

Supplementary Materials

Supplementary Methods

Docking parameters

GOLD 5.0.1 (CCDC): For docking and virtual screening, default parameters were set for GOLD Fitness, ChemScore and ASP scores. "Allow early termination" and soft potentials were turned off, and 200% search efficiency was employed to allow maximal exploration of ligand conformation. We used 20 genetic algorithm (GA) runs with internal energy offset. For pose reproduction analysis, the radius of the binding pocket was set as the maximal atomic distance from the geometrical center of the ligand plus 3Å. The top 10 ranked docking poses were retained for the 3D cumulative success rate analysis, cross-docking, and virtual screening studies. To perform the native pose ranking and RMSD-score correlation study, we found that the GOLD:GOLD Fitness combination with 100 population and 1000 maxops could help us obtain high diversity and quality of the conformational decoys. Therefore, GOLD:GOLD Fitness was employed to generate 100 conformational decoys for each target. Rescoring was conducted with the GOLD rescore option, in which poses would be optimized by the program.

Glide 5.6 (Schrödinger): Default parameters were employed for both Glide standard precision (SP) and extra precision (XP) docking. Both GlideScore and Emodel score were evaluated. Multiple starting conformations were prepared with LigPrep2.0. The binding site was defined as a box centered on the geometrical center of the bound ligand with each length equivalent to the maximal atomic distance from the center of the ligand plus 3Å. Flexible hydroxyl groups involved in the ligand binding were selected. The ligand internal energy offset option was turned on. The top 10 ranked poses were minimized and retained. Rescoring was performed by choosing "Refine (do not dock)" option. The decoys with no valid poses after minimization were excluded in RMSD-score correlation analysis, but included in other evaluations as bad poses (GlideScore or Emodel=10000).

Surflex 2.415 (Tripos): The binding pockets were defined by the area around the experimentally determined ligand structure. The protomol_bloat=5 was set for pocket identification. We used 4 additional starting poses and explored the best spin density parameter using 3, 5 and 10. Self_scoring option was turned on. We kept 10 final poses for analysis, and rescoring was performed by "-opt" flag.

rDock 2006.2: Radius of binding pocket was maximal atomic distance from the geometrical center of the ligand plus 3Å, and site searching scoring function was RbtCavityGridSF. Default parameters from "dock.prm" (standard scoring function) and "dock_solv.prm" (scoring function with solvation term) were used for scoring. We performed 200 separate runs for each docking exercise in order to cover enough conformational space. Top 10 ranked poses were retained. Rescoring was performed using the parameter in "minimise.prm" and "minimize_solv.prm" for rescoring with and without the solvation term, respectively.

AutoDock 4.1: The definition of grid box was the same as that of Glide with 0.2Å grid spacing. Lamarckian Genetic Algorithm (LGA) was used to perform 100 GA runs. Other parameters, such as 200 individuals in populations, 500,000 maximum energy evaluations, and 30,000 maximum generations were employed for LGA. The top 10 clusters were retained for analysis. Rescoring was performed using

AutoDockTools4 using optimized parameters.

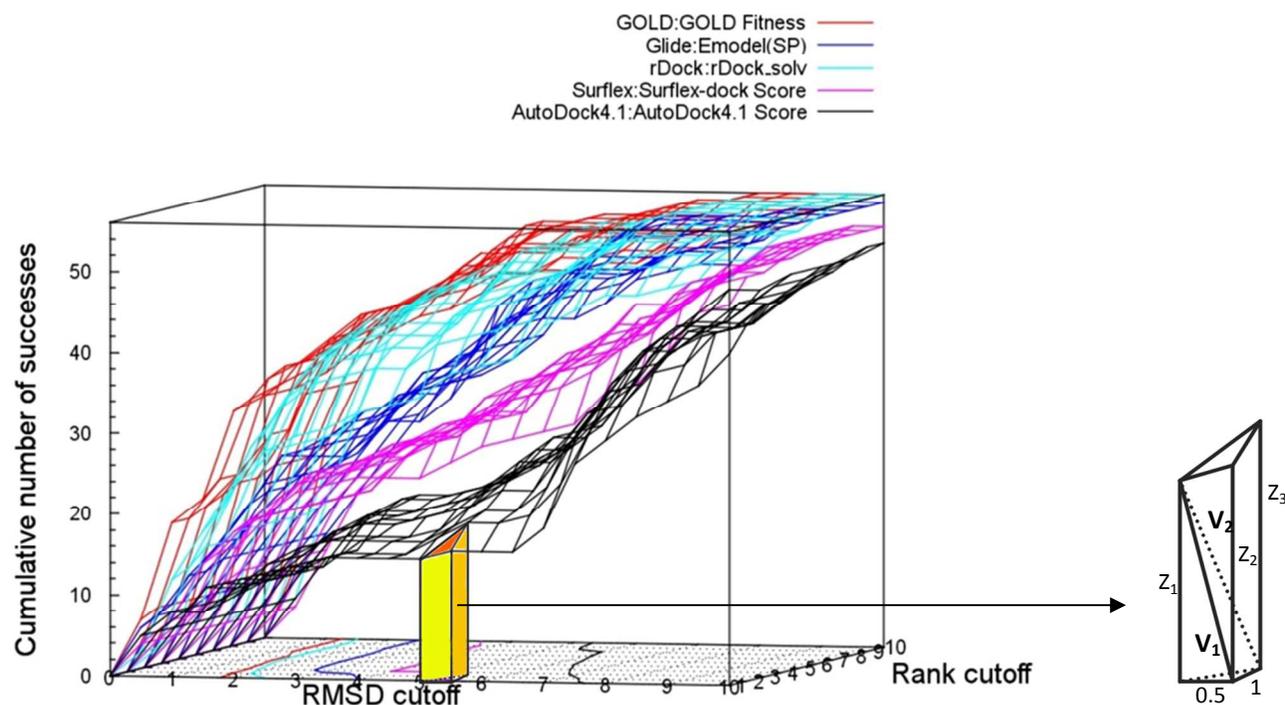
Volume under the surface (VUS) calculation

VUS were estimated as the sum of the volume of all triangular prism units under the surface, therefore

$$VUS = \sum (V_{\text{triangular_prism}})$$

The volume of each triangular prism unit ($V_{\text{triangular prism}}$) was calculated by the following equation. Each triangular prism unit was broken down into a tetrahedron (V_1) and a tetragonal pyramid (V_2), as illustrated below. Z_1 , Z_2 and Z_3 were the Z coordinates of triangle vertices on the 3D cumulative success rate surface, and we assume $Z_1 \leq Z_2 \leq Z_3$. Therefore,

$$\begin{aligned} V_{\text{triangular_prism}} &= V_1 + V_2 \\ &= \left(\frac{1}{6} \times 1 \times 0.5 \times Z_1\right) + \left(\frac{1}{3} \times \frac{Z_2 + Z_3}{2} \times 1 \times 0.5\right) \\ &= \frac{1}{12} (Z_1 + Z_2 + Z_3) \end{aligned}$$



Estimation of the binding site flexibility

The high-resolution crystal structures of bacterial rRNA A-site (PDB ID: 1J7T) and lysine riboswitch (PDB ID: 3DIL) were obtained from Protein Data Bank database. The active site was defined as the nucleotides 4Å around the ligand (paromomycin or lysine). To quantify the binding site flexibility, we compared the B-factors (or Debye-Waller factor) of active site with those of other non-terminal nucleotides. We averaged the B-factors of all atoms in one nucleotide to represent the flexibility of the given nucleotide. *P-values* were calculated with two-tailed student t-test. Normal model analyses were performed with oGNM web server, which evaluates the flexibility of RNA in a Gaussian network model. We employed the three-node-per-nucleotide representation (P, C4' and C2) with the r_p cutoff 10Å to construct the Gaussian network. Then, we compared the predicted atomic fluctuations of active site nucleotides with other non-terminal nucleotides using the five low-frequency modes. oGNM is available at http://ignm.cccb.pitt.edu/GNM_Online_Calculation.htm.

Supplementary Tables

Supplementary Table S1. List of 56 PDBs used in binding mode reproduction study.

Supplementary Table S2. Experimental binding free energy values (kJ/mol) used in score-binding affinity correlation study and scoring function optimization study.

Supplementary Table S3. Summary of the statistics from the binding mode reproduction study.

Supplementary Table S4. Detailed docking scores and RMSDs from the binding mode reproduction study. This includes 12 different docking & scoring combinations evaluated in our study.

Supplementary Table S5. Statistics from native pose ranking study.

Supplementary Table S6. Rankings of the cognate ligand from cross-docking study.

Supplementary Table S7. The performance of RNA ensemble docking/scoring.

Supplementary Table S8. The correlation between the dihedral angle ϵ from U23 (1LVJ) and native pose ranking of the ligand PMZ.

Supplementary Figures

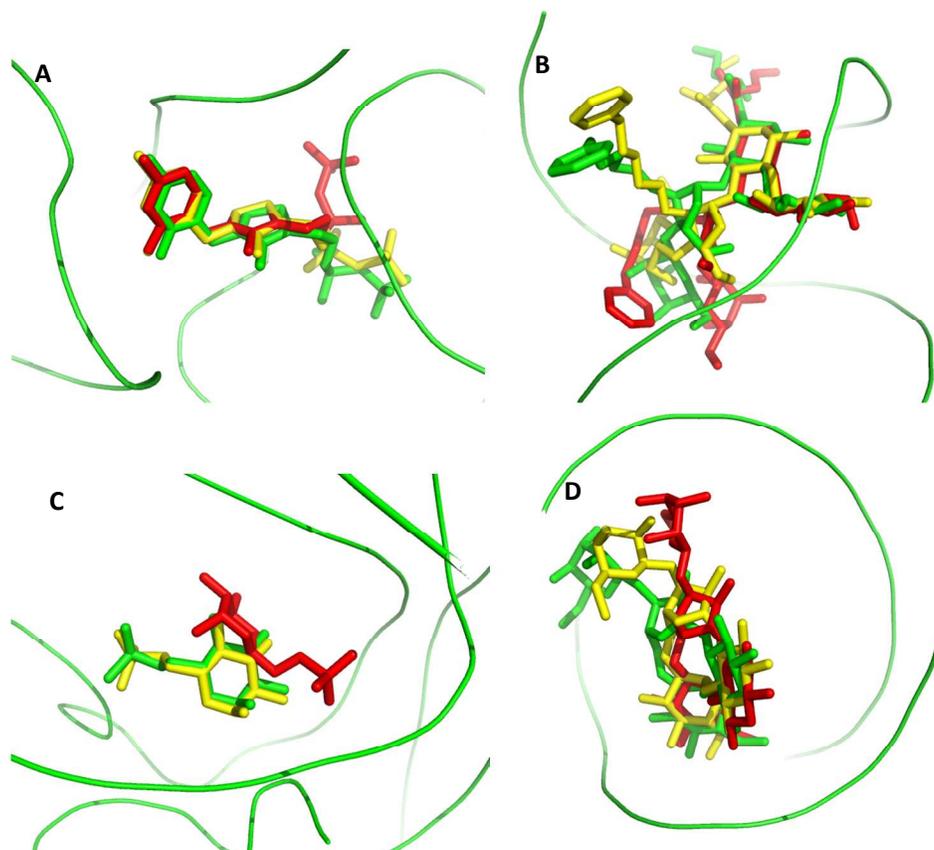
Supplementary Figure 1. The improvement of pose reproduction by the optimization with ASP scoring function.

Supplementary Figure 2. Correlation between scores and binding affinities for ASP, GOLD Fitness, AutoDock4.1 Score (default). Three common outliers, 1TOB, 2TOB and 1LVJ, were highlighted in rectangles.

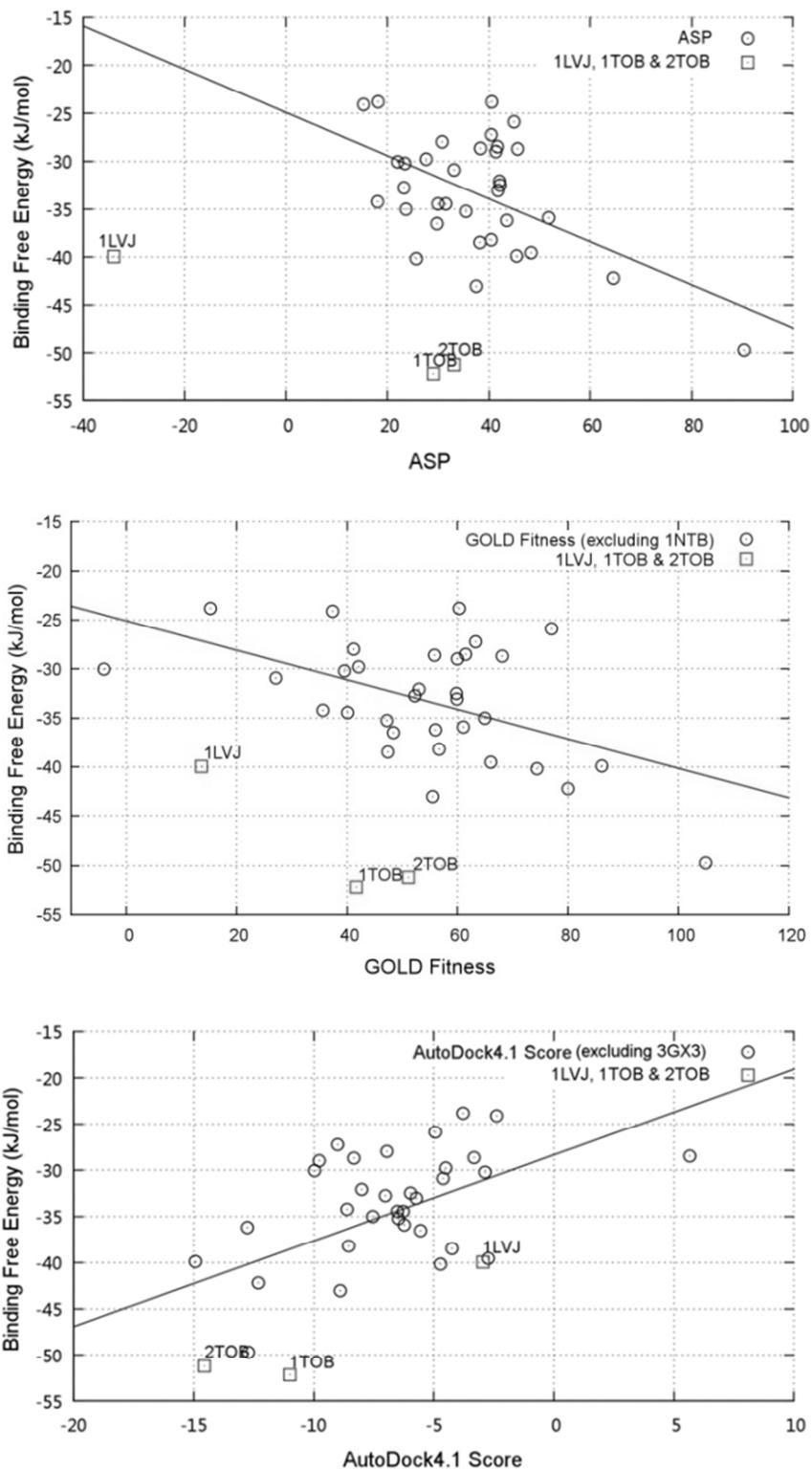
Supplementary Figure 3. Comparisons of the AutoDock4.1 predicted conformations (AutoDock4.1: iMDLScore2) with X-ray crystal structures.

Supplementary Figure 4. Comparison of the B-factors and normal modes from the rigid active site (lysine riboswitch) and flexible active site (16S rRNA A-site).

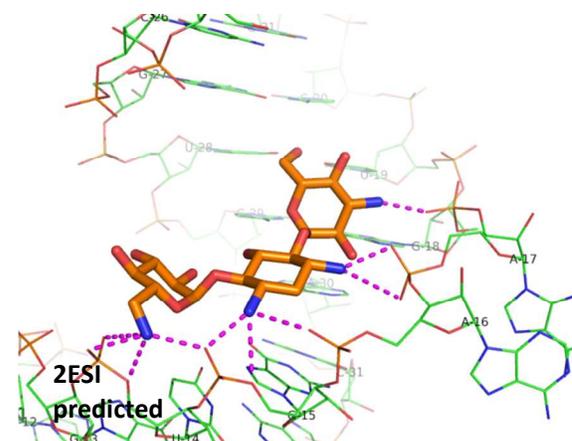
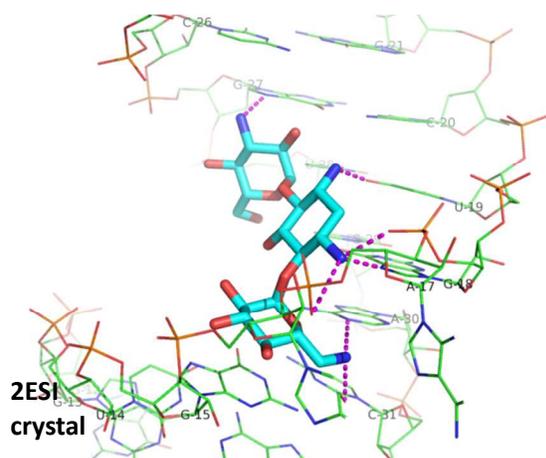
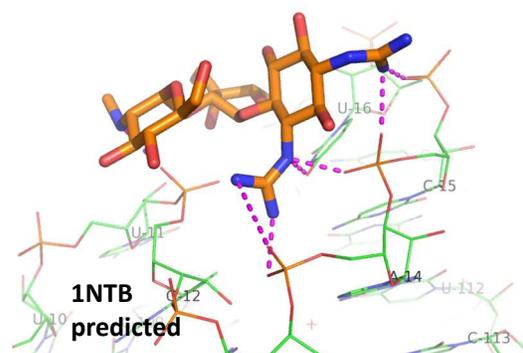
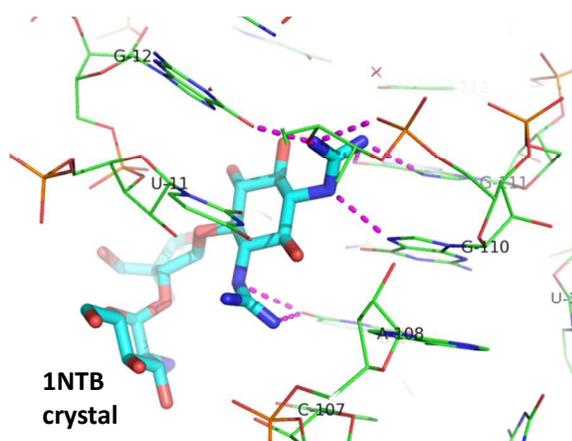
Supplementary Figure 1. The improvement of pose reproduction by the optimization with ASP scoring function. Experimental structures were in green (RNAs in ribbons, ligands in sticks). Only the docking conformation with the lowest RMSD selected from the top five-scored poses were shown. GOLD:GOLD Fitness poses were colored red, while ASP rescored poses are colored yellow. **(A)**. 2GDI; **(B)**. 2PWT; **(C)**. 2Z74; **(D)**. 1ZZ5.

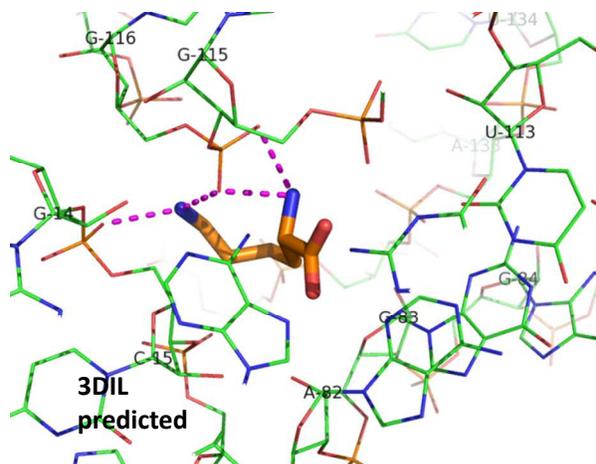
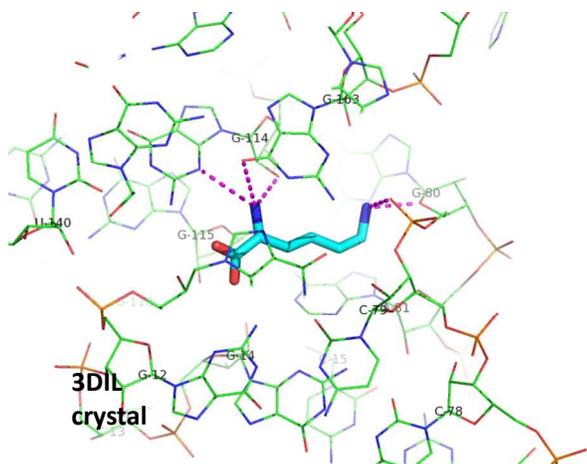


Supplementary Figure 2. Correlation between scores and binding affinities for ASP, GOLD Fitness, AutoDock4.1 Score (default). Three common outliers, 1TOB, 2TOB and 1LVJ, were highlighted in rectangles.

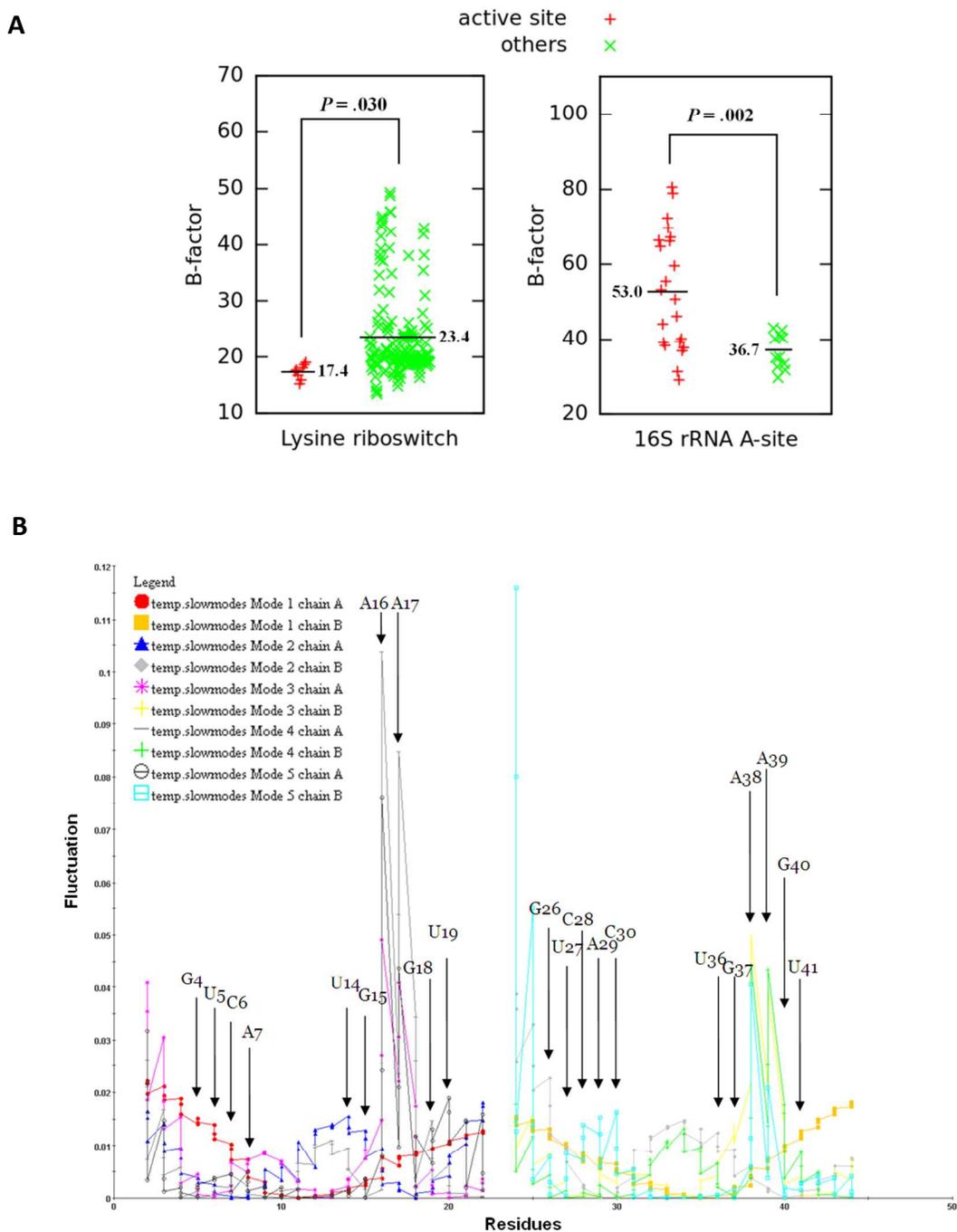


Supplementary Figure 3. Comparisons of AutoDock4.1 predicted conformations (AutoDock4.1: iMDScore2) with X-ray crystal structures. 1NTB, 2ESI and 3DIL were used as the examples to demonstrate the overestimation of polar interactions with RNA phosphate for initial pose generation purpose. RNA receptors were shown in green lines, while experimentally determined binding modes are shown in cyan sticks. AutoDock4.1 generated pose with the best RMSD were in orange sticks. The interactions between basic guanidinium/amine groups with RNA atoms were labeled with magenta dashes. We could observe that these basic groups were predicted dominantly to form interaction with the backbone phosphates; actually, the H-bonds with RNA base atoms and cation- π interactions were more favorable.





Supplementary Figure 4. (A) Comparison of the B-factors from the rigid active site (lysine riboswitch) and flexible active site (16S rRNA A-site). The bars indicate the average of B-factor each group. **(B)** Atomic fluctuations of 16S rRNA A-site predicted by five lowest-frequency normal modes using oGNM. Arrows indicate the active site residues. **(C)** Atomic fluctuations of lysine riboswitch predicted by five lowest-frequency normal modes using oGNM. Arrows indicate the active site residues.



C

