Biophysical Journal
Computational Tool

Biophysical Society

Bayesian-Estimated Hierarchical HMMs Enable Robust Analysis of Single-Molecule Kinetic Heterogeneity

Jason Hon[1] and Ruben L. Gonzalez, Jr.[1,*]
[1]Department of Chemistry, Columbia University, New York, New York

ABSTRACT Single-molecule kinetic experiments allow the reaction trajectories of individual biomolecules to be directly observed, eliminating the effects of population averaging and providing a powerful approach for elucidating the kinetic mechanisms of biomolecular processes. A major challenge to the analysis and interpretation of these experiments, however, is the kinetic heterogeneity that almost universally complicates the recorded single-molecule signal versus time trajectories (i.e., signal trajectories). Such heterogeneity manifests as changes and/or differences in the transition rates that are observed within individual signal trajectories or across a population of signal trajectories. Because characterizing kinetic heterogeneity can provide critical mechanistic information, we have developed a computational method that effectively and comprehensively enables such analysis. To this end, we have developed a computational algorithm and software program, hFRET, that uses the variational approximation for Bayesian inference to estimate the parameters of a hierarchical hidden Markov model, thereby enabling robust identification and characterization of kinetic heterogeneity. Using simulated signal trajectories, we demonstrate the ability of hFRET to accurately and precisely characterize kinetic heterogeneity. In addition, we use hFRET to analyze experimentally recorded signal trajectories reporting on the conformational dynamics of ribosomal pre-translocation (PRE) complexes. The results of our analyses demonstrate that PRE complexes exhibit kinetic heterogeneity, reveal the physical origins of this heterogeneity, and allow us to expand the current model of PRE complex dynamics. The methods described here can be applied to signal trajectories generated using any type of signal and can be easily extended to the analysis of signal trajectories exhibiting more complex kinetic behaviors. Moreover, variations of our approach can be easily developed to integrate kinetic data obtained from different experimental constructs and/or from molecular dynamics simulations of a biomolecule of interest.

INTRODUCTION

The kinetic mechanism of a biomolecular process is typically described by specifying the number of states that the biomolecular system samples, the order in which these states are sampled, and the rates of transitions between the sampled states. Over the past 20 years, single-molecule kinetic experiments have emerged as a powerful tool for elucidating such mechanisms (1,2). This is because the signal versus time trajectories (i.e., signal trajectories) that are recorded in such experiments report on the real-time transitions between the states sampled by an individual biomolecule and are therefore free of the population averaging that frequently confounds the analysis of ensemble kinetic experiments. Despite the mechanistically unique and valuable information they provide, single-molecule signal trajectories generally exhibit kinetic heterogeneity, a phenomenon that complicates trajectory analysis and can result in elucidation of incomplete or incorrect kinetic mechanisms (2–4). Kinetic heterogeneity in a single-molecule kinetic experiment manifests as stochastic, abrupt changes in the rates of transitions observed in individual signal trajectories (i.e., dynamic heterogeneity) and/or as differences in the rates of transitions observed between distinct subpopulations of signal trajectories (i.e., static heterogeneity) (3,5–12). These effects arise because the signal trajectories recorded in a typical single-molecule kinetic experiment directly detect transitions along only one dimension (i.e., the directly detected dimension) of the complex, multidimensional, free-energy landscape that generally governs a biomolecular process (13). Consequently, transitions along dimensions other than the directly detected dimension (i.e., the indirectly detected dimensions) are projected onto the directly detected dimension, where they materialize

Submitted August 30, 2018, and accepted for publication February 13, 2019.

*Correspondence: rlg2118@columbia.edu

Editor: Tamar Schlick.
https://doi.org/10.1016/j.bpj.2019.02.031

indirectly as changes and/or differences in the rates of transitions that are observed in the signal trajectories (Fig. 1).

Detecting the presence of transitions along the indirectly detected dimensions of a free-energy landscape and modeling the kinetics of these transitions is of great mechanistic interest. This is because doing so allows identification and characterization of states and subpopulations of a biomolecular system and/or pathways of a biological process that would otherwise be excluded from the kinetic mechanism that is elucidated (e.g., (5,6,12,14)). Despite its importance, however, detecting and modeling the kinetics of such transitions remains one of the most significant challenges in the analysis and interpretation of single-molecule kinetic experiments (3,5,6,12,14–16). This is due to inherent limitations (17–23) in the conventional hidden Markov model (HMM)-based approaches that are widely used to analyze signal trajectories recorded using all of the currently available experimental methods, including single-molecule patch-clamp (24), fluorescence resonance energy transfer (FRET) (1,25,26), force spectroscopy (1,27), and field-effect transistor (7,28–31) experiments. Because the noisy, discretely sampled signal trajectories that are recorded in such experiments (2,32,33) can be described, to a good approximation, as discrete-time Markov chains (17,34,35), HMMs have become useful tools for the analysis of these experiments. Nonetheless, it is important to note that the kinetic model employed by an HMM explicitly assumes that the signal trajectory being analyzed only contains transitions that occur along the single, directly detected dimension of a one-dimensional free-energy landscape (36). This assumption consequently renders HMMs inadequate for the analysis of signal trajectories that additionally contain projections of transitions that occur along the indirectly detected dimensions of a multidimensional free-energy landscape.

To rigorously address this problem, here we have adapted a class of inference tools based on a subclass of Markov chains, known as hierarchical Markov chains, to develop a hierarchical hidden Markov model (HHMM) (37–39)-based approach, which we call hFRET, for the analysis of single-molecule signal trajectories. HHMMs allow signal trajectories to be modeled as though they contain transitions along an arbitrary number of direct and indirectly detected dimensions of a free-energy landscape. Thus, hFRET can be used to identify and characterize kinetic heterogeneity and, correspondingly, to describe biomolecular processes using a hierarchical kinetic mechanism. Moreover, hFRET uses the variational approximation to Bayesian inference (36,40,41) to estimate the parameters of the HHMMs (i.e., the signal amplitudes of the states and the rates of transitions between states), a method we have previously and successfully used to estimate the parameters of HMMs (20–22). Because such variational Bayesian methods provide a powerful way to control model complexity (20–22), hFRET provides a principled approach for selecting the simplest hierarchical kinetic mechanism that best describes the data.

We begin this article by describing the theory underlying hFRET. Using computer-simulated single-molecule signal trajectories derived from a known hierarchical kinetic model and set of parameters, we then assess the accuracy with which hFRET can select the correct hierarchical kinetic model and infer the correct model parameters. Building from the analysis of computer-simulated data derived from a known model, we next use hFRET to analyze experimentally recorded single-molecule FRET (smFRET) data that report on the conformational dynamics of the ribosome, the biomolecular machine that is universally responsible for protein biosynthesis. Our analyses unambiguously reveal the presence of kinetic heterogeneity in single-molecule

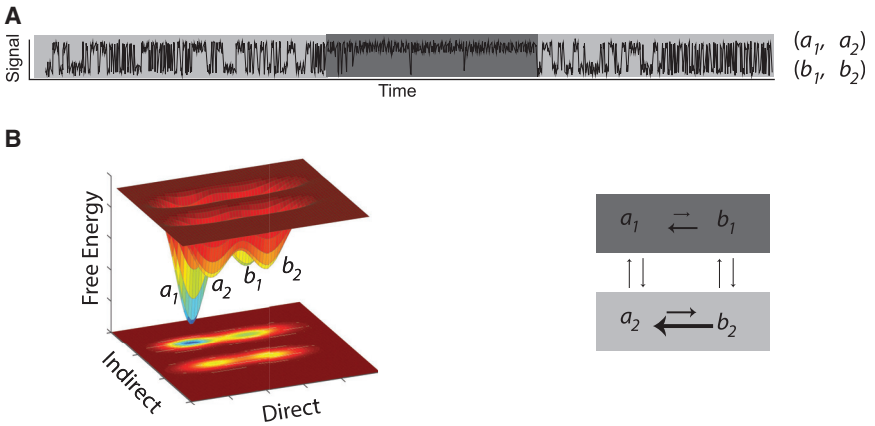

FIGURE 1 Manifestation and origin of kinetic heterogeneity in single-molecule signal trajectories. (*A*) A simulated single-molecule signal trajectory composed of contiguous periods, denoted by the variable grayscale backgrounds, is shown, in which the rates of transitions between two observable signal amplitudes, denoted as a_n and b_n, alternate between two distinct kinetic regimes (where the subscript n denotes the signal amplitudes associated with each kinetic regime). (*B*) The two-dimensional free-energy landscape (*left*) and corresponding kinetic mechanism (*right*) used to generate the simulated single-molecule signal trajectory shown in (*A*) are shown. A biomolecule governed by this free-energy landscape can undergo transitions along both a directly detected dimension, denoted by $a_1 \rightleftarrows b_1$ and $a_2 \rightleftarrows b_2$ transitions, and an indirectly detected dimension, denoted by $a_1 \rightleftarrows a_2$ and $b_1 \rightleftarrows b_2$ transitions. We note that a slightly more complex free-energy landscape and corresponding kinetic mechanism would allow simulation of a single-molecule trajectory for a biomolecule that can additionally undergo transitions between the directly and indirectly detected dimensions (i.e., $a_1 \rightleftarrows b_2$ and $a_2 \rightleftarrows b_1$ transitions). Although the approach described in this work allows simulation and/or modeling of such single-molecule trajectories, for illustrative purposes, we have nonetheless opted to use the relatively simpler free-energy landscape and kinetic model shown in (*B*) to simulate the single-molecule trajectory presented in (*A*). This figure is available in color online.

signal trajectories recorded on several ribosomal complexes and demonstrate that the extent of the heterogeneity depends on the composition of the ribosomal complexes. The approach we present here not only enables researchers to use single-molecule kinetic experiments to develop hierarchical kinetic models describing biological processes of interest, but it also paves the way for the development of closely related approaches that can further expand the data analysis capabilities of the field. Straightforward extensions of the approach presented here, for example, should allow multiple populations of signal trajectories that have been recorded on the same biological process, but using different signals, to be simultaneously analyzed within the context of a single, hierarchical kinetic model. Further extensions should enable the results of single-molecule kinetic experiments on a biological process of interest to be connected to the results of MD simulations of the same biological process.

MATERIALS AND METHODS

Theory

hFRET makes use of hierarchical Markov chains, which are a subclass of Markov chains whose states are parameterized in terms of multiple dimensions as opposed to a single dimension (37) (Fig. 2). The hierarchical Markov chain describing the states and transitions of a biological process, or the "system," obeys a Kolmogorov-Chapman equation (42) that propagates a state possessing D dimensions $\{z_{nt}^{d}\}$ for the nth trajectory at time t into the next state in the subsequent time point $\{z_{n,t+1}^{d}\}$, giving rise to the following likelihood function, L, for a given population of N mutually independent trajectories, each of length T_n:

$$L = \prod_{n=1}^{N} p\left(\{z_{nT_n}^{d}\}\ldots\{z_{n1}^{d}\}\right) = \prod_{n=1}^{N} p\left(\{z_{n1}^{d}\}\right)\prod_{t=2}^{T_n} p\left(\{z_{nt}^{d}\}\,\middle|\,\{z_{n,t-1}^{d}\}\right). \quad (1)$$

We separate these dimensions into 1) the directly detected dimension, denoted z_{nt}^{1} and also referred to as the production level of the state space, which specifies the distribution of observed signal emissions; 2) the first indirectly detected dimension, denoted z_{nt}^{2}, which specifies the distribution of kinetic regimes on the directly detected dimension; and 3) the arbitrarily higher-order indirectly detected dimensions, denoted z_{nt}^{d}, each of which specifies the distribution of kinetic regimes in the indirectly detected dimension that lies directly below it, z_{nt}^{d-1}. These dimensions are given natural number values that abstractly distinguish the dimensions of a free-energy landscape. The nested, conditional dependencies of this state-space coordinate system may be visualized as a tree of points (see *Models 0–5* in Fig. 3) (37–39), which may be thought of as enumerating the order in which the dimensions of a free-energy landscape are specified. Using this convention, the likelihood of the hierarchical Markov chain L can be decomposed (38), beginning with

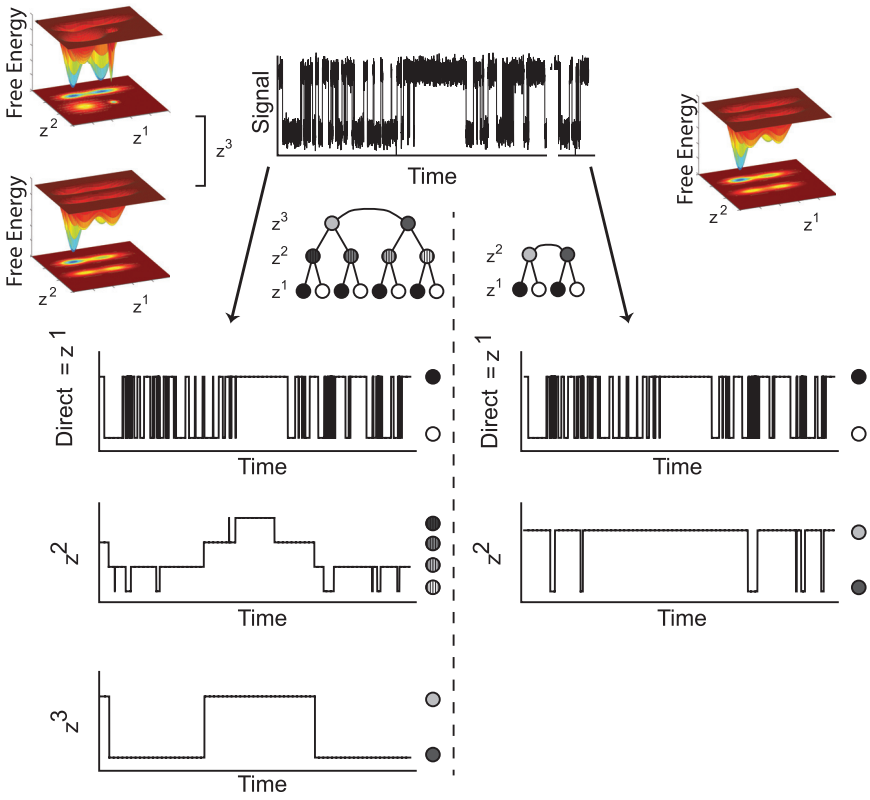

FIGURE 2 Simulation and analyses of single-molecule signal trajectories generated using a hierarchical Markov chain. A representative, 1000-time-point, single-molecule signal trajectory simulated using a hierarchical Markov chain composed of a directly detected dimension containing two states and two indirectly detected dimensions in which each dimension contains two states (*top center*) is shown. The free-energy landscape and variational Bayesian inference-based analysis generated using the correct HHMM comprised of a directly detected dimension, z^1, and two indirectly detected dimensions, z^2 and z^3 (*left*), are compared to the free-energy landscape and hFRET analysis of a less complex HHMM comprised of a directly detected dimension, z^1, and only one indirectly detected dimension, z^2 (*right*). Both models describe the same transitions along the directly detected dimension but differ in the transitions along and interpretation of indirectly detected dimensions. This figure is available in color online.

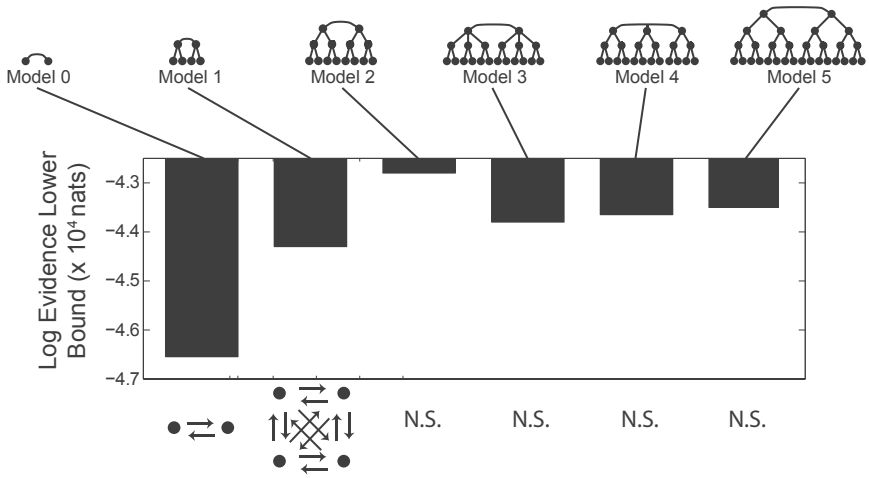

FIGURE 3 Selection among distinct HHMMs using variational Bayesian inference. Plot of the log evidence lower bound in natural units of information obtained from the variational Bayesian inference-based analysis of six HHMMs, denoted as Models 0–5, is given. Each HHMM is composed of one directly detected dimension containing two states and n indirectly detected dimension(s) (where n is a number between 0 and 5, as specified by the model numbers denoted along the *top* of the plot) in which each indirectly detected dimension also contains two states. The tree of points corresponding to each HHMM is depicted along the top of the plot, and the kinetic schemes corresponding to the HHMMs associated with Models 0 and 1 are depicted along the bottom of the plot. In the interest of presenting the simplest and clearest figure possible, the relatively complex, multidimensional kinetic schemes corresponding to the HHMMs associated with Models 2–5 are not shown and are instead labeled "NS" along the bottom of the plot.

the directly detected dimension and iteratively specifying the abstract values associated with the system on the indirectly detected dimensions:

$$L = \prod_{n=1}^{N}\left[\prod_{d=1}^{D}\pi_{d,z_{n1}^{d}}\right] \times \left[\prod_{t=2}^{T_n-1}\prod_{d=1}^{D-1}A_{d,z_{nt}^{d},exit}^{\delta_{z_{nt}^{d+1},z_{n,t+1}^{d+1}}}A_{d,z_{nt}^{d},z_{n,t-1}^{d}}^{\delta_{z_{nt}^{d},z_{n,t+1}^{d}}\left(1-\delta_{z_{nt}^{d+1},z_{n,t+1}^{d+1}}\right)}\pi_{d,z_{n,t+1}^{d}}\right] \times \left[\prod_{d=1}^{D}A_{d,z_{nT_n}^{d},exit}\right], \tag{2}$$

where we have introduced the standard notation

$$\delta_{ij} = \left\{\begin{array}{l} 1,\ if\ i = j \\ 0,\ if\ i \neq j \end{array}\right\},$$

$$p\left(z_{n1}^{d} = i\right) = \pi_{di},$$

$$p\left(z_{nt}^{d} = i \,\middle|\, z_{n,t+1}^{d} = j\right) = A_{dij},$$

$$p\left(z_{nt}^{d+1} \neq z_{n,t+1}^{d+1}, z_{nt}^{d} = i\right) = A_{di,exit}.$$

The final statement above represents the probability that the value associated with the system along the indirectly detected dimension d, which specifies the distribution of kinetic regimes in the indirectly detected dimension $d-1$, has transitioned to a new value. We note that a hierarchical kinetic model for static heterogeneity may be derived directly from Eq. 2 by simply limiting the equation such that it contains only one indirectly detected dimension and that transitions between kinetic regimes within that indirectly detected dimension are not allowed (presented in greater detail in Supporting Materials and Methods, Section S1).

Equation 2 specifies the hierarchical kinetic model. We use the variational approximation to Bayesian inference (36,40,41) to specify both the emission distributions and the computational algorithm for estimating the parameters of the HHMM, a procedure that we summarize here and present in greater detail in Supporting Materials and Methods, Section S2. Briefly, we seek to maximize the lower bound of the log probability, denoted as the "evidence," of a set of parameter distributions, denoted as θ, and a set of observations, denoted as $\{x_{nt}\}$, given prior information, denoted as ψ_0:

$$\ln p\left(\{x_{nt}\},\left\{z_{nt}^{d}\right\},\theta \,\middle|\, \psi_0\right) \geq \int d\theta \sum_{n}\sum_{z_{nt}^{d}} p\left(\left\{z_{nt}^{d}\right\},\theta|\{x_{nt}\},\psi_0\right) \times \ln\frac{p\left(\{x_{nt}\},\left\{z_{nt}^{d}\right\},\theta|\psi_0\right)}{p\left(\left\{z_{nt}^{d}\right\},\theta|\{x_{nt}\},\psi_0\right)}. \tag{3}$$

The variational approximation assumes that the coordinates do not depend on the parameter distributions such that the joint probability may be written as

$$p\left(\left\{z_{nt}^{d}\right\},\theta \,\middle|\, \{x_{nt}\},\psi_0\right) = q\left(\left\{z_{nt}^{d}\right\} \,\middle|\, \psi_0\right)q(\theta \,|\, \psi_0). \tag{4}$$

Although here we will assume that the emission distributions are normal distributions, this assumption can be generalized as necessary. Inference of the parameters of an HHMM then proceeds by iteratively locating parameters that optimize a lower bound for the evidence. Iterations proceed by optimizing $q(\{z_{nt}^{d}\} \,|\, \psi_0)$, then optimizing $q(\theta|\psi_0)$, and finally calculating the evidence lower bound. Convergence is achieved when the evidence lower bound remains virtually unchanged between iterations.

By utilizing the variational approximation, we can factorize the joint distribution of the kinetic model as follows. First, we simplify the hierarchical Markov chain likelihood in Eq. 2 in terms of its transition counts:

$$L = \prod_{d=1}^{D}\prod_{i=1}^{\Omega_d}\pi_{di}^{b_{di}}A_{di,exit}^{e_{di}}\prod_{j=1}^{\Omega_d}A_{dij}^{n_{dij}}, \tag{5}$$

where Ω_d denotes the number of distinct values of the system along the indirectly detected dimensions in level d; b_{di} denotes the number of transitions resulting in $z_{nt+1}^{d} = i$, given that $z_{nt}^{d+1} \neq z_{n,t+1}^{d+1}$; e_{di} denotes the number of transitions out of $z_{nt}^{d} = i$, given that $z_{nt}^{d+1} \neq z_{n,t+1}^{d+1}$; and n_{dij} denotes the

number of transitions from $z_{nt}^{d} = i$ to $z_{n,t+1}^{d} = j$. Notably, normalizing L implies that the factored distributions over the kinetic parameters decompose into multinomial distributions:

$$
\begin{aligned}
&q\left(\{b_{di}\},\{e_{di}\},\{n_{dij}\}\,\middle|\,\{\pi\},\{A\},\psi_0\right)\\
&\quad=\prod_{d=1}^{D} Mult(\{b_{di}\}\,|\,\pi_d,d,\psi_0\,)\\
&\qquad\times\prod_{i=1}^{\Omega_d} Mult\left(\{e_{di}\},\{n_{dij}\}\,\middle|\,A_{dij},i,\psi_0\right)\\
&\quad=\prod_{d=1}^{D} q(\{b_{di}\}\,|\,\pi_d,d,\psi_0\,)\prod_{i=1}^{\Omega_d} q\left(\{e_{di}\},\{n_{dij}\}\,\middle|\,A_{dij},i,\psi_0\right).
\end{aligned}
\tag{6}
$$

Therefore, considering the state space as a tree of points that are directionally interconnected by conditional relationships, each point can be considered as an independently operating Markov chain, and to infer the parameters and parameter distributions of the hierarchical kinetic model, it is sufficient to calculate the transition counts specified above. From these parameters, we calculate transition rates between the various free-energy minima (see Supporting Materials and Methods, Section S3) and therefore quantitatively specify the hierarchical kinetic model.

Generation of simulated signal trajectories using a specific hierarchical kinetic model

One thousand signal trajectories composed of 1000 time points each were simulated by randomly drawing each signal trajectory from a hierarchical Markov chain. This hierarchical Markov chain was composed of a directly detected dimension containing two states characterized by two distinct signal amplitudes and two indirectly detected dimensions in which each dimension contained two distinct states (i.e., the hierarchical Markov chain shown in the *left-hand side* of Fig. 2). Gaussian-distributed noise to a final signal/noise ratio of 5:1 was then added to each of the simulated signal trajectories. The source code for generating simulated signal trajectories, as well as sample simulated signal trajectories, can be found together with the hFRET source code, graphical user interface, and user manual at https://github.com/GonzalezBiophysicsLab/hFRET.

Collection and analysis of smFRET data

The L1-tRNA smFRET data that was analyzed using hFRET in this study consists of data sets that had been previously collected, analyzed using a different maximal-likelihood-estimated HMM approach (18), and interpreted and reported by Fei et al. (10). Briefly, the L1-tRNA smFRET signal was generated by preparing PRE complexes carrying an OH-(Cy3)tRNA^{Phe} within the P site and (Cy5)L1 within the L1 stalk of the 50S subunit. OH-(Cy3)tRNA^{Phe} was prepared by site-specifically labeling the dihydrouridine at position 47 of OH-tRNA^{Phe} (Sigma) with Cy3. (Cy5)L1 was prepared by site-specifically labeling an introduced cysteine at position 202 in a recombinantly overexpressed and purified single-cysteine variant of *Escherichia coli* ribosomal protein L1 with Cy5. (Cy5)L1-labeled 50 subunits were subsequently generated by reconstituting (Cy5)L1 into 50S subunits that had been purified from an *E. coli* strain lacking the gene encoding ribosomal protein L1.

As discussed in more detail elsewhere (10), three PRE complexes were assembled onto mRNAs containing a biotin moiety at the 5′ terminus. PRE^{-}, which carried an OH-(Cy3)tRNA^{Phe} at the P site and a vacant A site, was prepared by delivering puromycin, a ribosome-targeting inhibitor of protein synthesis, to the A site of a ribosomal elongation complex carrying an fMet-Phe-tRNA^{Phe} at the P site and a vacant A site. Puromycin is an analog of the 3′-terminal residue of aminoacyl-tRNA that binds to the A site of the peptidyl transferase center of the 50S subunit, acts as an acceptor substrate in the peptidyl transfer reaction with the fMet-Phe-(Cy3)tRNA^{Phe} at the P site, and rapidly dissociates from the 50S subunit, thereby generating a PRE complex containing a deacylated OH-(Cy3)tRNA^{Phe} in the P site and a vacant A site (i.e., PRE^{-}) (43). PRE^{fMFK}, which carried an OH-(Cy3)tRNA^{Phe} at the P site and an fMet-Phe-Lys-tRNA^{Lys} at the A site, was prepared by delivering a ternary complex composed of elongation factor (EF) Tu, GTP, and Lys-tRNA^{Lys} (EF-Tu(GTP)Lys-tRNA^{Lys}) to the A site of a ribosomal elongation complex carrying an fMet-Phe-(Cy3)tRNA^{Phe} at the P site and a vacant A site. Once accommodated into the A site, Lys-tRNA^{Lys} acts as an acceptor substrate in the peptidyl transfer reaction, thereby generating a PRE complex carrying an OH-(Cy3)tRNA^{Phe} at the P site and an fMet-Phe-Lys-tRNA^{Lys} at the A site (i.e., PRE^{fMFK}). PRE^{K}, which carried an OH-(Cy3)tRNA^{Phe} at the P site and an Lys-tRNA^{Lys} at the A site, was prepared by delivering EF-Tu(GTP)Lys-tRNA^{Lys} to the A site of a PRE complex identical to PRE^{-} and thereby carrying an OH-(Cy3)tRNA^{Phe} at the P site and a vacant A-site PRE^{-}. Although it accommodates into the A site, the lack of a peptidyl moiety on the OH-(Cy3)tRNA^{Phe} at the P site prevents Lys-tRNA^{Lys} from acting as an acceptor substrate in the peptidyl transfer reaction, thereby generating a PRE complex carrying an OH-(Cy3)tRNA^{Phe} at the P site and an Lys-tRNA^{Lys} at the A site (i.e., PRE^{K}).

As previously described in greater detail (10,44), a laboratory-built, prism-based, wide-field single-molecule TIRF microscope was used to image the three PRE complexes in a Tris-Polymix imaging buffer composed of 50 mM tris(hydroxymethyl)aminomethane acetate, 100 mM potassium chloride, 5 mM ammonium acetate, 0.5 mM calcium acetate, 15 mM magnesium acetate, 6 mM β-mercaptoethanol, 5 mM putrescine dihydrochloride, and 1 mM spermidine (free base) at a $pH_{25^{\circ}C}$ of 7.5 that was supplemented with an oxygen-scavenging system (1% β-D-glucose, 25 units/mL glucose oxidase, and 250 units/mL catalase) (10). Briefly, each PRE complex was tethered to the surface of a microfluidic TIRF microscopy observation flowcell that had been passivated with a mixture of polyethylene glycol (PEG) and biotinylated PEG and had been treated with streptavidin. Cy3 fluorophores were directly excited using a 532-nm laser excitation source (CrystaLaser, Reno, NV) and fluorescence emissions from both Cy3 and Cy5 were collected using a 1.2 numerical aperture/60× objective (Nikon, Tokyo, Japan), wavelength separated using a two-channel imaging system (Photometrics, Tucson, AZ), and recorded as a video using a back-illuminated, electron-multiplying, charge-coupled device camera (Photometrics) operating at an acquisition time of 50 ms per frame.

As detailed in our previous report (10), individual pairs of Cy3 and Cy5 fluorescence intensity versus time trajectories (i.e., intensity trajectories) reporting on the conformational dynamics of single PRE complexes were generated using laboratory-written, MATLAB (The MathWorks, Natick, MA)-based image-analysis software. Each pair of Cy3- and Cy5 fluorescence intensity trajectories was 1) truncated at the first time point at which either fluorophore "photobleached" (i.e., underwent an apparently irreversible loss of fluorescence intensity), 2) baseline corrected by subtracting the average intensity of the last 10 time points after the photobleaching event of either the Cy3 or Cy5 fluorophore, and 3) spectral-cross-talk-corrected by decreasing the Cy5 fluorescence intensity at each time point by 7% of the Cy3 fluorescence to account for the 7% bleed-through of Cy3 fluorescence emission into the Cy5 channel. The photobleaching-, baseline-, and spectral-cross-talk-corrected pairs of Cy3- and Cy5 fluorescence intensity trajectories were then used to generate individual E_{FRET} trajectories by using the Cy3 intensity at each timepoint, $I_{Cy3}(t)$, and the Cy5 intensity at each timepoint, $I_{Cy5}(t)$, to calculate the E_{FRET} at each time point as $E_{FRET}(t) = I_{Cy5}(t)/(I_{Cy5}(t) + I_{Cy3}(t))$. Assuming the quantum yield of the Cy3 fluorophore and the rotational motion of the Cy3 and/or Cy5 fluorophores are constant throughout the experiment, changes in E_{FRET} are proportional to changes in the distance between the Cy3 and Cy5 fluorophores as given by $E_{FRET}(R) = 1/(1 + (R/R_0)^6)$, where R is the distance between the Cy3 and Cy5 fluorophores and R_0 is the Förster radius, which is ~54 Å under our conditions (25,45). The resulting E_{FRET} trajectories were analyzed

using hFRET and visualized using state re-entry plots as discussed in the Results. The hFRET source code, graphical user interface, and user manual can be found together with source code for generating simulated signal trajectories and sample simulated signal trajectories at https://github.com/GonzalezBiophysicsLab/hFRET.

RESULTS

hFRET enables accurate hierarchical kinetic model selection

We first sought to calibrate hFRET by investigating whether it could be used to accurately select among hierarchical kinetic models similar to that shown in Fig. 1, butof systematically increasing complexity. Statistically rigorous model selection is important in data analyses that incorporate indirectly detected dimensions because statistical techniques are the only means for counting and distinguishing among alternative models. To test the model selection accuracy of hFRET, we used a specific hierarchical kinetic model to generate a set of simulated signal trajectories, used hFRET to infer the most probable parameters for several models of increasing complexity, and selected the most probable model by calculating and comparing the lower bound of the evidence (see Eq. 4). Although the lower bound of the evidence is generally not a rigorous metric for de novo model selection (41), it can be used to compare models as a function of increasing complexity of their state spaces along indirectly detected dimensions. Indeed, when comparing models with an equivalent number of directly detected dimensions but a varying number of indirectly detected dimensions via the difference between their evidence lower bound (see Supporting Materials and Methods, Section S2), the contribution of only the indirectly detected dimensions to the difference between the evidence lower bounds and the evidence per se is equivalent (22,36).

To generate the set of simulated signal trajectories, we used a hierarchical Markov chain composed of one directly detected dimension, two indirectly detected dimensions, and two signal amplitudes, thereby specifying a free-energy landscape consisting of eight distinct free-energy minima, shown in Fig. 2 (simulation parameters and sample data provided online at https://github.com/GonzalezBiophysicsLab/hFRET). Because it would be difficult for any model, including the HHMM we describe here, to confidently distinguish between free-energy minima that overlap along both the direct and indirect dimensions, we took care to choose simulation parameters that would not lead to such overlap. We subsequently used hFRET to analyze the set of simulated signal trajectories and infer the most probable parameters for six different models (Fig. 3). Each of these six models incorporates one directly detected dimension, anywhere from zero to three indirectly detected dimensions, and two signal amplitudes. In addition to allowing transitions along the directly and indirectly detected dimensions, all six models also allowed simultaneous transitions between the directly and indirectly detected dimensions (see the kinetic schemes depicted at the *bottom* of Fig. 3). We have denoted these models as Models 0–5, where the increasing numbers represent increasing model complexity. We then calculated the lower bound of the evidence for all six models (see Supporting Materials and Methods, Section S2). As expected, the evidence lower bound for the correct model (i.e., Model 2, the model that was used to simulate the data) was significantly larger than that of the incorrect, simpler models (i.e., Models 0–1) as well as the incorrect, more complex models (i.e., Models 3–5). Collectively, the results of these analyses demonstrate that selecting the model with the largest lower bound of the evidence can be used to specify the most parsimonious hierarchical kinetic model that is consistent with the data.

hFRET allows quantitative characterization of the hierarchical structural dynamics of the ribosome

To demonstrate how hFRET-based analyses of single-molecule kinetic data can be used to characterize the hierarchical dynamics of biomolecular systems, we have used hFRET to analyze signal trajectories from smFRET experiments reporting on the structural dynamics of the bacterial ribosome. The ribosome is the universally conserved, two-subunit, ribonucleoprotein complex that uses aminoacyl-transfer RNA (tRNA) substrates to translate the triplet-nucleotide codon sequence of messenger RNA (mRNA) templates into proteins (Fig. 4 *A*). During the elongation stage of translation, addition of each amino acid to the nascent polypeptide chain that is being synthesized by the ribosome proceeds through an elongation cycle composed of three steps: 1) aminoacyl-tRNA selection, 2) peptidyl transfer, and 3) translocation (46) (Fig. 4 *B*).

After undergoing peptidyl transfer but before undergoing translocation, bacterial ribosomal pre-translocation (PRE) complexes undergo thermally driven, reversible fluctuations between at least two major conformational states that we refer to as global state 1 (GS1) and global state 2 (GS2), establishing a GS1 $\rightleftarrows$ GS2 equilibrium (10) (Fig. 4 *C*). In GS1, the ribosomal subunits occupy their "nonrotated" intersubunit orientation, the L1 stalk element of the 50S subunit occupies its "open" conformation, and the ribosome-bound tRNAs occupy their "classical" configurations. Contrasting with this, in GS2, the ribosomal subunits occupy their "rotated" intersubunit orientation, the L1 stalk element of the 50S subunit occupies its "closed" conformation, and the ribosome-bound tRNAs occupy their "hybrid" configurations.

Previously, we have designed and developed an L1-tRNA smFRET signal by preparing PRE complexes carrying a cyanine (Cy) 3 FRET donor-fluorophore-labeled, deacylated, phenylalanine-specific tRNA (OH-(Cy3)tRNA^{Phe}) within the ribosomal peptidyl-tRNA binding (P) site and a Cy5 FRET acceptor-labeled ribosomal protein L1 ((Cy5)

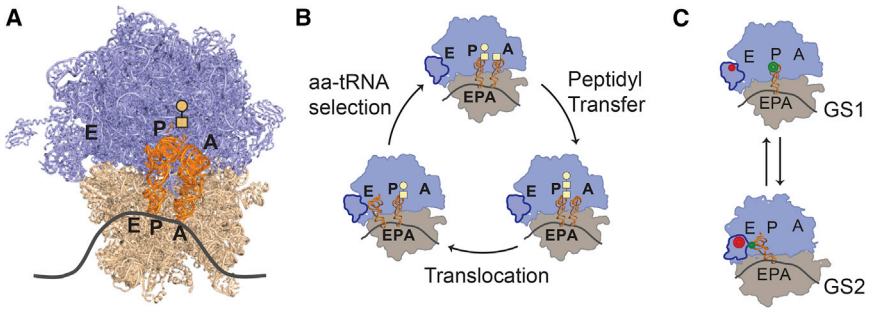

FIGURE 4 Structure of ribosomal complexes, the translation elongation cycle, and smFRET studies of PRE complexes. (*A*) X-ray crystallographic structure of an *E. coli* ribosomal complex (Protein Data Bank [PDB]: 4V51) (62). The large, or 50S, ribosomal subunit is shown in light blue; the small, or 30S, subunit is shown in tan; the A, P, and E sites on the 50S and 30S subunits are denoted as black letters; the path of the mRNA is denoted as a dark gray curve; the A- and P-site tRNAs are shown in orange; and the location of a hypothetical nascent dipeptide on the A-site tRNA is denoted as yellow shapes. (*B*) The translation elongation cycle is composed of aminoacyl-tRNA selection, peptidyl transfer, and translocation. (*C*) The equilibrium between the GS1 and GS2 conformational states of a PRE complex. The L1-tRNA smFRET signal that is used to report on transitions between GS1 (E_{FRET} of ~0.18) and GS2 (E_{FRET} of ~0.81) in the PRE complex data that are analyzed in this study is generated using an OH-(Cy3)tRNA^{Phe} within the P site and a (Cy5)L1 within the L1 stalk of the 50S subunit. The Cy3 and Cy5 fluorophores are depicted as green and red circles, respectively.

L1) within the L1 stalk (Fig. 4 *C*) (10,47). Using a total internal reflection fluorescence (TIRF) microscope, we have imaged these PRE complexes at single-molecule resolution, collecting Cy3 and Cy5 fluorescence intensity (I_{Cy3} and I_{Cy5}, respectively) versus time trajectories (hereafter referred to as I_{Cy3} and I_{Cy5} trajectories, respectively) for individual PRE complexes and using these intensities to calculate the corresponding FRET efficiency (E_{FRET}) versus time trajectory (hereafter referred to as the E_{FRET} trajectory), where $E_{FRET} = I_{Cy5}/(I_{Cy3} + I_{Cy5})$. The resulting E_{FRET} trajectories were observed to spontaneously and stochastically fluctuate between two E_{FRET} signal amplitudes, one centered at an E_{FRET} of ~0.18 and the other centered at an E_{FRET} of ~0.81 (see Table 1 for the specific E_{FRET} signal amplitudes corresponding to particular PRE complexes). Consistent with the interpretation that PRE complexes undergo thermally driven, reversible fluctuations between GS1 and GS2, the E_{FRET} signal amplitudes centered at E_{FRET}s of ~0.18 and ~0.81 could be assigned to GS1 and GS2, respectively. These assignments were made using structural models of PRE complexes in GS1 (48) and GS2 (49) and the known relationship between E_{FRET} and the distance between the donor and acceptor fluorophores $E_{FRET} = 1/[1 + (R/R_0)^6]$, where R is the distance between the donor and acceptor fluorophores and R_0, which is also known as the Förster radius, is the distance at which a specific donor and acceptor fluorophore pair exhibit a half-maximal E_{FRET} (i.e., an $E_{FRET} = 0.50$) (25). Further details regarding the design and development of the L1-tRNA smFRET signal and the collection, analysis, and interpretation of L1-tRNA smFRET data can be found in the Materials and Methods.

TABLE 1 Observed E_{FRET}s for each PRE Complex

PRE Complex	GS1 E_{FRET}	GS2 E_{FRET}
PRE^{-A}	0.168 ± 0.003	0.809 ± 0.003
PRE^{fMFK}	0.198 ± 0.002	0.818 ± 0.002
PRE^{K}	0.198 ± 0.003	0.811 ± 0.003

Error bars represent 99% confidence intervals estimated using hFRET.

Notably, the L1-tRNA smFRET signal has been used to investigate how the absence, presence, and acylation status of the peptidyl-tRNA in the ribosomal aminoacyl-tRNA binding (A) site modulates the kinetics of GS1 → GS2 and GS2 → GS1 transitions (10). Specifically, smFRET data were collected on PRE complexes containing a vacant A site (PRE^{-}), carrying an fMet-Phe-Lys-tRNA^{Lys} peptidyl-tRNA in the A site (PRE^{fMFK}), and carrying a Lys-tRNA^{Lys} aminoacyl-tRNA in the A site (PRE^{K}). Fig. 5, *A–C* present plots of representative I_{Cy3}, I_{Cy5}, and E_{FRET} trajectories corresponding to PRE^{-}, PRE^{fMFK}, and PRE^{K}, respectively. As expected, the E_{FRET} trajectories corresponding to all three PRE complexes fluctuate between two E_{FRET} signal amplitudes centered at E_{FRET}s of ~0.18 and ~0.81 and corresponding to GS1 and GS2, respectively. hFRET-based analyses of the entire population of E_{FRET} trajectories corresponding to either PRE^{-}, PRE^{fMFK}, or PRE^{K} demonstrated that, for each PRE complex, the most parsimonious hierarchical kinetic model is one in which the directly detected dimension is composed of fluctuations between two directly detected states that are characterized by distinct E_{FRET} signal amplitudes centered at E_{FRET}s of ~0.18 and ~0.81 and one indirectly detected dimension that is composed of fluctuations between two indirectly detected states, which are characterized by distinct rates of transitions between the two signal amplitudes centered at E_{FRET}s of ~0.18 and ~0.81. Although the hFRET-based analyses also revealed transitions between directly and indirectly detected dimensions, such transitions were exceedingly rare. Specifically, these transitions comprise less than 4% of the total number of transitions within the entire population of E_{FRET} trajectories corresponding to each PRE complex. Given that the exceedingly low frequency with which these transitions are observed results in extremely wide confidence intervals on the rate constants that can be estimated using these transitions and the fact that these rare transitions might arise from a small number of rapid, two-step transitions in which the time-resolution limit of our detector prevented observation of one of the steps, these transitions were excluded from further analysis.

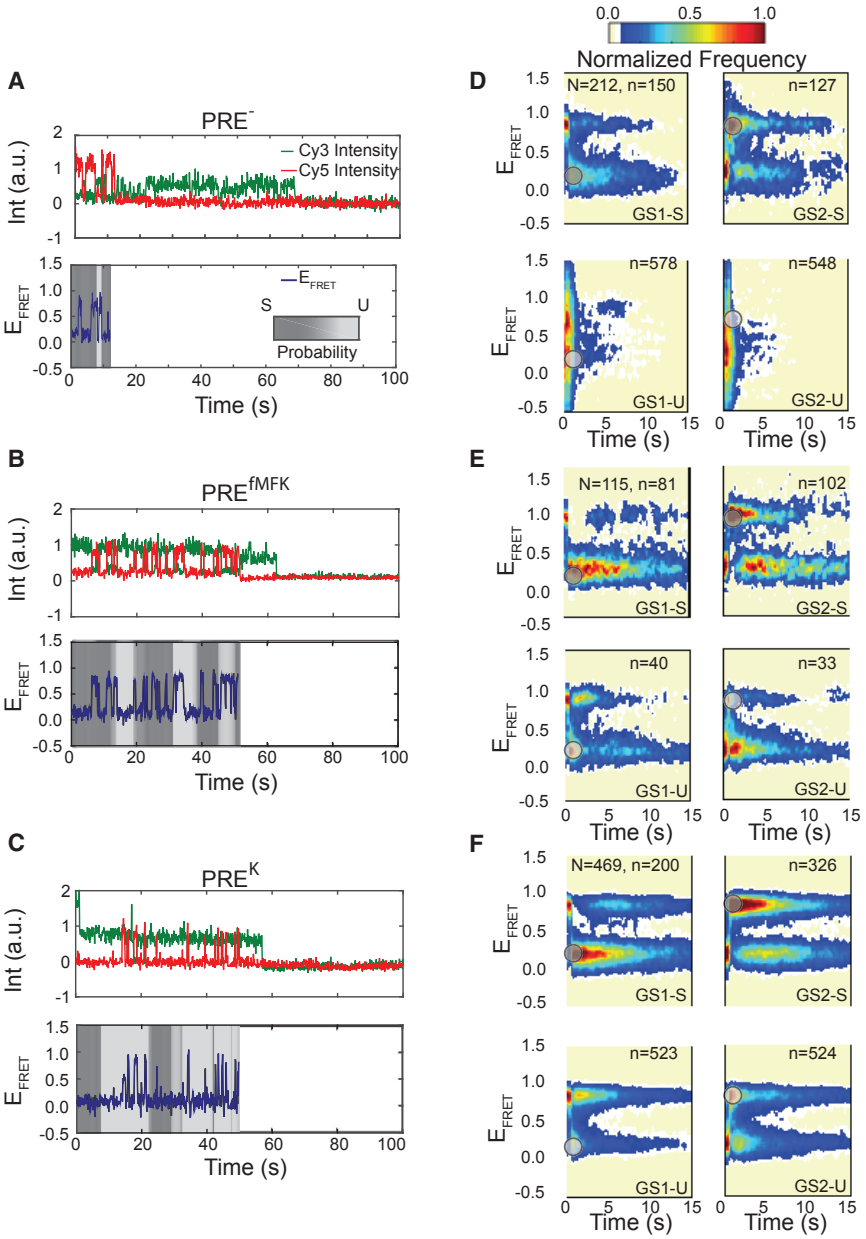

FIGURE 5 Kinetic heterogeneity of PRE complexes. Representative Cy3 and Cy5 fluorescence intensity trajectories (*top*) and corresponding E_{FRET} trajectories (*bottom*) are shown for (*A*) PRE^{-}, (*B*) PRE^{fMFK}, and (*C*) PRE^{K}. The backgrounds of the E_{FRET} trajectory plots are linearly grayscale-weighted by the probability that a time point belongs to either the S state, dark gray, or the U state, light gray. State re-entry plots (see Results) for the GS1-S state (*upper left-hand plot*, initiating at dark gray circle at E_{FRET} of ~0.18), GS2-S state (*upper right-hand plot*, initiating at *dark gray circle* at E_{FRET} of ~0.81), GS1-U state (*lower left-hand plot*, initiating at *light gray circle* at E_{FRET} of ~0.18), and GS2-U state (*lower right-hand plot*, initiating at *light gray circle* at E_{FRET} of ~0.81) of (*D*) PRE^{-}, (*E*) PRE^{fMFK}, and (*F*) PRE^{K}. For the purpose of generating these plots, time points with a probability of belonging to the S state that is greater than or equal to 50% are discretely assigned to the S state, and time points with a probability of belonging to the U state that is greater than or equal to 51% are discretely assigned to the U state. The "N" denoted at the top of the GS1-S state re-entry plot for each PRE complex specifies the number of distinct E_{FRET} trajectories that were used to generate the four state re-entry plots for the corresponding PRE complex. Similarly, the "n" denoted at the top of each state re-entry plot specifies the number of state subtrajectories that were used to generate the corresponding state re-entry plot.

Consistent with previous work from our group (9,10,47) and others (50–54), we interpret the fluctuations between the two directly detected states with distinct E_{FRET} signal amplitudes centered at E_{FRET}s of ~0.18 and ~0.81, consistent with the GS1 and GS2 states of the PRE complex, respectively, as reporting on the GS1 → GS2 and GS2 → GS1 transitions. Moreover, based on a visual inspection of the analyzed E_{FRET} trajectories as well as a full kinetic analysis that is presented further below, we interpret the fluctuations between the two indirectly detected states with distinct rates of fluctuations between GS1 and GS2 as reporting on transitions between an indirectly detected state in which excursions to GS1 and GS2 are relatively long-lived and stable, denoted as the stable (S) state of the PRE complex, and a state in which excursions to GS1 and GS2 are relatively short-lived and unstable, denoted as the unstable (U) state of the PRE complex.

To visually assess the extent to which the individual E_{FRET} trajectories recorded for each PRE complex fluctuate between the indirectly detected S and U states (i.e., exhibit dynamic heterogeneity), we generated "state re-entry plots" for the GS1-S, GS2-S, GS1-U, and GS2-U states of each PRE complex (Fig. 5, *D–F*). The goal of these plots is to illustrate the time that it takes and the states that are sampled when an E_{FRET} signal first enters a specific indirectly detected state (e.g., the GS1-S state); transitions to additional

states, at least one of which must be a different indirectly detected state (i.e., the GS1-U or GS2-U states); and ultimately re-enters the original indirectly detected state (i.e., the GS1-S state). To generate these plots, we first divided each E_{FRET} trajectory in the entire population of E_{FRET} trajectories corresponding to each PRE complex into GS1-S, GS2-S, GS1-U, and GS2-U subtrajectories, in which, as a specific example, GS1-S subtrajectories were defined as those that begin when the E_{FRET} signal enters GS1-S and undergo at least one transition to one of the U states (i.e., GS1-U or GS2-U) before returning to GS1-S. GS2-S, GS1-U, and GS2-U subtrajectories were analogously defined. For each PRE complex, we then plotted a postsynchronized surface contour plot of the time evolution of population FRET for each of the GS1-S, GS2-S, GS1-U, or GS2-U subtrajectories. As an example, the GS1-S contour plot for each PRE complex was generated by postsynchronizing the GS1-S subtrajectories from that PRE complex, such that the first time point that transitions into GS1-S was assigned to the 0 s time point on the plot, and then generating a surface contour plot that effectively superimposes all of the postsynchronized transitions into the GS1-S at the 0 s time point. GS2-S, GS1-U, and GS2-U contour plots were analogously generated. These plots demonstrate that the vast majority of E_{FRET} trajectories recorded for each PRE complex fluctuate reversibly between the S and U states and exhibit dynamic heterogeneity. In addition, comparative analyses of these plots suggest that the re-entry times for the U states are particularly sensitive to the presence of an A-site tRNA (compare the two *lower plots* in Fig. 5 *D* to those in Fig. 5, *E* and *F*), a qualitative observation that can be more quantitatively characterized using a full kinetic analysis, as described in the next paragraph.

The hierarchical kinetic model described above possesses four distinct free-energy minima corresponding to GS1-S, GS2-S, GS1-U, and GS2-U that are connected by eight rate constants (Fig. 6). Four of these rate constants connect the states along the directly detected dimension, thus corresponding to the rates of transitions between GS1 and GS2 within the S or U state, denoted as $k_{1,S \rightarrow 2,S}$, $k_{2,S \rightarrow 1,S}$, $k_{1,U \rightarrow 2,U}$, and $k_{2,U \rightarrow 1,U}$ (where the subscripts 1 and 2 denote GS1 and GS2, respectively, and the S and U subscripts denote the S and U states, respectively). The remaining four rate constants connect the states along the indirectly detected dimension, thus corresponding to the rates of transitions between the S and U states within GS1 or GS2, denoted as $k_{1,S \rightarrow 1,U}$, $k_{1,U \rightarrow 1,S}$, $k_{2,S \rightarrow 2,U}$, and $k_{2,U \rightarrow 2,S}$. Using this kinetic model, Fig. 7 reports the values of these rate constants for the three PRE complexes we have characterized. In all three cases, $k_{1,S \rightarrow 2,S}$ was between ~12- and 54-fold smaller than $k_{1,U \rightarrow 2,U}$, and $k_{2,S \rightarrow 1,S}$ was between ~13- and ~17-fold smaller than $k_{2,U \rightarrow 1,U}$, demonstrating that, as qualitatively observed in the individual E_{FRET} trajectories (Fig. 5), GS1 and GS2 within the S state are significantly more stable than they are within the U state.

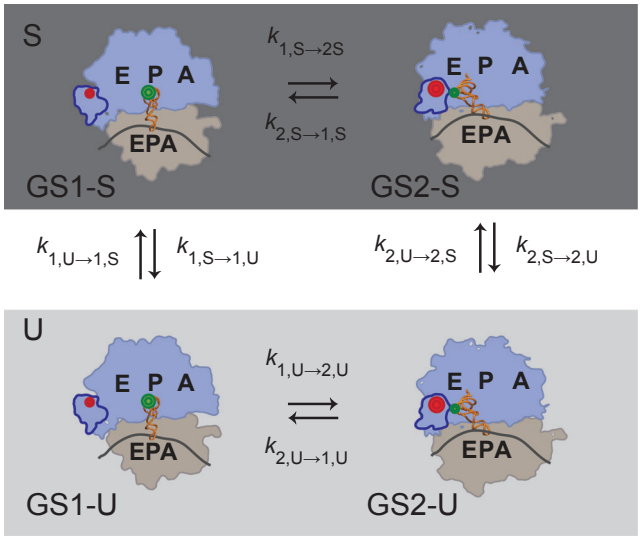

FIGURE 6 Hierarchical kinetic mechanism of PRE complex dynamics. Hierarchical kinetic mechanism describing the rates of transition between GS1-S, GS2-S, GS1-U, and GS2-U. Transitions between GS1 and GS2 within the S state constant, GS1-S ⇄ GS2-S, are enclosed within a box that is shaded in dark gray, and transitions between GS1 and GS2 within the U state, GS1-U ⇄ GS2-U, are enclosed within a box that is shaded in light gray.

Moreover, comparison of these rate constants for PRE^{-} and PRE^{fMFK} demonstrates that the PRE complex carrying an fMet-Phe-Lys-$tRNA^{Lys}$ in the A site exhibits a $k_{1,S \rightarrow 1,U}$ that is ~6-fold smaller than that of the PRE complex with an empty A site. Together with slightly smaller, ~4-fold decreases in $k_{1,U \rightarrow 1,S}$, $k_{2,U \rightarrow 1,U}$, and $k_{2,S \rightarrow 1,S}$, these data demonstrate that the presence of a peptidyl-tRNA in the A site of a PRE complex can modulate rates of transitions along more than one dimension of the multidimensional free-energy landscape of a PRE complex. As can be seen by comparing the two lower plots in Fig. 5 *D* to those in Fig. 5 *E*, these kinetic effects act to increase the re-entry times that are observed for the U states in PRE^{fMFK} versus PRE^{-}. Similarly, comparison of the rate constants for PRE^{K} and PRE^{fMFK} demonstrates that the PRE complex carrying a Lys-$tRNA^{Lys}$ in the A site exhibits a $k_{2,U \rightarrow 2,S}$ that is ~3-fold larger than that of the PRE complex carrying an fMet-Phe-Lys-$tRNA^{Lys}$ in the A site. Together with a slightly larger, ~2-fold increase in $k_{1,S \rightarrow 1,U}$, these data demonstrate that the acylation status of the A-site tRNA contributes to the ability of the A-site peptidyl-tRNA to modulate the rates of transitions along the indirectly detected dimensions of the free-energy landscape of a PRE complex.

DISCUSSION

This work demonstrates a rigorous approach for analyzing single-molecule kinetic data that exhibit kinetic heterogeneity. Specifically, hFRET can be used to identify and

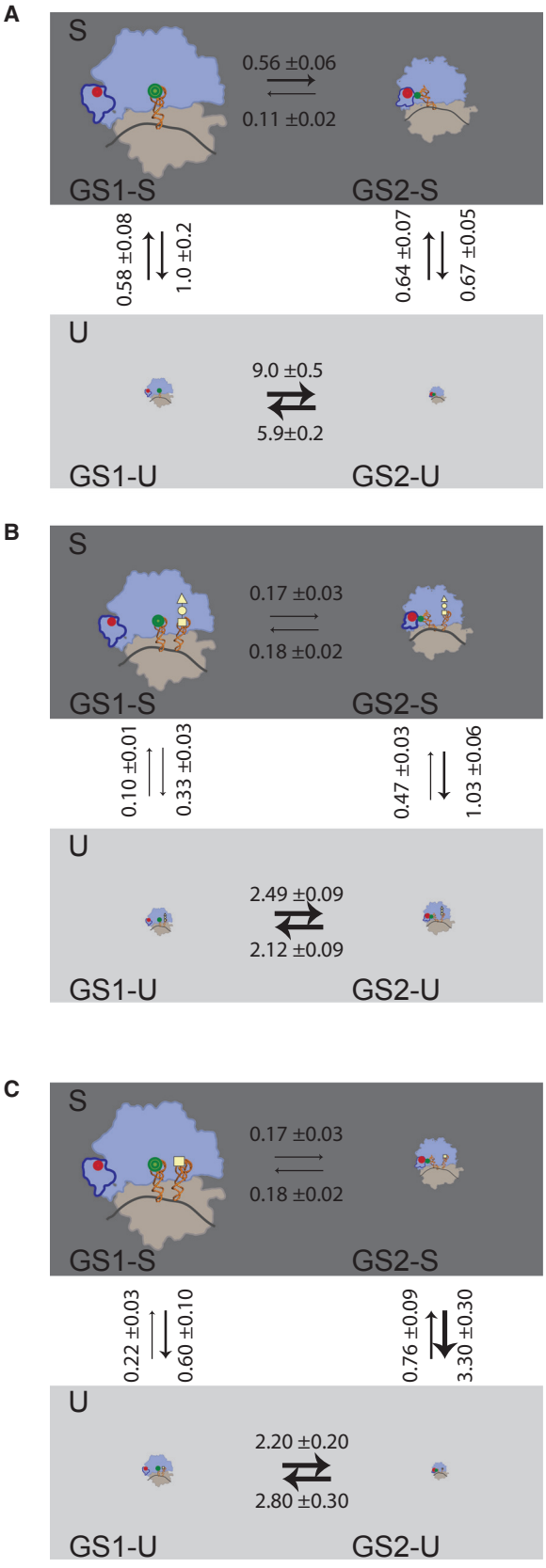

FIGURE 7 Quantified hierarchical kinetic mechanism of PRE^{-}, PRE^{fMFK}, and PRE^{K} dynamics. Fully quantified hierarchical kinetic mechanism describing the rates of transitions between GS1-S, GS2-S, GS1-U, and GS2-U for (*A*) PRE^{-}, (*B*) PRE^{fMFK}, and (*C*) PRE^{K}. Note that all rates are in units of s^{-1}, error values are 95% confidence intervals, and the relative sizes of the PRE complexes are proportional to the relative occupancies of GS1-S, GS2-S, GS1-U, and GS2-U. The boxes enclosing the various transitions are shaded as in Fig. 6.

kinetically characterize transitions observed in single-molecule signal trajectories whose underlying kinetic model can be best described by a hierarchical Markov chain with transitions between relatively well-separated free-energy minima. This kinetic model describes signal trajectories in which the rates of fluctuations between signal amplitudes that are observed along the dimension of a free-energy landscape that is directly detected in the experiment are modulated by the diffusion of the biomolecule along additional dimensions of the landscape that are not directly detected. hFRET, which uses the variational approximation to optimize the evidence lower bound, enables estimation of the parameters describing a set of hierarchical Markov chains that are consistent with a population of signal trajectories, as well as selection of the hierarchical Markov chain corresponding to the most probable kinetic model that is required to describe the population of signal trajectories. Complementing existing methods for analyzing single-molecule signal trajectories exhibiting kinetic heterogeneity arising from fluctuations between indirectly detected dimensions of a free-energy landscape (22,55–59), including a method described as this manuscript was in the final stages of preparation (60), hFRET enables experimentalists to directly quantify and select between kinetic models generated from hierarchical Markov chains of arbitrary complexity. Thus, hFRET uniquely provides experimentalists with a flexible, comprehensive, and statistically robust method for quantitatively characterizing alternative subpopulations of a biomolecule or biomolecular complex of interest and/or alternative pathways of a biological process of interest (5,6,12,14).

To demonstrate the ability of hFRET to analyze real-world, single-molecule signal trajectories, we have used it to analyze E_{FRET} trajectories reporting on fluctuations between the GS1 and GS2 states of ribosomal PRE complexes. The results of our analyses demonstrate that the most parsimonious hierarchical kinetic model is one that is composed of four states, denoted GS1-S, GS2-S, GS1-U, and GS2-U, in which fluctuations between GS1-S and GS2-S or GS1-U and GS2-U report on transitions along the directly detected dimension of the corresponding free-energy landscape, and fluctuations between GS1-S and GS1-U or GS2-S and GS2-U report on transitions along the indirectly detected dimension of the free-energy landscape. Interestingly, we find that in all three of the PRE complexes that we investigated, the majority of the E_{FRET} trajectories fluctuate reversibly between the S and U states, thereby exhibiting dynamic heterogeneity.

Moreover, our data provide insights into the physical origins of the dynamic heterogeneity of PRE complexes. For example, assuming that the dominant contributions to the energy barriers that separate the S states from the U states in PRE complexes are enthalpic, the fact that $k_{1,S \rightarrow 2,S}$ and $k_{2,S \rightarrow 1,S}$ are more than one order of magnitude smaller than $k_{1,U \rightarrow 2,U}$ and $k_{2,U \rightarrow 1,U}$, respectively, for all three PRE complexes strongly suggests that PRE complexes in the S states are able to form more and/or stronger intramolecular interactions than they can form in the corresponding U states. Interpreted within the context of PRE^{-}, which carries a tRNA only at the P site, this observation suggests the possibility that the aminoacyl-acceptor stem of the deacylated tRNA at the P site of PRE complexes can stochastically and reversibly sample at least two conformations. In one of these conformations, the aminoacyl-acceptor stem makes relatively more and/or stronger interactions with the 50S subunit P and E sites when the deacylated tRNA is in its "classical" and "hybrid" configurations, respectively, thereby giving rise to the GS1-S and GS2-S states. In the other conformation, the aminoacyl-acceptor stem makes relatively fewer and/or weaker interactions with the 50S subunit P and E sites when the deacylated tRNA is in its "classical" and "hybrid" configurations, respectively, thereby giving rise to the GS1-U and GS2-U states. Alternatively, or in addition, it is possible that the heterogeneity we observe in PRE^{-} originates from the ribosome itself. It is possible, for example, that the 30S (and/or 50S) subunit component of an intersubunit interaction can stochastically and reversibly sample at least two conformations that modulate the number and/or strength of the interactions that it could potentially make with its corresponding 50S (and/or 30S) subunit component while in the nonrotated and rotated intersubunit orientations, thereby giving rise to the S and U states. Thus, our findings strongly suggest that the deacylated tRNA at the P site and/or the ribosome itself are major contributors to the dynamic heterogeneity of PRE complexes.

Moreover, our observation that the rates of transitions between the S and U states, particularly $k_{1,U \rightarrow 1,S}$ and $k_{1,S \rightarrow 1,U}$, are sensitive to the presence of a peptidyl-tRNA in the A site of a PRE complex (compare Fig. 7, *A* and *B*) demonstrates that the peptidyl-tRNA at the A site also contributes to the dynamic heterogeneity of PRE complexes. Specifically, we hypothesize that presence of a peptidyl-tRNA at the A site can modulate the strength of the interactions that the tRNA at the P site can make with the 50S subunit P and E sites and/or the interactions that the 30S (and/or 50S) subunit component of an intersubunit interaction can make with its corresponding 50S (and/or 30S) subunit component in the nonrotated and rotated intersubunit orientations. This hypothesis is further supported and extended by the observation that the rates of transitions between the S and U states, particularly $k_{2,S \rightarrow 2,U}$ and $k_{1,S \rightarrow 1,U}$, are sensitive to the acylation status of the tRNA at the A site (compare Fig. 7, *B* and *C*). Thus, our findings strongly suggest that the presence and likely the identity, post-transcriptional modification status, length and sequence of the covalently attached peptide, etc., of the peptidyl-tRNA at the A site can modulate the P-site tRNA- and/or ribosome-mediated dynamic heterogeneity of PRE complexes.

Beyond the characterization of static and dynamic heterogeneity in single-molecule signal trajectories, the variational Bayesian HHMM-based approach that underlies hFRET can be adapted to address several other challenges in the field of single-molecule biophysics. Consider, for example, biological processes of greater complexity, in which a sample might exhibit static heterogeneity and each static subpopulation may or may not also exhibit dynamic heterogeneity. Currently, there are no computational methods for analyzing signal trajectories from such samples. Nonetheless, it should be possible to use the variational Bayesian inference-based approach that we have described here together with a set of kinetically non-interconverting hierarchical Markov chains to develop an hFRET-like algorithm that can be used to analyze signal trajectories of such complexity. In a second example, we note that there are currently no standard computational methods available for elucidating the single, most parsimonious kinetic model that is fully consistent with the signal trajectories from multiple data sets in which the signal trajectories from each data set report on the conformational dynamics of a distinct structural element of a biomolecule or biomolecular complex of interest (e.g., the various smFRET data sets associated with the different FRET donor and acceptor pairs that have been developed and used to investigate the intersubunit, L1 stalk, and tRNA dynamics of PRE complexes (13)). Using a variational Bayesian HHMM-based approach that builds on hFRET so as to treat the signal trajectories from each data set as reporting on a different dimension of the same free-energy landscape should allow the single, most parsimonious kinetic model that is fully consistent with all of the data sets to be selected. A final example is really an extension of the second example, in which the experimental signal trajectories from one data set are replaced by signal trajectories derived from molecular dynamics (MD) simulations. Such an approach would provide a robust method for integrating experimental and MD simulation data into a single kinetic model. Although the sampling period of the detectors that are used in most single-molecule biophysical experiments (~1–100 ms sampled for ~1–100 s) is currently much longer than the maximal time of an MD simulation (~10 ms (61)), rapid improvements in the detectors and increases in computational power will soon allow the timescales accessible to single-molecule experiments and MD simulations to be bridged.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.02.031.

AUTHOR CONTRIBUTIONS

J.H. and R.L.G. designed the research, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank Dr. Kelvin Caban, Dr. Colin Kinz-Thompson, and Dr. Thorsten Hugel for their comments on the manuscript.

This work was supported by a National Science Foundation CAREER award (MCB 0644262); National Institutes of Health grants under award numbers GM107417, GM084288, and GM119386; and an American Cancer Society grant under award number RSG-09-053-01-GMC to R.L.G.

REFERENCES

1. Tinoco, I., Jr., and R. L. Gonzalez, Jr. 2011. Biological mechanisms, one molecule at a time. *Genes Dev.* 25:1205–1231.
2. Kinz-Thompson, C. D., N. A. Bailey, and R. L. Gonzalez, Jr. 2016. Precisely and accurately inferring single-molecule rate constants. *Methods Enzymol.* 581:187–225.
3. Colquhoun, D., and A. G. Hawkes. 1982. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 300:1–59.
4. Hawkes, A. G., A. Jalali, and D. Colquhoun. 1992. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 337:383–404.
5. Tan, E., T. J. Wilson, …, T. Ha. 2003. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. USA.* 100:9308–9313.
6. Solomatin, S. V., M. Greenfeld, …, D. Herschlag. 2010. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature.* 463:681–684.
7. Sorgenfrei, S., C. Y. Chiu, …, K. L. Shepard. 2011. Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat. Nanotechnol.* 6:126–132.
8. Olsen, T. J., Y. Choi, …, G. A. Weiss. 2013. Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* 135:7855–7860.
9. Fei, J., J. E. Bronson, …, R. L. Gonzalez, Jr. 2009. Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc. Natl. Acad. Sci. USA.* 106:15702–15707.
10. 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.
11. Lee, J. Y., B. Okumus, …, T. Ha. 2005. Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA.* 102:18938–18943.
12. English, B. P., W. Min, …, X. S. Xie. 2006. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.* 2:87–94.
13. Frank, J., and R. L. Gonzalez, Jr. 2010. Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.* 79:381–412.
14. Rinaldi, A. J., P. E. Lund, …, N. G. Walter. 2016. The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nat. Commun.* 7:8976.
15. Bruno, W. J., J. Yang, and J. E. Pearson. 2005. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. *Proc. Natl. Acad. Sci. USA.* 102:6326–6331.
16. Kienker, P. 1989. Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B Biol. Sci.* 236:269–309.
17. Andrec, M., R. M. Levy, and D. S. Talaga. 2003. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden Markov models. *J. Phys. Chem. A.* 107:7454–7464.
18. McKinney, S. A., C. Joo, and T. Ha. 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* 91:1941–1951.
19. Qin, F., A. Auerbach, and F. Sachs. 2000. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* 79:1915–1927.
20. Bronson, J. E., J. Fei, …, C. H. Wiggins. 2009. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 97:3196–3205.
21. van de Meent, J. W., J. E. Bronson, …, C. H. Wiggins. 2013. Hierarchically-coupled hidden Markov models for learning kinetic rates from single-molecule data. *JMLR Workshop Conf. Proc.* 28:361–369.
22. van de Meent, J. W., J. E. Bronson, …, R. L. Gonzalez, Jr. 2014. Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.* 106:1327–1337.
23. Chen, Y., K. Shen, …, S. C. Kou. 2016. Analyzing single-molecule protein transportation experiments via hierarchical hidden markov models. *J. Am. Stat. Assoc.* 111:951–966.
24. Hille, B. 2001. Ion Channel Excitable Membranes. Sinauer Associates, Inc., Sunderland, MA.
25. Lakowicz, J. R. 2006. Principles of Fluorescence Spectroscopy. Springer, New York.
26. Roy, R., S. Hohng, and T. Ha. 2008. A practical guide to single-molecule FRET. *Nat. Methods.* 5:507–516.
27. A. Noy, ed 2008. Handbook of Molecular Force Spectroscopy Springer Science, New York.
28. Choi, Y., I. S. Moody, …, P. G. Collins. 2012. Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science.* 335:319–324.
29. Bouilly, D., J. Hon, …, C. Nuckolls. 2016. Single-molecule reaction chemistry in patterned nanowells. *Nano Lett.* 16:4679–4685.
30. Vernick, S., S. M. Trocchia, …, K. L. Shepard. 2017. Electrostatic melting in a single-molecule field-effect transistor with applications in genomic identification. *Nat. Commun.* 8:15450.
31. He, G., J. Li, …, X. Guo. 2016. Direct measurement of single-molecule DNA hybridization dynamics with single-base resolution. *Angew. Chem. Int. Ed. Engl.* 55:9036–9040.
32. Schuler, B., and W. A. Eaton. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18:16–26.
33. Rosenstein, J. K., S. G. Lemay, and K. L. Shepard. 2015. Single-molecule bioelectronics. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 7:475–493.
34. Colquhoun, D., and A. G. Hawkes. 1981. On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B Biol. Sci.* 211:205–235.
35. Kinz-Thompson, C. D., and R. L. Gonzalez, Jr. 2018. Increasing the time resolution of single-molecule experiments with bayesian inference. *Biophys. J.* 114:289–300.
36. Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Springer, New York.
37. Fine, S., Y. Singer, and N. Tishby. 1998. The hierarchical hidden Markov model : analysis and applications. *Mach. Learn.* 32:41–62.
38. Wakabayashi, K., and T. Miura. 2012. Forward-backward activation algorithm for hierarchical hidden Markov models. *In* Advances in Neural Information Processing Systems 25. F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. Curran Associates, Inc., pp. 1493–1501.
39. Murphy, K. P., and M. A. Paskin. 2001. Linear time inference in hierarchical HMMs. *In* Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic.

T. G. Dietterich, S. Becker, and Z. Ghahramani, eds. MIT Press, pp. 833–840.

40. Winn, J. M., and C. M. Bishop. 2005. Variational message passing. *J. Mach. Learn. Res.* 6:661–694.
41. Blei, D. M., A. Kucukelbir, and J. D. McAuliffe. 2016. Variational inference: a review for statisticians. *arXiv*, arXiv:1601.00670 http://arxiv.org/abs/1601.00670.
42. Todorovic, P. 1992. An Introduction to Stochastic Processes and Their Applications, Volume 26. Springer, New York.
43. Traut, R. R., and R. E. Monro. 1964. The puromycin reaction and its relation to protein synthesis. *J. Mol. Biol.* 10:63–72.
44. MacDougall, D. D., J. Fei, and R. L. Gonzalez, Jr. 2011. Single-molecule fluorescence resonance energy transfer investigations of ribosome-catalyzed protein synthesis. *In* Molecular Machines in Biology: Workshop of the Cell. J. Frank, ed. Cambridge University Press, pp. 93–116.
45. Murphy, M. C., I. Rasnik, …, T. Ha. 2004. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* 86:2530–2537.
46. Voorhees, R. M., and V. Ramakrishnan. 2013. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* 82:203–236.
47. Ning, W., J. Fei, and R. L. Gonzalez, Jr. 2014. The ribosome uses cooperative conformational changes to maximize and regulate the efficiency of translation. *Proc. Natl. Acad. Sci. USA.* 111:12073–12078.
48. Frank, J., and R. K. Agrawal. 2000. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature.* 406:318–322.
49. Jin, H., A. C. Kelley, and V. Ramakrishnan. 2011. Crystal structure of the hybrid state of ribosome in complex with the guanosine triphosphatase release factor 3. *Proc. Natl. Acad. Sci. USA.* 108:15798–15803.
50. Horan, L. H., and H. F. Noller. 2007. Intersubunit movement is required for ribosomal translocation. *Proc. Natl. Acad. Sci. USA.* 104:4881–4885.
51. Valle, M., A. Zavialov, …, J. Frank. 2003. Locking and unlocking of ribosomal motions. *Cell.* 114:123–134.
52. Trabuco, L. G., E. Schreiner, …, K. Schulten. 2010. The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. *J. Mol. Biol.* 402:741–760.
53. Cornish, P. V., D. N. Ermolenko, …, T. Ha. 2009. Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl. Acad. Sci. USA.* 106:2571–2576.
54. Cornish, P. V., D. N. Ermolenko, …, T. Ha. 2008. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell.* 30:578–588.
55. Rosales, R. A. 2004. MCMC for hidden Markov models incorporating aggregation of states and filtering. *Bull. Math. Biol.* 66:1173–1199.
56. Li, C. B., and T. Komatsuzaki. 2013. Aggregated markov model using time series of single molecule dwell times with minimum excessive information. *Phys. Rev. Lett.* 111:058301.
57. Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. Biol. Sci.* 264:375–383.
58. Blanco, M. R., J. S. Martin, …, N. G. Walter. 2015. Single Molecule Cluster Analysis dissects splicing pathway conformational dynamics. *Nat. Methods.* 12:1077–1084.
59. Schmid, S., M. Götz, and T. Hugel. 2016. Single-molecule analysis beyond dwell times: demonstration and assessment in and out of equilibrium. *Biophys. J.* 111:1375–1384.
60. Hwang, W., I. B. Lee, …, C. Hyeon. 2016. Decoding single molecule time traces with dynamic disorder. *PLoS Comput. Biol.* 12:e1005286.
61. Lindorff-Larsen, K., P. Maragakis, …, D. E. Shaw. 2016. Picosecond to millisecond structural dynamics in human ubiquitin. *J. Phys. Chem. B.* 120:8313–8320.
62. Selmer, M., C. M. Dunham, …, V. Ramakrishnan. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science.* 313:1935–1942.