HHS Public Access

Author manuscript

Methods. Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

Methods. 2016 May 01; 100: 61–67. doi:10.1016/j.ymeth.2016.02.007.

Continuous Changes in Structure Mapped by Manifold Embedding of Single-Particle Data in Cryo-EM

Joachim Frank[*],

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

Department of Biological Sciences, Columbia University, New York, NY 10027

Abbas Ourmazd[*]

Department of Physics, University of Wisconsin Milwaukee, 3135 N. Maryland Ave, Milwaukee, WI 53211

Abstract

Cryo-electron microscopy, when combined with single-particle reconstruction, is a powerful method for studying macromolecular structure. Recent developments in detector technology have pushed the resolution into a range comparable to that of X-ray crystallography. However, cryo-EM is able to separate and thus recover the structure of each of several discrete structures present in the sample. For the more general case involving continuous structural changes, a novel technique employing manifold embedding has been recently demonstrated. Potentially, the entire work-cycle of a molecular machine may be observed as it passes through a continuum of states, and its free-energy landscape may be mapped out. This technique will be outlined and discussed in the context of its application to a large single-particle dataset of yeast ribosomes.

Keywords

classification; heterogeneity; machine learning; molecular machines; ribosome

Introduction

Thanks to recent advances in detector technology, single-particle cryo-EM has become a competitor for X-ray crystallography in the high-resolution determination of molecular structure [1]. It has recently achieved resolutions below 3Å, that allow *ab initio* chain tracing and identification of individual Mg^{++} ions, water molecules, and sites of RNA modification [2,3]. Arguably the most intriguing aspect is the capability of this technique, thanks to the development of powerful classification methods, to recover the entire inventory of structures

[*]Corresponding authors, jf2192@cumc.columbia.edu and ourmazd@uwm.edu.

co-existing in the same sample ("Story in the Sample" [4]). This aspect, and its expansion into the recovery of a continuum of states, will be the focus of this review.

From mathematical and computational points of view, classification of heterogeneous data in single-particle reconstruction of biological molecules poses a challenge, which has invited many approaches, and continues to inspire new solutions. The nature of the data is such that: (1) each molecule is represented by one projection image only; (2) the image has typically a very low signal-to-noise-ratio (≤0.1); and, (3) in the dataset formed by such images, changes in the image can stem from changes in the orientation of the molecules on the grid, structural heterogeneity in the molecules themselves, or both (see [5,6]). In the earliest applications of single-particle reconstruction, heterogeneity implicit in point (3) was ignored, often leading to suboptimal, partially blurred reconstructions dominated by the features of the majority population.

The origins of structural heterogeneity are well known: rarely does biochemical purification result in a 100% compositionally pure sample, and even if the sample were chemically pure, the molecules can have multiple degrees of conformational freedom in the thermal bath. The problems posed by the two types of heterogeneity, compositional and conformational, can be exemplified by the ribosome during its work cycle: tRNAs and elongation factors are periodically bound and released (*compositional* heterogeneity) while the ribosome undergoes conformational changes as the two ribosomal subunits and their individual domains alter their relative positions (*conformational* heterogeneity).

Initial approaches to classification were reference-based (e.g. [7]) and, in the language of machine learning, *supervised*. In such approaches, the similarity of the particle images to projection images of two or more 3D references is quantified by the cross-correlation coefficient, and the images are assigned to classes and projection angles based on the 3D reference in the orientation producing the highest cross-correlation coefficient. Although the utility of this method has been demonstrated in many cases, it has obvious flaws, the most serious of which is the need to know and select appropriate references, a choice which is quite subjective. Many different approaches were developed to overcome these limitations (among these [5, 8-15]). Most popular are maximum likelihood-based, "unsupervised" methods of classification [5,14,16]. Here the entire reconstruction problem is explicitly treated as a problem of finding K density maps best describing the observed heterogeneous dataset in the presence of noise, presumed to be Gaussian and additive. One incarnation of this algorithm is now widely used in cryo-EM under the name RELION [16,17]. The capability to retrieve several structures from the same sample even in the absence of specific reference structures has become one of the celebrated feats of single-particle cryo-EM (e.g. [18-23]).

It is not necessary to go into the specifics of maximum likelihood methods to see their limitations. For one, the number of classes, K, needs to be guessed. When this number is *underestimated*, some classes with merged features are formed, while when it is *overestimated*, meaningless small classes are split off at unnecessary computational expense. Also, as an iterative method without guaranteed convergence, it requires intervention based on heuristic criteria, that is, on practical criteria based on experience (e.g. [24,25]). Quite

often, the practical application involves several successive rounds in a stepwise, hierarchical scheme. A good example is provided by our study of the eukaryotic preinitiation complex [20],in which only ~4% of the data had the expected complete composition of factors bound to the 40S subunit and could be used in the final biological analysis (Fig. 1).

Another, more serious limitation of maximum likelihood methods and other methods that implicitly assume discrete classes (e.g. [12] is the fact that they do not address changes in structure that are by their very nature continuous. An example, out of many, is provided by the orientation of the IRES-eIF3 complex bound to the 40S ribosomal subunit [26]. One could argue that by specifying a large enough number of classes, one might be able to capture fine subdivisions in a continuum of states. But this solution immediately faces two problems, one technical, the other more fundamental. The technical limitation is due to the large increase in computational expense, which becomes prohibitive already when the number of classes exceeds 10-15. The more fundamental one is the fact that discrete classes cannot be meaningfully defined when the underlying changes are continuous. If a continuum is sampled by a finite number of observations (i.e., molecule images), then random fluctuations in the sampled density of states will produce arbitrary boundaries of meaningless “classes”. Another example for continuous changes, arguably with more biological significance, is provided by the virtual continuum of states of the ATPase molecule in its 360-degree cycle of rotation during which ATP is synthetized [27].

In the following, we highlight a radically different method, based on manifold embedding, that maps out the entire sampled continuum of states in a coordinate system optimally tailored to the data [28]. Only after that mapping has been accomplished, decisions on the discrete vs. continuous nature of the data and – in the former case -- on the number of discrete classes are made on the basis of rigorous statistical criteria.

It is not possible, within the scope of this article, to describe the method in detail, since it involves somewhat unfamiliar mathematical and computational techniques (outlined in the Supplementary Material of ref. [28]). Instead, we provide a conceptual outline, and describe the results obtained from the application of the method to a large ribosome dataset in greater detail than in the original article. Another review that includes a brief outline of the manifold approach is found in [29]. A detailed treatment spanning both conceptual and technical aspects is planned for publication elsewhere (Ourmazd and Frank, in preparation).

Mapping of heterogeneity by Manifold Embedding

Our approach requires a collection of sightings of the object of interest in different orientational and conformational states. A sufficiently fine coverage makes it possible to find, for any image in the data set, a set of close neighbors – closeness here being defined by some measure of similarity, such as the Euclidian distance. Of course, comparisons of images involving scalar pixel-wise comparison of two or more arrays are possible only when the images are aligned with one another.

To illustrate the importance of alignment, let us consider two copies of a molecule lying in the same orientation, presenting the same view. The molecule's projections will still appear

in an arbitrary position within the frame of the image, hence, in a coordinate system affixed to the molecule, the same pixel appears in two different points of the two recorded images. The objective of alignment is to rotate and translate one image with respect to the other image, such that the two copies of the molecule come into precise registry. Only then, after alignment, the value of any pixel in one image can be meaningfully compared with the value of a pixel having the same index in another image, as it refers to the same pixel in the molecule's projection.

Early in the development of the single-particle method, the concept of aligned molecule images as vectors in the multidimensional space spanned by the pixels intensities was introduced [30,31]. For instance, an image of dimensions 100 × 100 would be represented by a vector in a space of dimension 10,000. This concept was quite useful, as it allowed principal component analysis (PCA) or related methods to be directly applied to a set of aligned images. Following PCA, a given set of images can be described in a much smaller "factor space," the factors being the eigenvectors in the PCA decomposition. As will become clear in the following, the method to be outlined here starts with the same concept of representing aligned images as vectors in "pixel space," but then goes well beyond it.

As a preliminary, it should be noted that conformational changes are typically manifested in the movements of several domains of a molecule, with each domain moving independently. We speak of multiple "degrees of freedom" exercised by the molecule as it performs its function. In addition to this *conformational* variability -- as already noted in the introduction -- we can also have *compositional* variability, created by the binding or release of a small ligand. The ligand itself may assume multiple orientations on the target molecule, acting in essence as an additional flexible domain while it is bound.

Let us first consider the set of molecule images in a single projection direction or – in practice – a very small range of projection directions forming a "cone" around the projection direction (PD) of interest. Because of the existence of multiple degrees of freedom and noise, the projection images of the molecule contained within the PD cone form a cloud of points in a (high-dimensional) "pixel space". In this cloud, the images are arranged in a complex topology according to the pattern of their similarities, with their arrangement reflecting the conformational states occupied by the molecule as seen in the given PD. Here it is necessary to select a sufficiently narrow orientational "cone", such that within it, image variability due to changes in orientation is small compared to image variability due to conformational and compositional heterogeneity [32]. From an information-theoretic point of view, the cone is sufficiently narrow when the angle it subtends is equal to, or smaller than one orientational Shannon angle [33,34]. With this in mind, we can proceed to process the data in two successive steps: in the first step, all projection images are aligned by conventional methods and grouped according to orientation, ignoring any conformational or compositional heterogeneity. In the second step, the individual subsets, one for each PD, are analyzed to characterize their conformational and/or compositional heterogeneity.

To illustrate the validity of above procedure in a practical situation, we can consider looking at a mixture of ribosomes in all stages of the elongation cycle of protein synthesis. Adding the mass of EF-G (~70kD) to the 70S *E. coli* ribosome (~2.5 MD), which happens during

this cycle, is in the order of 3%, and has no noticeable effect on the angle assigned to the projection image. However, there may well be situations where the assumption is invalid, for instance, in the case of a small molecule undergoing very large conformational changes. In this case, one would begin with conformational analysis, followed by orientational assignment of the sorted dataset [35]. To date, the case where the effects of orientational and (continuous) conformational heterogeneity are comparable has not been considered in detail.

For any PD, then, the cloud of points may be considered as a hypersurface or "manifold," a term (*"Manningfaltigkeit"*) Bernard Riemann introduced in 1854 (published posthumously: [36]) as a topological entity with an interesting property: that at any point of the manifold, a local Euclidean geometry can be defined. Dashti and coworkers [28] set themselves first the goal of describing the multidimensional manifold created by single-particle images within the cone representing any projection direction (PD). In a subsequent step, he manifolds in different PDs are related to one another, such that they form a final map describing the continuously varying conformations of the three-dimensional structure. It is important to note that, by the Shannon-Nyquist sampling theorem, a sufficiently dense sampling of the changes in the object can lead to a *continuous* description of the conformational changes.

An efficient description of the manifold requires the construction of a coordinate system tailored to the manifold's specific geometry. This step, called "manifold embedding," reveals the intrinsic dimensionality of the manifold, and uses a diagonalization procedure to find a set of Euclidean coordinates ψ_i, ideally suited to describing the manifold in question. The intrinsic dimensionality of the manifold corresponds to the number of degrees of freedom exercised by the system during observation. The number of relevant Euclidean coordinates to use can be deduced from the spectrum of eigenvalues. Given a sufficiently high signal-to-noise ratio, when the examined particles consist of different discrete "species", each produces a separate manifold, or cluster. When the sample consists of objects undergoing continuous conformational changes, the embedding procedure provides a description of the manifold as a continuous hypersurface, sampled at points corresponding to the experimental observations (snapshots) of the objects. Thus, in the absence of noise, and given a sufficiently large number of observations for the analysis to converge, the manifold embedding procedure provides a *continuous description* of continuous conformational changes, and no "binning" is needed. In the presence of noise, of course, the number of meaningfully distinct conformational states is determined by the ability to distinguish them from each other with sufficient confidence. Thus the number of meaningful conformational states is, in practice, limited by the signal-to-noise ratio. In the case of the ribosome work-cycle considered here, 50 conformational states separated by three standard deviations (99.7% confidence) could be identified.

Given a description of the manifold in terms of its "natural" coordinate system consisting of a small (typically ~5) Euclidean coordinates ψ_i, one performs principle components analysis (or, more precisely "nonlinear singular value decomposition"[37]) along each of the coordinates to determine the conformational changes corresponding to each axis ψ_i.

So far we have outlined what needs to be done to describe the manifold associated with a *single* projection direction. The step that follows -- the most complicated step of the whole procedure -- is the consolidation of the information gleaned from all individual PD's into one final map. This entails identifying “isoconformational contours” connecting the points representing the same conformation in different projection directions. If the rates of conformational change over the conformational manifold were the same in all projection directions, this would be a trivial matter. However, the rates of change are, in general, unknown over higher-dimensional manifolds. This difficulty is overcome by an elaborate technique, in which the rates of conformational change over the conformational manifolds in different PD's are related to one another. (For details, see [28])

The actual conformational changes displayed by the particles can, of course, be highly anisotropic in 3D space. In each PD, the component of conformational changes along the projection direction is invisible in the snapshots. However, the complete conformational spectrum can be recovered by “correcting” all conformational spectra by histogram equalization, whereby the conformational spectrum measured in each PD is normalized to the union of all spectra. This approach is valid to the extent that the conformation of a molecule does not depend on the molecule's orientation on the supporting substrate, an assumption that requires confirmation on a case by case basis. Comparison of angular distributions of particles falling in distinct classes with different conformations obtained by RELION is a possible way to ensure its validity.

Recall that the intrinsic dimensionality of the conformational manifold reflects the number of degrees of freedom exercised by the system during observation. Thus, a complete description of the conformational changes may require, say, five dimensions, with the appropriate number determined by embedding analysis. In practice, one selects a subset, say two, of the most important dimensions (as determined by their eigenvalues). In this case, the conformational changes along the dimensions not considered are projected onto the selected plane.

Results obtained for the ribosome

Ribosomes purified from a cell extract may be in any of the states of initiation, elongation, termination, and recycling. Since addition of each amino acid to a polypeptide – often running into hundreds of amino acids -- requires one round of the elongation cycle, the states along this cycle are most highly populated among the single-molecule projections, compared to states related to initiation, termination, and recycling combined.

The chart in Fig. 2 depicts the elongation cycle, valid in its outline for all kingdoms of life. Going clockwise in the diagram, changes in the conformation of the actively translating ribosome from one state to the next have been documented in several studies of *in vitro* systems by cryo-EM, and some are known from X-ray crystallography of nontranslating ribosomes (for detailed references see below). These conformational changes are largely combinations of domain movements: intersubunit rotation, small subunit head closure and swivel, L1 movement on the large subunit. Typically, in such an experiment of visualization, the ribosome is “frozen” in a particular state by a chemical intervention – by an antibiotic

or by the use of a nonhydrolyzable GTP analog. As the cryo-EM reconstructions in Fig. 2 show, one is able to see the elongation factors engaged with the ribosome at such points of intervention.

In contrast to such *in vitro* experiments, occupancy of the ribosome with its ligands (tRNA and mRNA, and EF-G, EF-Tu in case of eubacteria and archaea; or eEF2, eEF1α in case of eukaryotes) in a cell extract depends on the method of extraction and purification used. Milder treatments may retain some of the ligands, while more aggressive treatments leave the ribosome largely empty, or "naked."

It is with such an ensemble of *mostly empty* ribosomes from yeast that we tried the manifold mapping method. What would be expected in such a case? The ribosomes are in an undefined, nonfunctional state, and any variations in structure would be due to the innate flexibility of the multi-domain molecule and its immersion in the thermal bath at room temperature – the temperature just before the fast step of plunge-freezing.

As an example, Fig. 3 shows the results obtained for a typical projection direction. For data belonging to this or other projection directions, the manifold proved to have five significant dimensions, indicating that the empty ribosome has this number of degrees of freedom. The projection direction chosen in this figure happens to depict the ribosome in the front view (or, with reference to the small subunit, solvent view), which shows the small subunit on top of the large subunit (see also Fig. 4a, middle). It is readily apparent that the largest variation among the images, expressed by ψ_1, relates to the change in the orientation of the small relative to the large subunit, i.e., the intersubunit rotation first observed by comparing cryo-EM maps of the *E. coli* ribosome with and without EF-G bound to it [38,39].

The fact that the empty ribosome exists in multiple conformations, all related by rotation of the small subunit, is not surprising, given the background of experiments that implicate Brownian motion in mRNA-tRNA translocation [18, 40-42] . However, by looking at the other domain movements picked up by the analysis (Fig. 4), we gain an appreciation of the larger role Brownian motion plays in the functional dynamics of the translating ribosome. Figure 4 shows the final, consolidated map (Fig. 4b), along with a reconstruction of the 80S ribosome from yeast, on which the directions of observed movements are indicated (Fig. 4a). First of all, the two-dimensional colored map (Fig. 4b) is a free-energy landscape, which was computed from the map of occupancies by the Stefan Boltzmann relationship (for first use in cryo-EM context, see [18,43]). Where the map of occupancies shows peaks or a continuous mountain range, the energy landscape shows minima or a continuous valley. Going clockwise from one minimum to the next, the maps reconstructed from data in their vicinities show combinations of domain movements indicated by a combination of symbols. For instance, going from the minimum marked "14" to "21", the combination +R, +HC, -HS indicates that positive intersubunit movement, positive head closure, and negative head swivel are observed. This is a combination of movements encountered in the elongation cycle (Fig. 2) after tRNA selection and transfer of the peptide bond, as observed. Similar equivalence in the combination of movements can be garnered by comparing the two diagrams.

It is interesting to relate our findings to studies directed at ribosome *dynamics* (see, e.g., Noel et al., and references therein). Our approach uses the Boltzmann relation to derive the free-energy landscape from an ensemble of ribosomes in *equilibrium*. Specifically, the number of sightings of a given conformation compared to another yields the energy difference between the two conformations through the relation $\frac{n_1}{n_2} = \exp\left(-\frac{\Delta G}{k_B T}\right)$, with $n_{1,2}$ signifying the number of sightings of two different conformations, and ΔG the free-energy difference between them. In order to relate the free energy to dynamical factors such as transition rates, knowledge of the diffusion coefficient is needed. Such information cannot be obtained from equilibrium measurements.

As to the choice of the reaction coordinates used to describe the energy landscape, these flow directly and exclusively from the observed conformational changes, without templates or other a priori information. We cannot, from equilibrium measurements, identify the paths over the landscape leading to specific reactions, other than by compiling movies of the conformational changes. As mentioned above, the movies compiled along the triangular path of minimum energy are strongly reminiscent of the translation work cycle. This suggests that the conformational states observed correspond to those of the elongation cycle, and that they might be ordered along that path as in a functioning translating ribosome.

On the face of it, this association seems implausible, since, as noted before, the ribosomes in the specimen, as products of a purification from a cell extract, were not functionally active. However, by making reconstructions from subsets along the path, we could confirm that all those conformational changes do take place that are expected when going through successive steps of aminoacyl-tRNA selection and mRNA-tRNA translocation (conformational changes accompanying peptidyl-transfer, the step in between, would not be observable at the resolution of the study). The knowledge on conformational changes associated with the elongation cycle for both bacterial and eukaryotic translation comes from numerous 3D visualizations by cryo-EM and X-ray crystallography (see [44,45] and from single-molecule FRET experiments (see [46]) Specifically, cognate aa-tRNA selection is known to go hand in hand with a closing of the head position (“nodding”) [47] while translocation involves a combination of intersubunit rotation [38,39], L1 stalk rotation [39,48], and small-subunit head rotation [49].

A compelling interpretation of these findings, which has to be confirmed with large datasets from ribosomes engaged in translation, is that all these functionally necessary movements of domains are facilitated by Brownian motion (in conjunction with pawl-type interventions by factors, requiring the expenditure of energy). Thus, the Brownian motion paradigm, previously only experimentally confirmed for intersubunit rotation [28,40-42], may have more general validity, encompassing domain-wise motions, as well. Future investigations will tell to what extent some of these motions are coupled or geared, as observed for the motion of the L1 stalk in conjunction with the intersubunit rotation of the *E. coli* ribosome [28,39]

Conclusions: Implications for future studies of biological macromolecules

The demonstration of the continuous conformational mapping for ribosomes suggests the general utility of this technique in characterizing multiple states of molecular machines freely equilibrating in a sample. There are many examples of large molecular machines where current knowledge of mechanism is sketchy, because the examination of states has relied entirely on chemical trapping, for example ATP synthase, RNA polymerase, various chaperones, and the proteasome. The results with the ribosome suggest that a continuum of states can be captured without chemical intervention, provided that the number of conformational samples (i.e., projections collected in the experiment) is sufficiently large. User-friendly programs and a graphical user interface are currently being developed by the authors and their collaborators as a tool set, intended for free distribution.

Acknowledgments

We thank Ali Dashti and Peter Schwander for discussions. Assistance by Melissa Thomas-Baum (Buckyball Design) with the preparation of some of the figures is gratefully acknowledged. This work was supported by HHMI and NIH R01 GM 55440 (to J.F.), the US Dept. of Energy, Office of Science, Basic Energy Sciences under award DE-FG02-09ER16114, and the US National Science Foundation under awards STC 1231306 and 1551489 (to A.O.).

References

1. Nogales E, Scheres SHW. Cryo-EM: A unique tool for the visualization of macromolecular complexity. Mol. Cell. 2015; 58: 677–689. [PubMed: 26000851]
2. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H. Structure of the E. coli ribosome–EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. Nature. 2015; 520: 567–570. [PubMed: 25707802]
3. Liu Z, Wei J, Gutierrez-Vargas C, Sun M, Espina N, Madison-Antenucci S, Tong L, Frank J. Cryo-EM structure of the Trypanosoma cruzi ribosome at 2.5Å. 2016.
4. Frank J. Story in a sample—the potential (and limitations) of cryo-electron microscopy applied to molecular machines. Biopolymers. 2013; 99: 832–836. [PubMed: 23640776]
5. Scheres SHW, Gao H, Valle M, Herman GT, Eggermont PPB, Frank J, Carazo J-M. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods. 2007; 4: 27–29. [PubMed: 17179934]
6. Sigworth FJ. From cryo-EM, multiple protein structures in one shot. Nat. Methods. 2007; 4: 20–21. [PubMed: 17195021]
7. Valle M, Sengupta J, Swami NK, Grassucci RA, Burkhardt N, Nierhaus KH, Agrawal RK, Frank J. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. EMBO J. 2002; 21: 3557–3567. [PubMed: 12093756]
8. Leschziner AE, Nogales E. Visualizing flexibility at molecular resolution: Analysis of heterogeneity in single-particle electron microscopy reconstructions. Ann. Rev. Biophys. Biomol. Struct. 2007; 36: 43–62. [PubMed: 17201674]
9. Elad N, Clare DK, Saibil HR, Orlova EV. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. 2008; 162: 108–120. [PubMed: 18166488]
10. Herman GT, Kalinowski M. Classification of heterogeneous electron microscopic projections into homogeneous subsets. Ultramicroscopy. 2008; 108: 327–338. [PubMed: 17574340]
11. Zhang W, Kimmel M, Spahn CMT, Penczek PA. Heterogeneity of large macromolecular complexes revealed by 3-D cryo-EM variance analysis. Structure. 2008; 16: 1770–1776. [PubMed: 19081053]

12. Elmlund D, Davis R, Elmlund H. Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states. Structure. 2010; 18: 777–786. [PubMed: 20637414]
13. Shatsky M, Hall RJ, Nogales E, Malik J, Brenner SE. Automated multi-model reconstruction from single-particle electron microscopy data. J. Struct. Biol. 2010; 170: 98–108. [PubMed: 20085819]
14. Lyumkis D, Brilot AF, Theobald DL, Grigorieff N. Likelihood-based classification of cryo-EM images using FREALIGN. J. Struct. Biol. 2013; 183: 377–388. [PubMed: 23872434]
15. Behrmann E, Loerke J, Budkevich TV, Yamamoto K, Schmidt A, Penczek PA, Vos MR, Bürger J, Mielke T, Scheerer P, Spahn CM. Structural snapshots of actively translating human ribosomes. Cell. 2015; 161: 845–857. [PubMed: 25957688]
16. Scheres SHW. A Bayesian view on cryo-EM structure determination. J. Mol. Biol. 2012a; 415: 406–418. [PubMed: 22100448]
17. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 2012b; 180: 519–530. [PubMed: 23000701]
18. Agirrezabala X, Liao HY, Schreiner E, Fu J, Ortiz-Meoz RF, Schulten K, Green R, Frank J. Structural characterization of mRNA-tRNA translocation intermediates. Proc. Natl. Acad. Sci. USA. 2012; 109: 6094–6099. [PubMed: 22467828]
19. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature. 2013; 504: 107–112. [PubMed: 24305160]
20. Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Hellen CUT, Pestova TV, Frank J. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell. 2013; 153: 1108–1119. [PubMed: 23706745]
21. Amunts A, Brown A, Toots J, Scheres SHW, Ramakrishnan V. The structure of the human mitochondrial ribosome. Science. 2015; 348: 95–98. [PubMed: 25838379]
22. Li W, Liu Z, Koripella RK, Langlois R, Sanyal S, Frank J. Activation of GTP hydrolysis in mRNA-tRNA translocation by Elongation Factor G. Science Advances. 2015; 1: e1500169. [PubMed: 26229983]
23. Zalk R, Clarke OB, des Georges A, Grassucci RA, Reiken S, Mancia F, Hendrickson WA, Frank J, Marks AR. Structure of a mammalian ryanodine receptor. Nature. 2015; 517: 44–49. [PubMed: 25470061]
24. Chen B, Shen B, Frank J. Particle migration analysis in iterative classification of cryo-EM single-particle data. J. Struct. Biol. 2014; 188: 267–273. [PubMed: 25449317]
25. Shen, B, Chen, B, Liao, H, Frank, J. Quantitative analysis in iterative classification schemes for cryo-EM applications.. In: Herman, GT, Frank, J, editors. In: Computational Methods for Three-dimensional Microscopy Reconstruction. Birkhauser; Basel: 2014. 67–95.
26. Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Pestova TV, Hellen CUT, Frank J. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. Nature. 2013; 503: 539–543. [PubMed: 24185006]
27. Zhao J, Benlekbir S, Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature. 2015; 521: 241–245. [PubMed: 25971514]
28. Dashti A, Schwander P, Langlois R, Fung R, Li W, Hosseinizadeh A, Liao HY, Pallesen J, Sharma G, Stupina VA, Simon AE, Dinman JD, Frank J, Ourmazd A. Trajectories of the ribosome as a brownian nanomachine. Proc. Natl. Acad. Sci. USA. 2014; 111: 17492–17497. [PubMed: 25422471]
29. Chen B, Frank J. Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. Microscopy. 2015; 2015: 1–11.
30. van Heel M, Frank J. Use of multivariates statistics in analysing the images of biological macromolecules. Ultramicroscopy. 1981; 6: 187–194. [PubMed: 7268930]
31. Frank J, van Heel M. Correspondence analysis of aligned images of biological particles. J. Mol. Biol. 1982; 161: 134–137. [PubMed: 7154073]
32. Fu J, Gao H, Frank J. Unsupervised classification of single particles by cluster tracking in multi-dimensional space. J. Struct. Biol. 2007; 157: 226–239. [PubMed: 16931050]
33. Fung R, Shneerson, Saldin DK, Ourmazd A. Structure from fleeting illumination of faint spinning objects in flight. Nature Physics. 2009; 5: 64–67.

34. Moths B, Ourmazd A. Bayesian algorithms for recovering structure from single-particle diffraction snapshots of unknown orientation: a comparison. Acta Cryst. 2011; A76: 481–486.
35. Schwander P, Fung R, Ourmazd A. Conformations of macromolecules and their complexes from heterogeneous datasets. Phil. Trans. Roy. Soc. 2014; B369: 20130567.
36. Riemann B. Űber die Hypothesen, welche der Geometrie zu Grunde liegen. Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen. 1867; 13
37. Giannakis D, Majda A. Nonlinear Laplacian spectral analysis. Proc. Natl. Acad. Sci. USA. 2012; 109: 2222–2227. [PubMed: 22308430]
38. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature. 2000; 406: 318–322. [PubMed: 10917535]
39. Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. Locking and unlocking of ribosomal motions. Cell. 2003; 114: 123–134. [PubMed: 12859903]
40. Cornish PV, Ermolenko DN, Noller HF, Ha T. Spontaneous intersubunit rotation in single ribosomes. Mol. Cell. 2008; 30: 578–588. [PubMed: 18538656]
41. Agirrezabala X, Lei J, Brunelle JL, Ortiz-Meoz RF, Green R, Frank J. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. Mol. Cell. 2008; 32: 190–197. [PubMed: 18951087]
42. Julian P, Konevega AL, Scheres SH, Lazaro M, Gil D, Wintermeyer W, Rodnina MV, Valle M. Structure of ratcheted ribosomes with tRNAs in hybrid states. Proc. Natl. Acad. Sci. USA. 2008; 105: 16924–16927. [PubMed: 18971332]
43. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature. 2010; 466: 329–333. [PubMed: 20631791]
44. Korostelev A, Ermolenko DN, Noller HF. Structural dynamics of the ribosome. Curr. Opin. Chem. Biol. 2008; 12: 674–683. [PubMed: 18848900]
45. Dunkle JA, Cate JHD. Ribosome structure and dynamics during translocation and termination. Ann. Rev. Biophys. 2010; 39: 227–244. [PubMed: 20192776]
46. Frank J, Gonzalez RL. Structure and dynamics of a processive Brownian motor: the translating ribosome. Ann Rev Biochem. 2010; 79: 381–412. [PubMed: 20235828]
47. Ramakrishnan V. Ribosome structure and the mechanism of translation. Cell. 2002; 108: 557–572. [PubMed: 11909526]
48. Fei J, Kosuri P, MacDougall DD, Gonzalez RL Jr. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. Mol. Cell. 2008; 30: 348–359. [PubMed: 18471980]
49. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, Dönhöfer A, Connell SR, Fucini P, Mielke T, Whitford PC, Onuchic JN, Yanan Yu Y, Sanbonmatsu KY, Hartmann RK, Penczek PA, Wilson DN, Spahn CMT. Head swivel on the ribosome facilitates translocation via intra-subunit tRNA hybrid sites. Nature. 2010; 468: 713–716. [PubMed: 21124459]
50. Frank, J. The dynamics of the ribosome as inferred by cryo-EM: Induced and self-organized motions.. In: Cheng, RH, Hammar, L, editors. World Scientific; Singapore: 2004. Conformational Proteomics of Macromolecular Architecture. 291–306.

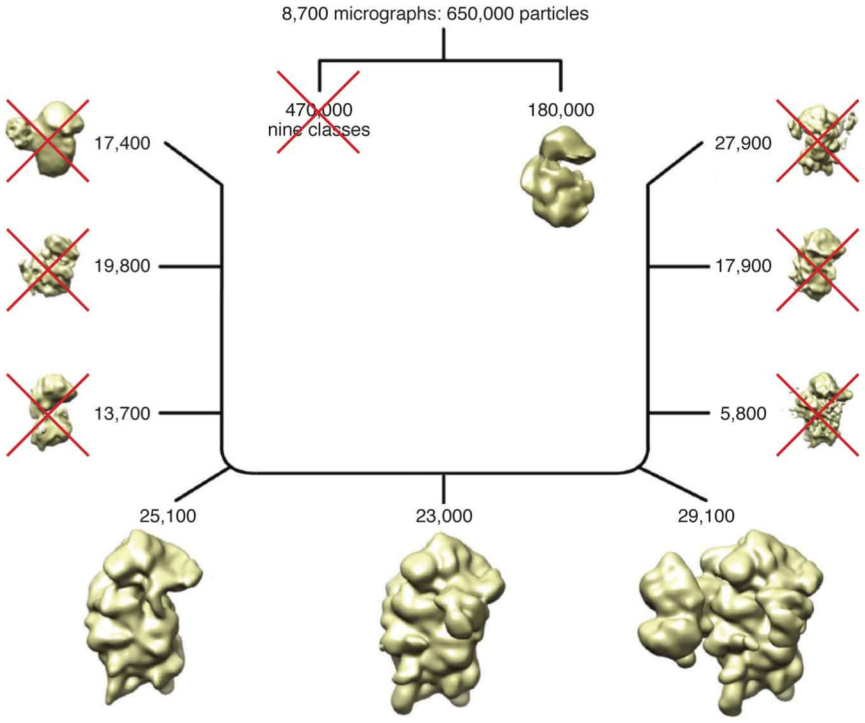

Fig. 1.

Example for the classification of molecule images (the 43S translation pre-initiation complex) using Relion. Initially 650,000 particles were classified into 10 classes. Of these, only one class of 180,000 was found to contain intact particles recognizable as 40S subunits. Another step of classification, this time into 9 classes, yielded three high-quality maps (enlarged, at the bottom), of which only one (on the right) displays all expected components of the eukaryotic pre-initiation complex. (Adapted from[20]).

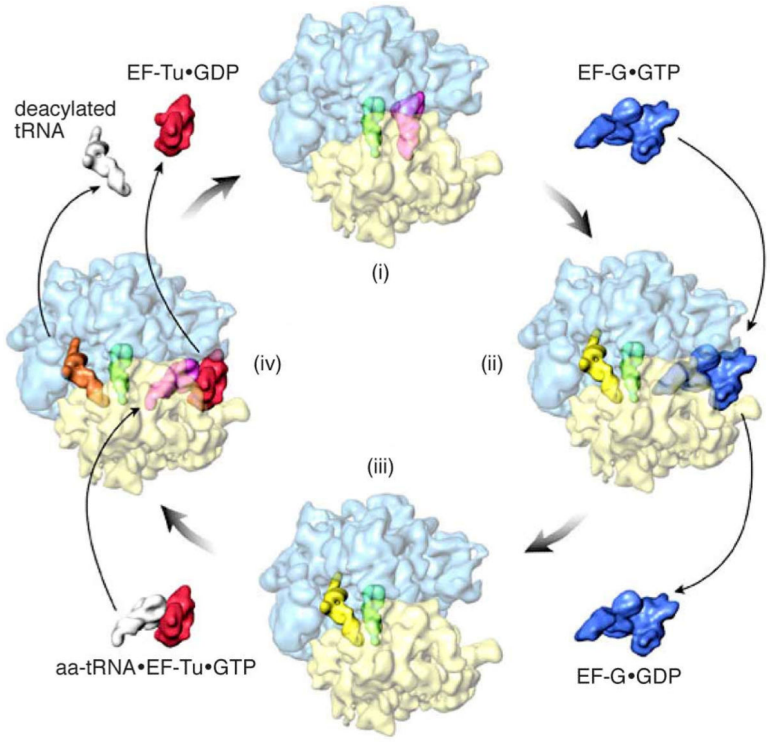

Fig. 2.

Schematic diagram of the translation elongation cycle in bacteria. In each clockwise round, a new amino acid is added to the nascent polypeptide as specified by the genetic code on the mRNA. In principle, in a sample of ribosomes purified from cell extract, all states of the ribosome during the elongation cycle should be present, but the purification often eliminates the bound ligands: tRNA, mRNA, EF-G, and the aa-tRNA•EF-Tu•GTPbn ternary complex. (In addition, states associated with initiation, termination, and recycling should also be present in such a sample, but they occur with less frequency). (Reproduced with permission from [50]).

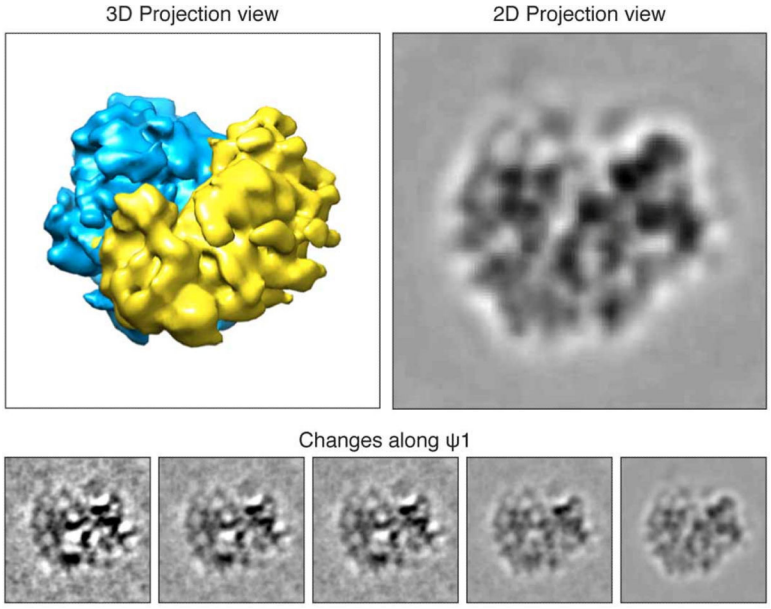

Fig. 3.

Analysis of data falling into a single projection direction by manifold embedding. This particular projection direction shows the side view of the 80S ribosome (surface representation in upper left). The changes picked out by the first eigenvector, ψ_1, can be recognized as the result of intersubunit rotation (see lower 5 panels which, from right to left, show an increasing redistribution of mass consistent with the increasing rotation of the small subunit with respect to the large subunit). In other words, through manifold embedding, the projection data were sorted according to changes in angle between the two subunits. The other eigenvectors, ψ_2 through ψ_5, depict more subtle changes in conformation.

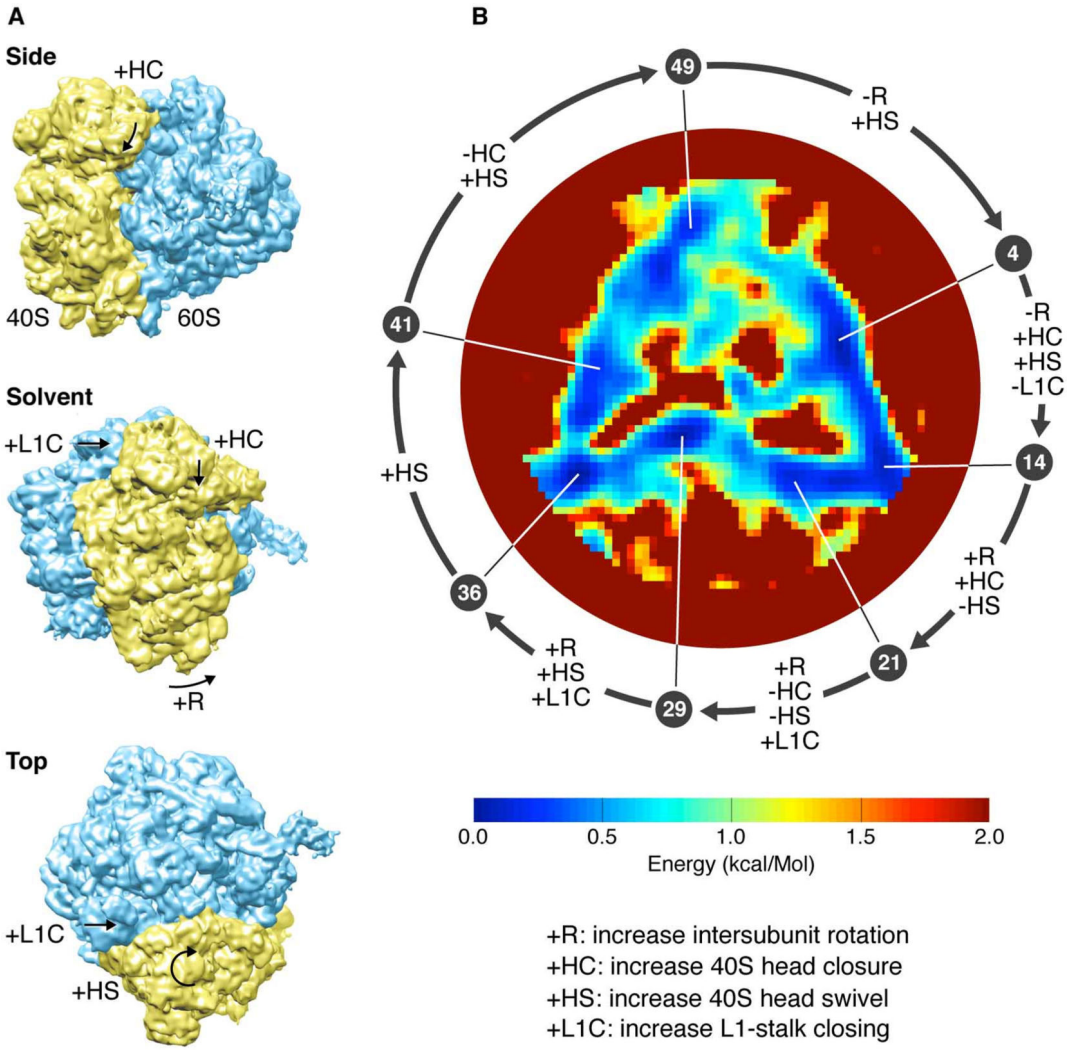

Fig. 4.

Conformational variability and energy landscape of the ribosome. (a) Three views of a cryo-EM map of the 80S ribosome from yeast, with arrows and symbols indicating four prominent conformational changes associated with the elongation work cycle of the ribosome (see key on the bottom right). The three views are orthogonal to one another. “Solvent” refers to the solvent view with respect to the small subunit, also called “front view” of the 80S ribosome. (b) The energy landscape constructed by the manifold embedding technique of Dashti et al. (2014), showing the preferred path followed by the ribosome. Horizontal and vertical axes are the first two eigenvectors ψ_1 and ψ_2, respectively. The color bar shows the energy scale. The energy range has been truncated at 2 kcal/mol to show details of the roughly triangular, closed minimum free-energy trajectory. The error in energy determination along the trajectory is 0.05 kcal/mol. The trajectory has been divided into 50 states. The pointers indicate 7 selected minima, each identified by its position along the sequence of the 50 states. Arrows along circle between successive

minima indicate combinations of observed conformational changes explained on the left. (Reproduced from [28] with permission)