Article

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, Houston, Texas; [2]Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, Brazil; and [3]Department of Physics, Northeastern University, Boston, Massachusetts

ABSTRACT To reveal the molecular determinants of biological function, one seeks to characterize the interactions that are formed in conformational and chemical transition states. In other words, what interactions govern the molecule's energy landscape? To accomplish this, it is necessary to determine which degrees of freedom can unambiguously identify each transition state. Here, we perform simulations of large-scale aminoacyl-transfer RNA (aa-tRNA) rearrangements during accommodation on the ribosome and project the dynamics along experimentally accessible atomic distances. From this analysis, we obtain evidence for which coordinates capture the correct number of barrier-crossing events and accurately indicate when the aa-tRNA is on a transition path. Although a commonly used coordinate in single-molecule experiments performs poorly, this study implicates alternative coordinates along which rearrangements are accurately described as diffusive movements across a one-dimensional free-energy profile. From this, we provide the theoretical foundation required for single-molecule techniques to uncover the energy landscape governing aa-tRNA selection by the ribosome.

INTRODUCTION

During biological functioning, molecular assemblies interconvert between energetically competing configurations where the kinetics are determined by the relative stabilities of the endpoints and the scale of the intervening free-energy barriers. Thus, quantifying the energetic and structural content of the transition state ensembles (TSEs) is central to understanding the mechanisms that govern biological regulation. This has long been recognized in the area of protein dynamics, where decades of theoretical and experimental efforts have aimed to describe the underlying energy landscapes associated with folding, binding, and conformational rearrangements (1–13). These studies have provided the field of biophysics with a rich set of general methods that are grounded in statistical mechanics, which may now be applied to quantitatively probe a wide range of biological processes. Accordingly, here, we have adopted these tools to elucidate the character of the energy landscapes associated with conformational transitions in the ribosome (Fig. 1).

The ribosome is a large (2.4 MDa) RNA-protein assembly, and it is the sole producer of proteins in the cell. To translate messenger RNA (mRNA) sequences into protein sequences, each aminoacyl-transfer RNA (aa-tRNA) molecule is delivered to the ribosome by the prokaryotic EF-Tu (i.e., elongation factor-thermo unstable), after which it may add an amino acid to the growing protein chain. During initial association of the ternary complex (aa-tRNA•EF-Tu•GTP), codon-anticodon interactions are formed between the aa-tRNA and mRNA. Through thermodynamic discrimination alone, correct (cognate) aa-tRNA molecules can be selected over incorrect (near/noncognate) ones by a factor of ~100. However, the fidelity of aa-tRNA selection is roughly 1:3000 to 1:10,000, thus implicating the presence of a subsequent free-energy barrier that serves as a proofreading step (14).

The putative conformational change responsible for proofreading is known as aa-tRNA accommodation (15), where the incoming amino acid is displaced ≈100 Å from the periphery of the ribosome to the peptidyltransferase center (Fig. 1). The biological significance of this process has motivated many investigations with cryo-electron microscopy (16,17), x-ray crystallography (18,19), biochemical kinetic techniques (20,21), single-molecule methods (22,23), simulations (24,25), and bioinformatics (26,27). These studies have elucidated the structural content of the endpoints and overall rates of interconversion, and have provided qualitative descriptions of accessible pathways. Despite this progress, it is still unclear precisely where the TSE is that governs the dynamics of accommodation, and what coordinates most accurately describe it. By addressing these points, experimental approaches will be able to more precisely probe the physical-chemical interactions that govern this biological process.

With an energy landscape perspective, one may systematically describe the structural and energetic characteristics of

Submitted August 8, 2014, and accepted for publication October 16, 2014.
*Correspondence: p.whitford@neu.edu
Editor: Matthias Rief.

0006-3495/14/12/2881/10 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.10.022

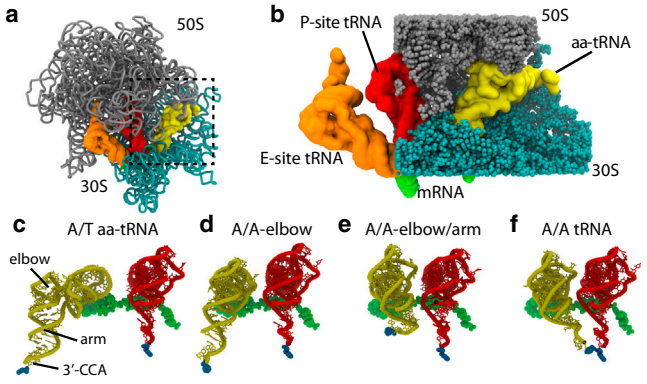

FIGURE 1 Structural description of accommodation. (*a*) The large subunit rRNA (*gray*) and small subunit rRNA (*cyan*) are shown in cartoon representation, with the aa-tRNA (*yellow*), P-site tRNA (*red*), and E-site tRNA (*orange*) shown in surface representation. (*Green*) mRNA. (*Dashed box*) Accommodation corridor. (*b*) Close-up perspective of the A/T (preaccommodation) configuration of aa-tRNA; (*spheres*) accommodation corridor atoms (rRNA and ribosomal proteins). (*c*) aa-tRNA and P-site tRNA shown in the A/T configuration. View rotated relative to panel *b*. (*c–f*) During aa-tRNA accommodation, the incoming amino acid is delivered to the PTC. Previous theoretical and experimental studies have implicated a multistep accommodation process, where intermediate elbow-accommodated (*d*) and arm-accommodated (*e*) configurations are transiently populated. In this study, we probe elbow accommodation, because it is monitored by single-molecule methods.

the TSEs associated with ribosomal conformational transitions. However, to do so, one should first ask:

1. When is it appropriate to describe aa-tRNA movement as diffusion across an energy landscape?
2. Which coordinates most accurately describe movement along the dominant (i.e., lowest free-energy) pathway/s?

One way to answer these questions is to analyze the properties of the transition paths associated with each rearrangement. In the diffusive regime, the dynamics is Markovian and configurations in the TSE are equally likely to continue to the product or reactant states. Quantitatively, this may be phrased as: The conditional probability of being on a transition path (TP), expressed as $P(TP|\vec{x})$ (where $\vec{x}$ is the configuration of the system in the full phase space), is equal to 0.5 for configurations that are in the TSE. While biological dynamics are described in the full $6N$-dimensional phase space (coordinates and velocities), when an appropriate set of reaction coordinates $\{\rho_i\}$ is chosen, $P(TP|\{\rho_i\})$ will also be equal to 0.5 for configurations in the TSE. Reducing the dimensionality is often advantageous/required in experimentation and computation. In single-molecule experiments, it is necessary to have accurate low-dimensional representations, because it is only possible to probe a small number of degrees of freedom simultaneously (22,28–31).

From a theoretical perspective, the choice of a particular coordinate can allow for a more intuitive description to emerge, where common low-dimensional projections are based on atomic distances (7,32–34) or collective variables (35–43). Regardless of the technique, it is important that the true dynamics are not masked when projected onto a specific coordinate. To determine whether there is an experimentally accessible one-dimensional coordinate that captures the essential features of aa-tRNA accommodation, we use long molecular dynamics simulations of aa-tRNA accommodation movement, and from them calculate $P(TP|\rho)$, i.e., the conditional probability of being on a transition path as a function of each reaction coordinate ρ. As already mentioned, when a coordinate accurately measures movement over the free-energy barrier, $P(TP|\rho)$ will reach a peak value of 0.5 (44,45). Accordingly, we have used $P(TP|\rho)$ as an unbiased metric to measure each coordinate's ability to accurately capture barrier-crossing events, while also pinpointing the location of the TSE. From this analysis, we have identified numerous coordinates that can be used in experiments or simulations to more precisely characterize the free-energy landscape of aa-tRNA accommodation.

MATERIALS AND METHODS

Energy function

We simulated the movement of aa-tRNA elbow accommodation using an all-atom structure-based SMOG model (46,47). All nonhydrogen atoms were explicitly represented, and the A/A configuration (PDB:3I8F (48)) was defined as the global potential energy minimum, similar to our earlier simulations of the ribosome (25). Input files for the model were generated using the SMOG-SERVER.ORG webtool (47). For completeness, we provide a detailed description of the energy function used here, including all parameters necessary to reproduce the model. All calculations employed reduced units, where each atom is represented as a single bead of unit mass. The covalent geometry is maintained through harmonic interactions that maintain bond lengths, bond angles, improper dihedral angles, and planar dihedral angles. Nonbonded atom pairs that are in contact in the native configuration (in this case, the A/A configuration) between residues i and j, where $i > j + 3$ for proteins and $i \neq j$ for RNA, are given a 6-12 potential, where the minimum corresponds to the distance in the A/A configuration. All other nonlocal interactions between atoms that do not interact through bonded terms are repulsive. Contacts were defined according to a 4 Å cutoff criterion (49). The functional form of the potential is

$$\begin{aligned} V = & \sum_{\text{bonds}} \epsilon_r (r_i - r_{i,o})^2 + \sum_{\text{angles}} \epsilon_\theta (\theta_i - \theta_{i,o})^2 \\ & + \sum_{\text{impropers}} \epsilon_{\chi_{\text{imp}}} (\chi_i - \chi_{i,o})^2 + \sum_{\text{planar}} \epsilon_{\chi_{\text{planar}}} (\chi_i - \chi_{i,o})^2 \\ & + \sum_{\text{backbone}} \epsilon_{BB} F_D(\phi_i) + \sum_{\text{side chains}} \epsilon_{SC} F_D(\phi_i) \\ & + \sum_{\text{contacts}} \epsilon_C \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r}\right)^{6} \right] \\ & + \sum_{\text{noncontacts}} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12}, \end{aligned} \tag{1}$$

where

$$\begin{aligned} F_D(\phi) = & \left[1 - \cos\left(\phi_i - \phi_{i,o}\right)\right] \\ & + 1/2\left[1 - \cos\left(3\left(\phi_i - \phi_{i,o}\right)\right)\right] \end{aligned} \tag{2}$$

and

$$\epsilon_r = 50\epsilon_0,\ \epsilon_\theta = 40\epsilon_0,\ \epsilon_{\chi_{\text{imp}}} = 10\epsilon_0,\ \epsilon_{\chi_{\text{planar}}} = 40\epsilon_0,\ \text{and}\ \epsilon_{NC} = 0.1\epsilon_0$$

($\epsilon_0 = 1$ for all calculations).

$$r_{i,o},\ \theta_{i,o},\ \chi_{i,o},\ \phi_{i,o},\ \text{and}\ \sigma_{i,j}$$

are given the values found in the native state, $\sigma_{NC} = 2.5$ Å. To assign dihedral interaction weights, dihedrals are first grouped if they share a common middle bond. In a protein backbone, there are up to four dihedral angles that possess the same $C - C_\alpha$ bond as the middle bond. Each dihedral group is given a summed weight of ϵ_{BB} or ϵ_{SC}. $R_{BB/SC}$ was defined as $\epsilon_{BB}/\epsilon_{SC}$ and was set to 1 for nucleic acid dihedrals and 2 for protein dihedrals. The weights of protein and nucleic acid backbone dihedrals were set to be equal. Finally, dihedral strengths and contact strengths were scaled such that the ratio of total contact energy to total dihedral energy is

$$R_{C/D} = \frac{\sum \epsilon_C}{\sum \epsilon_{BB} + \sum \epsilon_{SC}} = 2,$$

and the total stabilizing energy is

$$\sum \epsilon_C + \sum \epsilon_{BB} + \sum \epsilon_{SC} = N\epsilon_0,$$

where N is the number of atoms in the system.

After generating the default structure-based model, there was one atypical modification that was necessary to obtain sufficient statistics of elbow-accommodation events. Specifically, intermolecular contacts in the A/A configuration between aa-tRNA and the ribosome (rRNA and proteins) were rescaled by a factor of $F_{A/A}$. The precise value of $F_{A/A}$ was determined phenomenologically. That is, single-molecule experiments indicate that the aa-tRNA elbow undergoes reversible fluctuations between A/T and A/A-like configurations before full aa-tRNA entry into the ribosome (23).

To describe these fluctuations, multiple values of $F_{A/A} < 1.0$ were employed in shorter trajectories to identify a value for which reversible fluctuations were observed in the simulation. Using a factor $F_{A/A} < 1$ can be interpreted in terms of effective energetics. Specifically, structure-based models define the crystal structure as the lowest effective-energy configuration, based on the observation that the crystallized structure necessarily represents a stable configuration of the complex. However, because tRNA molecules only transiently associate with the ribosome and undergo rearrangements between binding sites, the tRNA-ribosome interactions are effectively less stable than interactions that maintain the secondary and tertiary structures of the ribosome. By rescaling tRNA-ribosome interactions by a factor of $F_{A/A} < 1$, our model mimics this difference in effective stabilities.

Simulation details and boundary conditions

By using modified weights for the bond angle and bond lengths, relative to our earlier implementation of the model (46,47), we were able to use a timestep of 0.002 (reduced units) to be used without the introduction of numerical artifacts, whereas earlier simulations employed a timestep of 0.0005. Control simulations were performed for small proteins to ensure that the dynamics are not sensitive to these modifications. To reduce computational demand, only atoms near the accommodation corridor were included in each calculation. To avoid the introduction of boundary effects, isotropic spatial harmonic restraints were imposed on each atom i possessing native interactions with atoms in the full ribosome which were not included in the truncated system. The value of the restraint on each atom k_i was iteratively tuned until the mean-squared fluctuations of each atom i in the truncated system was consistent with the values obtained from simulations of a complete ribosome (see the Supporting Material for details). Each simulation was performed using the GROMACS Ver. 4.6.1 software package (50,51), and constant temperature was maintained through the use of Langevin dynamics protocols. Each simulation was performed using 128 computer cores of the Texas Advanced Computing Center STAMPEDE Cluster and the Northeastern University Discovery Cluster.

To test the robustness of the analysis of reaction coordinates, we performed two independent simulations at different temperatures (0.42 and 0.5, in reduced units). Earlier calculations showed that the scale of atomic fluctuations at these temperatures are consistent with fluctuations observed in all-atom explicit-solvent simulations (25). Because there is a decrease in configurational entropy as the aa-tRNA accommodates, increasing the temperature stabilizes the A/T ensemble in our model, relative to the A/A ensemble. To compensate for this and ensure that the A/A and A/T ensembles have comparable free energies, $F_{A/A}$ was larger for the higher-temperature simulation. For the lower-temperature simulation ($T = 0.42$), $F_{A/A}$ was set to 0.25, and 2×10^{10} timesteps were simulated (95 barrier crossing events). For the higher temperature simulation ($T = 0.5$), $F_{A/A}$ was set to 0.375, and 1.5×10^{10} timesteps were simulated (38 barrier crossing events). All analysis was performed for both parameter-temperature sets, which showed robustness of the results.

RESULTS

Using an all-atom structure-based model (46,47) of the ribosome in complex with aa-tRNA^{Phe}, we simulated the first step of aa-tRNA accommodation (Fig. 1) and generated a trajectory with ≈ 100 spontaneous barrier crossing events. To identify whether any atom-atom distances $R_{i,j}$ accurately capture the accommodation TSE, we calculated the probability of being on a transition path $P(TP|R_{i,j})$ for a wide range of atom pairs i,j (Fig. 2). We focused on distances between Uracil residues because these residues are often accessible for labeling with Förster resonance energy transfer (FRET) dyes.

Before discussing the results, it is valuable to first consider the energetic model that serves as a foundation for the simulations. In this study, all nonhydrogen atoms are explicitly represented, and the potential energy landscape has a global minimum corresponding to the A/A configuration (i.e., the endpoint of accommodation; Fig. 1). Because the global minimum is assigned based on a crystallographic structure, the models are commonly referred to as "structure-based energy functions" (52). In this model, electrostatics and solvation effects are implicitly accounted for, insofar as the stabilizing interactions in the model describe the effective atomic interactions after averaging over all interactions that stabilize the A/A configuration.

This effective description is consistent with the observation that the crystallographic structure represents a stable configuration (i.e., free-energy minimum) on the ribosome's landscape, allowing the tRNA to act as a molecular spring (53) that accumulates and releases strain energy during the elongation cycle. Because our model does not include nonspecific energetic roughness, the primary contributors to the simulated dynamics are molecular flexibility and steric effects. As described below, these two properties are

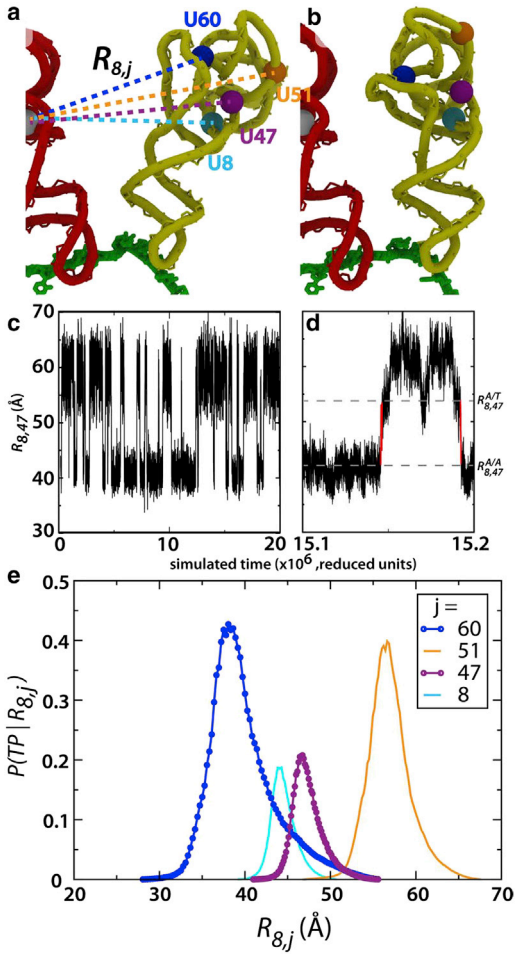

FIGURE 2 Different atomic distances $R_{i,j}$ capture the transition state with varied accuracy. (*a*) The A/T configuration of the aa-tRNA, shown with four Uracil residues (*spheres*). Several $R_{i,j}$ distances, where i is the P-site tRNA residue index (= 8) and j is the aa-tRNA residue index (*dotted lines*). The O3′ of U8 on the P-site tRNA (*white sphere*). (*b*) The elbow-accommodated configuration, shown with the same representation as panel *a*. By comparing the dynamics projected along different coordinates $R_{i,j}$, one may assess the ability of each to capture the transition state associated with aa-tRNA elbow accommodation. (*c*) $R_{8,47}$, which is monitored in our smFRET studies, is shown as a function of time for one-half of the simulated trajectory. There is a visible distinction between the A/T ($R_{8,47} \approx 60$ Å) and elbow-accommodated ($R_{8,47} \approx 40$ Å) ensembles, where transitions appear to be two-state in character. (*d*) Same as panel *c*, with the timescale dilated. (*Dashed lines*) $R^{A/T}_{8,47}$ and $R^{A/A}_{8,47}$. Two transition paths (*red*), as identified with $R_{8,47}$. (*e*) The conditional probability of being on a transition path as a function of each interatomic distance, $P(TP|R_{i,j})$, is shown for distances between U8 of the P-site tRNA and different aa-tRNA residues j. Of the 169 coordinates studied, $R_{8,60}$ most precisely captures transition paths and performs significantly better than the experimentally employed coordinate $R_{8,47}$. While $P(TP|R_{8,47})$ reaches a maximum value $P^{TP}_{\max}$ of only 0.2, $P(TP|R_{8,60})$ exceeds 0.4 (*blue*).

sufficient to ensure that there is a single dominant sequence of events during accommodation. In addition, because the representation of sterics and flexibility is consistent between this model and all-atom explicit-solvent simulations (see the Supporting Material of Whitford et al. (25)), this model provides an overall description that is consistent with more computationally demanding methods.

Using an all-atom structure-based (SMOG (46,47)) model, our earlier simulations of the ribosome found a specific ordering of conformational events during aa-tRNA accommodation (25), and the A/A-ensemble was only marginally more stable than the A/T ensemble. Both features are supported by observations from experimentation, as described below. Although the model describes the potential energy as having a global minimum at the A/A (aa-tRNA accommodated) configuration, steric and entropic contributions give rise to distinct free-energy barriers. By characterizing the dynamics for a range of parameters, it was shown that the ordering of conformational events is model-independent and follows a specific sequence (Fig. 1):

1. aa-tRNA elbow accommodation,
2. aa-tRNA arm accommodation, and
3. 3′-CCA entry into the peptidyltransferase center (PTC).

This ordering was also observed in simulations that employed targeted MD protocols with an explicit-solvent force field (24). Physically, the late entry of the 3′-CCA tail can be understood as being the result of molecular flexibility. Specifically, as the 3′-CCA tail enters the ribosome, a large decrease in configurational entropy serves as a penalizing factor that delays entry of the tail. Consistent with the flexibility exhibited by this model, flexibility of the aa-tRNA 3′-CCA end has been noted in crystallographic studies, where it has elevated B-factors (54) (or is not resolved (55)), and cryo-EM reconstructions typically have lower densities for the tail (56), suggesting conformational heterogeneity within each identified state. Evidence of sequential motion during accommodation also comes from single-molecule FRET experiments, where changes in the amino-acid composition impact elongation rates while having negligible impact on elbow movement (57). Other single-molecule (sm) FRET measurements suggest that the elbow-accommodated conformation is only slightly more stable than the A/T ensemble, which leads to reversible accommodation-like fluctuations of the elbow before incorporation of the amino acid (23), as also seen in simulations with the SMOG model. Together, these theoretical and experimental studies strongly implicate a sequential accommodation process that begins with elbow movement into the A site, consistent with the description provided by this simple model.

With evidence implicating a separation of aa-tRNA elbow and 3′-CCA accommodation, we will focus our attention on the former. One reason for doing so is that because smFRET experiments can monitor elbow displacements (23,58,59), insights from the presented analysis have immediate experimental utility. In contrast, we are unaware of single-molecule approaches that can directly monitor movement of the 3′-CCA end. An additional reason for first focusing on elbow movement is that simulations have shown there are

at least three accessible routes for 3′-CCA entry into the PTC (25). Although it may be possible to extend the methods discussed here and construct collective coordinates that can distinguish between these routes, it is unclear how such coordinates would aid the design of more precise experiments.

In the following sections, we systematically compare the ability of 256 different atomic distances ($R_{i,j}$) to capture transition events associated with aa-tRNA elbow accommodation. By subjecting each coordinate to several tests, we identified 21 high-performing coordinates that may be explored using single-molecule methods. The following criteria were used as metrics to identify optimal coordinates:

1. The atomic distances in the A/T and elbow-accommodated configurations differ by at least 10 Å, which focuses the discussion on coordinates that are likely to yield detectable changes in FRET signal;
2. The number of transition paths is minimized (i.e., the number of false-positives is minimized);
3. $P(TP|R_{i,j})$ is maximized; and
4. Movement along $R_{i,j}$ exhibits diffusive characteristics.

We will close the discussion by providing a brief structural description of the TSE, as identified by the optimal coordinate.

Identifying transition paths

The first criterion that we used to compare coordinates was that an optimal coordinate should maximally separate the endpoints (i.e., minimize the number of identified transitions N_T). For a poorly chosen coordinate, nonreactive (i.e., nonbarrier crossing) fluctuations may be perceived to be transitions between the endpoints. To detect these events, we use simulations to directly count the number of projected transitions along each coordinate. To monitor for barrier-crossing events, we first defined the endpoints by two reference configurations: 1), the A/T configuration (60), and 2), a structural model of an A/A-elbow configuration (Fig. 1). Each interatomic pair distance $R_{i,j}$ (distance between the O3′ atom of residue i on the P-site tRNA and residue j of the aa-tRNA) was then calculated for the endpoint configurations ($R_{i,j}^{A/T}$ and $R_{i,j}^{A/A}$). Next, we calculated the distance as a function of simulated time $R_{i,j}(t)$.

A transition path (TP) was defined as having occurred when $R_{i,j}(t)$ moved from $R_{i,j}^{A/T}$ to $R_{i,j}^{A/A}$ without recrossing $R_{i,j}^{A/T}$, or when it moved from $R_{i,j}^{A/A}$ to $R_{i,j}^{A/T}$ without recrossing $R_{i,j}^{A/A}$. It is important to note that the results presented below were not sensitive to minor (10–20%) changes in the endpoint distances used for TP analysis. To further confirm robustness of the results, we repeated the analysis for a second simulation, which was performed at a temperature that was elevated by 20%. Changing temperature significantly perturbs the free-energy landscape, because it alters the balance between energy and configurational entropy. Even though entropic changes have been implicated during accommodation (25), this major alteration to the simulation parameters did not lead to notable differences in our evaluation of coordinates.

The TP analysis was performed for distances between all inter-tRNA U-U pairs. With 16 U residues in each $tRNA^{Phe}$, 256 U-U distances were considered. Of these, $\left|R_{i,j}^{A/T}-R_{i,j}^{A/A}\right|>10$ Å for 169 pairs. For those, we calculated $R_{i,j}(t)$, $R_{i,j}^{A/T}$, and $R_{i,j}^{A/A}$ to evaluate the number of transition paths detected with each. According to our analysis, 40 coordinates minimized the number of crossing events, with $N_T = 95$ (see Table S1 in the Supporting Material). To provide evidence that $N_T = 95$ represents the number of true barrier crossing events, and does not include false positives, we employed a more stringent measure of transitions that is based on all 169 coordinates.

Specifically, we counted the number of times that the tRNA transitioned from a configuration where $R_{i,j}(t)<R_{i,j}^{A/A}$ simultaneously for all 169 coordinates and then later adopted a configuration where $R_{i,j}(t)>R_{i,j}^{A/T}$ for all coordinates (reverse transitions were also counted). Using this strict definition of what constitutes a transition, which describes the collective movements of U residues that are present throughout the elbow and arm regions of the tRNA molecules (see Fig. S5 in the Supporting Material), we found $N_T = 95$, indicating that all 95 transition events apparent with $R_{8,60}$ are common to every coordinate.

Perhaps surprising is that despite its intuitive positioning, the residue pair commonly used in smFRET experiments (U8–U47 (23,58,59)) overestimated the N_T value by >10% ($N_T = 107$). Although one may argue that this casts doubts on the interpretation of earlier experiments, timescale considerations suggest it is unlikely that false positives would have resulted purely from the use of $R_{8,47}$ as a coordinate. Because single-molecule experiments provide measures that are averaged over finite time intervals (typically milliseconds), we performed TP analysis for time-averaged values of $R_{i,j}(t)$. We averaged each $R_{i,j}(t)$ over $N = 2^m$ sampled frames ($M = 1, 2,\ldots,12$) and recounted the number of apparent transitions. For $R_{8,60}$, $N_T = 95$ for $M < 11$ (see Fig. S4). For $M \geq 11$, almost all coordinates capture <95 transitions, because this window is approaching the average dwell time of the endpoint ensembles.

Because there is a range of averaging windows for which $N_T = 95$, if experimentally accessible timescales are not far below the timescale of accommodation ($1/k_{acc}$), our analysis suggests false positives are unlikely to arise from the use of $R_{8,47}$. However, because quantitative methods are being developed that can remove features arising from instrumentation noise and time-averaging (61,62), single-molecule methods may soon begin to probe timescales well below $1/k_{acc}$, at which point artifacts arising from the use of less

optimal coordinates are expected to significantly impact the measured statistics.

Isolating the transition state ensemble

To determine how well each coordinate can capture the TSE, we calculated the conditional probability of being on a transition path as a function of each coordinate $P(TP|R_{i,j})$, which may be related to the splitting probabilities $\phi_A(R_{i,j})$ and $\phi_B(R_{i,j})$ (44,45). For a coordinate describing a transition between states A and B (with $R_{i,j}$ values of $R_{i,j}^A$ and $R_{i,j}^B$), $\phi_B(R_{i,j})$ is the probability that given a particular value of $R_{i,j}$, the system will reach $R_{i,j}^B$ before returning to $R_{i,j}^A$.

When the movement in the full phase space is Markovian and the coordinate accurately captures the underlying barrier, $P(TP|R_{i,j}) = 2\phi_A(R_{i,j})\phi_B(R_{i,j})$. Then, at the TSE ($R_{i,j} \equiv R_{i,j}^{TSE}$), $\phi_B(R_{i,j}^{TSE}) = \phi_A(R_{i,j}^{TSE}) = 0.5$, and $P(TP|R_{i,j}^{TSE}) = 0.5$. When the coordinate does not accurately capture the TSE, nonreactive fluctuations can lead to reduced values of $P(TP|R_{i,j}^{TSE})$. In protein folding studies this property has found significant utility, where maximizing $P(TP|R_{i,j}^{TSE})$ has been used as a strategy for constructing collective reaction coordinates for coarse-grained models (45) and for all-atom explicit-solvent simulations (35,63).

To determine the extent to which a coordinate can uniquely identify the TSE, we calculated $P(TP|R_{i,j})$ for each. We found that for 10 coordinates $P_{\max}^{TP} > 0.4$, whereas most others are relatively poor indicators of when the aa-tRNA is on a transition path, with $P(TP|R_{i,j})$ values <0.3. A striking observation is that out of the top 10 performing coordinates, eight measure the position of U60 in the aa-tRNA ($R_{i,60}$: i = 8, 16, 20, 45, 47, 50, 59, 60). In fact, only three different aa-tRNA residues (j) were implicated in the top 21 coordinates: U51, U59, and U60. This demonstrates that the choice of an appropriate coordinate is not arbitrary, and that results are strongly dependent upon which aa-tRNA residue is being monitored. In stark contrast, there was not a single P-site tRNA residue that consistently performed best. Although the most relevant coordinates are narrowly distributed across the aa-tRNA, there is more degeneracy in the choice of P-site tRNA residue. This finding is critical for the design of more accurate single-molecule experiments. Nonetheless, it is also rather intuitive, because accommodation is associated with movement of the aa-tRNA, and the P-site tRNA is treated as a point of reference.

The dependence of $P(TP|R_{i,j})$ on the choice of aa-tRNA residue being monitored is depicted in Fig. 2, which shows $P(TP|R_{8,j})$ for four different atomic distances involving U8 of the P-site tRNA. In Fig. 2, only the choice of the aa-tRNA residue is varied. As already noted, $R_{8,60}$ is a relatively accurate indicator of when aa-tRNA is on a transition path, where $P_{\max}^{TP} = 0.43$ (*red*). In contrast, movement along $R_{8,47}$ is a not a strong indicator of when the system is on a transition path, with $P_{\max}^{TP} \approx 0.2$. Together, our analysis indicates that $R_{8,47}$ only overestimates the number of crossing events by 10%, making it a reasonable metric for discerning between the endpoints. However, although this allows $R_{8,47}$ to have utility in experiments, its low $P_{\max}^{TP}$ value indicates that it is unlikely to provide precise information regarding the TSE. To resolve this limitation in the experiments, the presented analysis suggests that by labeling aa-tRNA at U60, observed signals will be more strongly correlated with barrier-crossing events.

Because experimental measurements provide time-averaged coordinates ($\overline{R}_{i,j}$), it is interesting to consider the impact of time averaging on the presented calculation of $P(TP|R_{i,j})$. We found that the peak value of $P(TP|\overline{R}_{i,j})$ increases for all coordinates when time-averaging is employed, although this should not necessarily be interpreted as suggesting improved performance of a particular coordinate. That is, time-averaging can introduce memory into the projected dynamics. For example, consider a time-averaged coordinate $\overline{R}$ that initially adopts a value corresponding to state A, followed by a value corresponding to the TSE. In this case, if the time-averaging window is long relative to the timescale of an individual barrier crossing event (which is not to be confused with the mean first passage time), when a TSE value of $\overline{R}$ is observed, it is likely that within a single time window the molecule will have initially resided in state A and transited fully to state B. Under these conditions, the peak value of $P(TP|\overline{R})$ may exceed 0.5. Unfortunately, when averaging over such long times, the dynamics along $\overline{R}$ will not be Markovian, making a diffusive description inappropriate. Nonetheless, we computed $P(TP|\overline{R}_{i,j})$ for relatively short time intervals ($N < 100$) and found that the maximal value was always larger for $\overline{R}_{8,60}$ than for $\overline{R}_{8,47}$, consistent with our analysis of nonaveraged dynamics.

Diffusive versus subdiffusive dynamics at the TSE

Analysis of $P(TP|R_{i,j})$ suggests that motion along some coordinates is likely to be diffusive at the TSE. To directly evaluate the degree to which motion is diffusive, we calculated the displacement squared as a function of the lag time $\langle \delta R_{i,j}^2(\tau) \rangle$ for each coordinate. To specifically characterize the TSE, we only included starting frames (i.e., $\tau = 0$) for which $R_{i,j}(t) = R_{i,j}^{TSE}$. We fit each curve to ${R_0}^2\tau^\alpha$, where the value of α indicates whether the motion is subdiffusive (< 1), diffusive ($= 1$), or superdiffusive (> 1). For $R_{8,47}$ (the coordinate measured in our smFRET experiments), we found a distinctly subdiffusive character, with $\alpha \approx 0.75$, which is consistent with the relatively low value of $P(TP|R_{8,47})$.

In contrast, when the dynamics was projected along $R_{8,60}$, $\alpha \approx 1$, providing further evidence of diffusive dynamics (Fig. 3). When $\alpha = 1$, the diffusion coefficient D is given

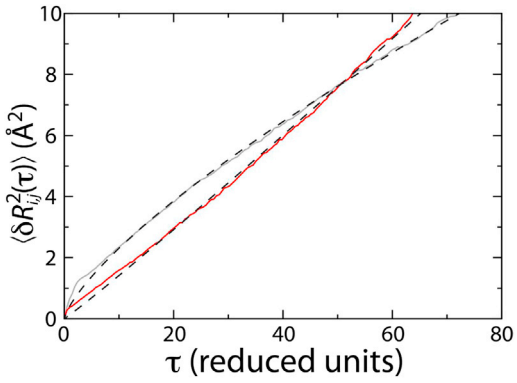

FIGURE 3 The dynamics is diffusive along $R_{8,60}$ at the TSE. To assess the degree to which motion at the TSE is diffusive, the mean-squared displacement was calculated as a function of lag time $\langle \delta R^2_{i,j}(\tau) \rangle$ for $R_{8,47}$ (*gray*) and $R_{8,60}$ (*red*). Each curve was fit to $R^2{}_0\tau^\alpha$ (*dashed lines*). For $R_{8,47}$, the dynamics is subdiffusive, with $\alpha \approx 0.75$. In contrast, the dynamics along $R_{8,60}$ is diffusive ($\alpha \approx 1$), consistent with $P(TP|R_{8,60})$ approaching a value of 0.5 (see Fig. 2).

by $D = R_0^2/2$. Although there is a correlation between $P^{TP}_{\max}$ and α (see Fig. S1), very few coordinates maximize both quantities. Further, some coordinates that do not minimize the number of transitions also have relatively high values of α and $P^{TP}_{\max}$. By considering all three metrics (N_T, $P^{TP}_{\max}$, and α), this analysis strongly supports $R_{8,60}$ as an appropriate coordinate for describing accommodation as diffusive movement across a one-dimensional free-energy barrier.

Kinetics and rate-limiting barriers

Because the projected dynamics along $R_{8,60}$ suggest it is an accurate indicator of when the aa-tRNA is on a transition path, and movement about the TSE is diffusive, we will use it to describe the kinetics of accommodation. Under these conditions, it is possible to interconvert between the free-energy profile $G(\rho)$ and rate k according to (64)

$$1/k = \langle \tau \rangle = \int_{\rho_{A/T}}^{\rho_{A/A}} d\rho \int_{\infty}^{\rho} d\rho' \frac{\exp[(G(\rho) - G(\rho'))/k_B T]}{D_\rho(\rho)}, \tag{3}$$

where $D_\rho(\rho)$ is the diffusion coefficient along ρ as a function of ρ and $\langle \tau \rangle$ is the mean first passage time between points $\rho_{A/T}$ and $\rho_{A/A}$. In our previous study of aa-tRNA diffusion during accommodation, we estimated the diffusion coefficient of the aa-tRNA elbow from explicit-solvent simulations (65). We then used Eq. 3 to infer a relationship between the rates and barriers, under the assumption that the employed coordinate was appropriate. In that study, $R_{8,47}$ was measured. Because we now see that $R_{8,47}$ performs more poorly than $R_{8,60}$, we can illustrate the strength of $R_{8,60}$ by comparing the projected free-energy profiles and observed mean first passage time. From Eq. 3, we calculated the mean first passage time $\langle \tau^{\text{calc}}_{i,j} \rangle$ using the free energy and diffusion coefficient along $R_{8,60}$. For this calculation D_ρ was assumed constant and the value obtained for the TSE was used, because the value at the TSE provides the largest contribution to the integral. For an appropriately chosen coordinate, $\langle \tau^{\text{calc}}_{i,j} \rangle$ will be consistent with the observed time for interconversion between states: $\langle \tau^{\text{obs}} \rangle$. We find that $R_{8,60}$ provides an accurate description of the overall rate with

$$\frac{\left\langle \tau^{\text{calc}}_{8,60} \right\rangle}{\langle \tau^{\text{obs}} \rangle} \approx 0.5 - 1,$$

further supporting its use when describing aa-tRNA accommodation.

Differences between the projections along each coordinate, including reduced values of $P(TP|R_{8,47})$ and subdiffusive dynamics along $R_{8,47}$, can be further rationalized by plotting the two-dimensional potential of mean force: PMF ($R_{8,47}$, $R_{8,60}$) (Fig. 4). In the PMF, there is a visible separation of the basins along $R_{8,60}$ at the TSE

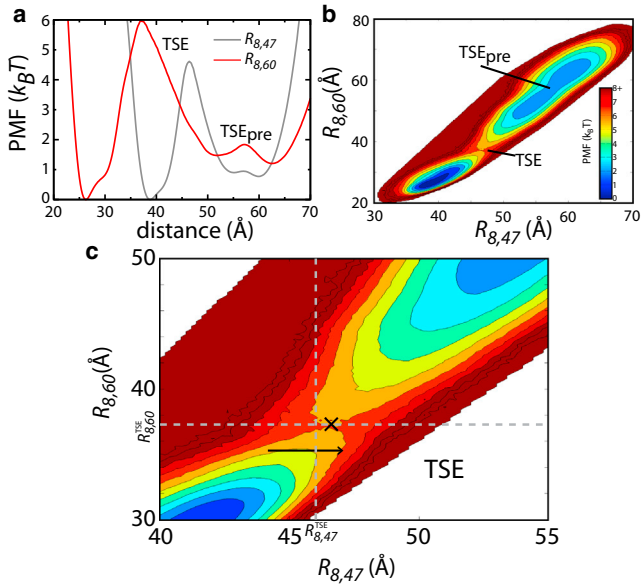

FIGURE 4 The projected free-energy barrier is larger when a more appropriate coordinate is used. (*a*) The PMF, or free energy, as a function of $R_{8,47}$ (*gray*) and $R_{8,60}$ (*red*). The accommodation-related TSE is indicated. Interestingly, there is an additional TSE (TSE_{pre}) that is apparent in the PMF, although it falls outside the bounds of the A/T and A/A configurations, and it therefore does not impact the presented analysis. (*b*) The PMF plotted against both coordinates. (*c*) Zoomed-in perspective of panel *b*. (*Dotted lines*) Positions of the maximum free energy $R^{\text{TSE}}_{8,j}$ when plotted against each coordinate individually, as shown in panel *a*. Here, $R^{TSE}_{8,60} = 37.3$ Å and $R^{TSE}_{8,47} = 46.5$ Å. The reduced value of the free-energy barrier and lower values of $P(TP|R_{8,47})$ is due to the presence of large-scale nonreactive fluctuations (i.e., nonbarrier crossing events) along $R_{8,47}$ that are capable of crossing the apparent TSE without moving to the other free-energy basin. (*Arrow*) Example of such a fluctuation. In contrast, fluctuations along $R_{8,60}$ are more likely to be associated with barrier-crossing transitions. The point of lowest free energy along $R^{TSE}_{8,60}$ (i.e., the most probable point of crossing) is marked with an *X*.

($R^{TSE}_{8,60} = 37.3$ Å). When the aa-tRNA adopts the value of $R_{8,47}$ that corresponds to the TSE ($R^{TSE}_{8,47} = 46.5$ Å), the tRNA is not necessarily at a saddle point on the free-energy surface, but it may be in either of the endpoint basins (A/T, or accommodated elbow basins). As already discussed, this leads to an underestimate of the free energy when projected along $R_{8,47}$, as compared to $R_{8,60}$. To further illustrate this, the values of $R^{TSE}_{8,60}$ and $R^{TSE}_{8,47}$ are indicated by dotted lines in Fig. 4 *c*. The $R^{TSE}_{8,47}$ line visibly intersects the endpoint basins, where the lowest free-energy points along $R^{TSE}_{8,47}$ correspond to the endpoint basins. In contrast, $R_{8,60} = R^{TSE}_{8,60}$ visibly separates the basins and the lowest free-energy point along the line (at $R_{8,47} \approx 46.7$ Å) corresponds to a saddle point on the surface, consistent with it being a free-energy TSE.

Structural composition of the TSE

In Fig. 5 we show two representative structural snapshots taken from the TSE, as identified by $R_{8,60}$. In the TSE, the aa-tRNA minor groove, or major groove, closely interacts with the stem loop of H89, consistent with previous biochemical and simulation studies (24,25,66). In contrast to earlier simulation studies, the identification of the appropriate coordinate and the associated TSE here was based entirely on the properties of the projected dynamics, and not construction of coordinates that are crafted for illustrative purposes. Visual inspection of earlier simulated trajectories, and the use of $R_{8,47}$ as a coordinate, generally suggested that H89 acts as a steric barrier during elbow accommodation (25). However, as shown in Fig. 4, for TSE configurations identified with $R_{8,60}$ (= 37 Å), there are multiple minima along $R_{8,47}$ (ranging from 43 to 50 Å). This inherent uncertainty that is introduced by $R_{8,47}$ is alleviated by moving to an alternate coordinate, $R_{8,60}$.

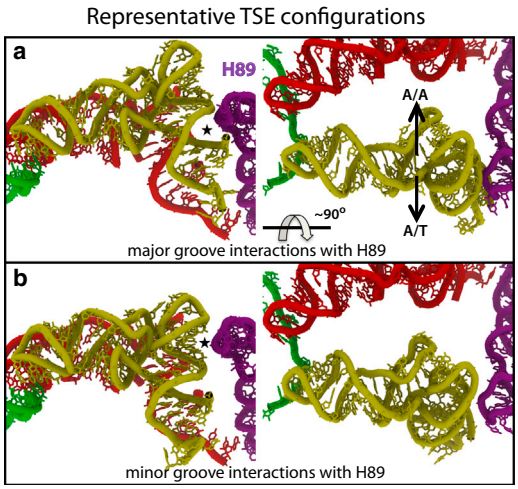

FIGURE 5 Two representative configurations from the TSE. Both configurations correspond to the point of lowest free energy along $R^{TSE}_{8,60}$, marked by an *X* in Fig. 4 *c*. (*a*) A TSE configuration where the major groove of aa-tRNA (*yellow*) interacts with H89 (*purple*). P-site tRNA (*red*) and mRNA (*green*). (*b*) A TSE configuration where the aa-tRNA minor groove contacts H89. (*Star*) The relevant aa-tRNA-H89 interaction. The same configurations are shown in the right panels, each rotated by ≈90°. (*Arrows*) Relative positions of the aa-tRNA in the A/A and A/T configurations.

DISCUSSION

As we seek to develop deeper insights into the origins of the energy landscapes that govern biological function, theoretical concepts will prove invaluable. In the case of the ribosome, the theoretical framework for interpreting and analyzing its dynamics is nascent, relative to other biological processes such as protein folding. Fortunately, the theoretical tools developed for folding are grounded in the statistical mechanics of biomolecular dynamics, allowing them to be adapted and applied to other biological phenomena. Here, we have demonstrated how the analysis of projected dynamics can be used to identify the kinetically and thermodynamically relevant coordinates for describing the ribosome's energy landscape. These results provide a clear foundation, upon which to devise new experiments that can be unambiguously compared to theoretical and computational predictions. This integration of experimental and theoretical tools will be mutually beneficial, and will allow the field to precisely describe the energetic properties of increasingly complex biological phenomena.

SUPPORTING MATERIAL

One table, five figures and additional supplemental information are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)01073-X.

We thank Ryan Hayes at Rice University for discussions on the use of longer timesteps. We also acknowledge generous support provided by the Northeastern University Discovery Cluster.

The authors declare no competing financial interests.

This work was supported in part by a National Science Foundation CAREER Award (grant No. MCB-1350312) and by the Center for Theoretical Biological Physics, sponsored by the National Science Foundation (under grants No. PHY-1427654 and No. NSF-MCB-1214457). This work was supported in part by the Welch Foundation (grant No. C-1792). V.B.P.L. has been supported by Fundação de Amparo à Pesquisa do Estado de São Paulo and Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil). This research was supported in part by the National Science Foundation through XSEDE resources provided by the Texas Advanced Computing Center under grant No. TG-MCB110021.

REFERENCES

1. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
2. Thirumalai, D., and G. H. Lorimer. 2001. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 30:245–269.
3. Hyeon, C., and D. Thirumalai. 2011. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* 2:487.

4. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
5. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* 14:76–88.
6. Gopich, I. V., and A. Szabo. 2007. Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B.* 111:12925–12932.
7. Gopich, I. V., D. Nettels, …, A. Szabo. 2009. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* 131:095102.
8. Nymeyer, H., S. Gnanakaran, and A. E. García. 2004. Atomic simulations of protein folding, using the replica exchange algorithm. *Methods Enzymol.* 383:119–149.
9. Gnanakaran, S., H. Nymeyer, …, A. E. García. 2003. Peptide folding simulations. *Curr. Opin. Struct. Biol.* 13:168–174.
10. Chan, H. S., Z. Zhang, …, Z. Liu. 2011. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* 62:301–326.
11. Zhang, Z., and H. S. Chan. 2012. Transition paths, diffusive processes, and preequilibria of protein folding. *Proc. Natl. Acad. Sci. USA.* 109:20919–20924.
12. Olsson, U., and M. Wolf-Watz. 2010. Overlap between folding and functional energy landscapes for adenylate kinase conformational change. *Nat. Commun.* 1:111.
13. Grant, B. J., A. A. Gorfe, and J. A. McCammon. 2010. Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.* 20:142–147.
14. Hopfield, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA.* 71:4135–4139.
15. Pape, T., W. Wintermeyer, and M. Rodnina. 1999. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* 18:3800–3807.
16. Valle, M., J. Sengupta, …, J. Frank. 2002. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* 21:3557–3567.
17. Valle, M., A. Zavialov, …, J. Frank. 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat. Struct. Biol.* 10:899–906.
18. Schmeing, T. M., and V. Ramakrishnan. 2009. What recent ribosome structures have revealed about the mechanism of translation. *Nature.* 461:1234–1242.
19. Demeshkina, N., L. Jenner, …, G. Yusupova. 2012. A new understanding of the decoding principle on the ribosome. *Nature.* 484:256–259.
20. Johansson, M., M. Lovmar, and M. Ehrenberg. 2008. Rate and accuracy of bacterial protein synthesis revisited. *Curr. Opin. Microbiol.* 11:141–147.
21. Mittelstaet, J., A. L. Konevega, and M. V. Rodnina. 2011. Distortion of tRNA upon near-cognate codon recognition on the ribosome. *J. Biol. Chem.* 286:8158–8164.
22. Petrov, A., G. Kornberg, …, J. D. Puglisi. 2011. Dynamics of the translational machinery. *Curr. Opin. Struct. Biol.* 21:137–145.
23. Geggier, P., R. Dave, …, S. C. Blanchard. 2010. Conformational sampling of aminoacyl-tRNA during selection on the bacterial ribosome. *J. Mol. Biol.* 399:576–595.
24. Sanbonmatsu, K. Y., S. Joseph, and C.-S. Tung. 2005. Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA.* 102:15854–15859.
25. Whitford, P. C., P. Geggier, …, K. Y. Sanbonmatsu. 2010. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA.* 16:1196–1204.
26. Eargle, J., A. A. Black, …, Z. Luthey-Schulten. 2008. Dynamics of recognition between tRNA and elongation factor Tu. *J. Mol. Biol.* 377:1382–1405.
27. Chen, K., J. Eargle, …, Z. Luthey-Schulten. 2010. Functional role of ribosomal signatures. *Biophys. J.* 99:3930–3940.
28. Ermolenko, D. N., P. C. Spiegel, …, H. F. Noller. 2007. The antibiotic viomycin traps the ribosome in an intermediate state of translocation. *Nat. Struct. Mol. Biol.* 14:493–497.
29. Fei, J., P. Kosuri, …, R. L. Gonzalez, Jr. 2008. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol. Cell.* 30:348–359.
30. Munro, J. B., R. B. Altman, …, S. C. Blanchard. 2010. Spontaneous formation of the unlocked state of the ribosome is a multistep process. *Proc. Natl. Acad. Sci. USA.* 107:709–714.
31. Nettels, D., I. V. Gopich, …, B. Schuler. 2007. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA.* 104:2655–2660.
32. Best, R. B., E. Paci, …, O. K. Dudko. 2008. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B.* 112:5968–5976.
33. Dudko, O. K., T. G. Graham, and R. B. Best. 2011. Locating the barrier for folding of single molecules under an external force. *Phys. Rev. Lett.* 107:208301.
34. Hyeon, C., G. Morrison, …, D. Thirumalai. 2009. Refolding dynamics of stretched biopolymers upon force quench. *Proc. Natl. Acad. Sci. USA.* 106:20288–20293.
35. Best, R. B., G. Hummer, and W. A. Eaton. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA.* 110:17874–17879.
36. Cho, S. S., Y. Levy, and P. G. Wolynes. 2006. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA.* 103:586–591.
37. Krivov, S. V., and M. Karplus. 2008. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. USA.* 105:13841–13846.
38. Das, P., M. Moll, …, C. Clementi. 2006. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. USA.* 103:9885–9890.
39. Tama, F., M. Valle, …, C. L. Brooks, 3rd. 2003. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA.* 100:9319–9323.
40. Kurkcuoglu, O., Z. Kurkcuoglu, …, R. L. Jernigan. 2009. Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. *Proteins.* 75:837–845.
41. Rohrdanz, M. A., W. Zheng, …, C. Clementi. 2011. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* 134:124116.
42. Hinczewski, M., Y. von Hansen, …, R. R. Netz. 2010. How the diffusivity profile reduces the arbitrariness of protein folding free energies. *J. Chem. Phys.* 132:245103.
43. Bock, L. V., C. Blau, …, H. Grubmüller. 2013. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat. Struct. Mol. Biol.* 20:1390–1396.
44. Hummer, G. 2004. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* 120:516–523.
45. Best, R. B., and G. Hummer. 2005. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. USA.* 102:6732–6737.
46. Whitford, P. C., J. K. Noel, …, J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins.* 75:430–441.
47. Noel, J. K., P. C. Whitford, …, J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38:W657–W661.
48. Jenner, L., N. Demeshkina, …, M. Yusupov. 2010. Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat. Struct. Mol. Biol.* 17:1072–1078.

49. Noel, J. K., P. C. Whitford, and J. N. Onuchic. 2012. The Shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B.* 116:8692–8702.

50. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.

51. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

52. Whitford, P. C., K. Y. Sanbonmatsu, and J. N. Onuchic. 2012. Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep. Prog. Phys.* 75:076601.

53. Frank, J., J. Sengupta, …, M. Ehrenberg. 2005. The role of tRNA as a molecular spring in decoding, accommodation, and peptidyl transfer. *FEBS Lett.* 579:959–962.

54. Shi, H., and P. B. Moore. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA.* 6:1091–1105.

55. Byrne, R. T., A. L. Konevega, …, A. A. Antson. 2010. The crystal structure of unmodified tRNAPhe from *Escherichia coli*. *Nucleic Acids Res.* 38:4154–4162.

56. Fischer, N., A. L. Konevega, …, H. Stark. 2010. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature.* 466:329–333.

57. Effraim, P. R., J. Wang, …, V. W. Cornish. 2009. Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. *Nat. Chem. Biol.* 5:947–953.

58. Blanchard, S. C., R. L. Gonzalez, …, J. D. Puglisi. 2004. tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.* 11:1008–1014.

59. Lee, T.-H., S. C. Blanchard, …, S. Chu. 2007. The role of fluctuations in tRNA selection by the ribosome. *Proc. Natl. Acad. Sci. USA.* 104:13661–13665.

60. Schmeing, T. M., R. M. Voorhees, …, V. Ramakrishnan. 2009. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science.* 326:688–694.

61. von Hansen, Y., A. Mehlich, …, R. R. Netz. 2012. Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev. Sci. Instrum.* 83:095116.

62. Hinczewski, M., J. C. M. Gebhardt, …, D. Thirumalai. 2013. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl. Acad. Sci. USA.* 110:4500–4505.

63. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.

64. Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.

65. Whitford, P. C., J. N. Onuchic, and K. Y. Sanbonmatsu. 2010. Connecting energy landscapes with experimental rates for aminoacyl-tRNA accommodation in the ribosome. *J. Am. Chem. Soc.* 132:13170–13171.

66. Petrov, A. N., A. Meskauskas, …, J. D. Dinman. 2008. Yeast ribosomal protein L10 helps coordinate tRNA movement through the large subunit. *Nucleic Acids Res.* 36:6187–6198.

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.

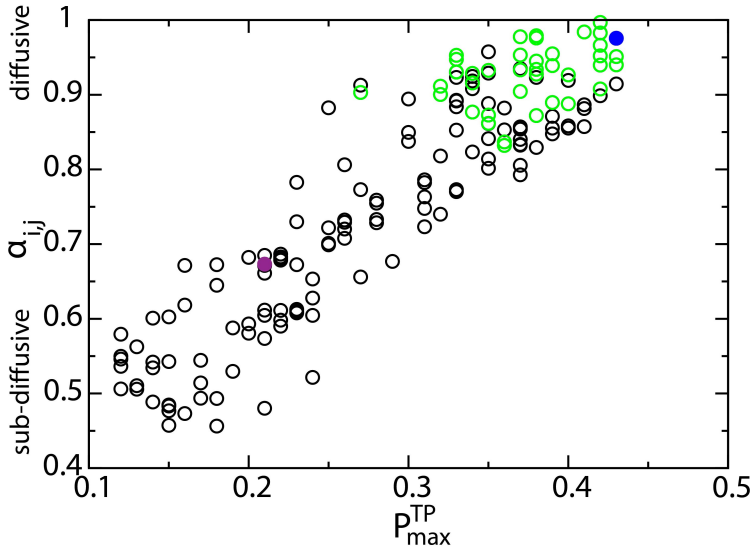

Figure S1: For all 169 coordinates for which $|R^{AA}_{i,j} - R^{AT}_{i,j}| > 10$ Å, the maximum value of $P(TP|R_{i,j})$ (P^{TP}_{max}) and the mean squared displacement was calculated. The mean squared displacement was fit to $R_0^2\tau^{\alpha}$, where the value of α implicates diffusive, subdiffusive 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.

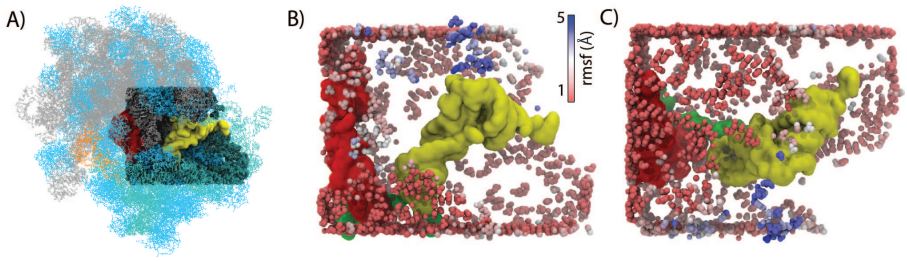

Figure S2: A) Structure of an A/T configuration of aa-tRNA, in the context of the full ribosome. For the simulations here, a subset of atoms was included, as depicted by spheres. B) To ensure that boundary effects were not introduced by imposing overly restrictive boundary conditions, atoms at the edges 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.

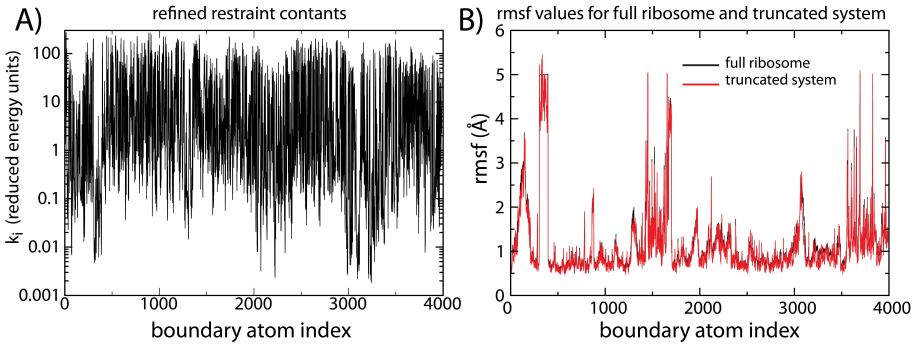

Figure S3: A) Values of the harmonic spatial restraint constants k_i imposed on boundary atoms after all refinement steps were completed. Refinement of boundary restraints resulted in k_i values that span four orders of magnitude, which highlights the need for refinement protocols. B) Comparison of spatial 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.

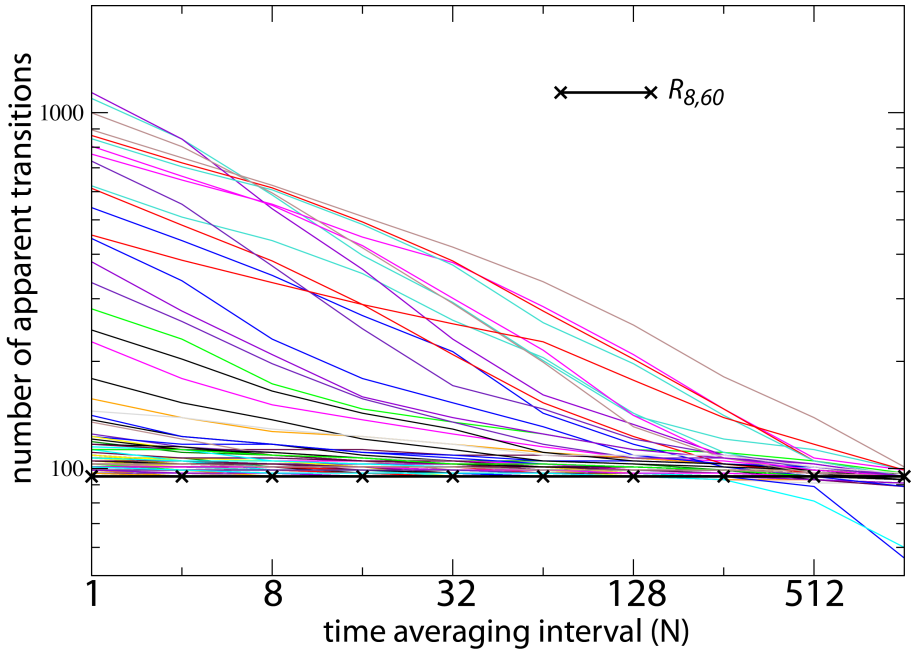

Figure S4: Number of apparent transition events N_T for all 169 coordinates $R_{i,j}(t)$, after time averaging over $N = 2^M$ frames. Each line represents a different coordinate. The number of transitions is robust for $R_{8,60}$ (marked line), whereas the number of false positives identified with alternate 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.

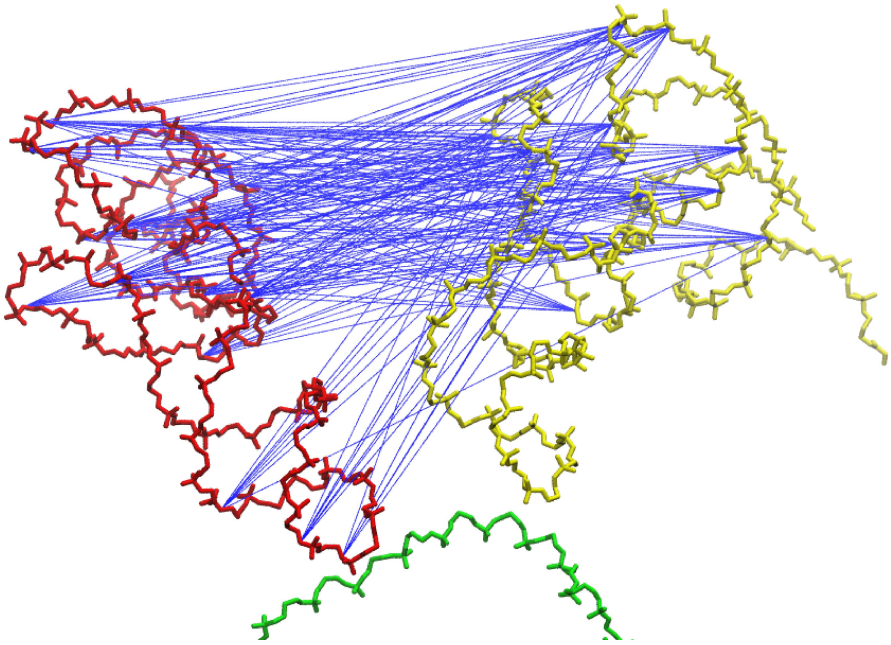

Figure S5: Structural depiction of all 169 coordinates evaluated. Each blue line indicates a unique distance that was calculated and characterized. For clarity, only the tRNA and mRNA backbone atoms are shown.

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 ($N_T = 95$)