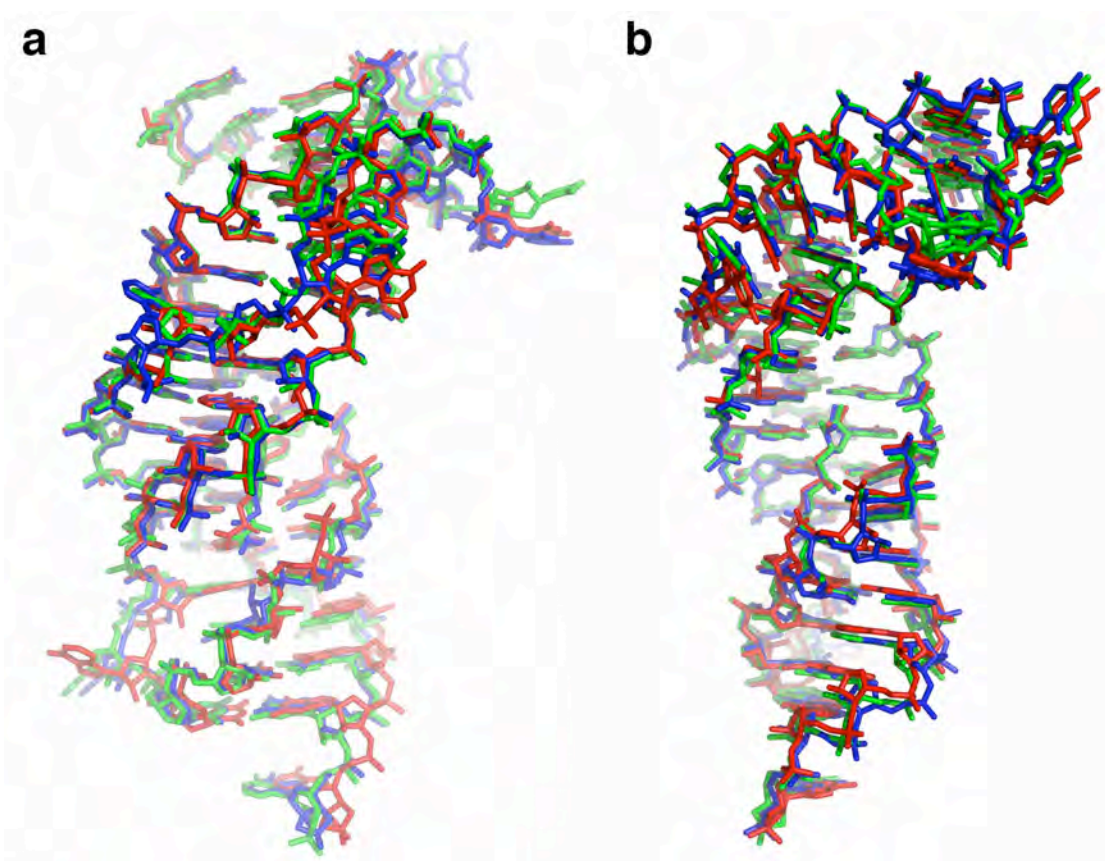
Supplementary Figure 2 SAM binding affinity of wild type (left) and crystallized (right) SAM-II RNA determined by Isothermal Titration Calorimetry (ITC).



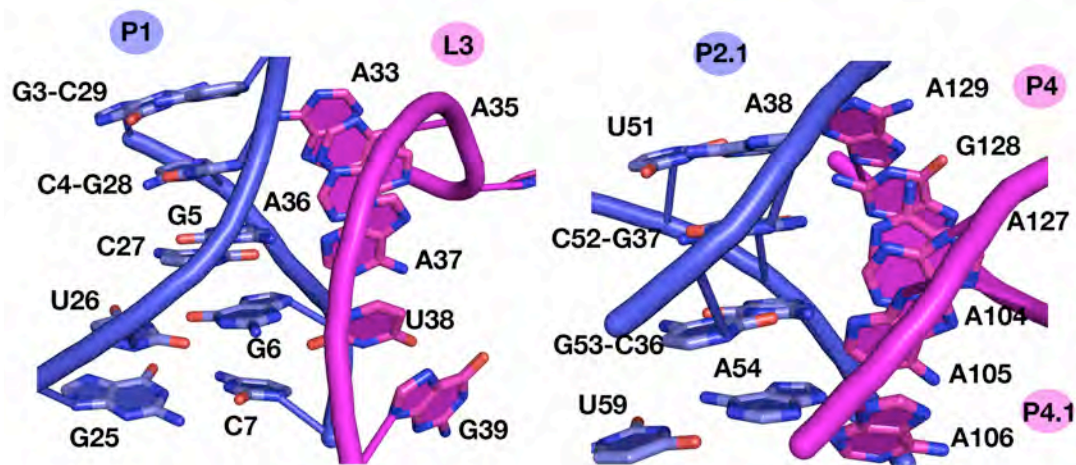
Supplementary Figure 3 Electron density maps around the SAM binding site protomer A of the SAM-II riboswitch contoured at 1σ (orange cage). The final model is superimposed upon the density (green, RNA; magenta, SAM). (a) Solvent flattened experimental electron density map. (b) Final $2F_o - F_c$ electron density map. (c, next page) Final $2F_o - F_c$ electron density map in wall-eyed stereo. Image is centered on the P1-P2b junction and shows the location of a cesium atom (purple) bound to an engineered phasing module in the RNA.

C

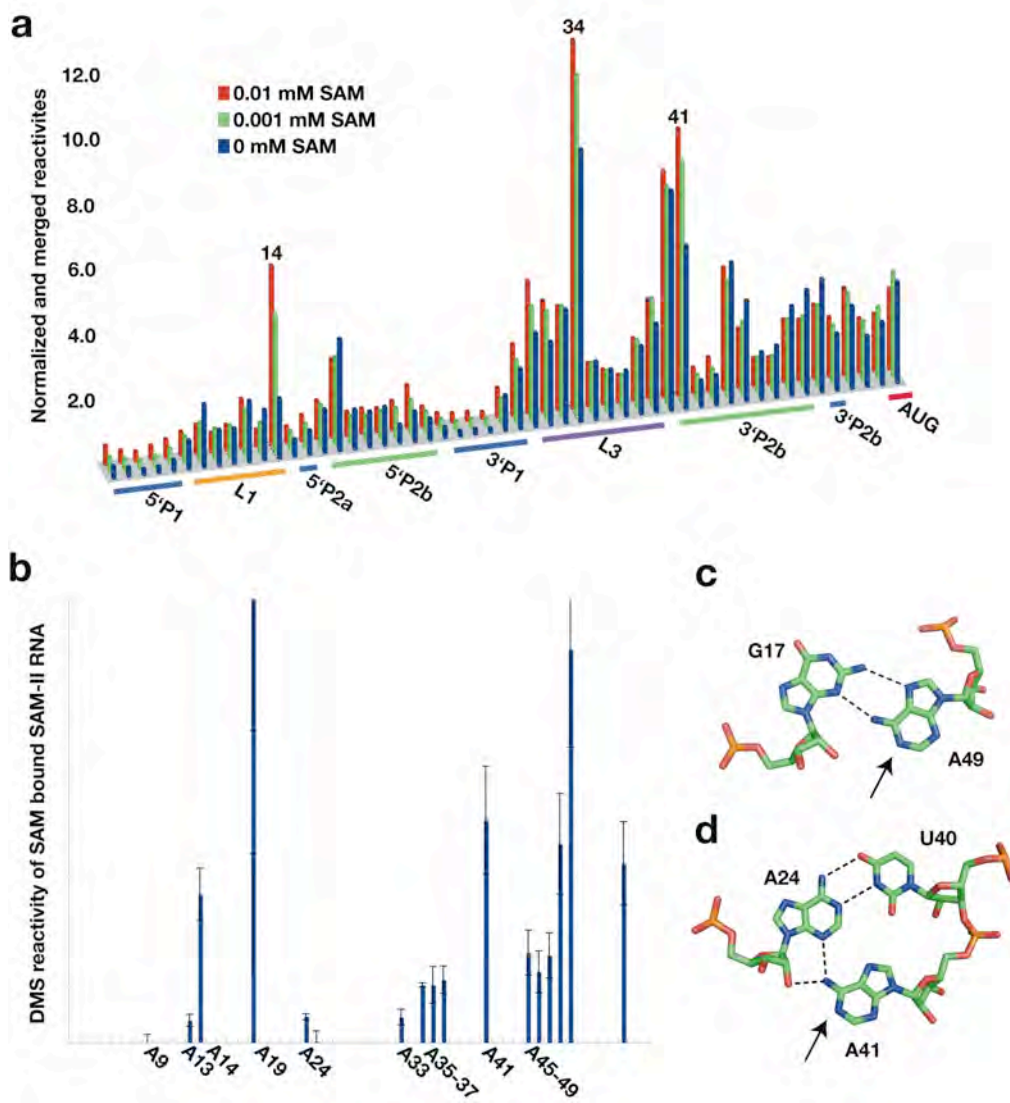




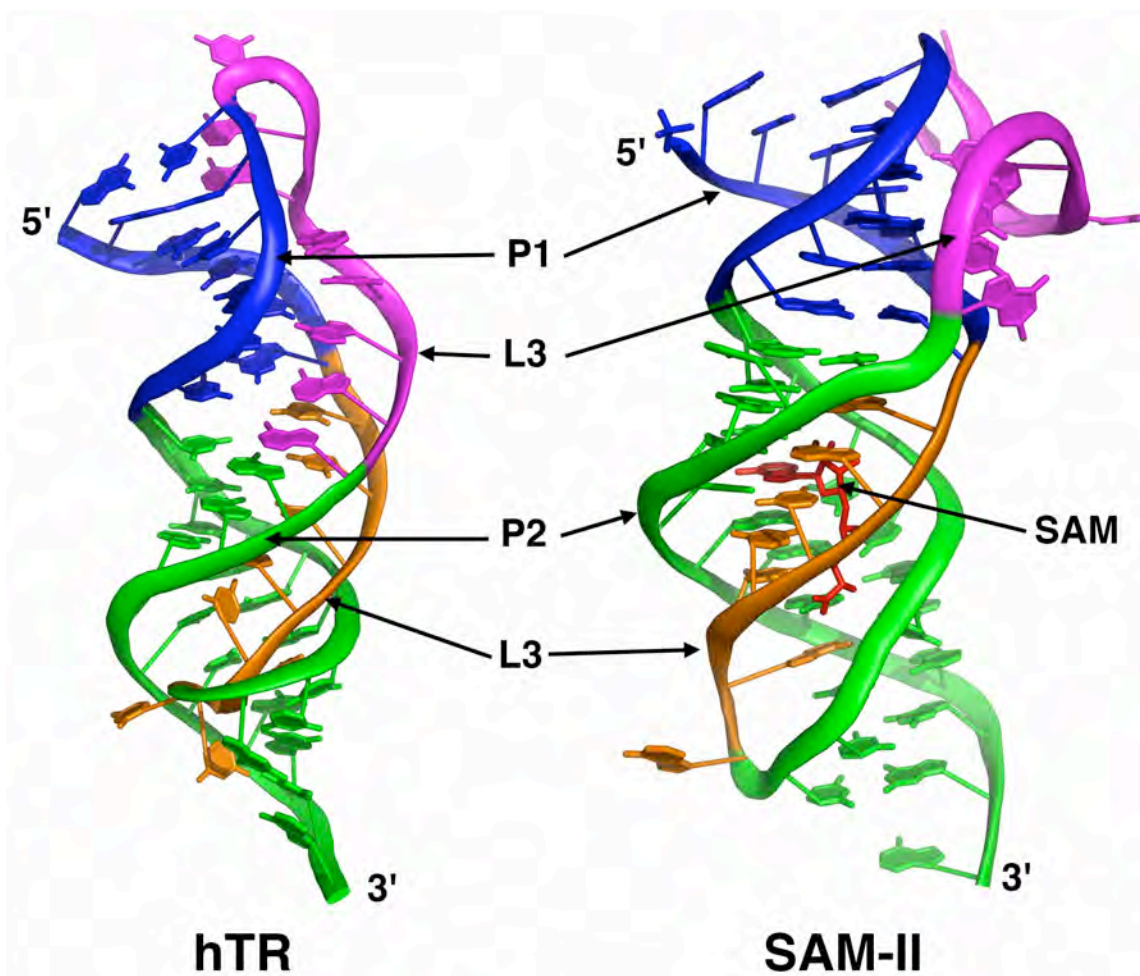
Supplementary Figure 4 Superposition of the three protomers in the asymmetric unit that were built and refined individually. The standard pairwise r.m.s.d. for all atoms in the RNA and SAM is 1.26 Å and the maximum likelihood r.m.s.d. for all atoms is 0.19 Å, as calculated using the program THESEUS²⁰. The two perspectives correspond to (a) Figure 1b and (b) Figure 1c (Figure 1b, c shows protomer A). Colors correspond to: red, molecule A; blue, molecule B; green, molecule C, as defined in the PDB coordinate file.



Supplementary Figure 5 Comparison of A-minor twist motifs in SAM-II riboswitch (left) and *glmS* ribozymes (right)^{18,19}. Watson-Crick paired helical regions are colored in blue. A-minor twist interaction is colored in magenta. The *glmS* model and numbering shown was created from the Glc6P-bound 2'-deoxy precleavage structure of the *glmS* ribozyme from *T. tencongensis* (PDB ID 2h0z)¹⁹.



Supplementary Figure 6 Raw data obtained from chemical probing experiments as well as stick representation of adenosine residue expected to be modified by DMS in the ligand bound structure. (a) Normalized and merged NMIA probing data. Peak height reflects the average percent reactivity of each residue over three experiments. The ligand-dependent changes in NMIA reactivity were calculated as the peak height difference between no SAM and 10 μ M SAM (blue-red). Numbered residues are flipped out in the crystal structure and are expected to be increasingly modified by NMIA in the presence of SAM. (b) DMS reactivity of the SAM bound RNA. Blue bars represent total average radioactivity counts over three experiments at each adenosine position minus a (-) DMS reaction. (c) A49-G17 sheared base pair that exposes the Watson-Crick face of A49 in the ligand bound structure. (d) A41•A24-U40 base triple at the junction between the P1 and P2b helices. The Watson-Crick face of A41 is thrust into solution upon ligand binding, a result of the coaxial stacking of the two helices as P2b becomes ordered. Arrows in (c) and (d) point to expected sites of DMS modification.



Supplementary Figure 7. Side-by-side comparisons of the pseudoknot from human telomerase RNA²¹ (hTR, left) and SAM-II/SAM complex (right). The colors reflect the secondary structures of the RNA (blue, P1; green, P2; orange, L1; magenta, L3); the coloring pattern of SAM-II is slightly different from Figure 1 to make a clearer comparison between the two RNAs. The hTR structure shown is a model from the family of structures derived from NMR constraints (PDB ID 1YMO).

METHODS

RNA library synthesis and purification. A series of RNAs corresponding to secondary structure and sequence variations of the SAM-II RNA observed across phylogeny was constructed according to the length of the P1 and P2a helices of the minimal riboswitch (**Supplementary Fig. 1**)¹. These helices vary between 5-8 base pairs and 2-6 base pairs, respectively. Combinations of different helix length resulted in an initial library containing 13 representative RNAs that included the *metA* RNA previously characterized¹. RNAs were constructed by standard PCR methods using overlapping DNA oligonucleotides (Integrated DNA Technologies)². The resulting DNA fragment, that contained *EcoRI* and *NcoI* restriction sites at the 5'- and 3'- ends, respectively, as well as sequences coding for a T7 RNA polymerase promoter and an 3'- H δ V ribozyme was ligated into pRAV12³. The resulting plasmid was verified by sequence analysis. Template DNA for in vitro transcription was generated by PCR from each individual plasmid using primers directed against the T7 RNA polymerase promoter at the 5' end and the 3' side H δ V ribozyme (5'- GCGCGCGAATTCTAATACGACTCACTATAG; 3'- GCACAGTCTAGAGGTCCCATTCGCCATGCCGAAGCATGTTG). RNA was transcribed in a 12.5 mL reactions containing 40 mM Tris-HCl (pH 8.0), 10 mM DTT, 0.01 % Triton X-100, 2 mM spermidine-HCl, 6 mM of both ATP and GTP, 4 mM of both UTP and CTP, 36 mM MgCl₂, 25 mg/mL T7 RNA polymerase, 1 mL of ~0.5 μ M template³, and 1 unit/mL inorganic pyrophosphatase. The reaction was incubated for 2 hours at 37 °C. RNA was precipitated in 70% EtOH, pelleted, and resuspended in load buffer containing 4 M urea, 100 mM Na-EDTA, pH 8.0, 25%

formamide, xylene cyanol, and bromophenol blue, and purified by denaturing 12% PAGE. Gel slices containing target RNA were excised from the gel, electroeluted in 1x TBE buffer, collected, exchanged against 3 x 15 ml aliquots of buffer containing 10 mM K⁺-HEPES, pH 7.5 using a 10,000 MWCO centrifugal filter device (Amicon, Ultra-15), and concentrated to ~500 μ L. RNA concentrations were determined from the magnitude of the UV absorbance at 260 nm and the calculated extinction coefficient ($703,040 \text{ M}^{-1} \text{ cm}^{-1}$) of the individual RNA's base composition.

RNA crystallization. SAM-II riboswitch crystals were obtained by the hanging drop/ vapor diffusion method in which the RNA solution was mixed in a 1:1 ratio with mother liquor. The initial library of RNAs was screened versus the PEG-Ion, Crystal Screen, Natrix, and Nucleic Acid Mini-screen (Hampton Research). RNAs and promising conditions were further refined based on crystal morphology and size, as well as diffraction quality and space group. The final RNA construct was engineered to contain a heavy-ion binding phasing module in the P1 helix that did not alter any of the residues that are conserved across phylogeny (starred,

Supplementary Fig. 1b)⁴. The final conditions that yielded diffractions quality crystals contained 8 mM cobalt hexammine chloride, 640 mM ammonium acetate, 10% PEG 1K, 10 mM barium chloride, 50 mM Na⁺-cacodylate, pH 6.0, 25 °C. Single crystal growth required cat whisker micro-seeding from a solution containing microcrystals from previously grown, but polymorphic, SAM-II crystals suspended in mother liquor plus 8% (2*R*,3*R*)-butanediol. Crystals reached their maximum size (~200 x 200 x 200 μ m, cube-like) in 3-5 days, and were subsequently backsoaked for

10-30 minutes in 30 μ l of one of two heavy-atom derivative solutions. Crystals designated for isomorphous replacement were backsoaked in mother liquor containing 2 mM magnesium chloride, 10 mM SAM, and 200 mM cesium chloride for \sim 10 minutes. Crystals designated as native were backsoaked for 10-30 minutes in mother liquor that only contained 320 mM ammonium acetate (all other components remaining the same) with the addition of 2 mM magnesium chloride, 10 mM SAM, and 600 mM lithium chloride also for \sim 10 minutes. This was followed by a 10-30 minute exchange into the same solutions containing the addition of 8 % (2*R*,3*R*)-butanediol. Crystals were looped and flash-frozen in liquid nitrogen. Diffraction data was collected on a home X-ray source (Rigaku MSC) using CuK_α radiation. Anomalous diffraction data was collected by an inverse-beam experiment and was integrated, indexed, and scaled using HKL2000⁵.

Phasing and structure determination. An initial heavy atom solution was calculated using data from a single anomalous dispersion (SAD) experiment at CuK_α with cesium as the anomalous scatterer. The solution containing four heavy atoms (FOM = 0.27, Z-score = 17) was generated in SOLVE⁶. The heavy atom solution was then used to calculate experimental phases by single isomorphous replacement with anomalous scattering (SIRAS) experiment and diffraction data extending to 2.5 Å resolution for the native data set and 2.8 Å resolution for the same cesium derivative set. Phases were calculated using CNS⁷ in which a cesium was treated as the heavy atom derivative, ($f' = -1.16$, $f'' = 8.80$), while a lithium back soaked crystal was treated as the native data set. Following density modification in CNS⁷, the figure

of merit improved to 0.35. An experimental electron density map was of sufficient quality to follow the trace of the phosphate backbone and identify regions of base-pairing for two of the three molecules in the asymmetric unit. Using model and experimental phases, a combined electron density map was calculated (FOM = 0.74) and the location of the third protomer was identified. All nucleotides in the three RNAs, as well as the locations of SAM, and three Cesium atoms bound to the engineered phasing module were eventually built through subsequent rounds of iterative model building in PyMOL⁸ and refinement in CNS⁷ while following the improvement of R_{free} . Solvent molecules were added using two rounds of water-picking in CNS⁷. The 5'- and 3'- ends of the RNAs contained functional groups that required additional refinement at the end of the building process. The addition of a 5'ppGTP and a 2',3'-cyclic phosphate at the RNA termini were supported by the electron density in these regions. A Mg^{2+} ion was also identified in the region where protomers A and B form a "head-to-head" stack causing the two 5'-diphosphate groups to come into close proximity. The final R and R_{free} are 20.6% and 26.3%, respectively. The refined model had a cross-validated Luzzati coordinate error of 0.50 Å.

Chemical modification and structure probing. An RNA corresponding to the SAM-II riboswitch found upstream of the putative *metX* gene from a Sargasso Sea metagenomic environmental sequence (COG2021; Accession number/position AACY0107673.1/130-200)¹ was constructed by standard PCR methods using overlapping nucleotides. In addition, the RNA contained 5' and 3' structure probing

cassettes that facilitate easy analysis of all the nucleotide positions⁹. The resulting PCR fragment was further amplified using a 5' primer containing a T7 promoter (5'-CTAATACGACTCACTATAGGCCTTCGGGCCAA) and a 3' primer corresponding to the terminus of the 3'- structure cassette (3'-GAACCGGACCGAAGCCCGATTTGGATCCGGCGAACCGGATC). RNA was transcribed and purified as described above. The final concentration of the RNA was calculated from the UV absorption at 260 nm and an extinction coefficient of $1,316,100 \text{ M}^{-1}\text{cm}^{-1}$.

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) using the chemical reagent *N*-methylisatoic acid (NMIA) was performed essentially as previously described by the Weeks laboratory⁹. Briefly, 2 pmols of SAM-II RNA in 11 μl of 0.5x TE buffer was subjected to heat denaturation by incubating at 90 °C for 3 minutes followed by refolding on ice for one minute at which point 1 μl of different concentrations of SAM (typically serial stock concentration dilutions of 200 μM , 100 μM , 20 μM , and 10 μM) and 6 μl of folding buffer (333 mM K^+ -HEPES, pH 8.0, 20 mM MgCl_2 , 333 mM NaCl) were added. Reactions were allowed to incubate on ice for a full 10 minutes. The final concentrations of SAM with respect to the RNA ranged from 20x to 1x while the final concentration of MgCl_2 was 6 mM. Prior to incubation with NMIA, RNA samples were separated into two 9 μl reactions in thin-walled PCR tubes and allowed to equilibrate at 25 °C for five minutes. A 65 mM NMIA stock was prepared in anhydrous DMSO (Sigma) immediately prior to its addition to the reaction as a chemical probing reagent. NMIA was added in 1 μl amounts to each reaction, while the same amount of DMSO was added to each

negative control. Reactions were allowed to incubate for 125 minutes at 25 °C (five half-lives of the NMIA reagent in solution) before being stored at -20 °C for up to one week.

Dimethyl sulphate (DMS) modification for probing the Watson-Crick faces of adenosines was performed as previously described¹⁰. 10 pmols of renatured RNA prepared as described above was added to 200 µl total volume of DMS modification buffer (70 mM HEPES-KOH, pH 7.8, 10 mM MgCl₂, 270 mM KCl) on ice.

Reactions containing ligand had a final concentration of 100 µM SAM, while the same volume of water was added to the unliganded reactions to maintain consistent salt concentrations. A 1:2 dilution of DMS (Sigma) in ethanol was added to each reaction in 2 µl amounts while the same volume of EtOH was added to negative controls. The reactions were incubated at 30 °C for 5 minutes, followed by the immediate addition of 50 µl of DMS termination buffer (1 M Tris-acetate, pH 7.5, 1 M β-mercaptoethanol, 1.5 M NaOAc, 0.1 mM EDTA) and 750 µl of ice cold EtOH. RNAs were precipitated by centrifugation in a microcentrifuge at 16,000xg for 30 minutes at 4 °C and redissolved in 20 µl of 0.5x TE buffer. Reactions were stored at -20 °C for up to one week.

Following chemical modification, reverse transcription with a radiolabeled probe was performed to analyze the extent of modification at each nucleotide. To 10 µl of each reaction was added 2 µl of ³²P 5'-end labeled primer (~0.3 µM, 5'-GAACCGGACCGAAGCCCG) and incubated at 65 °C for five minutes to promote the melting of secondary structure. A second incubation at 35 °C for 5 minutes was sufficient for primer annealing. Following an initial one minute incubation at 52 °C,

6 μ l of SHAPE enzyme mix (250 mM KCl, 167 mM Tris-HCl, pH 8.3, 16.7 mM DTT, 1.67 mM of each dNTP, and 0.33 units of Superscript III Reverse Transcriptase (Invitrogen, Inc.)) was added and incubated at 52 °C for an additional five minutes. The reverse transcriptase reaction was stopped by adding 1 μ l of 4 M NaOH and transferring to a 90 °C sand bath for five minutes. Finally, the addition of 29 μ l of acid stop mix (4:25 (v/v) mix of 1 M unbuffered Tris-HCl and stop dye (85% Formamide, 0.5x TBE buffer, 50 mM EDTA, pH 8.0, with bromophenol blue and xylene cyanol added)). Samples were loaded onto a 12% denaturing PAGE gel and run for 5-6 hours at a constant 55 W. Gel images were obtained by scanning on a Typhoon phosphorimager (Molecular Dynamics).

Band intensity was quantified using the SAFA program available from MatLab¹¹. DMS data was averaged across three separate experiments run out on the same gel. Band intensities from residues 50-53 corresponding to the series of four guanosines immediately 5' to the AUG that were observed to be similarly modified in all reactions were used to normalize band intensities in each lane representing different reaction conditions. Data handling, calculation, and graphing were done in Microsoft Excel. All calculated and graphed data were subsequently correlated to observations made on the original gel.

NMIA data were also averaged across three experiments. Because no position in the RNA could be considered equally modified in all of the reaction conditions tested, and differences in loading, exposure time, and amount of radioactivity over multiple experiments run on multiple gels increased the variability in each measurement, we used a normalizing method based on the total radioactivity counts

per lane¹². Normalizing to differences in the sum total of radioactivity counts per lane accounted for these systematic errors. As a control, we observed that a significant percentage (>40%) of the nucleotides analyzed were largely unmodified by NMIA or unchanged by the presence of SAM. We were able to calculate a percentage reactivity for each band in each lane and compare the percent across multiple experiments, correlating this with our observations. Changes in reactivity upon SAM binding that were considered significant (greater in magnitude than the dashed lines in **Fig. 4a**) are >0.5 change in percentage reactivity or <-0.5 change in percent reactivity (for the negative values) from the unliganded RNA. Changes in reactivity are considered significant if they are greater than one standard deviation from the mean of the percent change in NMIA reactivity of the lowest 15% of the residues counted (all in the P1 and P2a helices). For comparison, we normalized our data according to data analysis procedures reported for previous RNA structure probing experiments^{9,13,14} and observed that the overall trend in activity differences of the RNA in the presence and absence of SAM remained the same. However, the method we used prevented false positives that were inconsistent across reaction conditions and experiments. Further, it effectively eliminated the need to consider the (-) NMIA lanes.

Isothermal titration calorimetry. Experiments of RNA binding to SAM were carried out as previously described. RNA was dialyzed in 6-8 kDa MWCO dialysis tubing overnight at 4 °C against 50 mM K⁺-Hepes, pH 7.5, 100 mM KCl, and 10 mM MgCl₂. To ensure an exact buffer match in the experiment, SAM (Sigma-Aldrich)

was dissolved fresh in water to a stock concentration of ~100 mM and then readjusted in dialysis buffer to within 7-10x the concentration of RNA. Concentrations of SAM ($\epsilon_{260} = 13,400 \text{ M}^{-1} \text{ cm}^{-1}$) and the RNA ($\epsilon_{260} = 703,040 \text{ M}^{-1} \text{ cm}^{-1}$) were determined by UV absorbance on a spectrophotometer. RNA concentrations were all ~ 15-20 μM . Samples were degassed for 10 minutes at 25 °C prior to titration experiments. ITC experiments were carried out at 30 °C, a reference power of 5 $\mu\text{cal sec}^{-1}$, an initial delay of 60 seconds, and titration of 32 ten μl injections at an injection rate of 0.5 $\mu\text{l sec}^{-1}$; individual injections were spaced 210 seconds apart.

Data from each experiment was fit to a single-site binding model using Origin ITC software (Microcal Software, Inc). Values for the association constant K_a were extrapolated using the equation

$$q = v \cdot \Delta H \cdot [\text{RNA}] \cdot \left(\frac{K_a [L]_i^n}{1 + K_a [L]_i^n} - \frac{K_a [L]_{i-1}^n}{1 + K_a [L]_{i-1}^n} \right)$$

where q is the directly measured amount of heat released during each injection, v is the volume of the reaction, L_i is the ligand concentration at the i^{th} injection¹⁵. The disassociation constant, K_d (in μM) was calculated from the inverse of the K_a .

References.

1. Corbino, K.A. et al. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol* **6**, R70 (2005).
2. Sambrook, J. & Russel, D.W. *Molecular Cloning: A laboratory manual*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
3. Kieft, J.S. & Batey, R.T. A general method for rapid and nondenaturing purification of RNAs. *RNA* **10**, 988-995 (2004).
4. Keel, A.Y., Rambo, R.P., Batey, R.T. & Kieft, J.S. A general strategy to solve the phase problem in RNA crystallography. *Structure* **15**, 761-72 (2007).
5. Otwinowski, Z. & Minor, D. Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307-326 (1997).
6. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11**, 49-52 (2004).
7. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21 (1998).
8. DeLano, W.L. The PyMOL Molecular Graphics System. (2002).
9. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-6 (2006).
10. Kjems, J., Egebjerg, J. & Christiansen, J. *Analysis of RNA-protein complexes in vitro*, (Elsevier, 1998).
11. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D. & Altman, R.B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11**, 344-54 (2005).
12. Kieft, J.S., Costantion, D.A., Filbin, M.E., Hammond, J. & Pfingsten, J.S. *Structural methods for studying IRES function*, (Elsevier, 2007).
13. Ryder, S.P. & Strobel, S.A. Nucleotide analog interference mapping. *Methods* **18**, 38-50 (1999).
14. Vicens, Q., Gooding, A.R., Laederach, A. & Cech, T.R. Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA* **13**, 536-48 (2007).
15. Wiseman, T., Williston, S., Brandts, J.F. & Lin, L.N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131-7 (1989).
16. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, D121-4 (2005).
17. Venter, J.C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74 (2004).
18. Cochrane, J.C., Lipchock, S.V. & Strobel, S.A. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chem Biol* **14**, 97-105 (2007).

19. Klein, D.J. & Ferre-D'Amare, A.R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752-6 (2006).
20. Theobald, D.L. & Wuttke, D.S. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**, 2171-2 (2006).
21. Theimer, C.A., Blois, C.A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **17**, 671-82 (2005).