Trajectories of the ribosome as a Brownian nanomachine

Ali Dashti[a,1], Peter Schwander[a,1], Robert Langlois[b], Russell Fung[a], Wen Li[b], Ahmad Hosseinizadeh[a], Hstau Y. Liao[b], Jesper Pallesen[c,2], Gyanesh Sharma[b,3], Vera A. Stupina[d], Anne E. Simon[d], Jonathan D. Dinman[d], Joachim Frank[b,c,4], and Abbas Ourmazd[a,1,4]

[a]Department of Physics, University of Wisconsin, Milwaukee, WI 53211; [b]Department of Biochemistry and Molecular Biophysics, and [c]Howard Hughes Medical Institute, Columbia University, New York, NY 10032; and [d]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742

Contributed by Joachim Frank, October 8, 2014 (sent for review September 10, 2014)

A Brownian machine, a tiny device buffeted by the random motions of molecules in the environment, is capable of exploiting these thermal motions for many of the conformational changes in its work cycle. Such machines are now thought to be ubiquitous, with the ribosome, a molecular machine responsible for protein synthesis, increasingly regarded as prototypical. Here we present a new analytical approach capable of determining the free-energy landscape and the continuous trajectories of molecular machines from a large number of snapshots obtained by cryogenic electron microscopy. We demonstrate this approach in the context of experimental cryogenic electron microscope images of a large ensemble of nontranslating ribosomes purified from yeast cells. The free-energy landscape is seen to contain a closed path of low energy, along which the ribosome exhibits conformational changes known to be associated with the elongation cycle. Our approach allows model-free quantitative analysis of the degrees of freedom and the energy landscape underlying continuous conformational changes in nanomachines, including those important for biological function.

cryo-electron microscopy | elongation cycle | manifold embedding | nanomachines | translation

Ideally, one would like to "see" the conformational changes of a Brownian machine as it traverses its work cycle trajectory over the energy landscape. This information is particularly relevant for a biologically important molecular machine such as the ribosome, which is responsible for protein synthesis in all living cells. During the so-called elongation process, the ribosome repeatedly links an amino acid carried in by transfer RNA (tRNA) to the nascent polypeptide chain, with the choice of amino acid in each cycle dictated by the genetic message on the mRNA. In the eukaryotic ribosome, this process is facilitated by elongation factors eEF1A and eEF2, both GTPases.

It is believed that many intermediate conformational states must be involved in the elongation cycle of the ribosome (1), but the evidence is inferred, albeit from an impressive array of experimental techniques. Both cryogenic electron microscopy (cryo-EM) (2) and X-ray crystallographic approaches (3) have been used to determine the structures of several biochemically "trapped" states along the conformational trajectory. However, it has been pointed out that these likely represent only a fraction of the relevant conformational states, that each biochemically trapped state may correspond to more than one conformational state, and that the observed intermediate structures may be affected by the trapping process itself (1). Powerful algorithms (4, 5) have been used to sort cryo-EM snapshots into a small number of discrete classes, each presumed to represent an intermediate state (6). In some cases, however, snapshots of major ribosomal regions with large conformational flexibility have defied classification into discrete states altogether, even by the most advanced analytical methods (7). Single-molecule FRET experiments have yielded evidence for discrete conformational changes in single,

freely equilibrating pretranslocational ribosomes, and provided ensemble averages for such changes, but have been unable to provide data for short-lived intermediates (8, 9).

In a groundbreaking study, Fischer and coworkers (10) used cryo-EM to determine the structure and occupancy of different ribosomal conformational states as a function of time, obtaining the free-energy landscape through the Boltzmann factor (11). The specific process studied was that of back-translocation, a slow process (on the order of 30 min), in which the elongation cycle is partially reversed through interaction with another GTPase: LepA. Cryo-EM snapshots were classified in a hierarchical series of supervised (reference-based) steps to yield multiple conformations, differing mainly in specific preselected features: degree of intersubunit rotation, tRNA positions, and degree of head swivel of the small subunit—all changes known to be associated with the elongation work cycle of the ribosome.

Significance

Many functions in the cell are performed by Brownian machines, macromolecular assemblies that use energy from the thermal environment for many of the conformational changes involved in their work cycles. Here we present a new approach capable of mapping the continuous motions of such nanomachines along their trajectories in the free-energy landscape and demonstrate this capability in the context of experimental cryogenic electron microscope snapshots of the ribosome, the nanomachine responsible for protein synthesis in all living organisms. We believe our approach constitutes a universal platform for the analysis of free-energy landscapes and conformational motions of molecular nanomachines and their dependencies on temperature, buffer conditions, and regulatory factors.

There is general agreement, however, that the value of any approach capable of revealing conformational changes along specific trajectories is determined by the extent to which it provides unbiased quantitative insight into the conformational degrees of freedom, the ensemble kinetics, and the free-energy landscape traversed by the system under examination (1, 12, 13).

Here, we present a new analytical approach capable of mapping continuous conformational changes of nanomachines along any trajectory in the energy landscape, without timing information, supervision, or templates. Our unbiased approach determines the number of degrees of freedom exercised during the observations, the energy landscape explored by the nanomachine, the energy trajectory traversed during the work cycle, and the continuous conformational changes associated with the work cycle. These novel capabilities constitute a powerful platform for quantitative study of the conformational and energy trajectories of nanomachines, including those engaged in a wide range of important biological processes. As reported below, the first application of this approach to an experimental cryo-EM dataset of ribosomes purified from yeast not engaged in translation reveals a closed trajectory of minimum energy along which the ribosomes appear to idle with conformational changes reminiscent of those observed during the protein elongation cycle.

Conceptual Outline

A flow diagram is shown in Fig. 1, with a more technical description following the conceptual outline. Details are presented in the *SI Text*. As used, for example, in early approaches to image classification in electron microscopy (14), a snapshot can be represented as a vector in high-dimensional space by regarding the intensity value at each pixel of the snapshot as a component of the vector. Distance is a measure of similarity between the snapshots in this space. A collection of snapshots produces one or more data clouds, with discrete conformations producing separate clouds. We have previously shown that this approach is able to determine 3D structure from a collection of ultra-low signal 2D snapshots of unknown orientation (15–18), and, in the presence of orientational heterogeneity and defocus variations, distinguish

between discrete conformations with best-in-class performance (17).

Here, we are concerned with continuous conformational changes, which produce correlations among the points representing the snapshots. These correlations define a hypersurface—a manifold—a concept found useful also in a previous approach to heterogeneity in cryo-EM (19). Using advanced machine-learning methods, we have demonstrated the ability to identify such manifolds in the presence of overwhelming noise and defocus variations (17).

Our approach begins with determining the orientations of the snapshots by any of a number of approaches (17), in this instance by a standard method used in electron microscopy (20). As shown previously (17, 19), this can be achieved without taking conformational heterogeneity into account, because the effects of orientational change dominate. Snapshots lying within a tight orientational aperture are then selected, and the manifold spanned by them is determined. This step in the analysis yields the conformational manifold (spectrum) in the selected viewing (projection) direction. Of course, the conformational changes may be governed by more than one parameter. In general, therefore, the conformational manifold is best described by a set of orthogonal coordinates. Such a description can be achieved by one of many well-established machine-learning techniques, which also reveal the intrinsic dimensionality of the manifold, and hence the number of degrees of freedom exercised by the system under observation.

Unfortunately, it is not possible to determine the conformational changes from such a description, because the local rates of change in multidimensional manifolds obtained from machine-learning techniques are, in general, unknown (21, 22) and cannot be easily related to the underlying changes in the system under observation (15, 23). To overcome this well-known difficulty, we introduce an additional step, in which the cloud of points is mapped to another coordinate system, where the local rates of change can be determined exactly, and related to the underlying conformational changes. This step leads to a representation of conformational change in terms of a universal parameter

Fig. 1. Flowchart representation of the approach used to determine the free-energy landscape, work-cycle trajectory, and associated continuous conformational changes from experimental snapshots of nanomachines in unknown orientational and conformational states.

(metric). The density of points in this space can now be related to the energy landscape sampled by the system through the Boltzmann factor $e^{-\Delta G/k_B T}$ (10), with ΔG denoting the change in the Gibbs free energy, k_B the Boltzmann constant, and T the temperature. The locus of minimum energy in this landscape represents the trajectory traversed by the machine during its work cycle. In each viewing direction, 2D movies can be compiled to reveal the conformational changes along the path of minimum energy or indeed any chosen trajectory. 3D movies can be compiled by stepping along such a trajectory and, in each step, performing a 3D reconstruction by integrating the information from many different viewing directions.

Analytical Procedure

The analysis is illustrated below in more technical terms with reference to a set of 849,914 experimental cryo-EM snapshots of 80S ribosomes from yeast, obtained in the course of a study of translational initiation by a plant virus (*SI Text*). The procedure of purification rendered the ribosome free of mRNA and most of the tRNAs.

Manifolds defined by data clouds are, in general, nonlinear: the data lie on intrinsically curved rather than flat hypersurfaces. Such manifolds can be identified and described (embedded) by graph-theoretic machine-learning techniques often used for dimensionality reduction. The so-called diffusion map embedding algorithm used in our work yields a description of a curved manifold in terms of the orthogonal eigenfunctions (more precisely eigenvectors) of known operators, specifically the Laplace–Beltrami operator (21, 22, 24) (*SI Text*). We use a specially developed kernel to deal with the substantial defocus variations encountered in cryo-EM data (17) (*SI Text*). In Fig. 2, we show representative cryo-EM snapshots (Fig. 2*A*) and a typical data manifold (Fig. 2*B*) described in terms of the first two eigenvectors of the Laplace–Beltrami operator. [In general, there is no simple correspondence between the eigenvectors obtained from this nonlinear analysis, and those obtained with linear approaches, such as principal component analysis (PCA) and singular value decomposition (SVD).] The manifold produced by the experimental snapshots is, in fact, nonlinear, with an intrinsic dimensionality of five (*SI Text*).

Fig. 2 describes the data in terms of the eigenfunctions of the Laplace–Beltrami operator with respect to an unknown metric (22). The absence of information on the metric precludes a consistent description of the conformational changes in terms of a known universal parameter. We solve this problem by mapping the manifold to another space, in which the eigenfunctions are known exactly. For this, we use an approach used in nonlinear Laplacian spectral analysis (NLSA) (25), a technique capable of performing SVD on nonlinear manifolds (nonlinear SVD for short) (*SI Text*). Briefly, one considers a collection of supervectors formed by concatenating snapshots falling within a window moving over the data vectors. The snapshots within each supervector are ordered according to the projections of the points representing them on a line through the origin, making an angle θ with, say, the horizontal axis (Fig. 2 and *SI Text*). By virtue of this projection, the arrangement picks out the conformational evolution along the selected line characterized by θ, with the random ordering of conformational changes along other directions assuming the character of noise. Nonlinear SVD (25) is then used to extract characteristic images (topos) and their evolutions (chronos) from these supervectors (*SI Text*). Each topo/chrono pair constitutes an element of a biorthogonal decomposition of the conformational changes along the selected line. Noise-reduced snapshots can be reconstructed from the topo/chrono pairs with significant (above-noise) singular values and embedded to obtain the manifold characteristic of the conformational changes along the selected line (*SI Text*). By construction, this manifold is one-dimensional, described in terms of known eigenfunctions, viz. $\cos(k\pi\tau)$, $k = 1, 2, 3, \ldots$, and governed by the single parameter τ (*SI Text*).

Given a sufficiently dense collection of radial lines, each making an angle θ with the horizontal axis, the conformational changes can be described in any direction in the multidimensional space of conformations by the parameter $\tau(\theta)$. The mapping from a space of unknown metric and hence eigenfunctions to one characterized by known eigenfunctions governed by a single parameter $\tau(\theta)$ allows a consistent description of the conformational changes (*SI Text* and Fig. S1). This description of conformational changes in a given projection direction can be related to descriptions in other projection directions by assuming that the same conformational spectrum is viewed in all projection directions. In other words, the histograms of occupancy vs.
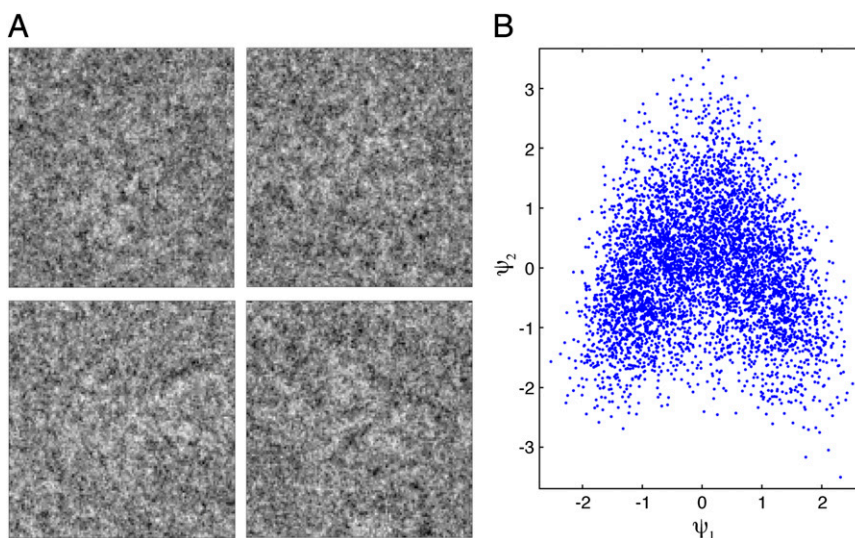

Fig. 2. (*A*) Representative cryo-EM snapshots, and (*B*) 2D view of a typical conformational manifold. The manifold is derived from a measure of similarity among ~1,500 cryo-EM snapshots of ribosome particles viewed within a tight orientational aperture. The axes ψ_1, ψ_2 represent the first two eigenvectors obtained by the diffusion map algorithm with a kernel able to deal with defocus variations.

conformational parameter $\tau(\theta)$ for different projection directions all represent the same spectrum of conformations and can thus be equalized. This equalization allows a universal description of the conformational spectrum across all projection directions. For any selected projected direction, the energy landscape along any line can be determined from the density of points along that line via the Boltzmann factor. Energy landscapes of higher dimension (with the dimensionality given by the number of eigenvectors) can be reconstructed by a tomographic extension of this approach (Fig. 3 and *SI Text*). Here, we concentrate on the first two eigenvectors, which contain some of the most interesting biological information (see below).

With this information in hand, the structural evolution of the system, including 2D movies of the conformational changes and the system's thermodynamic properties, can be quantitatively investigated in any viewing direction (*SI Text* and Movie S1). For any point on the energy landscape, a 3D structure map can be compiled by integrating the 2D information from many viewing directions into a 3D representation (Fig. 4) (*SI Text* and Movies S2–S4). Thus, 3D movies can be compiled by stepping along any selected trajectory in the energy landscape and reconstructing a 3D structure map in each step. The conformational changes along the closed minimum-energy trajectory of Fig. 3B are summarized in Fig. 3 and exemplified in Fig. 4. The associated movies (Movies S2–S4) present the ribosome as it evolves, with each movie corresponding to a different viewing direction selected for optimal visibility of the biologically relevant domain

motions (Fig. 3A). We note that, by construction, these movies are based on similarity rather than time. Accordingly, identical trajectories in opposite directions cannot be distinguished.

Results

The trough of minimum free energy is seen to form a closed, roughly triangular path (Fig. 3B), with variations of <0.9 kcal/mol ($1.5k_BT$ at room temperature) in the energy positions of the deepest points. The energy difference between the lowest and highest points of the entire landscape, corresponding to states with the highest and lowest nonzero occupancy in the experiment, amount to 3.8 ± 0.65 kcal/mol ($6.5 \pm 1.0k_BT$ at room temperature), with the uncertainty stemming chiefly from the low occupancy of high-energy states. The energy range covered nonetheless indicates that so-called transition states with energies several times the thermal energy have been probed in our experiment.

We examine the 3D movies showing the conformational changes along the closed triangular trajectory, analyzing in detail seven structure maps along the way (*SI Text* and Movie S5). Using rigid-body fitting of domains while making use of a published cryo-EM map of the yeast ribosome (26) for reference, we observe combinations of four motions previously described in the literature and associated with the elongation cycle, as follows: (*i*) ratchet-like intersubunit rotation, a counterclockwise rotation of the small subunit about an axis normal to the subunit interface (Fig. 3A, solvent view) (27, 28); (*ii*) rotation (closing movement) of
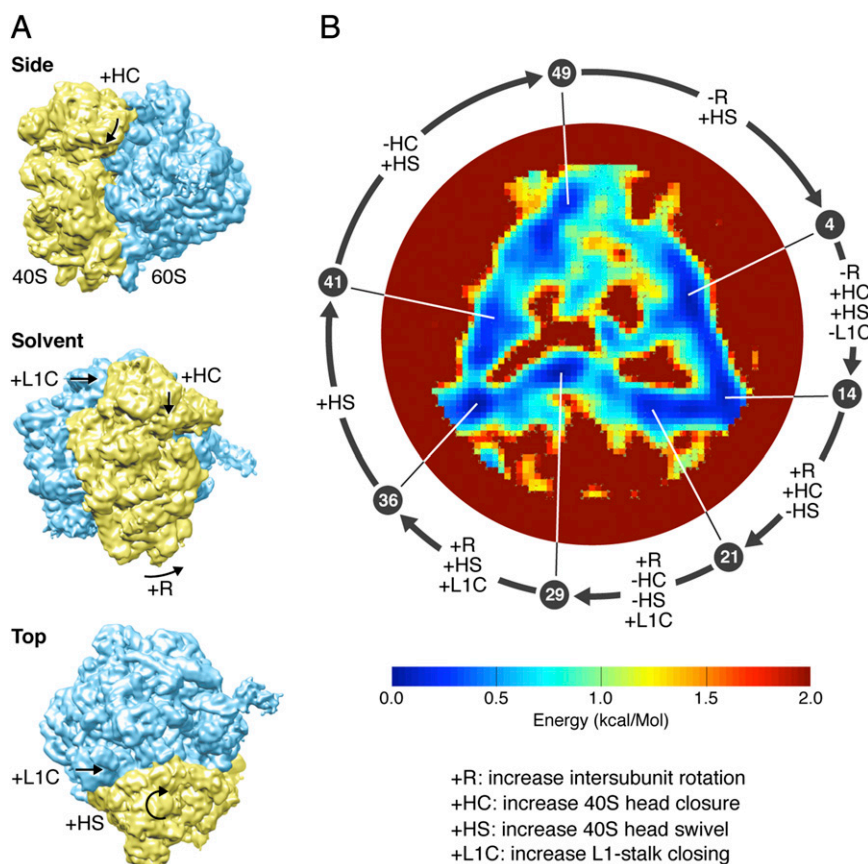

Fig. 3. (A) Three views of a cryo-EM map of the 80S ribosome from yeast (32), with arrows indicating four key conformational changes associated with the elongation work cycle of the ribosome. (B) The energy landscape traversed by the ribosome. The color bar shows the energy scale. The energy range has been truncated at 2 kcal/mol to show details of the triangular trough. The error in energy determination along the closed triangle is 0.05 kcal/mol. The roughly triangular minimum free-energy trajectory is divided into 50 states. The arrows indicate the structural changes between 7 selected states, each identified by its place in the sequence of 50 states.
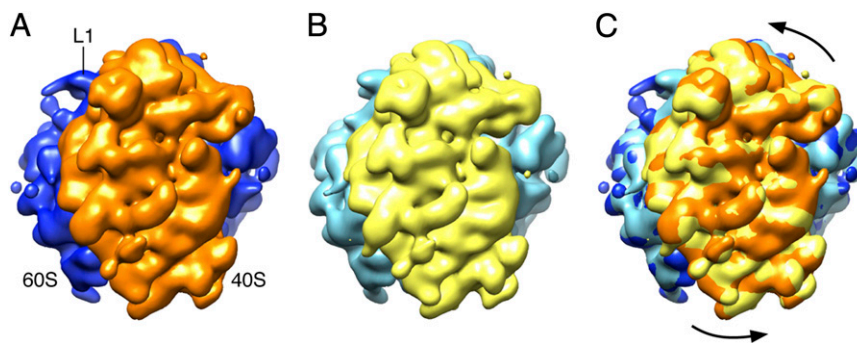
Fig. 4. Example of conformational changes along the trajectory: ratchet-like motion. (*A*) Unrotated ribosome, map 14 in Fig. 3*B*. (*B*) Maximally rotated ribosome, map 36. (*C*) Superposition of the two maps. The full set of frames showing continuous conformational changes is shown in three common viewing directions in Movies S2–S4.

the L1 stalk toward the intersubunit space (Fig. 3*A*, solvent and top views) (29); (*iii*) small subunit head swivel, a rotation of the subunit head about its long axis (Fig. 3*A*, top view) (30); and (*iv*) small subunit head closure, a "nodding" motion of the head (Fig. 3*A*, side and solvent views) (31). The first three motions are correlated and known to be associated with the process of mRNA–tRNA translocation (30, 32), whereas the fourth is known to accompany the selection of cognate aminoacyl tRNA during the process of decoding (31).

Proceeding clockwise along the minimum free-energy pathway, these motions occur in the sequence encountered during the elongation cycle: starting at the apex of the triangle and moving to the right side of the base (Fig. 3*B*), we see the small subunit head closure as a prominent motion, corresponding to the tRNA selection step. Next, proceeding along the base of the triangle, three coordinated movements of small subunit body, head swivel, and L1 stalk occur—all expected during translocation—while the small subunit head remains in the closed position. Finally, moving up from the left side of the base to the apex, we observe the reversal of head closing (the conformational change associated with the disengagement of the mRNA from the decoding center), as well as the reversal of the intersubunit motion.

Discussion

The energy landscape constructed by the analysis of the ribosome ensemble shows multiple interconnected paths of relatively low energy. Of these, a path of roughly triangular shape clearly stands out, encompassing roughly one third of the molecules. Comparison of 3D reconstructions from data at selected points along this path reveals conformational changes previously observed in the elongation cycle of fully programmed, translating ribosomes. Because the ribosomes captured in our snapshots were not engaged in protein synthesis, lacking mRNA and aminoacylated tRNAs, we must assume that their idle motions in the thermal environment sample the conformational space permitted by the degrees of freedom, thus exhibiting the conformational changes that would be productive in the presence of the ligands of the translational machinery (mRNA, aa-tRNA, eEF2, and eEF1A). Indeed, idling of the pretranslocational ribosome in the absence of elongation factors along the direction of the most prominent conformational changes (intersubunit motion and opening/closing of L1 stalk) has been previously observed by single-molecule FRET (8, 9) and cryo-EM (33). Idling of the empty ribosome along the same path has also been inferred from a series of X-ray structures (34). It remains to be established by the approach outlined in this paper whether fully programmed ribosomes follow the same path identified here. As to the other paths traversing the free-energy landscape (Fig. 3*B*), it is tempting to speculate that

they may represent alternative routes of the molecular machine under different buffer and temperature conditions.

We now discuss the salient features of our approach. First, the usefulness of movies of Brownian machines has been rightly questioned, because each trajectory of a single machine is strongly influenced by stochastic factors and thus is unique (1). The movies presented here, however, integrate information from a large ensemble of snapshots, each stemming from an object viewed only once. By using manifolds to capture the properties of the entire dataset, and nonlinear singular value analysis to suppress noise, our approach offers an efficient means for extracting the ensemble kinetics, i.e., the information common to a collection of objects, each viewed in an initially unknown orientational and conformational state.

Most successful experimental studies of the conformational spectra of biological machines have been hitherto restricted to sorting snapshots into a small number of classes (10, 11), using templates in some form, relying on timing information, or a combination of these tools (10). In contrast, our approach naturally yields detailed conformational trajectories and energy landscapes without a priori information or assumptions and at moderate computational expense (*SI Text*).

Of course, the maximum number of detectable conformational states is limited by various factors. These factors include the frequency (via the Shannon–Nyquist theorem) with which the conformational spectrum has been sampled, the information content of the individual snapshots, and, ultimately, the atomic nature of the object itself. In our analysis, the conformational signal stems primarily from the most heavily sampled projection directions, in each of which the conformational spectrum is randomly sampled by up to 5,000 snapshots. The number of meaningfully distinct conformational states is governed by the signal-to-noise ratio, which determines the statistical confidence with which neighboring states can be distinguished. In this study, the requirement for neighboring conformational states to be separated by 3 SDs (3 σ) means that ~50 conformational states can be distinguished (*SI Text*). This number is about an order of magnitude larger than previously achieved without timing information or templates. In combination with the recent availability of large cryo-EM datasets with near-atomic resolution (7, 35), our approach promises the possibility to extract conformational information with unprecedented detail.

In recent years, increasing computing power has fueled efforts to simulate the conformational trajectories of molecular machines, particularly the ribosome (36), by molecular dynamics. Experimental results obtained with the new approach presented here offer the promise to guide these efforts and provide the means for verifying important modeling assumptions.

Conclusions

Given our increasingly detailed knowledge of the structure of biological machines in general, and the ribosome in particular, it has been suggested that the "heroic age" of biostructure determination is drawing to a close and that further progress requires the study of ensemble kinetics and conformational energy landscapes (1). As shown by the results presented here, our approach offers a powerful platform for the quantitative study of continuous conformational motions over the energy landscapes traversed by biological machines in the course of their operation. This development may significantly accelerate the dawn of the predicted new era in biology. More generally, the way to the quantitative study of the structure, kinetics, and operational cycles of nanomachines is now open.

1. Moore PB (2012) How should we think about the ribosome? *Annu Rev Biophys* 41(1):1–19.
2. Villa E, et al. (2009) Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci USA* 106(4):1063–1068.
3. Korostelev A, Ermolenko DN, Noller HF (2008) Structural dynamics of the ribosome. *Curr Opin Chem Biol* 12(6):674–683.
4. Scheres SHW (2010) Maximum-likelihood methods in cryo-EM. Part II: Application to experimental data. *Methods Enzymol* 482:295–320.
5. Scheres SH (2012) A Bayesian view on cryo-EM structure determination. *J Mol Biol* 415(2):406–418.
6. Frank J (2012) Intermediate states during mRNA-tRNA translocation. *Curr Opin Struct Biol* 22(6):778–785.
7. Amunts A, et al. (2014) Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343(6178):1485–1489.
8. Fei J, Kosuri P, MacDougall DD, Gonzalez RL, Jr (2008) Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol Cell* 30(3):348–359.
9. Cornish PV, Ermolenko DN, Noller HF, Ha T (2008) Spontaneous intersubunit rotation in single ribosomes. *Mol Cell* 30(5):578–588.
10. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H (2010) Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466(7304):329–333.
11. Agirrezabala X, et al. (2012) Structural characterization of mRNA-tRNA translocation intermediates. *Proc Natl Acad Sci USA* 109(16):6094–6099.
12. Frank J, Gonzalez RL, Jr (2010) Structure and dynamics of a processive Brownian motor: The translating ribosome. *Annu Rev Biochem* 79(9):381–412.
13. Munro JB, Sanbonmatsu KY, Spahn CMT, Blanchard SC (2009) Navigating the ribosome's metastable energy landscape. *Trends Biochem Sci* 34(8):390–400.
14. van Heel M, Frank J (1981) Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* 6(2):187–194.
15. Giannakis D, Schwander P, Ourmazd A (2012) The symmetries of image formation by scattering. I. Theoretical framework. *Opt Express* 20(12):12799–12826.
16. Schwander P, Giannakis D, Yoon CH, Ourmazd A (2012) The symmetries of image formation by scattering. II. Applications. *Opt Express* 20(12):12827–12849.
17. Schwander P, Fung R, Ourmazd A (2014) Conformations of macromolecules and their complexes from heterogeneous datasets. *Philos Trans R Soc Lond B Biol Sci* 369(1647):20130567.
18. Hosseinizadeh A, et al. (2014) High-resolution structure of viruses from random diffraction snapshots. *Philos Trans R Soc Lond B Biol Sci* 369(1647):20130326.
19. Fu J, Gao H, Frank J (2007) Unsupervised classification of single particles by cluster tracking in multi-dimensional space. *J Struct Biol* 157(1):226–239.
20. Penczek PA, Grassucci RA, Frank J (1994) The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53(3):251–270.
21. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396.
22. Coifman RR, et al. (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc Natl Acad Sci USA* 102(21):7426–7431.
23. Coifman RR, Shkolnisky Y, Sigworth FJ, Singer A (2010) Reference free structure determination through eigenvectors of center of mass operators. *Appl Comput Harmon Anal* 28(3):296–312.
24. Coifman RR, Kevrekidis IG, Lafon S, Maggioni M, Nadler B (2008) Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Model Simulation* 7(2):842–864.
25. Giannakis D, Majda AJ (2012) Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc Natl Acad Sci USA* 109(7):2222–2227.
26. Taylor DJ, et al. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17(12):1591–1604.
27. Frank J, Agrawal RK (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406(6793):318–322.
28. Budkevich T, et al. (2011) Structure and dynamics of the mammalian ribosomal pre-translocation complex. *Mol Cell* 44(2):214–224.
29. Valle M, et al. (2003) Locking and unlocking of ribosomal motions. *Cell* 114(1):123–134.
30. Ratje AH, et al. (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* 468(7324):713–716.
31. Ogle JM, Carter AP, Ramakrishnan V (2003) Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci* 28(5):259–266.
32. Frank J, Gao H, Sengupta J, Gao N, Taylor DJ (2007) The process of mRNA-tRNA translocation. *Proc Natl Acad Sci USA* 104(50):19671–19678.
33. Agirrezabala X, et al. (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol Cell* 32(2):190–197.
34. Zhang W, Dunkle JA, Cate JHD (2009) Structures of the ribosome in intermediate states of ratcheting. *Science* 325(5943):1014–1017.
35. Liao M, Cao E, Julius D, Cheng Y (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 504(7478):107–112.
36. Whitford PC, Blanchard SC, Cate JHD, Sanbonmatsu KY (2013) Connecting the kinetics and energy landscape of tRNA translocation on the ribosome. *PLOS Comput Biol* 9(3):e1003003.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Supporting Information

Dashti et al. 10.1073/pnas.1419276111

SI Text

Cryo-EM Dataset.

Description. The dataset consisted of 849,914 projection images ("particles") of the yeast 80S ribosome embedded in a thin layer of ice, obtained with a transmission electron microscope, as detailed below.

Ribosome preparation. Purification of active yeast 80S ribosomes using cysteine-charged sulfolink columns was performed as described in Leshin et al. (1).

Electron microscopy. A carbon-coated Quantifoil 2/4 grid (Quantifoil Micro Tools) was prepared following standard cryo-EM procedures (2). Grids were glow-discharged for 25 s in an H_2, O_2 atmosphere, with 15 W and 25 s using a plasma cleaner (Gatan Solarus Model 950 Advanced Plasma System; Gatan) to render them hydrophilic. An aliquot of 4 μL containing the sample at a concentration of 50 nM was pipetted onto the carbon-coated grid. The grid was blotted in 100% humidity at 4 °C for 5 s and plunge-frozen into liquid ethane cooled by liquid nitrogen using the Vitrobot (FEI). Images were recorded with a 4k × 4k Tietz CCD detector mounted on an FEI Tecnai Polara electron microscope operating at 300 kV and a nominal magnification of 80,000× under low-dose conditions (∼20 e⁻/Å²). The resulting pixel size was 1.5 Å on the object scale. For automated data collection, the program AutoEMation (3) was used.

Image preprocessing and orientation determination. After visual inspection and evaluation of the micrographs and their power spectra, ∼4,700 micrographs were selected for further analysis. The total number of computer-selected particle candidates was ∼1,100,000, brought down to 849,914 by manual verification. Orientations of particles were determined by iterative projection matching using the pySPIDER (4) software package with an angular step size decreasing from 15° to 0.5°. Electron Microscopy Data Bank (EMDB) 1067 (5), low pass-filtered to 70Å, was used as the initial reference. Contrast transfer functions (CTFs) were determined for each micrograph in the standard way (6). Before further analysis, the snapshots within a small aperture over the two-sphere of projection directions were brought into in-plane orientational registry.

Manifold Embedding by Diffusion Map.

Outline. Common to all manifold embedding algorithms is the representation of manifolds in a low-dimensional Euclidean space (7–12). The family of diffusion maps, introduced by Coifman and coworkers (11, 12), establishes a rigorous link between the eigenfunctions (more precisely eigenvectors) ψ_i of the Laplace–Beltrami operator with respect to the Riemannian metric g

$$\Delta_g = -\frac{1}{\sqrt{\det(g)}}\partial_i\left[\sqrt{\det(g)}g^{ij}\partial_j\right] \quad i,j=1\ldots n, \quad \text{[S1]}$$

and the similarity between snapshots, as measured by the diffusion distance between them. The intrinsic dimensionality of the manifold, determined by well-established procedures (13), corresponds to the number of degrees of freedom exercised by the system under observation.

Defocus-tolerant kernel. The information content of a cryo-EM snapshot depends on the defocus at which it is obtained. The effect of defocus on the similarity measure between two otherwise identical snapshots must be eliminated. We achieve this by a double-filtering kernel, which corrects the Euclidean distance D_{ij}, as follows:

$$D_{ij}^2 = \sum_{pixels} |PSF_j \otimes I_i - PSF_i \otimes I_j|^2$$
$$= \sum_{pixels} |PSF_j \otimes (PSF_i \otimes P_i) - PSF_i \otimes (PSF_j \otimes P_j)|^2. \quad \text{[S2]}$$
$$= \sum_{pixels} |PSF_i \otimes PSF_j \otimes \{P_i - P_j\}|^2$$

For each snapshot i, I_i represents the image intensity distribution, PSF_i the microscope point-spread function, P_i the projected potential, and $\otimes$ the convolution operator. This scheme ensures a zero Euclidean distance between two snapshots differing in defocus only. For computational efficiency, the distances are calculated in Fourier space so that convolution becomes multiplication. Let $\tilde{I}_i$ be the Fourier transform of the image and CTF_i the Fourier transform of the point-spread function. The application of Parseval's theorem (14) yields

$$D_{ij}^2 = \sum_q |CTF_j \cdot \tilde{I}_i - CTF_i \cdot \tilde{I}_j|^2. \quad \text{[S3]}$$

One-dimensional manifolds. Snapshots of a system exercising only one degree of freedom give rise to a 1D Riemannian manifold described in terms of the eigenfunctions of the operator

$$\Delta_g = -\frac{1}{\sqrt{g}}\partial\left(\frac{1}{\sqrt{g}}\partial\right). \quad \text{[S4]}$$

For such manifolds, it is always possible to select a coordinate system in which the metric $g(\tau) = 1$. The eigenvalue equation then assumes the simple form

$$\frac{\partial^2 \psi(\tau)}{\partial \tau^2} = -\lambda \psi(\tau), \quad \text{[S5]}$$

whose normalized eigenfunctions satisfying the Neumann boundary conditions are of the form $\sqrt{2}\cos(k\pi\tau)$. It can be easily seen that a representation of a bounded 1D manifold in terms of the first two eigenfunctions ($k = 1, 2$) is a parabola. Unbounded 1D manifolds are characterized by eigenfunctions of the form $\sqrt{2}\cos(2k\pi\tau)$ and $\sqrt{2}\sin(2k\pi\tau)$. The representation of an unbounded 1D manifold in terms of the first two eigenfunctions is thus a circle. For both types of 1D manifold, each snapshot is associated with a value of τ (modulo 2π for the unbounded type), regardless of the dimensionality of the representation (number of eigenfunctions used).

SVD on Nonlinear Manifolds.

Outline. SVD is a data-analytical approach capable of extracting characteristic images (topos) and their time evolutions (chronos) from large noisy datasets (15, 16). Like PCA, SVD assumes a linear structure in the data. Geometrically, this is tantamount to the data lying on a flat hyperplane (16). Often, however, the data have a nonlinear intrinsic structure and thus lie on a curved hyperplane (nonlinear manifold) (7), precluding a straightforward application of SVD. This limitation has been overcome by a machine-learning technique (NLSA), which is able to perform SVD on nonlinear manifolds (17, 18).

Concatenation. NLSA analysis requires a series of snapshots ordered according to a single parameter. In our approach, the snapshot order is determined by the projected position of the embedded point representing it on a trajectory. This trajectory can constitute a straight line, for example, the ψ_1 or ψ_2 axis above, or any other trajectory of interest. The analysis proceeds

as follows. First, ordered (time-lagged) embedding (19–22) is used to obtain c-fold concatenated supervectors X from a dataset consisting of vectors x, with a typical supervector taking the form

$$X_t = (x_t, x_{t-\delta t}, \dots x_{t-(c-1)\delta t}).$$ [S6]

The timestamp assigned to each supervector corresponds to the mean of the timestamps of its constituent vectors. Second, graph-based machine learning (9, 11, 23, 24), here the diffusion map algorithm (11, 12), is used to identify the nonlinear data manifold formed by the collection of supervectors. This analysis provides a Euclidean description of the manifold in terms of the eigenfunctions of the Laplace–Beltrami operator, collectively referred to as Ψ below.

Nonlinear SVD. The supervectors X_t are projected onto the manifold to obtain the matrix A

$$A = \mathbf{X}\mu\Psi,$$ [S7]

with $\mathbf{X}$ representing the matrix of supervectors X_t, μ the Riemannian measure of the manifold, and the (empirical) eigenfunction Ψ is a truncated set of the eigenfunctions of the Laplace–Beltrami operator on the manifold (17). This Euclidean description of the nonlinear manifold allows one to analyze the matrix A by standard SVD.

Reconstruction and parameter extraction. The modes (specifically the chronos) obtained from SVD are projected from the space spanned by Ψ back to the time domain, and the topos corresponding to the supervectors are unwrapped to obtain individual snapshots (17). These snapshots reflect the conformational changes along the selected trajectory (e.g., a line in embedding space making an angle θ with the horizontal axis). The embedding of these snapshots by diffusion map produces an excellent 1D manifold, accurately described by eigenfunctions of the form $\sqrt{2}\cos(k\pi\tau)$. The conformational parameter $\tau(\theta)$ is extracted from the 1D manifold using the first three eigenvectors of the 1D manifold (Fig. S1).

The mapping from a description of the data in terms of unknown eigenfunctions of the Laplace–Beltrami operator ψ_i to one in terms of known eigenfunctions of the type $\cos(k\pi\tau)$ is a key feature of our analysis. The resulting conformational parameterization in terms of τ allows a consistent approach to reconstructing the energy landscape and compiling movies along any trajectory, including the minimum-energy path.

Free-Energy Landscape. We use a tomographic approach to retrieve the 2D free-energy landscape from snapshots obtained by NLSA. The dimensionality here refers to the number of eigenfunctions ψ_i considered in the analysis. The density of points $n(\tau, \theta)$ on the 1D manifold stemming from concatenation along a straight line at an angle θ to the horizontal axis is the result of projecting the data cloud onto that line. Thus, $n(\tau, \theta)$ is a measure of the density of points perpendicular to the line at a point characterized by τ. Thus, $n(\tau, \theta)$ is, in effect, the Radon transform of the embedded data cloud in the direction $(\theta + \pi/2)$. The set of $n(\tau, \theta)$ compiled over the range $(0 \leq \theta \leq \pi)$ thus represents a collection of Radon transforms of the cloud of points in the space of ψ_i. The inverse Radon transform reproduces the cloud of points, described in terms of the known eigenfunctions above.

This approach was implemented using 180 concatenation directions at $1°$ intervals. The projected density $n(\tau, \theta)$ was deduced from the $\tau(\theta)$ values as follows. For each direction, the $\tau(\theta)$ values were centered to zero-mean to align the different Radon transforms. The density was obtained by counting the number of points in each of 50 classes. Inversion was performed by filtered back-projection with a Shepp–Logan filter. The energy landscape was derived from the density of points via the Boltzmann factor, $n(\tau, \theta) = e^{-\Delta G/k_B T}$, with ΔG the change in Gibb's free energy and

the usual meaning for the other parameters. It can be readily shown that the error in this approach is determined by counting statistics. In our case, this is about 0.05 kcal/mol along the minimum-energy trajectory.

To retrieve n-dimensional energy landscapes (with n being the number of ψ_i included), concatenation is performed along directions defined by an $(n-1)$ sphere.

Compiling Movies and 3D Volumes.

2D movies. We now address the compilation of movies showing the continuous conformational changes as seen in a given projection direction (2D movies). These movies consist of a sequence of NLSA reconstructed snapshots, ordered along the minimum-energy path in the energy landscape (Movies S1–S5). To identify the raw snapshots in the neighborhood of this path, the trajectory is mapped back from space of known eigenfunctions to the space of ψ_i using Eq. S7 and the procedure described below. NLSA is then applied to the snapshots along the path to generate NLSA reconstructed snapshots as frames of the 2D movie.

The movies were compiled using the MATLAB video writer object, with the color axis for all frames normalized to the minimum and maximum intensities of the ensemble; 12.5 frames are shown per second to produce a 50-frame, 4-s movie.

Patching together information from different orientations. Because the minimum energy trajectory is a closed path (Fig. 3), embedding reconstructed snapshots along this path results in a circular manifold, characterized by a single angular parameter $0 \leq \tau < 2\pi$, which serves as a conformational parameter. In principle, the parameter τ can be used to establish equivalent conformational classes across projection (viewing) directions and thus reconstruct a 3D volume for each conformational class. However, the zero of the parameter τ in each direction is arbitrary. The resulting ambiguity can be resolved as follows:

i) At each point along a great circle, all snapshots spanning the conformational spectrum are concatenated into a supervector ordered according to increasing τ in that projection direction.

ii) The correlation coefficient between two consecutive supervectors along a great circle is calculated.

iii) The same correlation coefficient is calculated with the snapshot sequence for all possible starting (zero-τ) positions, with and without reversing the sense of the sequence.

iv) The largest correlation identifies the mutually consistent sense and starting point for adjacent conformational spectra. The results are propagated along the great circle by repeating the process above.

v) The conformational sense established along one great circle is propagated to other great circles via their points of intersection with the first.

The above approach is robust and computationally fast.

The rate at which the conformational parameter τ changes along the 1D manifold can vary slightly from projection direction to projection direction due to orientational inhomogeneity of the metric g (Eq. S4). We compensate for this effect by equalizing the histogram of the τ parameter in each projection direction to that corresponding to the union of 10 projection directions with the most active conformational spectra (measured by the magnitude of the singular values in the NLSA reconstruction along the selected trajectory). The efficacy of this approach has been validated by simulation.

3D volumes and movies. 3D movies were compiled using University of California, San Francisco (UCSF) Chimera (25), taking orthographic surface representations of output volumes for the 50 reconstructed classes used as input frames. Each volume was first segmented into large and small subunits using Chimera. A Python script was used to capture the sequence of frames from three standard points of view (side view, solvent view of the 40S

subunit, and top view; Movies S2–S4, respectively). Movie S1 shows the conformational changes between the seven maps of Fig. 3 as seen from the standard viewing directions.

Computational Details.

Hardware. Computations were performed on the following cluster: 16 CPU nodes, each consisting of two Intel Xeon E5620 CPUs @2.4 GHz with four cores; 48 GB of memory; and 200 GB of local hard drive.

Software. Snapshots were picked and preprocessed using SPIDER (4). The NLSA pipeline, energy landscape reconstruction, and compilation of 2D movies were implemented in MATLAB. 3D reconstruction was performed with the Arachnid software package (4). UCSF Chimera was used for visualization and compilation of 3D movies.

Computational expense and scaling. The NLSA pipeline, an embarrassingly parallelizable problem for different projection directions, was implemented using the MATLAB parallel processing toolbox. NLSA data analysis with 32 MATLAB workers requires ~10 h for 320 projection directions with a total of 184,500 snapshots. The number of snapshots in the most densely populated projection direction is ~6,500. Reconstruction of the energy landscape for all 320 projection directions requires ~7 d on the same cluster. 3D reconstruction of volumes for 50 different conformations can be completed in ~1 h using one cluster node. Due to the nonuniform density of snapshots over the two-sphere, the computational expense of the NLSA varies across projection directions. The data analysis process is thus asynchronous between nodes, with the projection direction with the largest number of snapshots determining the total computation time. Generation of a (k-nearest neighbor) k-NN graph of all snapshots in each projection direction is the most time-consuming part of NLSA and thus a good measure of algorithmic scaling. This task has $O(n_s^2 \times d)$ complexity, with n_s as the number of snapshots in each projection direction and d as the number of pixels in each snapshot.

Number of Conformational Classes. We consider the effect of SVD on noise and the number of distinguishable conformational classes with reference to the process used to extract continuous conformations:

i) Embed data by diffusion map. This step produces a cloud of points, viewed, e.g., in projection on the (ψ_1, ψ_2) plane.
ii) Reconstruct 2D energy landscape using NLSA.
iii) Reconstruct NLSA snapshots along a selected trajectory and embed to obtain a 1D manifold.

This process is designed to extract the conformational variations along the selected trajectory, e.g., the minimum-energy path.

The extent to which two conformations can be distinguished is determined by two factors: (*i*) the relevant conformational signal, as manifested in the extent of the data cloud along the trajectory, hereon designated Δ; and (*ii*) The noise as manifested by the distribution of the data cloud perpendicular to the trajectory, characterized, for a normal distribution, by an SD σ. The noise characterized by σ determines the confidence with which two conformations can be regarded as distinct. The number of distinct conformational states is the extent of the data cloud along the trajectory, measured in units of $p\sigma$, ($1 \leq p \leq 6$), depending on the desired level of confidence.

Consider now the noise-reduction capability of the SVD step. Henry and Hofrichter (16) discussed the effect of noise on SVD analysis of a matrix consisting of n vectors, each with m components. They show that with input data characterized by noise of variance σ_{input}^2, SVD produces an output, whose variance due to noise is $\left(\frac{1}{m} + \frac{1}{n}\right)\sigma_{input}^2$. When, e.g., $m \gg n$, this yields $\sigma_{output} = \sigma_{input}/\sqrt{n}$, a conceptually appealing scaling behavior.

Consider next a single-mode reconstruction, ignoring the singular value. As before, let the range spanned by the resulting conformations be Δ. Different conformational classes must be separated by $p\sigma_{output}$. The number of permissible conformational classes is therefore

$$N_{class} = \frac{\Delta}{p\sigma_{output}} = \frac{\Delta}{p\sigma_{input}\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}} \approx \frac{\Delta\sqrt{n}}{p\sigma_{input}}, \quad m \gg n, \quad \textbf{[S8]}$$

with p depending on the required confidence level. The parameter m is replaced by cm for c-fold concatenated snapshots. In the present work, $\Delta \approx 5$, $\sigma_{input} \approx 1$, $p = 3$, $n \approx 10^3$, $cm \gg n$, yielding $N_{class} \sim 50$.

The above analysis ignores noise reduction due to projection onto the eigenvectors of the Laplace–Beltrami operator. Taking this into account suggests that up to ~300 conformational classes can be distinguished, subject, of course, to the other limiting factors mentioned in the main text. In this paper, we use the more conservative estimate of 50 conformational classes based on the noise-reduction capability of SVD only.

Pseudocode.

End-to-end analysis.

   Input.

   Cryo-EM micrographs.

   Outputs.

   Spectrum of conformational changes along a selected trajectory.
   2D and 3D movies of conformation spectrum.

   Steps.

    i) Particle picking and orientation recovery (using SPIDER).
    ii) Locate n_{GC} great circles (one Shannon angle width) in the space of orientations that are highly populated with experimental snapshots.
    iii) Divide each great circle into n_P projections each spanning about one Shannon angle (the ratio of resolution to object diameter).
    iv) For great circle $i_{GC} = 1$ to n_{GC} do.
    v) For projection $i_P = 1$ to n_P do.
    vi) Identify snapshots that fall into great circle i_{GC} and projection i_P.
    vii) Bring snapshots into in-plane registry.
    viii) Embed with diffusion map with defocus-tolerant kernel to obtain eigenvectors ψ of Laplace–Beltrami operator.
    ix) Order snapshots along a line making an angle θ with the horizontal axis in ψ_1-ψ_2 space.
    x) Perform NLSA (see pseudocode below).
    xi) Extract the parameter $\tau(\theta)$ in 1° intervals for $0 \leq \theta < \pi$.
    xii) Reconstruct 2D energy landscape.
    xiii) Perform NLSA on a selected trajectory on the landscape.
    xiv) Embed NLSA reconstructed snapshots with diffusion map.
    xv) Extract conformation parameter τ (conformation spectrum).
    xvi) Establish consistent sense and starting points for all conformation spectra.
    xvii) Compile 2D movie along selected trajectory on energy landscape.
    xviii) endfor.
    xix) endfor.
    xx) Bin snapshots into N_{class} classes based on conformation parameter τ.
    xxi) Perform 3D reconstruction of each class using Arachnid.
    xxii) Compile 3D movie.

NLSA.

Input.

Set of n_s snapshots (each with d pixels) along the selected direction (ψ_1, ψ_2 or a line in between).

Output.

Set of NLSA noise-reduced reconstructed snapshots.

Steps.

i) Generate supervectors X of c-fold concatenated snapshots. $\mathbf{X}$ is a matrix with dimensions $cd \times (n_s - c + 1)$, $X[(k-1)d + 1: kd, j]$ is the $j + k - 1$th snapshot.

ii) Embed X with diffusion map to obtain truncated set of eigenvectors of the Laplace–Beltrami operator $\Psi = \{\psi_i | i = 1, \cdots, l\}$ and Riemannian measure μ.

iii) Project X onto $\Psi : A = \mathbf{X}\mu\Psi$.

iv) Perform SVD: $A = USV^T$.

v) Project V from manifold back to time domain: $V'^T = V^T \Psi^T$.

vi) Reconstruct with p modes: $\tilde{X} = \sum_1^P X_k$, where $X_k = U_k S_{kk} V'^T_k$ (U_k is the kth column of the matrix U; similarly, for V'_k, S_{kk} is the kth element of the diagonal matrix S).

vii) c reconstructions are available for each time point t: $\tilde{X}[(k-1)d+1 \ldots kd, t-k+1]$ $1 \leq k \leq c$. These reconstruc-

tions are averaged to obtain noise-reduced reconstructed snapshots.

Reconstruct NLSA snapshots along the selected trajectory.

Input.

Selected trajectory on the map of density of points in τ space.

Output.

Set of NLSA reconstructed snapshots along the trajectory.

Steps.

i) On the map of density of points, identify two lines passing through each point on the trajectory, with the first line connecting the point to the center of the density map and the second normal to the first.

ii) Find class of raw snapshots at the intersection of the two lines using Eq. **S7**.

iii) Repeat steps *i* and *ii* to identify all raw snapshots contributing to selected trajectory.

iv) Perform NLSA on the raw snapshots to obtain a 1D compact manifold (circle).

v) Extract universal conformational parameter τ for each NLSA reconstructed snapshot.

1. Leshin JA, Rakauskaitė R, Dinman JD, Meskauskas A (2010) Enhanced purity, activity and structural integrity of yeast ribosomes purified using a general chromatographic method. *RNA Biol* 7(3):354–360.
2. Grassucci RA, Taylor DJ, Frank J (2007) Preparation of macromolecular complexes for cryo-electron microscopy. *Nat Protoc* 2(12):3239–3246.
3. Lei J, Frank J (2005) Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J Struct Biol* 150(1):69–80.
4. Langlois R, et al. (2014) Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. *J Struct Biol* 186(1):1–7.
5. Spahn CM, et al. (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J* 23(5):1008–1019.
6. Shaikh TR, et al. (2008) SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* 3(12):1941–1974.
7. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319–2323.
8. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326.
9. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396.
10. Donoho DL, Grimes C (2003) Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci USA* 100(10):5591–5596.
11. Coifman RR, et al. (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc Natl Acad Sci USA* 102(21):7426–7431.
12. Coifman RR, Lafon S (2006) Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Appl Comput Harmon Anal* 21(1):31–52.
13. Coifman RR, Shkolnisky Y, Sigworth FJ, Singer A (2008) Graph Laplacian tomography from unknown random projections. *IEEE Trans Image Process* 17(10):1891–1899.
14. Champeney DC (1973) *Fourier Transforms and Their Physical Applications* (Academic Press, London), p 256.
15. Aubry N, Guyonnet R, Lima R (1991) Spatiotemporal analysis of complex signals: Theory and applications. *J Stat Phys* 64(3-4):683–739.
16. Henry E, Hofrichter J (1992) Singular value decomposition: Application to analysis of experimental data. *Methods Enzymol* 210:129–192.
17. Giannakis D, Majda AJ (2012) Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc Natl Acad Sci USA* 109(7): 2222–2227.
18. Giannakis D, Majda AJ (2012) Nonlinear Laplacian spectral analysis: Capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. *Stat Anal Data Mining* 6(3):1–15.
19. Packard N, Crutchfield J, Farmer J, Shaw R (1980) Geometry from a time series. *Phys Rev Lett* 45(9):712–716.
20. Takens F (1981) *Dynamical Systems & Turbulence, Warwick 1980* (Springer, Berlin).
21. Sauer T, Yorke JA, Casdagli M (1991) Embedology. *J Stat Phys* 65(3):579–616.
22. Deyle ER, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* 6(3):e18295.
23. Coifman R, Lafon S (2006) Diffusion maps. *Appl Comput Harmon Anal* 21(1):5–30.
24. Jones PW, Maggioni M, Schul R (2008) Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels. *Proc Natl Acad Sci USA* 105(6):1803–1808.
25. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
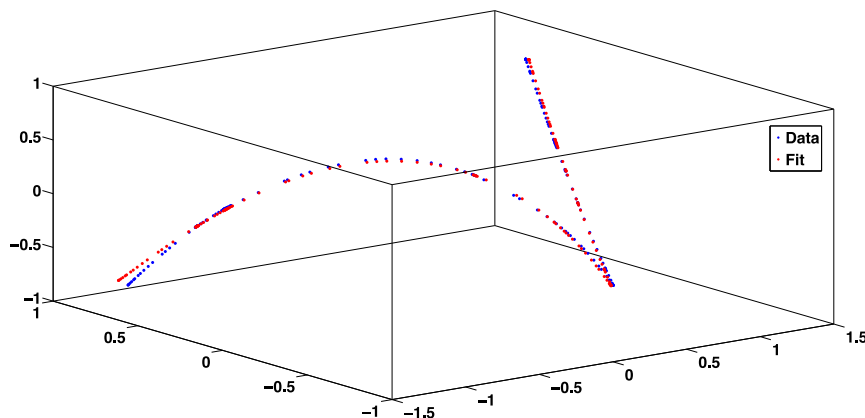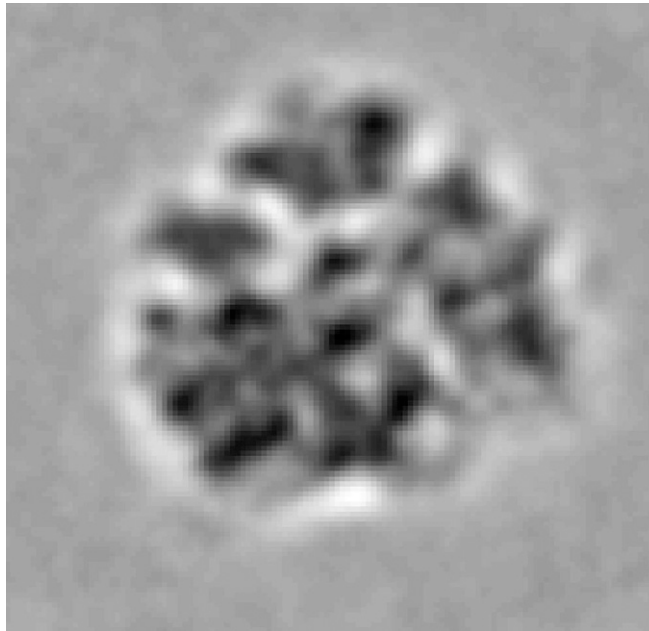
Fig. S1. The manifold produced by snapshots obtained by NLSA along a trajectory consisting of a straight line at an angle θ to the horizontal axis in the (ψ_1, ψ_2) plane.

Movie S1. Changes in reconstructed cryo-EM snapshots from a selected viewing direction, as the path of lowest energy is traversed. The position along the path is indicated by the purple dot in the right panel.
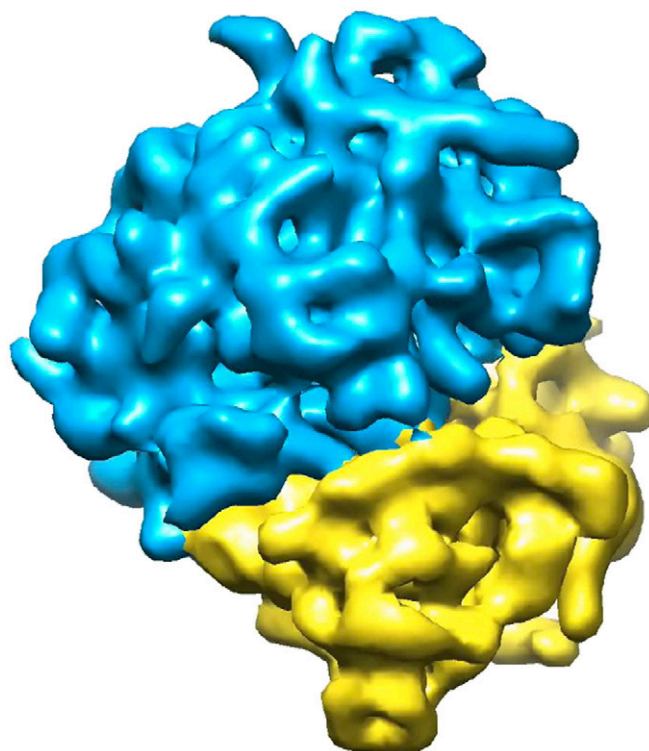
Movie S1
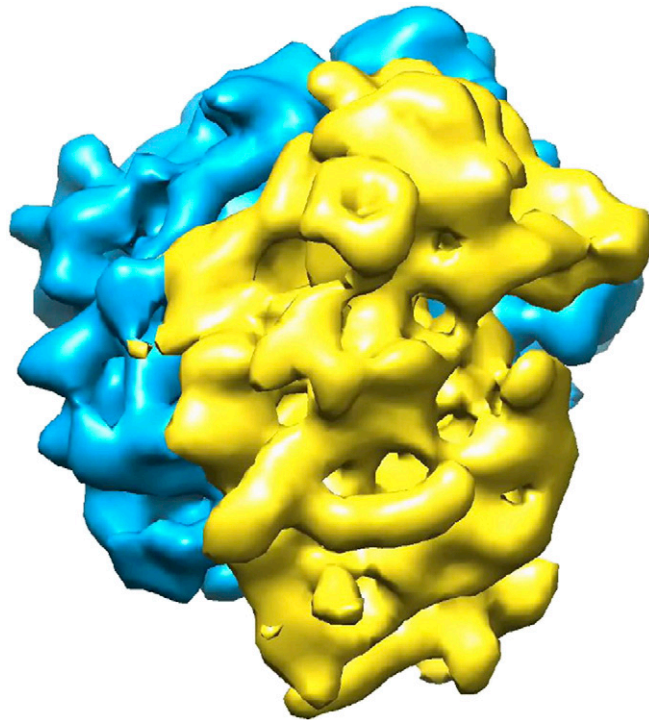
Movie S2. Variations in the ribosome structure along the path of lowest energy, as indicated by the position of the purple dot in the right panel. The sequence of 3D structures is identical in Movies S2 through S4, but is shown in three different standard viewing directions. Viewing direction here: from the factor-binding side.

Movie S2
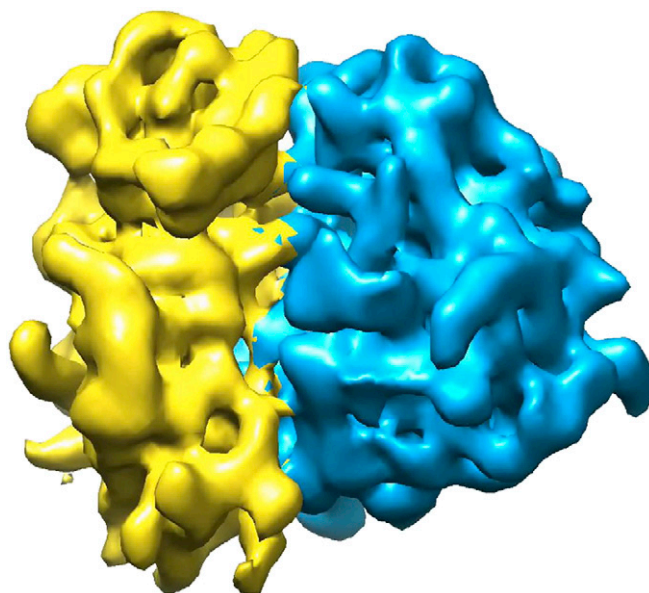
Movie S3. Variations in the ribosome structure along the path of lowest energy, as indicated by the position of the purple dot in the right panel. The sequence of 3D structures is identical in Movies S2 through S4, but is shown in three different standard viewing directions. Viewing direction here: from the solvent side.

Movie S3

Movie S4. Variations in the ribosome structure along the path of lowest energy, as indicated by the position of the purple dot in the right panel. The sequence of 3D structures is identical in Movies S2 through S4, but is shown in three different standard viewing directions. Viewing direction here: from the top.

Movie S4

Movie S5. The ribosome structure, in three standard views, as it varies between the seven successive points along the path of lowest energy indicated by the position of the red circle in the right panel.

Movie S5