Supporting Information: Capturing Transition Paths and Transition States for Conformational Rearrangements in the Ribosome

Jeffrey K. Noel[1], Jorge Chahine[2], Vitor B.P. Leite[2], and Paul Charles Whitford*[3]

[1]Center for Theoretical Biological Physics Rice University, 6500 Main St MS-654, Houston, TX 77054
[2]Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, 15054-000 São José do Rio Preto, Brazil;
[3]Department of Physics, Northeastern University Dana Research Center 123, 360 Huntington Ave, Boston, MA 02115

September 25, 2014

Imposing Boundary Restraints to Mimic the Full Ribosome

To reduce the computational demand of each simulation, we developed a strategy for simulating a subset of atoms in the ribosome that minimizes artificial boundary effects. The strategy is the following:

1. Construct a SMOG model for the full ribosome, where every non-hydrogen atom is explicitly represented (150,018 atoms).

2. Construct a SMOG model that is identical to the full ribosome model, except only a subset of atoms is included (i.e. the "truncated" system). In our case, the dimensions of the truncated system were chosen to include atoms near the accommodation corridor[1]. The truncated system included 23,888 non-hydrogen atoms (Fig. S2A)

3. Identify the set of boundary atoms ($A_{boundary}$) in the truncated system that have interactions (bonds, bond angles, dihedrals, or native contacts) with atoms in the full ribosome that are not included in the truncated system. For example, if atom 50 and atom 200 interact in the full ribosome, and atom 200 were excluded from the truncated system, atom 50 would then be considered a boundary atom. For the truncated system used here, $A_{boundary}$ included 3,989 atoms. (Fig. S2B)

4. Perform a simulation of the full ribosome, until the values of the spatial root mean-squared fluctuations of the $A_{boundary}$ atoms converge. These values are used as reference values, against which restraints are refined. This is physically warranted, since the structural fluctuations in SMOG models of the ribosome are consistent with estimates from explicit-solvent simulations and crystallographic B-factors[2].

5. In the truncated system, introduce isotropic harmonic spatial restraints of weight k_i on all $A_{boundary}$ atoms, where i is the atom index.

6. Iteratively adjust the value of k_i for each atom until the root mean-squared fluctuations of the $A_{boundary}$ atoms in the truncated system are consistent with those in the full ribosome. (Fig. S3)

7. After a set of k_i values is determined, the new truncated model with restraints is used for production calculations.

Below, we provide a detailed description of the construction of the spatial restraints.

Calculating the reference values for the msf of each atom

To construct the truncated simulation, we first established a reference system, against which the boundary restraints were refined in the truncated system. This was accomplished by constructing an all-atom structure-based model[3] of a full ribosome, where the A/A configuration (PDB ID: 3I8F[4]) was defined as the global potential energy minimum, consistent with our earlier simulations of aa-tRNA accommodation[2]. This included all non-hydrogen atoms, for a total of 150,018 atoms. Using a timestep of 0.002, we performed a 10 million timestep simulation of the full ribosome. From this, we discarded the first 500,000 steps for equilibration purposes, and then calculated the mean-squared fluctuations of each atom in $A_{boundary}$ using the remaining simulated frames. To verify convergence, we repeated the analysis using the first, or second, half of the trajectory and found indistinguishable values for the msf. The rmsf of the $A_{boundary}$ atoms in the full ribosome simulation is shown in Figure S3.

Refining the position restraints k_i

To refine the position restraints, we first calculated the msf of the $A_{boundary}$ atoms in the full ribosome simulation, msf_i^{full} (with an average $\langle msf_i^{full} \rangle$). The values obtained for the full ribosome served as target values, against which the restraints were refined. Note: any reference rmsf value that exceeded 5 Å was set to 5 Å for refinement purposes. The rationale for this is that motions occurring on larger scales are increasingly anisotropic, and isotropic restraints become increasingly inaccurate. The only region for which fluctuations are that large are atoms in the L11 stalk. Rather than attempt to reproduce the full range of stalk motion with isotropic restraints, by imposing a limit on the target values, the stalk was modeled as remaining in a relatively 'closed' configuration, as observed in many crystal structures. After obtaining this target set of msf values, the following iterative protocol was employed to refine the values of the position restraints.

1. Set all position restraints to a uniform value k.

2. Simulate the truncated system for 4×10^6 timesteps.

3. Calculate the msf of the $A_{boundary}$ atoms, including all frames between 5×10^5 and 4×10^6 timesteps.

4. Calculate the average msf for all $A_{boundary}$ atoms: $\langle msf_i^{truncated} \rangle$.

5. Rescale k by $\langle msf_i^{truncated} \rangle / \langle msf_i^{full} \rangle$

6. Return to step 2.

The protocol was repeated until $\langle msf_i^{truncated} \rangle$ was within 1 percent of $\langle msf_i^{full} \rangle$, which required 5 iterations. The values of k_i were then independently refined on a per-atom basis. When refining by atom, each k_i was rescaled (step 5, above) by $\frac{msf_i^{truncated}/msf_i^{full}+1}{2}$. This heterogeneous refinement was performed for 34 iterations, at which point there was visible agreement between the truncated and full systems (Figure S3B). The refined values of k_i span four orders of magnitude (Figure S3A), which emphasizes the need for careful refinement when restraints are imposed on boundary atoms in truncated systems.

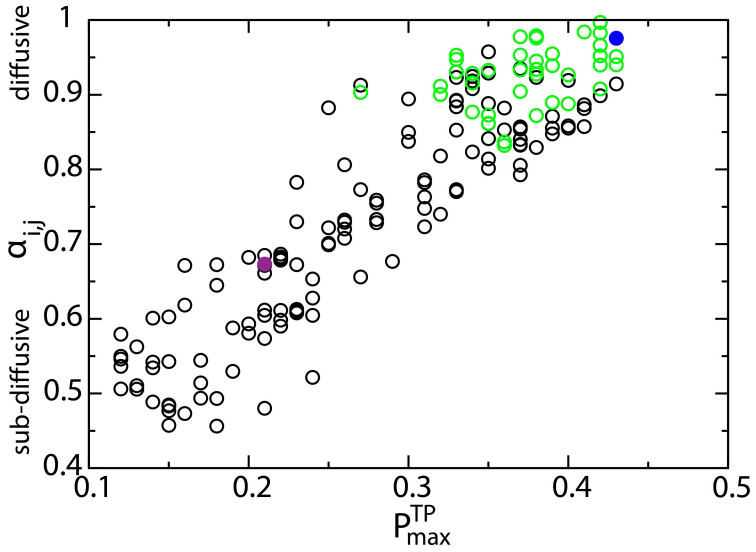

Figure S1: For all 169 coordinates for which $|R^{AA}_{i,j} - R^{AT}_{i,j}| > 10$ Å, the maximum value of $P(TP|R_{i,j})$ (P^{TP}_{max}) and the mean squared displacement was calculated. The mean squared displacement was fit to $R_0^2\tau^{\alpha}$, where the value of α implicates diffusive, subdiffusive or ballistic motion. Overall, there is a strong correlation between α and P^{TP}_{max}. Points shown in green correspond for coordinates that minimize the number of transitions N_T. The points corresponding to $R_{8,60}$ and $R_{8,47}$ are shown in solid blue and purple.

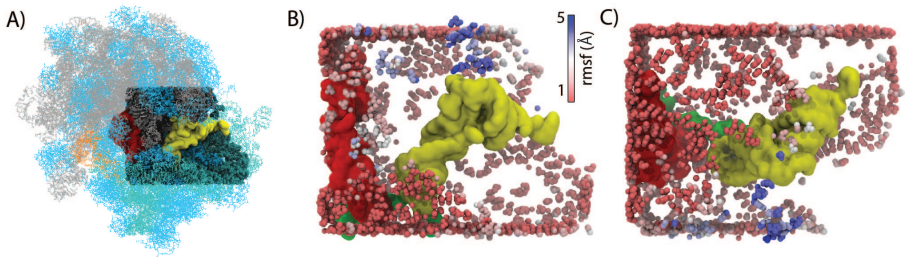

Figure S2: A) Structure of an A/T configuration of aa-tRNA, in the context of the full ribosome. For the simulations here, a subset of atoms was included, as depicted by spheres. B) To ensure that boundary effects were not introduced by imposing overly restrictive boundary conditions, atoms at the edges of the simulated system (shown as spheres) were restrained by isotropic harmonic restraints that were tuned to reproduce the scale of the fluctuations in the full ribosome. Boundary atoms are colored by their rmsf values in the full ribosome (red=1Å, blue>5Å). C) Rotated view of panel B.

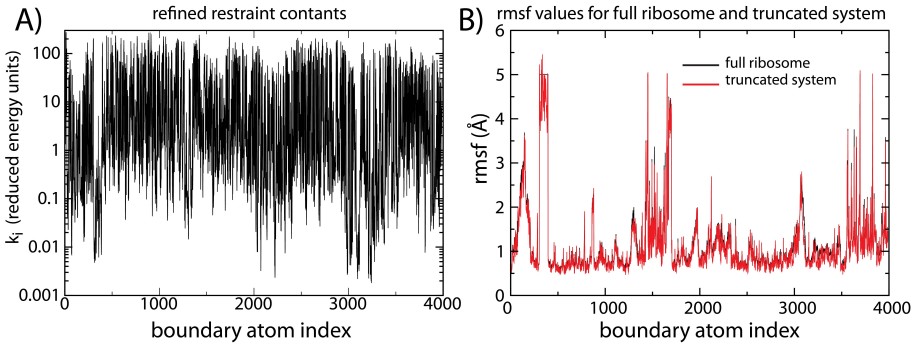

Figure S3: A) Values of the harmonic spatial restraint constants k_i imposed on boundary atoms after all refinement steps were completed. Refinement of boundary restraints resulted in k_i values that span four orders of magnitude, which highlights the need for refinement protocols. B) Comparison of spatial rmsf values for boundary atoms in a simulation of a full ribosome (black) and the truncated system with isotropic restraints on boundary atoms (red). The average values of the rmsf differ by less than 10 % and the overall character is consistent. Note: Out of a total of 23,888 simulated atoms, only 3,989 were restrained.

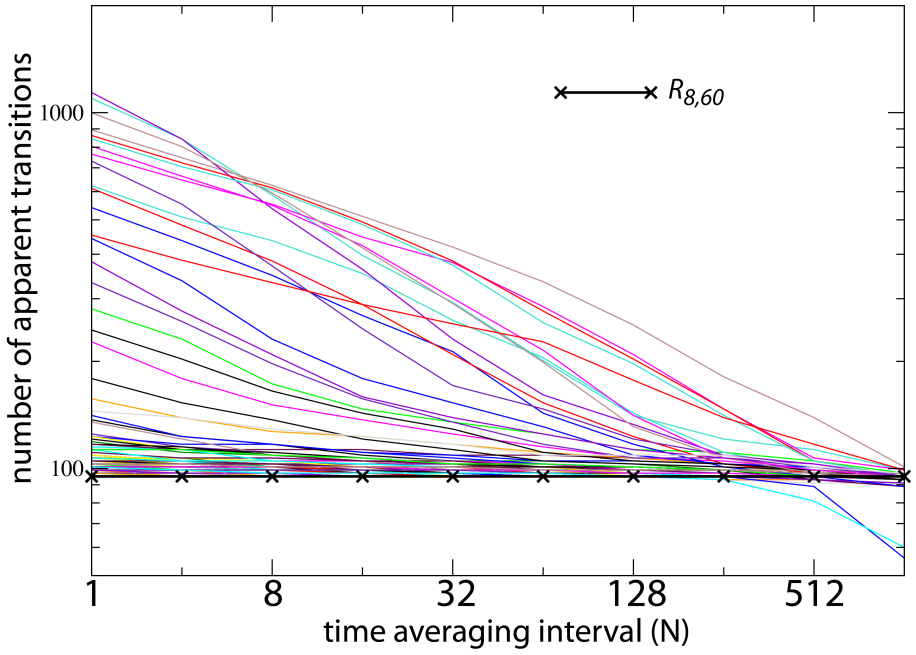

Figure S4: Number of apparent transition events N_T for all 169 coordinates $R_{i,j}(t)$, after time averaging over $N = 2^M$ frames. Each line represents a different coordinate. The number of transitions is robust for $R_{8,60}$ (marked line), whereas the number of false positives identified with alternate coordinates reduces when averaging is employed. For $R_{8,47}$, $N_T = 107$ when averaging is not employed. Only when $R_{8,47}$ is averaged over at least 256 sampled configurations, does $N_T = 95$. When averaging beyond 1024, almost every coordinate (including $R_{8,47}$) captures less than 95 transitions. Accordingly, when using suboptimal coordinates, there can be a narrow range of times for which false positives are not present, whereas more appropriate coordinates can be robust to averaging effects.

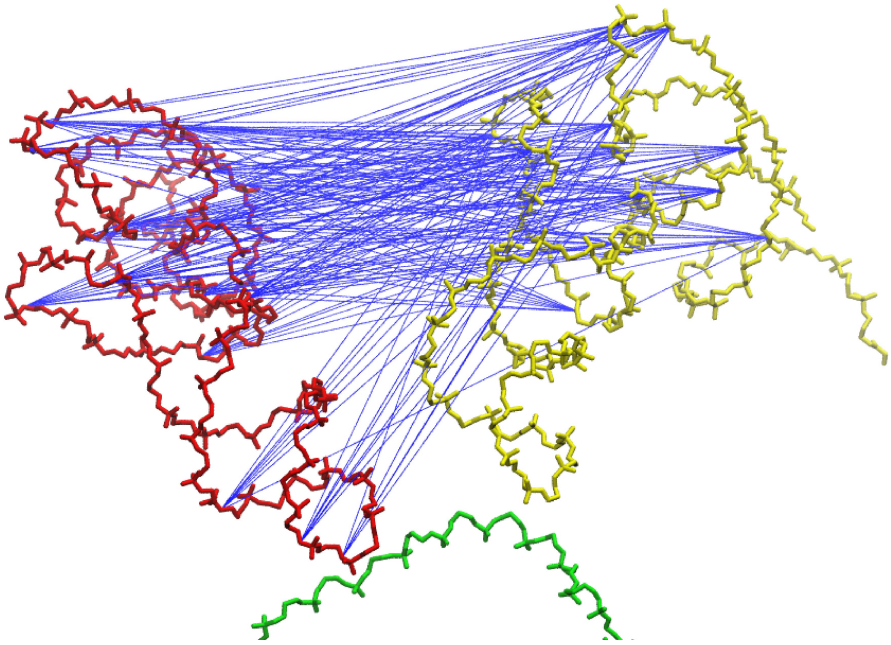

Figure S5: Structural depiction of all 169 coordinates evaluated. Each blue line indicates a unique distance that was calculated and characterized. For clarity, only the tRNA and mRNA backbone atoms are shown.

References

[1] K Y Sanbonmatsu, S Joseph, and C-S Tung. Simulating movement of tRNA into the ribosome during decoding. *Proc Nat Acad Sci USA*, 102(44):15854–9, 2005.

[2] P C Whitford, P Geggier, R B Altman, S C Blanchard, J N Onuchic, and K Y Sanbonmatsu. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA*, 16:1196–1204, 2010.

[3] P C Whitford, J K Noel, S Gosavi, A Schug, K Y Sanbonmatsu, and J N Onuchic. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Prot Struct Func Bioinfo*, 75:430–41, 2009.

[4] L Jenner, N Demeshkina, G Yusupova, and M Yusupov. Structural rearrangements of the ribosome at the trna proofreading step. *Nat Struct Mol Biol*, 17(9):1072–8, 2010.

P-site tRNA (i)	aa-tRNA (j)	$\max(P(TP\|R_{i,j}))$	α_{ij}
8	60	0.43	0.98
16	59	0.38	0.93
16	60	0.42	0.98
20	59	0.39	0.94
20	60	0.42	0.95
20	66	0.34	0.88
45	59	0.37	0.90
45	60	0.42	0.94
47	59	0.4	0.93
47	60	0.43	0.95
47	66	0.32	0.90
50	51	0.39	0.89
50	59	0.37	0.95
50	60	0.41	0.98
50	66	0.33	0.93
51	50	0.35	0.87
51	51	0.42	0.91
51	59	0.37	0.93
51	60	0.39	0.95
51	66	0.33	0.93
54	16	0.27	0.90
54	51	0.38	0.87
54	54	0.36	0.84
54	59	0.34	0.93
54	60	0.38	0.98
54	66	0.33	0.95
55	51	0.35	0.86
55	54	0.36	0.83
55	59	0.33	0.95
55	60	0.38	0.98
55	66	0.32	0.91
59	51	0.4	0.89
59	59	0.38	0.94
59	60	0.42	1.00
59	66	0.34	0.92
60	59	0.38	0.93
60	60	0.42	0.97
60	66	0.35	0.93
66	51	0.43	0.94
66	66	0.37	0.98

Table 1: Maximum value of $P(TP|R_{i,j})$ and α_{ij} for all coordinates that minimize the number of crossing events ($N_T = 95$)