Please cite this article in press as: Jin et al., Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes, Structure (2014), http://dx.doi.org/10.1016/j.str.2014.01.004

Structure
Resource

Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes

Qiyu Jin,[1,6] Carlos Oscar S. Sorzano,[2] José Miguel de la Rosa-Trevín,[2] José Román Bilbao-Castro,[3] Rafael Núñez-Ramírez,[4,7] Oscar Llorca,[4] Florence Tama,[5] and Slavica Jonić[1,*]
[1]IMPMC, Sorbonne Universités–CNRS UMR 7590, UPMC Paris 6, MNHN, IRD UMR 206, 4 Place Jussieu, 75005 Paris, France
[2]Biocomputing Unit, Centro Nacional de Biotecnología–CSIC, Madrid 28049, Spain
[3]Departamento de Informática, Universidad de Almería, Almería 04120, Spain
[4]Centro de Investigaciones Biológicas, CSIC, Madrid 28040, Spain
[5]RIKEN, Advanced Institute for Computational Sciences, Kobe, Hyogo 650-0047, Japan
[6]Present address: Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
[7]Present address: Instituto de Estructura de la Materia, CSIC, Madrid 28006, Spain
*Correspondence: slavica.jonic@impmc.upmc.fr
http://dx.doi.org/10.1016/j.str.2014.01.004

SUMMARY

This article presents a method to study large-scale conformational changes by combining electron microscopy (EM) single-particle image analysis and normal mode analysis (NMA). It is referred to as HEMNMA, which stands for hybrid electron microscopy normal mode analysis. NMA of a reference structure (atomic-resolution structure or EM volume) is used to predict possible motions that are then confronted with EM images within an automatic iterative elastic 3D-to-2D alignment procedure to identify actual motions in the imaged samples. HEMNMA can be used to extensively analyze the conformational changes and may be used in combination with classic discrete procedures. The identified conformations allow modeling of deformation pathways compatible with the experimental data. HEMNMA was tested with synthetic and experimental data sets of *E. coli* 70S ribosome, DNA polymerase Pol α and B subunit complex of the eukaryotic primosome, and tomato bushy stunt virus.

INTRODUCTION

Normal mode analysis (NMA) of structures relies on modeling complex dynamics by a linear combination of harmonic oscillations around an equilibrium structural conformation. NMA was extensively used in electron microscopy (EM), for predicting functional motions from non-atomic-resolution EM structures (Chacón et al., 2003; Ming et al., 2002; Tama et al., 2002) and for flexible fitting (Suhre et al., 2006; Tama et al., 2004a, 2004b) of atomic-resolution structures (e.g., from X-ray crystallography) into EM structures (EM density volumes). Normal modes of EM density volumes were shown to provide a good approximation of atomic-resolution normal modes in the low-frequency range that generally contains the modes reflecting experimentally observed large-scale conformational changes (Tama et al., 2002). They are thus useful for predicting conformational dynamics of complexes whose structure at atomic resolution is unavailable, but an intermediate-resolution structure can be obtained by EM (Chacón et al., 2003; Ming et al., 2002; Tama et al., 2002).

Single-particle analysis (SPA) EM techniques have been traditionally used in determining structures of large macromolecular complexes (diameter, 10–30 nm) from samples prepared biochemically to ideally contain all complexes in the same steady conformation. The high homogeneity of conformations is required for a high resolution of the structure reconstructed from images. As the homogeneous population is sometimes difficult to obtain, even with the most improved biochemical procedures, image-processing methods have been developed to separate images into a few classes and determine the ones corresponding to the steady conformations that are being studied (Elad et al., 2008; Grob et al., 2006; Penczek et al., 2006; Scheres et al., 2007; Simonetti et al., 2008). The goal of these standard approaches is thus to identify the most homogeneous classes of images in terms of conformations and compute the average structure from the classes. More accurate classification methods combined with the biochemical sample preparation to reduce the number of possible conformations should result in a higher homogeneity of classes and a higher resolution of computed structures. As the expected number of conformations is small in such cases, the standard methods are generally used with a small number of classes defined by the user initially. The standard methods thus allow the study of only a few discrete conformations of the complex, and new developments are needed for studying a large range of conformations with many intermediate conformations. This problem arises when studying a full dynamics of the complex that can freely change the conformation or when the biochemical stabilization of the complex in a few steady conformations is impossible. In such cases, the standard methods provide the dynamics information biased by the mentioned initial discretization into a small number of classes (the reconstructed structures are not necessarily discrete samples of the conformational change, but they may result from mixing different conformations into the selected number of classes).

Please cite this article in press as: Jin et al., Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes, Structure (2014), http://dx.doi.org/10.1016/j.str.2014.01.004

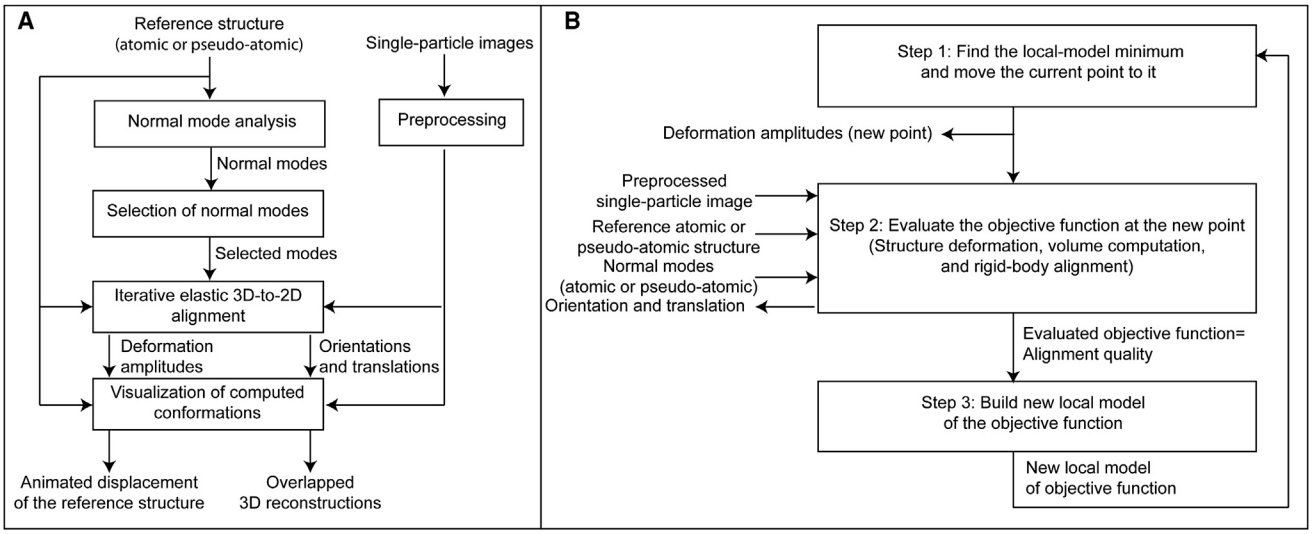

Figure 1. HEMNMA

Flowchart of the method for studying macromolecular dynamics (A) using a 3D-to-2D normal-mode-based alignment of atomic or pseudo-atomic structures with single-particle images (B).

In this article, we propose a method that was specifically designed for studying a large range of conformations with many intermediate conformations. It allows the analysis of highly heterogeneous populations of complexes with slow and continuous conformational changes (gradual conformational changes where a discrete number of states cannot fully describe the continuum of conformations). Given a reference structure (an atomic-resolution structure or a reduced representation of an EM structure), a set of normal modes of the structure, and a set of images, projection-matching-based elastic alignment of each single-particle image with the reference structure provides an overall view of the conformational distribution. A potential of combining NMA with EM image analysis was demonstrated by Brink et al. (2004), who compared several different conformations observed in images with the conformational changes predicted by NMA, through a visual evaluation of their consistency. Here, we describe an automatic method to determine conformations from images using normal modes. The proposed method is based on a procedure for iterative elastic projection matching. This procedure allows the processing of a high number of images to determine all actual confirmations and evaluate their pertinence, which is impossible to achieve with other EM methods or X-ray crystallography. The proposed method has the following four main features: (1) the conformational modeling is performed by displacing the reference structure along combinations of normal modes, and the amplitudes of the displacement are computed through the elastic image alignment; (2) each single-particle image may represent a unique conformation; (3) classification and 3D reconstruction are not mandatory; and (4) classic discrete EM methods may be used to compute the reference structure when it is unavailable at atomic resolution and/or to identify a subset of images to analyze. Thanks to these features, this method allows modeling of deformation pathways compatible with the experimental data and analyzing the conformational changes more extensively than using the classic discrete EM methods. The proposed methodology is referred to as hybrid EM NMA (HEMNMA). It was tested here using several synthetic and experimental data sets: synthetic and EM images of *E. coli* 70S ribosome (70S), EM images of DNA polymerase Pol α and B subunit complex (Pol α-B) of the eukaryotic primosome, and EM images of tomato bushy stunt virus (TBSV).

RESULTS

HEMNMA Design

The workflow (Figure 1A) comprises the following four steps: (1) NMA of a reference structure at atomic or pseudoatomic resolution (see Experimental Procedures for the volume-to-pseudoatoms conversion and NMA methods used here); (2) selection of normal modes; (3) iterative elastic 3D-to-2D alignment of single-particle images with the reference structure; and (4) visualization of computed conformations. The preprocessing block in Figure 1A may include any preprocessing of single-particle images (e.g., correction of the effects of the electron-microscope contrast transfer function [CTF]).

Selection of Normal Modes

The computation of normal modes takes negligible time with respect to the analysis of a long series of images using normal modes (e.g., 10,000 single-particle images). To reduce image analysis time while focusing on the most biologically relevant conformational changes, normal modes can be selected on the basis of some a priori knowledge about the conformational change (e.g., two known motions of 70S, ratchet-like and L1-stalk motions, were used in experiments 1 and 2). If no a priori knowledge is available, lowest frequency modes with the collectivity degree above a threshold can be selected, as highly collective low-frequency modes have been shown to be relevant to functional conformational changes (Delarue and Dumas, 2004; Suhre et al., 2006; Tama and Brooks, 2006; Tama et al.,

2004a; Tama and Sanejouand, 2001; Wang et al., 2004). For instance, the overlap between experimentally observed conformational changes (difference between two atomic conformations of the same complex) and the modes computed for one of the conformations of the complex has showed that 1–3 low-frequency modes with the highest overlap are highly collective and usually capture 60% to 70% of the conformational change (Tama and Sanejouand, 2001; Tama et al., 2002). The most relevant modes are generally among the 20 lowest frequency modes for low-symmetry structures (Tama and Sanejouand, 2001; Tama et al., 2002). They move to higher frequencies (due to degeneracy) for symmetric structures but remain among the 100 lowest frequency modes, even for highly symmetric structures such as icosahedral viruses (Tama and Brooks, 2005). The number of lowest frequency modes to compute can thus generally be fixed to 100. The collectivity degree is computed to count the number of (pseudo-)atoms that are significantly affected by the mode (Bruschweiler, 1995). Given the number of (pseudo-) atoms N, the collectivity degree is normalized between 1/N and 1. The collectivity degree approaches 1 for maximally collective movements, whereas for localized motions (few atoms move), it approaches 0. The most collective modes capturing functional conformational changes can generally be selected by setting the collectivity threshold to 0.5 (experiment 3). Higher values of the threshold should be chosen to select a reduced number of modes for highly symmetric structures (experiment 4). For instance, in the case of "swelling" icosahedral-symmetry viruses, many modes have collectivities above 0.5 (experiment 4). However, a few of them are usually enough to describe functional conformational changes (e.g., the mode describing best the experimentally observed conformational changes is related to a radially symmetric expansion of the capsid) (Tama and Brooks, 2005) and can be selected using the collectivity threshold of 0.75 (experiment 4).

Iterative Elastic 3D-to-2D Alignment Method

The amplitudes of the reference-structure displacement (deformation) along normal modes and the orientation and the position of the deformed structure are refined until a projection of the deformed-reference density volume is found that is the most similar to the analyzed single-particle image. The method is based on a local-search Powell's UOBYQA optimization (Powell, 2002) that uses quadratic approximation for building a local model of the objective function. It consists of repeating the following three steps iteratively until the change in elastic parameters (deformation amplitudes) and rigid-body parameters (orientation and position of the deformed structure) between two successive iterations becomes insignificant or a specified number of iterations is reached (Figure 1B):

Step 1: Find the minimum of the current local model of the objective function and move the current point (current vector of deformation amplitudes) to this minimum.

Step 2: Evaluate the objective function at the new vector of deformation amplitudes (new point). The structure is first deformed with the new deformation amplitudes, and the modified atomic or pseudoatomic coordinates are converted into a volume. The computation of volumes from pseudoatomic coordinates is based on the volume representation by a sum of Gaussian functions (Nogales-Cadenas et al., 2013). The computation of volumes from atomic coordinates uses electronic-form atomic factors (Peng et al., 1996). The volume is subsequently aligned (rigid-body) with the single-particle image using a rough alignment based on a discrete library of reference projections of the volume (Sorzano et al., 2004a) and a continuous refinement using a gradient-based approach (Jonić et al., 2005). The alignment quality determined by a measure of similarity between the single-particle image and the corresponding best matching volume projection is considered as the new evaluation of the objective function for the optimization.

Step 3: Reconstruct the local model of the objective function around the new point with the new evaluation. Go back to Step 1.

The source code of the method is available in the open-source software Xmipp (Sorzano et al., 2004b), starting from the version Xmipp 3.0.1. The current implementation takes, on average, around 2, 4, and 6 min to analyze an image of size 128 × 128 pixels using one, two, and three normal modes, respectively, on a core of a Dual Intel Xeon X5472 processor (3.00GHz). Note that the execution time was measured with one to three modes, but the program allows the use of any number of modes. An MPI-parallelized version of the method allows an efficient analysis of a large number of images (xmipp_mpi_nma_alignment whose tutorial is at http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/Nma_alignment_v3), which was tested on standard-size clusters and supercomputer centers.

Visualization of Computed Conformations

The computed conformations can be visualized by displacing the reference atomic or pseudoatomic structure along normal modes using the amplitudes of the displacement computed through the elastic image alignment (direct visualization). Also, the animated volumes can be obtained by converting displaced atoms or pseudoatoms into volumes, as in Step 2 of the elastic 3D-to-2D alignment method. Optionally, 3D reconstruction from images can be obtained by separating images into classes with similar conformations (similar values of the computed displacement amplitudes). However, for highly heterogeneous samples, the underlying dynamics can only be partially elucidated by analyzing these reconstructed structures due to a discretization of the conformational distribution (discrete number of classes) and averaging during the reconstruction. A direct visualization shall be a preferred choice in such cases, as we show with Pol α-B (experiment 3). For more details on the visualization methods, see the corresponding section of Supplemental Experimental Procedures S1 available online.

Experiment 1: Synthetic Continuous-type Conformational Change

In this experiment, we show robustness of HEMNMA to noise and to the CTF effects remaining after the phase inversion that is usually used for CTF correction. A variety of conformations were sampled from a synthetic continuous-type conformational change determined by deforming an atomic-resolution *E. coli* 70S structure (Figure S1A) referred to as 3I1OP (without ligands, in the conformation in which the 30S subunit is nonrotated with respect to the 50S subunit) (Zhang et al., 2009). Synthetic images [size, 128^2 pixels; pixel size, (3 Å)^2] were computed as 300

Please cite this article in press as: Jin et al., Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes, Structure (2014), http://dx.doi.org/10.1016/j.str.2014.01.004

random projections of the atomic structure displaced with random displacement amplitudes along two of its normal modes while preserving a constant ratio between the two displacement amplitudes (for details on the random image generation with a linear relationship between the contributions of the two modes, see Supplemental Experimental Procedures S2). The used modes were mode 11 as it describes the ratchet-like motion (rotation of 30S with respect to 50S visible in Movie S1) and mode 12 as it describes the L1-stalk motion. The modes responsible for these two motions have already been characterized computationally (Tama et al., 2003; Wang et al., 2004), and they were chosen here as the combined motion was already observed experimentally (Fu et al., 2011; Valle et al., 2003). The overall characteristics of the modes used in this article are similar to those reported previously, although the mode numbers are not the same because of different structures and methods used for calculations.

The amount of noise before and after CTF was adjusted for each tested defocus value (0.5 μm, 1 μm, and 2 μm) so that the images were obtained with the signal-to-noise ratio (SNR) of 0.3 or 0.1. The different image sets (one without noise and six for different defocus and SNR values; Figure S1D) were analyzed using only the mode capturing ratchet-like motion to see whether the proposed method can recover both synthesized motions: ratchet-like and L1-stalk motions. The elastic alignment of images was performed with respect to the given atomic structure and the pseudoatomic structure obtained from a synthetic volume of the atomic structure (Figure S1C). Movie S2 shows the animated pseudoatomic mode 13 (ratchet-like mode) used in image analysis. Details of the pseudoatomic structure and normal modes computations can be found in Supplemental Experimental Procedures S2.

The image alignment with a pseudoatomic structure is more challenging than the alignment with an atomic structure, as pseudoatomic and atomic normal modes have different resolutions and the same movement often appears at a different frequency in the two sets of modes (Tama et al., 2002). Here, a more difficult alignment in the pseudoatomic case was expected also because different structures were used to generate images (atomic structure) and to analyze them (pseudoatomic structure), while the same structure was used to generate and analyze images in the atomic case. However, such a difficulty is expected in practice, as the resolution of experimental images is usually higher than the resolution of the reference structure when using EM structures. The exploration of only one movement instead of two that are present in images is also challenging but close to realistic as, in practice, images are expected to be analyzed using only a few normal modes (sometimes, only one as in experiment 2). Interestingly enough, although the used mode was selected on the basis of previously observed 70S motions instead of using the modes collectivity criterion, this mode is actually the most collective one for each of the used 70S structures in experiments 1 and 2.

The results in the atomic and pseudoatomic cases show that the accuracy of estimation of rigid-body and elastic parameters decreases as SNR decreases and the defocus increases. Although the maximum value (over seven image sets) of the relative deformation-amplitude error doubled in the pseudoatomic case compared to the atomic case (8.9% for pseudoatoms with respect to 4.2% for atoms; Table S1A), the accuracy is globally good, with the absolute angular and shift errors that are similar in the two cases (below 0.6° and 0.15 pixels for pseudoatoms; below 0.5° and 0.15 pixels for atoms; Table S1B). In the most difficult case (SNR = 0.1, defocus = 2 μm), the root-mean-square deviation between the structures with computed and ground-truth deformation amplitudes along the mode describing the ratchet-like motion is about 0.46 Å for pseudoatoms and 0.22 Å for atoms, which is far below the commonly achievable resolution in EM. Table S1C summarizes how the key parameters of the volume-to-pseudoatoms conversion and NMA methods and the resulting alignment error depend on the volume approximation error (for more details on the influence of these parameters, see Supplemental Experimental Procedures S2).

Experiment 2: Mixed EM Images of *E. coli* 70S Ribosome with and without EF-G

In this experiment, HEMNMA was used to analyze cryo-EM images of a mixture of *E. coli* 70S ribosome with and without Elongation Factor G (EF-G) bound. These images were used in previous EM studies where ratchet-like and L1-stalk motions were observed on the reconstructions corresponding to the ribosome with and without EF-G (Elad et al., 2008; Scheres et al., 2007). Here, the images were analyzed with the ratchet-like mode and separated into the classes with and without EF-G using the computed deformation amplitudes along the mode. Indeed, our tests with synthetic data (generated using EM structures of three different 70S conformations) showed that this mode alone can be used to classify images of the ribosome at distinct states of the conformational change combining ratchet-like and L1-stalk motions and to reconstruct the structure at distinct states of the conformational change (for details of these tests, see Supplemental Experimental Procedures S3; Figure S2).

The data set contained 10,000 cryo-EM images [size, 130^2 pixels; pixel size, $(2.82\ Å)^2$] and a result of a supervised image classification into two subsets, with or without EF-G (5,000 images per subset). Ribosomes without EF-G contain three tRNAs in the classical A/A, P/P, and E/E positions and are in the normal (unratcheted) conformation. Ribosomes with EF-G contain a single tRNA in the hybrid P/E position and are ratcheted. The unratcheted-conformation EM structure EMD-5262 (Fu et al., 2011) was used to compute the pseudoatomic structure and its normal modes. Details of the computations are given in Supplemental Experimental Procedures, S3, as the same structure and normal modes were used there with synthetic images capturing ratchet-like and L1-stalk motions from authentic EM structures. Image analysis was conducted using the pseudoatomic structure and its mode 9 that describes ratchet-like motion (Supplemental Experimental Procedures S3; Movie S3).

The histogram of the computed amplitudes along the pseudoatomic mode 9 (green, in Figure 2A) may be interpreted as a mixture of at least two Gaussian-type distributions, as two peaks can be noticed. Figure 2A also shows the histogram portions corresponding to the image separation by supervised classification (with EF-G, red; without EF-G, blue). The bins containing one third of the images with the smallest amplitudes (mean amplitude, −401) are indicated by red asterisks, while the bins

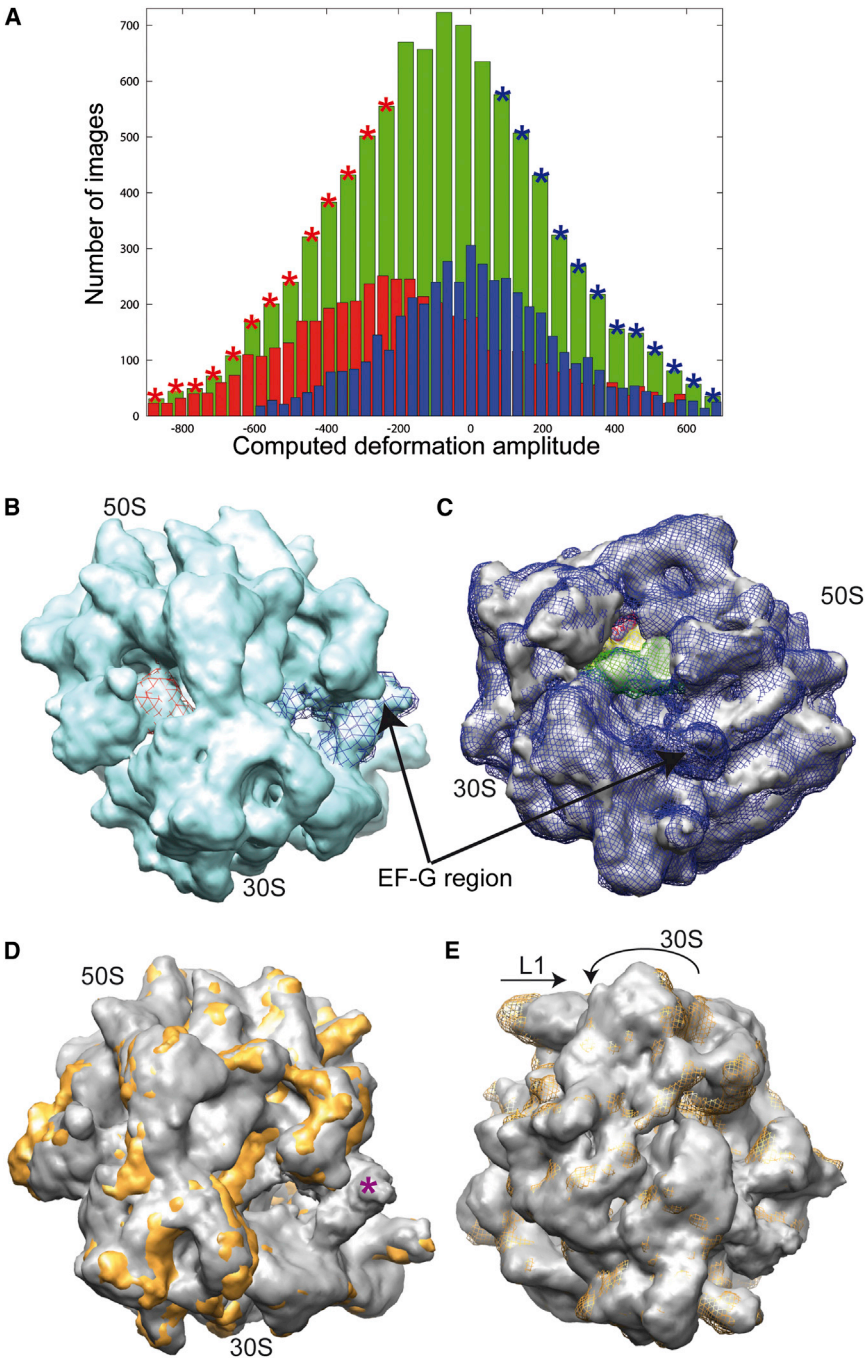

Figure 2. Experiment with EM Images of a Mixture of *E. coli* 70S Ribosome with and without EF-G Factor

(A) Histogram of computed deformation amplitudes (green) with the contribution of images classified by supervised classification as without EF-G (nonrotated conformation) and with EF-G (rotated conformation) indicated in blue and red, respectively. The bins corresponding to the images used for computing the two 3D reconstructions shown in (B)–(E) are indicated by two different colors of asterisks (red and blue).

(B) Structure reconstructed from the images corresponding to the bins indicated by red asterisks, with EF-G and P/E-position tRNA indicated with blue and red mesh, respectively.

(C) Overlapped structure shown in (B) (blue mesh) with the structure reconstructed from the images corresponding to the bins with blue asterisks (gray solid), with A/A-, P/P- and E/E-position tRNAs indicated with mesh in green, yellow, and red, respectively (one large part of the EF-G region is indicated by an arrow on the blue mesh, which helps to show that the region is occupied by no density from the gray solid structure).

(D) Overlapped structures from (C) showing different relative orientations of the 30S and 50S subunits in the two reconstructions, i.e., ratchet-like movement (gray, with EF-G; tan, without EF-G; magenta asterisk, density belonging to EF-G).

(E) Second view of the overlapped structures from (D), with arrows showing the 30S rotation and the L1 movement (gray solid, with EF-G; tan mesh, without EF-G). Although the L1 movement is rotational, it is shown as a shift to designate an inward direction of the movement.

See also Figure S2 and Movie S3.

containing one third of the images with the largest amplitudes (mean amplitude, 237) are indicated by blue asterisks (Figure 2A). The structures reconstructed from images in the red and blue bins clearly show two distinct conformations, with and without EF-G, respectively (Figures 2B and 2C). The structure from the midamplitude range (mean amplitude, −78) seemed to result from a mixture of different conformations with partially bound EF-G and was not further analyzed in this article.

The reconstructed structures are in excellent agreement with those obtained with other methods and the same data set (Elad et al., 2008; Scheres et al., 2007), although the overall quality of the obtained structures cannot be fully compared because of a different number of analyzed images or a different number of images used in reconstruction (e.g., around 90,000 particle images were available in the first paper using this data [Scheres et al., 2007], while only 10,000 images were available here). A particularly interesting result is that the mode related to the movement of the L1 stalk with respect to the 30S subunit was not explicitly explored through the image analysis, but this movement was recovered (Figure 2E) by classifying images based on the ratchet-like motion visible in Figures 2C and 2D.

These results show robustness of the proposed method to the use of a reference structure that comes from a different data set or a different composition of the studied complex (the number of tRNAs in the reference EMD-5262 structure is different from the number of tRNAs in the analyzed images, and the EF-G factor is present in some of the images but not in the EMD-5262 structure).

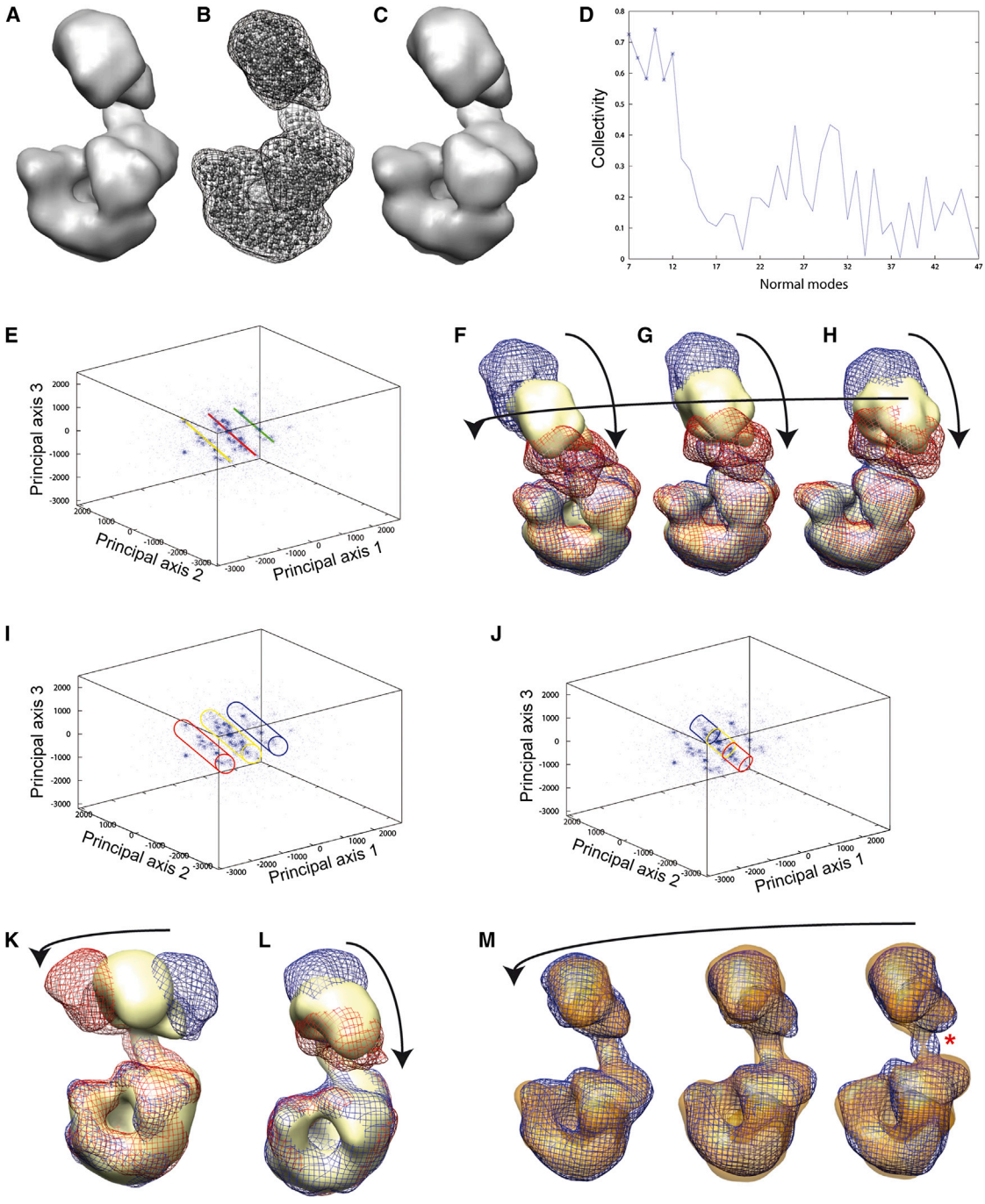

Figure 3. Experiment with EM Images of Pol α-B

(A) EM volume obtained by ML3D (from the largest class of images).

(B) Pseudoatomic structure (spheres) from the volume in (A) overlapped with the volume (mesh).

(C) Approximation of the volume shown in (A) by the pseudoatomic structure shown in (B).

(D) Collectivity degree of normal modes computed for the pseudoatomic structure (the first six modes related to rigid-body movements are not shown; for better visibility, only 40 modes are shown; the asterisks denote the modes with the collectivity degree above 0.5).

(E) Computed deformation amplitudes projected on to the three most important PAs (the points denote single-particle images and PA1, the most important PA), and linear regression lines (yellow, red, and green) through three clusters; see also (I).

(F–H) Volumetric form of the pseudoatomic structure displaced along the lines shown in (E) [(F), (G), and (H) for the yellow, red, and green lines, respectively], with the overlapped structures from two endpoints of the line (blue mesh and red mesh) and the midpoint (yellow solid).

(I) Clusters of images denoted by cylinders.

(J) Three equal-sized subsets of the yellow (central) cluster in (I).

(legend continued on next page)

Experiment 3: EM Images of Pol α-B

In the absence of a structure at atomic resolution, Pol α-B was previously studied using ML3D, an EM method based on 3D maximum likelihood (Scheres et al., 2007). Negative-stain single-particle images of the complex were analyzed, which yielded a structure composed of two lobes connected by a flexible linker where the larger and smaller lobes roughly correspond to the Pol α and B subunits, respectively (Klinge et al., 2009). More precisely, ML3D revealed three different classes of images that contained 42%, 30%, and 28% of the total number of images and yielded to the reconstructions at 22.9 Å, 24.5 Å, and 25.5 Å resolutions, respectively, each corresponding to a different degree of flexion of the linker (Klinge et al., 2009). The same images (12,000 single particles) were analyzed here using HEMNMA. The structure obtained by ML3D from the largest class of images was converted into a pseudoatomic structure that was used as the reference structure to analyze images. The volume and image sizes were 64^3 voxels [voxel size, $(3.8\ Å)^3$] and 64^2 pixels [pixel size, $(3.8\ Å)^2$], respectively. A view of the reference conformation is shown in Figures 3A–3C (a structure with 968 pseudoatoms was obtained for the EM volume approximation error of 5%, and the SD of Gaussian functions $\sigma = 3.8$).

From 100 computed normal modes, we selected six lowest frequency non-rigid-body modes, as their collectivity degree was above 0.5 (Figure 3D). These six modes were used for the image analysis with the proposed method. The computed deformation amplitudes along the modes were then analyzed by principal-component analysis (PCA) to reduce the complexity from six to three dimensions. To visualize how the conformation changes along a mode or a principal axis (PA), a volumetric form of the pseudoatomic structure (obtained by pseudoatoms-to-volume conversion) was displaced in positive and negative directions of each of six normal modes and six PAs (Figure S3). As the PAs are linear combinations of normal modes, a contribution of different normal modes to the PAs can be observed in Figures S3G–S3L. For example, the most important PA (PA1; Figure S3G) is mainly contributed by mode 7 (Figure S3A) (a similar movement of the upper lobe is visible in Figure S3G and Figure S3A), although some other modes contribute as well (e.g., the linker is moving in Figure S3G but not in Figure S3A). Similarly, the second most important PA (PA2; Figure S3H) is mostly contributed by mode 8 (Figure S3B). The remaining PAs are more difficult to interpret visually, and their quantitative analysis showed that they were more "mixed" than PA1 and PA2 (i.e., a normal mode has a similar contribution to a PA as the other modes).

The computed deformation amplitudes are shown in Figure 3E as projections on to PA1–PA3 (each image is represented by a point), together with linear regression lines through three identified clusters of points (Figures 3E and 3I). Displacements of the volumetric form of the pseudoatomic structure along each of the three lines (Figures 3F–3H) show a movement similar to that along PA2 (Figure S3H). The main difference between the displacements along the three lines is related to the degree of an additional movement similar to that along PA1 (Figure S3G). Movies S4 and S5 show the displacement of the pseudoatomic structure and its volumetric form along the red (central) line as the highest number of images was distributed along this line.

These results clearly show a high conformational heterogeneity of the complex, which suggests that any structure obtained by 3D reconstruction from these images would be of low resolution and partially inaccurate, as the actual flexibility would be limited by averaging of different conformations. Here, we present some reconstructed structures to show these effects. A structure was computed from each of the three clusters of images (Figures 3I and 3K). The overlapped structures (Figure 3K) show a movement similar to that along PA1 (Figure S3G). An overlap between the structures reconstructed along any of the three regression lines (Figure 3E) shows a movement similar to that along PA2 (Figure S3H). For example, Figure 3L shows an overlap of the structures reconstructed from three subsets (Figure 3J) of the yellow (central) cluster in Figure 3I. These reconstructed structures (Figure 3L) are similar to those shown for the displaced pseudoatomic structure (Figure 3G) along the corresponding central regression line in Figure 3E, but their resolution is low due to a variability of the incorporated conformations in the reconstruction.

As the reconstructed structures show movements similar to those along two most important PAs and these axes are mostly contributed by modes 7 and 8, we can conclude that these two modes are the most dominant modes for this data set. Interestingly enough, it has been shown that one to three modes can be sufficient to globally describe experimentally observed conformational changes (Tama et al., 2004b; Tama and Sanejouand, 2001). Our method, thus, can help to find the most dominant modes, which may be particularly useful when they are not among a few lowest frequency modes (e.g., experiment 4).

Regarding the reconstructions in Figure 3K, a similar movement was observed previously (Klinge et al., 2009), but with a smaller amplitude. To check if the smaller amplitude was due to an incorporation of different conformations into the same reconstructed structure, we designed a test involving three more reconstructions. We split the central cluster in Figure 3I into three subsets of images according to their coordinate on PA1 (each subset contained around 30% of images). Two structures were computed from two artificially mixed sets (each containing a side cluster and its neighboring central-cluster subset). The third structure was computed from the remaining central-cluster subset. Interestingly enough, these last three structures overlap quite well with the conformations obtained previously (Klinge et al., 2009) (similar movements with similar amplitudes can be observed in Figure 3M). The smaller amplitude of the movement observed previously can thus be explained by a

(K) Structures reconstructed from the clusters in (I) (the color of the structure is that of the cluster).

(L) Structures reconstructed from the subsets in (J), with the same coloring as that used in (K).

(M) Structures from mixed subsets (blue mesh, see the text for more details) overlapped with the structures obtained with ML3D (orange solid). The red asterisk denotes an absence of the complete mass in the linker region obtained by ML3D, which is an evidence of the conformational heterogeneity of the image set used to compute this structure. Arrows show movements.

See also Figure S3 and Movies S4 and S5.

Please cite this article in press as: Jin et al., Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes, Structure (2014), http://dx.doi.org/10.1016/j.str.2014.01.004

heterogeneity of conformations incorporated in the reconstructed average structures (the flexibility could not be fully explored with ML3D using a small number of classes, and the full exploration would require a much larger number of classes and images; Klinge et al., 2009). HEMNMA described more extensively the conformational heterogeneity that was detected by ML3D using the same data set. For example, it offered the possibility to additionally visualize the changes of the linker length coupled with the changes of the relative distance between the lobes (movements reflecting mainly displacements along PA2 (Figures 3G and 3L; Figure S3H).

Experiment 4: EM Images of TBSV

Swelling of many icosahedral plant viruses has been observed upon changes of pH and EDTA chelation of divalent cations on the interface between subunits (Ca^{2+} or Mg^{2+} that appear to play a critical role in virion stability), but the swelling mechanism is still not well understood (Witz and Brown, 2001). In a previous EM study of the TBSV swelling, a structure of the native (compact) TBSV form was obtained at 13 Å resolution, and a continuous particle size variation was found on micrographs of the virus after its dialysis in Tris-EDTA buffer (Aramayo et al., 2005). Also, a structure of the most swollen form was obtained at 19 Å resolution by an iterative procedure whose each iteration comprised image alignment, sorting of the largest particles (based on a threshold cross-correlation with the last reference structure), and 3D reconstruction to compute a new reference structure for the next iteration (Aramayo et al., 2005). HEMNMA was used here to analyze a set of 7,000 TBSV images.

The EM structure and image sizes were reduced to 128^3 voxels and 128^2 pixels, respectively [pixel size, $(3.2\ Å)^2$]. The pseudoatomic structure for NMA was obtained from the structure of the compact form (6,096 pseudoatoms for the volume approximation error of 5%, and for the SD of Gaussian functions $\sigma = 3.2$) (Figure 4A). From 100 computed normal modes, image analysis was conducted using four lowest frequency non-rigid-body modes whose collectivity degree was above 0.75 (modes 13, 24, 40, and 77, marked by asterisks in Figure 4B). No information was used about the potential symmetry of the conformations during the image analysis, and the used modes allowed for asymmetric geometrical transformations. In fact, only one of the used modes deforms the structure while perfectly preserving its symmetry (mode 24, which is related to the radially symmetric expansion). The other three modes result in asymmetric changes. The deformation amplitudes computed through image analysis for each of the four modes were then mapped on to the 3D PCA space (Figure 4C). The data were fitted with a linear regression line in the PCA space, and the reference structure was displaced along this line (blue asterisks in Figure 4C; Movie S6). Also, the data were split into four classes according to the data coordinate on PA1 (four equal segments of the axis), and 3D reconstruction was conducted from the images in each class (Figure 4D). The colors red, magenta, green, and yellow in Figures 4C and 4D correspond to the four classes with 21%, 25%, 24%, and 30% of all images, respectively. Although the image analysis was conducted with no a priori information about the conformational symmetry, the reconstructed structures were icosahedrally symmetrized to reduce noise, as it was done in the previous study (Aramayo et al., 2005). The difference between the diameter of the structure from the most swollen particles (yellow in Figure 4D) and that of the compact-form structure (gray in Figure 4D) is about 4 nm, which is consistent with the previously reported differences (Aramayo et al., 2005; Krüse et al., 1982). The difference in the diameter of the structure from the most and least swollen particles (yellow and red in Figure 4D, respectively) is about 3.2 nm. This indicates that the diameter of the least swollen particles in the analyzed images is larger than that of the compact-form particle and that the particles were all more or less swollen in the analyzed samples.

The conformational change can be mainly described as a combination of an increase in the radius of the particle (increase in the negative direction of PA1), a rotation of the subunits around the 5-fold symmetry axis, and an increase in the size of the intersubunit space at the subunits quasi-trimer (central part of the capsid in Figure 4D and Movie S6). These results are consistent with the models of the TBSV conformational change proposed in the literature (Aramayo et al., 2005; Krüse et al., 1982; Robinson and Harrison, 1982). Most important, the visualization of the full conformational distribution revealed many intermediate states of the change, which describes more extensively the conformational heterogeneity that was detected previously (Aramayo et al., 2005).

DISCUSSION

This article describes the method HEMNMA that allows analysis of the dynamics of large macromolecular complexes by experimental EM image analysis of possible conformational-change directions predicted by NMA. HEMNMA is particularly useful in the cases in which the reference structure at atomic resolution is unavailable but a structure can be obtained by EM.

Full Dynamics Can Be Studied

HEMNMA is ideally suited to study full dynamics, as it does not use classification during the image analysis. The classification of the computed deformation amplitudes can be performed at the last step to compute structures with similar conformations. We showed that the conformational transition conveyed by the reconstructed structures is consistent with the one conveyed by the reference structure displaced along the trajectories fitted through the computed deformation amplitudes. However, the full conformational distribution conveys additional information that is not contained in the reconstructions obtained from the classes. Therefore, when flexibility is described by a large range of conformations, a clear advantage of our method is in looking for a full range of conformations, contrary to the standard methods that only look for discrete and limited number of conformational states.

HEMNMA Robustness

In the ribosome experiments (experiments 1 and 2), we analyzed images with the proposed method using only the mode related to ratchet-like motion to see whether we could recover both ratchet-like and L1-stalk motions. The motions were successfully recovered, and robustness of the method was shown to different realistic factors (e.g., different SNRs, different CTFs, different resolutions and compositions of the reference structure with respect to those of the imaged complexes).

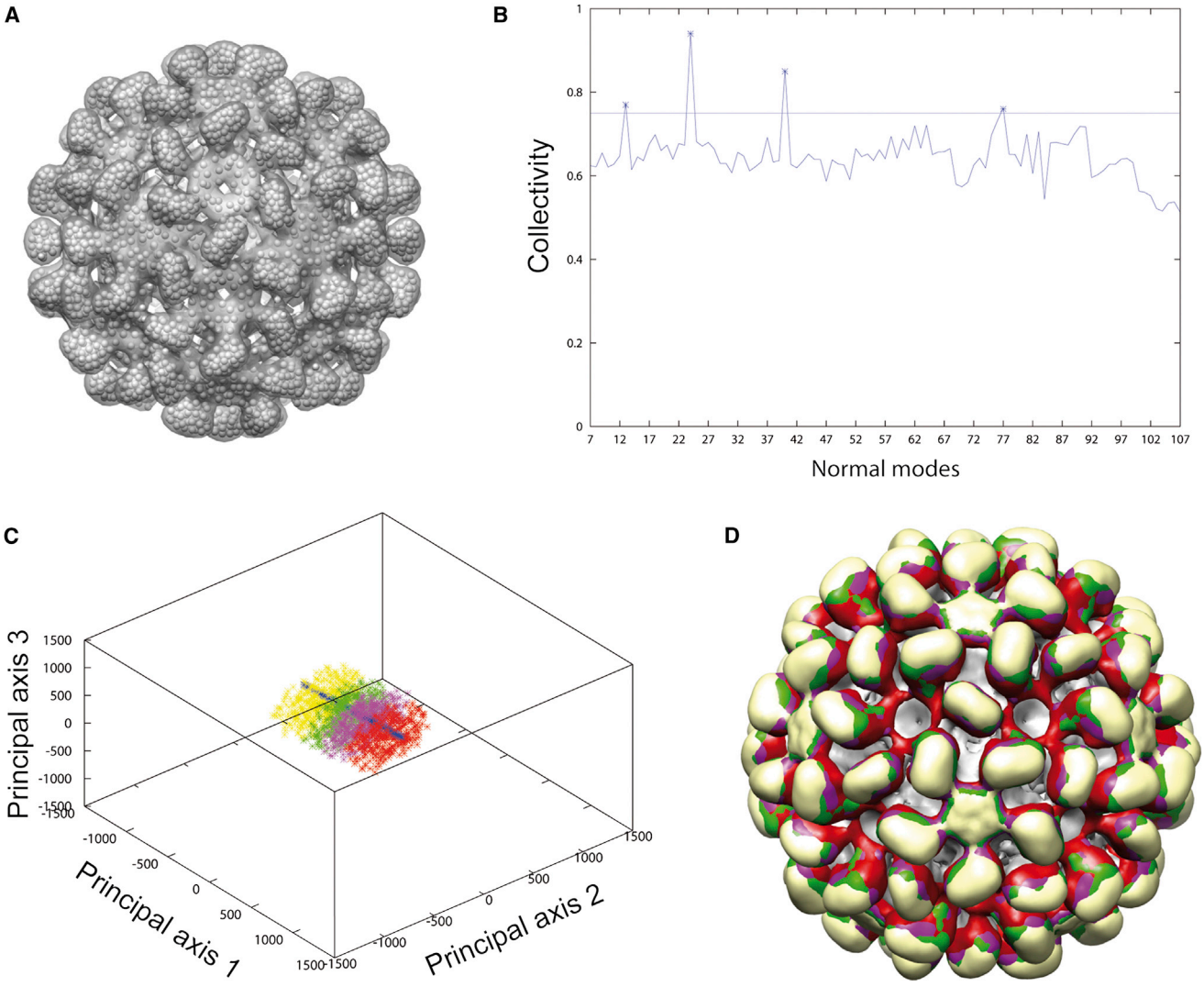

Figure 4. Experiment with EM Images of TBSV

(A) Pseudoatomic structure (spheres) computed using a compact-form TBSV structure from a previous EM study (the EM structure [Aramayo et al., 2005] is shown in transparent gray).

(B) Collectivity degree of the pseudoatomic normal modes. The first six modes related to rigid-body movements are not shown; the horizontal line corresponds to the collectivity of 0.75; the asterisks denote the modes with the collectivity degree above 0.75.

(C) Computed deformation amplitudes projected on to the three most important PAs (the points denote single-particle images and PA1, the most important PA), the linear regression line (blue), and four groups of images (red, magenta, green, and yellow) determined by splitting the full range of the data coordinates on the first PA into four equal-length intervals.

(D) Overlapped structures reconstructed from the four groups of images shown in (C). The volumes are indicated with the color of the corresponding group in (C). See also Movie S6.

Complexes of Different Composition Can Be Studied

The experiment with EM images of 70S ribosome (experiment 2) was principally used to show that the method can work with different compositions of complexes in the same sample (with EF-G, without EF-G, and with partially bound EF-G) that, at the same time, produce different conformations of the complexes. We refer to this kind of conformational variability to as "discrete-type conformational changes," because the sample is prepared so that a small number of conformational classes are expected and the problem can be solved by classical ("discrete") methods. This experiment shows that, in the case of such "discrete-type" problems, our method can reproduce the results obtained using classical methods and the same data (Elad et al., 2008; Scheres et al., 2007). The functional significance of these results was reported in the first paper in which this data set was analyzed (Scheres et al., 2007).

Functional Significance of Observed Motions

The experiments with Pol α-B and TBSV highlight the main characteristics of the methodology, showing that it can be used to explore truly continuous-type conformational changes where the complex cannot be stabilized in one or a few particular

Please cite this article in press as: Jin et al., Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes, Structure (2014), http://dx.doi.org/10.1016/j.str.2014.01.004

conformations. They also showed how the method works when no a priori knowledge about the conformational change is used for the selection of normal modes. The method can find contributions of the used modes to the PAs as well as their contributions to the principal trajectories that determine the most dominant motions. Moreover, the methodology is particularly useful in identifying the most dominant modes when they are not among a few lowest frequency modes (e.g., as in the TBSV case). These experiments are also interesting as they show that the method allows studying conformational pathways that may actually exist but may be difficult to solve by X-ray crystallography or classical EM methods that rely on a predefined initial number of classes.

Pol α-B is arranged as an elongated structure organized in two lobes connected by a flexible linker. This quaternary organization has been observed for other eukaryotic replicative polymerases such as DNA polymerase ε (Asturias et al., 2006) and DNA polymerase δ (Jain et al., 2009), which are responsible for the replication of the leading and lagging strands, respectively. In all these replicative complexes, the flexibility between the different lobes has been identified as a functional characteristic that allows the accommodation of different conformations of DNA during its replication and facilitates the interactions of these complexes with other components of the replication machinery. An EM analysis of the structure of the complete eukaryotic primosome, consisting of Pol α, the B subunit, and the two components of the primase (PriL and PriS) revealed a very flexible connection between the two lobes and that this complex was capable of adopting a large range of different conformations (Núñez-Ramírez et al., 2011). This flexibility was postulated as a mechanism for the transfer of the DNA substrate from the primase catalytic site placed in one of the lobes to the DNA polymerase catalytic site located in the other lobe. It is surprising that ML3D image classification of Pol α-B, a subcomplex, comprising only two of the components of the primosome, revealed a limited flexibility between the catalytic core of Pol α (the big lobe) and the B-CTD platform (the small lobe) (Klinge et al., 2009). The distinct behavior of the full primosome complex and the Pol α-B subcomplex was difficult to rationalize. Now, the analysis of the same data set with HEMNMA has revealed that the high degree of flexibility showed in the full complex is also present in the Pol α-B subcomplex. Intermediate conformations of the Pol α-B subcomplex were undetected by ML3D, probably because of the difficulty in assigning a large range of conformations to a discrete number of solutions. In addition, the classical refinement-based methods use an initial template and typically discard images that poorly correlate with the average 3D model obtained during the template refinement. Thus, if a dominant conformation was used to build the initial reference, we suspect that the particles corresponding to distant conformations would be removed during the refinement by classical methods. HEMNMA has provided a more precise description of the continuous range of conformations of Pol α-B complexes, allowing a better understanding of their structure-function relationship.

The visualization of the full conformational distribution of the TBSV helped to reveal many intermediate conformations that were previously hypothesized but impossible to visualize (Aramayo et al., 2005). Indeed, many intermediate conformations might explain a difficult crystallization of such swelling viruses (e.g., crystallography produced a swollen TBSV structure at resolution of only 8 Å; Robinson and Harrison, 1982). The Ca^{2+} ions are known to be present between the subunits forming the TBSV capsid in order to strengthen the links between them. Many different intermediate conformations may be explained by an incomplete removal of these ions by EDTA (sample preparation is described by Aramayo et al., 2005) that could lead to a partial opening of the viruses, partially held by remaining Ca^{2+} ions. Similar conformations may exist as mechanical intermediate states permitting the release of the viral RNA within the host cell. The genome of the TBSV consists of a single RNA filament shifting toward the capsid in an environment depleted of Ca^{2+} ions, and such intermediate conformations could be enough for the release of the single RNA filament in vivo.

Thus, HEMNMA seems highly suitable for macromolecular complexes displaying a continuous range of conformations. This may be the case of a large number of complexes in biology. In these conditions, HEMNMA seems significantly more powerful than the methods with a predefined initial number of classes, which would require larger data sets and larger computing times to evaluate the presence of a large number of possible conformers. However, the classification-based methods can be combined with HEMNMA to better analyze overall heterogeneity (e.g., by providing one or several reference structures as in the Pol α-B case).

EXPERIMENTAL PROCEDURES

Pseudoatomic Structures from EM Volumes

To obtain a reduced EM density representation, our recently published method for volume-to-pseudoatoms conversion (Nogales-Cadenas et al., 2013) was used as it allows controlling the volume approximation error and, thus, controlling quality of the projections for the elastic projection-matching-based image alignment. The used conversion method represents a volume by a collection of Gaussian functions whose number, positions, and amplitudes are adjusted for a given volume approximation error and a given Gaussian standard deviation. The Gaussian functions are referred to as pseudo-atoms although their positions do not generally coincide with the atomic positions, and the structure with pseudo-atoms is referred to as pseudo-atomic structure (Nogales-Cadenas et al., 2013).

NMA

To compute normal modes, a simplified elastic network representation of the potential energy function (Tirion, 1996) was used, as implemented on the ElNemo web server (Suhre and Sanejouand, 2004) for atomic structures and the 3DEM Loupe web server (Nogales-Cadenas et al., 2013) for EM structures. Before computing normal modes, 3DEM Loupe converts the input EM volume into a pseudoatomic structure that can be directly used as the reference structure for the method proposed here. The coordinates of the normal mode vectors are computed in angstroms, while the deformation amplitudes have no units. Six lowest frequency modes (modes 1–6) are not used as they are related to rigid-body movements.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, one table, and six movies and can be found with this article online at http://dx.doi.org/10.1016/j.str.2014.01.004.

ACKNOWLEDGMENTS

This work was supported by the French National Center for Scientific Research and the Spanish National Research Council (CSIC2009FR0015 and PICS 2011 to S.J. and C.O.S.S.); the Spanish Ministry of Economy and Competitiveness (ACI2009-10220, ACI2010-1088, and BIO2010-16566 to J.M. Carazo and

SAF2011-22988 to O.L.); the European Social Fund and the Ministerio de Educación y Ciencia ("Ramón y Cajal" fellowship to C.O.S.S. and J.R.B.-C.); the European Commission (DECI-6 to C.O.S.S. and S.J.); and the French National Research Agency (ANR-11-BSV8-010-04 to S.J.). We thank the GENCI-CINES/IDRIS for computing resources (2011-x2010072174 and 2012-x2010072174 to S.J.), J. Frank and H. Gao (Columbia University) for depositing the 70S/EF-G images on the EMDataBank web site, L. Pellegrini and S. Klinge (Cambridge University) for providing the Pol α-B samples, E. Larquet and R. Aramayo (IMPMC) for providing the TBSV data, and I. Callebaut (IMPMC) and J.M. Carazo (Centro Nacional de Biotecnología) for stimulating discussions.

Received: August 28, 2013
Revised: December 28, 2013
Accepted: January 2, 2014
Published: February 6, 2014

REFERENCES

Aramayo, R., Mérigoux, C., Larquet, E., Bron, P., Pérez, J., Dumas, C., Vachette, P., and Boisset, N. (2005). Divalent ion-dependent swelling of tomato bushy stunt virus: a multi-approach study. Biochim. Biophys. Acta *1724*, 345–354.

Asturias, F.J., Cheung, I.K., Sabouri, N., Chilkova, O., Wepplo, D., and Johansson, E. (2006). Structure of Saccharomyces cerevisiae DNA polymerase epsilon by cryo-electron microscopy. Nat. Struct. Mol. Biol. *13*, 35–43.

Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., and Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. Structure *12*, 185–191.

Bruschweiler, R. (1995). Collective protein dynamics and nuclear spin relaxation. J. Chem. Phys. *102*, 3396–3403.

Chacón, P., Tama, F., and Wriggers, W. (2003). Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. J. Mol. Biol. *326*, 485–492.

Delarue, M., and Dumas, P. (2004). On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc. Natl. Acad. Sci. USA *101*, 6957–6962.

Elad, N., Clare, D.K., Saibil, H.R., and Orlova, E.V. (2008). Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. *162*, 108–120.

Fu, J., Munro, J.B., Blanchard, S.C., and Frank, J. (2011). Cryoelectron microscopy structures of the ribosome complex in intermediate states during tRNA translocation. Proc. Natl. Acad. Sci. USA *108*, 4817–4821.

Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. Structure *14*, 511–520.

Jain, R., Hammel, M., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2009). Structural insights into yeast DNA polymerase delta by small angle X-ray scattering. J. Mol. Biol. *394*, 377–382.

Jonić, S., Sorzano, C.O.S., Thévenaz, P., El-Bez, C., De Carlo, S., and Unser, M. (2005). Spline-based image-to-volume registration for three-dimensional electron microscopy. Ultramicroscopy *103*, 303–317.

Klinge, S., Núñez-Ramírez, R., Llorca, O., and Pellegrini, L. (2009). 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. EMBO J. *28*, 1978–1987.

Krüse, J., Krüse, K.M., Witz, J., Chauvin, C., Jacrot, B., and Tardieu, A. (1982). Divalent ion-dependent reversible swelling of tomato bushy stunt virus and organization of the expanded virion. J. Mol. Biol. *162*, 393–414.

Ming, D., Kong, Y., Lambert, M.A., Huang, Z., and Ma, J. (2002). How to describe protein motion without amino acid sequence and atomic coordinates. Proc. Natl. Acad. Sci. USA *99*, 8620–8625.

Nogales-Cadenas, R., Jonic, S., Tama, F., Arteni, A.A., Tabas-Madrid, D., Vázquez, M., Pascual-Montano, A., and Sorzano, C.O.S. (2013). 3DEM Loupe: analysis of macromolecular dynamics using structures from electron microscopy. Nucleic Acids Res. *41*, W363–W367.

Núñez-Ramírez, R., Klinge, S., Sauguet, L., Melero, R., Recuero-Checa, M.A., Kilkenny, M., Perera, R.L., García-Alvarez, B., Hall, R.J., Nogales, E., et al. (2011). Flexible tethering of primase and DNA Pol α in the eukaryotic primosome. Nucleic Acids Res. *39*, 8187–8199.

Penczek, P.A., Frank, J., and Spahn, C.M. (2006). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. J. Struct. Biol. *154*, 184–194.

Peng, L.M., Ren, G., Dudarev, S.L., and Whelan, M.J. (1996). Robust parameterization of elastic and absorptive electron atomic scattering factors. Acta Crystallogr. *A52*, 257–276.

Powell, M.J.D. (2002). UOBYQA: unconstrained optimization by quadratic approximation. Math. Program. *92*, 555–582.

Robinson, I.K., and Harrison, S.C. (1982). Structure of the expanded state of tomato bushy stunt virus. Nature *297*, 563–568.

Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J., and Carazo, J.M. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods *4*, 27–29.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O., and Klaholz, B.P. (2008). Structure of the 30S translation initiation complex. Nature *455*, 416–420.

Sorzano, C.O.S., Jonić, S., El-Bez, C., Carazo, J.M., De Carlo, S., Thévenaz, P., and Unser, M. (2004a). A multiresolution approach to orientation assignment in 3D electron microscopy of single particles. J. Struct. Biol. *146*, 381–392.

Sorzano, C.O.S., Marabini, R., Velázquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., and Pascual-Montano, A. (2004b). XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. *148*, 194–204.

Suhre, K., and Sanejouand, Y.H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. *32* (Web Server issue), W610–W614.

Suhre, K., Navaza, J., and Sanejouand, Y.H. (2006). NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. Acta Crystallogr. D Biol. Crystallogr. *62*, 1098–1100.

Tama, F., and Sanejouand, Y.H. (2001). Conformational change of proteins arising from normal mode calculations. Protein Eng. *14*, 1–6.

Tama, F., and Brooks, C.L., 3rd. (2005). Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. J. Mol. Biol. *345*, 299–314.

Tama, F., and Brooks, C.L. (2006). Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu. Rev. Biophys. Biomol. Struct. *35*, 115–133.

Tama, F., Wriggers, W., and Brooks, C.L., 3rd. (2002). Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J. Mol. Biol. *321*, 297–305.

Tama, F., Valle, M., Frank, J., and Brooks, C.L., 3rd. (2003). Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. Proc. Natl. Acad. Sci. USA *100*, 9319–9323.

Tama, F., Miyashita, O., and Brooks, C.L., 3rd. (2004a). Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. J. Mol. Biol. *337*, 985–999.

Tama, F., Miyashita, O., and Brooks, C.L., 3rd. (2004b). Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J. Struct. Biol. *147*, 315–326.

Tirion, M.M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys. Rev. Lett. *77*, 1905–1908.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003). Locking and unlocking of ribosomal motions. Cell *114*, 123–134.

Wang, Y., Rader, A.J., Bahar, I., and Jernigan, R.L. (2004). Global ribosome motions revealed with elastic network model. J. Struct. Biol. *147*, 302–314.

Witz, J., and Brown, F. (2001). Structural dynamics, an intrinsic property of viral capsids. Arch. Virol. *146*, 2263–2274.

Zhang, W., Dunkle, J.A., and Cate, J.H. (2009). Structures of the ribosome in intermediate states of ratcheting. Science *325*, 1014–1017.

Structure, Volume *22*

Supplemental Information

Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes

Qiyu Jin, Carlos Oscar S. Sorzano, José Miguel de la Rosa-Trevín, José Román Bilbao-Castro, Rafael Núñez-Ramírez, Oscar Llorca, Florence Tama, and Slavica Jonić

Supplemental Data

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Structure used to compute images with synthetic continuous conformational change of the *E. coli* 70S ribosome, the images, and structure used to analyze them. (A) Atomic-resolution structure of the 70S ribosome in the conformation in which the subunit 30S is not rotated with respect to the subunit 50S (unratcheted conformation). (B) Synthetic volume at the resolution of 15 Å from the structure shown in (A). (C) Pseudo-atomic structure (spheres) from the volume in (B) overlapped with the volume (mesh). (D) Synthetic images without noise and CTF effects (row 1), images with SNR=0.3 and the defocus of 0.5 μm, 1 μm, and 2 μm (rows 2-4, respectively), and images with SNR=0.1 and the defocus of 0.5 μm, 1 μm, and 2 μm (rows 5-7, respectively). See the main text (**Experiment 1)** and **Supplemental experimental procedure S2** for more details.

Figure S2: Experiment with synthetic discrete conformational change given by EM structures of three different *E. coli* 70S conformations: (A-C) EM structures of unratcheted (EMD-5262 (A)), intermediate (EMD-5267 (B)), and ratcheted (EMD-5265 (C)) conformations used for computing synthetic images. (D) Pseudo-atomic structure (spheres) overlapped with the structure shown in (A) (mesh). (E) Histogram of the estimated deformation amplitudes (green) for the most difficult of the seven tested cases (SNR=0.1, defocus=2μm), where blue, magenta, and red indicate the contribution of images of the structures shown in (A), (B), and (C), respectively, and the vertical tan and cyan lines indicate the separation of images with respect to the computed deformation amplitudes into three classes of equal size (300 images per class). (F) Overlap between the structure reconstructed from the class with the smallest deformation amplitudes (red mesh for the reconstruction from the images at the left side of the tan class-separation line in (E)), the structure reconstructed from the class with the intermediate values of the deformation amplitudes (grey solid for the reconstruction from the images at the right side of the tan class-separation line and at the left side of the cyan class-separation line in (E)), and the structure reconstructed from the class with the largest deformation amplitudes (blue mesh for the reconstruction from the images at the right side of the cyan class-separation line in (E)). (G) Overlap between the structure reconstructed from the class with the smallest deformation amplitudes (red solid in (G) is red mesh in (F)) and the structure reconstructed from the class with the largest deformation amplitudes (blue solid in (G) is blue mesh in (F)). Rotation of the 30S subunit and the movement of the L1 stalk are denoted with arrows (although the L1 movement is rotational, it is shown as a shift to designate an inward direction of the movement). See the main text (**Experiment 2, Figure 2**) and **Supplemental experimental procedure S3** for more details**.**

Figure S3: Displacement of a volumetric form of the Pol α – B pseudo-atomic structure along normal modes and along principal component axes. (A-F) Displacement along normal modes, from mode 7 (A) to mode 12 (F). (G-L) Displacement along principal component axes, from the most important principal axis (G) to the least important principal axis (L). The displacement amplitude was chosen large to show clearly the movement directions. Displacement amplitude +/-1000 (blue mesh and red mesh) and zero displacement amplitude (reference conformation, yellow). See the main text for more details (**Experiment 3**) as well as **Figure 3**.

SUPPLEMENTARY TABLE LEGENDS

Table S1: Results of the experiment in which the ground-truth synthetic continuous-type conformational change of *E. coli* 70S ribosome was determined by two atomic normal modes (modes 11 and 12) but it was estimated by analyzing images with only one atomic (mode 11 of the atomic reference structure (**Movie S1**)) or pseudo-atomic normal mode (mode 13 of

the pseudo-atomic synthetic reference structure (**Movie S2**)). (A,B) Error of the estimated elastic (A) and rigid-body (B) parameters with respect to the ground-truth values for no noise, SNR=0.3, SNR=0.1, and the defocus of 0.5 μm, 1 μm, and 2 μm. (C) Dependence of parameters of the volume-to-pseudoatoms conversion and NMA methods and that of the image alignment error on the volume approximation error and the size of pseudo-atoms (the standard deviation of the Gaussian functions), for the images without noise and CTF effects. See the main text (**Experiment 1**) and **Supplemental experimental procedure S2** for more details.

SUPPLEMENTARY MOVIE LEGENDS

Movie S1: Displacement along the atomic mode 11 related to the ratchet-like motion of the 70S ribosome (the atomic structure is 3I1OP (**Fig. S1A**)). The mode was used to analyze images in **Experiment 1**. See also **Supplemental experimental procedure S2**. **Movies S1-S3** were produced using large amplitudes of the displacement along the mode to better visualize the movements.

Movie S2: Displacement along the pseudo-atomic mode 13 related to the ratchet-like motion of the 70S ribosome, from a synthetic density volume of the atomic (3I1OP) structure (the pseudo-atomic structure (**Fig. S1C**) was obtained by converting the 15 Å-resolution 3I1OP density volume (**Fig. S1B**)). The mode was used to analyze images in **Experiment 1.** See also **Supplemental experimental procedure S2**.

Movie S3: Displacement along the pseudo-atomic mode 9 related to the ratchet-like motion of the 70S ribosome, from the EM structure EMD_5262 (the EM and pseudo-atomic structures are shown in **Fig. S2A,D**). The mode was used to analyze images in **Experiment 2**. See also **Supplemental experimental procedure S3**.

Movie S4: Animated displacements of the reference pseudo-atomic Pol α – B structure along the red line in **Figure 3E**. The amplitude of the displacement is a result of the image analysis in **Experiment 3**. The same view of the structure is showed during the displacement along the line.

Movie S5: Animated displacements of a volumetric form of the reference pseudo-atomic Pol α – B structure along the red line in **Figure 3E**. The amplitude of the displacement is a result of the image analysis in **Experiment 3**. The volume is rotated during the displacement along the line so that it can be viewed from different directions.

Movie S6: Animated displacements of the reference pseudo-atomic TBSV structure along the blue line in **Figure 4C**. The amplitude of the displacement is a result of the image analysis in **Experiment 4**. The same view of the structure is showed during the displacement along the line.

Supplemental experimental procedures

Supplemental experimental procedure S1 (related to Methodology)

Visualization of computed conformations

Three-dimensional reconstruction
The 3D-reconstructed structures shown throughout the paper were obtained using the computed deformation amplitudes (for the classification), the computed orientations and translations of particles (for the reconstruction), and a standard 3D reconstruction method (here, ART implemented in Xmipp (Sorzano et al., 2004b)), without any further refinement of

the alignment parameters, noise reduction, or resolution improvement (*e.g.*, by adding more images).

Visualization of pseudo-atomic models
Figures and movies of pseudo-atomic models were obtained by approximating Gaussian functions with spheres. Overlapping spheres were used in the movies (obtained with VMD (Humphrey et al., 1996)) to enhance the structure visually (to give impression of density without converting the pseudo-atomic structure into a density volume) and to get closer to the reality as the Gaussian functions overlap (they have unlimited support).

Supplemental experimental procedure S2 (related to Experiment 1)

Atomic structure and atomic modes
The ground-truth synthetic conformational changes were obtained by deforming a reference atomic-resolution structure (**Fig. S1A**) that was a coarse-grained model (10204 atoms comprising only Cα and phosphate (P) atoms) of a 0.35 nm-resolution X-ray crystal structure of the *E. coli* 70S ribosome without ligands, in the conformation in which the small subunit of the ribosome (30S) is non-rotated with respect to the large subunit (50S) (a composite of the PDB entries 3I1O and 3I1P that will be referred to as 3I1OP (Zhang et al., 2009)). Twenty atomic lowest-frequency normal modes were computed for the atomic reference structure using ElNemo (Suhre and Sanejouand, 2004) web server.

Synthetic images
The synthetic images were computed as random projections of the atomic reference structure displaced with a combination of random displacement amplitudes along two atomic normal modes describing ratchet-like motion (mode 11) and L1-stalk motion (mode 12). The images were computed for the deformation amplitude along mode 11 that was random and had a uniform distribution (equal probability of different values) in the range [-500,0]. The amplitude along mode 12 was chosen to be 2.5 times smaller than the amplitude along mode 11, which means that mode 11 contributes 2.5 times more than mode 12 to the final motion. In this synthetic data experiment, the ratio of 2.5 was chosen to generate the total deformation (weighted linear combination of two modes with their deformation amplitudes as the weights) describing the structural rearrangements similar to those already observed for the 70S ribosome (30S rotates with respect to 50S by ~8° and the distal part of the L1 stalk moves by ~20 A° (Fu et al., 2011; Valle et al., 2003)).

Density volumes were computed from atomic coordinates using electronic-form atomic factors (Peng et al., 1996) and they were then projected onto the image plane using random values of Euler angles determining the volume orientation and using random values of the particle coordinates in the image (two out-of-plane rotations were distributed uniformly in the ranges [0°,360°] and [0°,180°], the in-plane rotation was 0°, and each of the two in-plane translations was distributed uniformly in the range [-5,5] pixels). The CTF and noise were simulated for a 200 kV microscope with a spherical aberration of 0.5 mm using the method of Velazquez-Muriel et al. (2003). The computed data thus comprised a set of 300 images without noise and CTF effects (image size: 128^2 pixels; pixel size: $(3\ \text{Å})^2$) and six versions of these images affected by different levels of noise (SNR=0.3 or SNR=0.1) and different CTF effects (defocus values of 0.5 μm, 1 μm, or 2 μm) (**Fig. S1D**). The chosen defocus and SNR values are commonly encountered in EM practice.

Pseudo-atomic structure and pseudo-atomic modes
The reference atomic-resolution structure (**Fig. S1A**) was converted into volumetric data (size: 128^3 voxels, voxel size: $(3\ \text{Å})^3$) that was low-pass filtered at 15 Å (**Fig. S1B**). This volume was converted into a pseudo-atomic structure with 6083 pseudo-atoms for the standard deviation of Gaussian functions $\sigma = 3$ and the volume approximation error of 2 % (**Fig. S1C**) and twenty lowest-frequency normal modes were computed for the pseudo-

atomic structure with the pseudo-atom interaction cutoff distance of 22 Å, using the methods of our 3DEM Loupe web server (Nogales-Cadenas et al., 2013) (the values of the volume-to-pseudoatoms conversion and NMA parameters were chosen as described in (Nogales-Cadenas et al., 2013)).

Image analysis
The best overlap between the atomic mode responsible for the ratchet-like motion (mode 11) and the set of pseudo-atomic modes was found for the pseudo-atomic mode 13 (the movies animating displacements along the atomic mode 11 and the pseudo-atomic mode 13 are given in **Movies S1-S2**). The elastic 3D-to-2D alignment of the synthetic images with the atomic reference structure was thus done using the atomic mode 11 and the one with the pseudo-atomic reference structure was done using the pseudo-atomic mode 13.

Influence of the quality of pseudo-atomic representation on image alignment error
Table S1C summarizes how the key parameters of the used volume-to-pseudoatoms conversion and NMA methods (see also **Experimental procedures**) and the resulting alignment error depend on the volume approximation error and the size of pseudo-atoms (the standard deviation of the Gaussian function). The results in **Table S1C** were obtained using images without noise and CTF effects and can be compared with the results in **Table S1A,B** for no-noise/no-CTF case (6083 pseudo-atoms, σ = 3, volume approximation error of 2 %, pseudo-atom interaction cutoff distance of 22 Å). We can notice that the number of pseudo-atoms decreases as the volume approximation error or the Gaussian standard deviation increases (*e.g.*, from 6083 pseudo-atoms for the target approximation error of 2 % to 450 pseudo-atoms for the target approximation error of 7 %, when using σ = 3 in both cases). Also, the pseudo-atom interaction cutoff distance increases as the number of pseudo-atoms decreases (from 22 Å for 6083 pseudo-atoms to 40 Å for 450 pseudo-atoms) and the alignment error increases as the volume approximation error increases (*e.g.*, the relative deformation-amplitude error of 4.3% (**Table S1A**) increases to 27.3% (**Table S1C**) when the volume approximation error increases from 2 % to 7 % for the same Gaussian standard deviation of 3). Similar results may be expected with noise- and CTF-affected images.

Supplemental experimental procedure S3 (related to Experiment 2)

EM structures
We analyzed synthetic images of EM structures representing the following three conformations of the *E. coli* 70S ribosome with two tRNAs, published by Fu et al. (2011): 1) unratcheted conformation EMD-5262 with resolution of around 13 Å (**Fig. S2A**) (the subunits 30S and 50S are not rotated with respect to each other and the A-site and P-site tRNAs are in the classical A/A and P/P positions, respectively, 2) ratcheted conformation EMD-5265 with resolution of around 14 Å (**Fig. S2C**) (the two subunits are fully rotated with respect to each other and the A-site and P-site tRNAs are in the A/A and P/E positions, respectively), and 3) structure with an intermediate relative rotation of the two subunits and an intermediate position of the P-site tRNA called position P-flip (EMD-5267, resolution of around 18 Å, **Fig. S2B**).

Synthetic images
The synthetic images were computed as random projections of the EM structures. The given EM volumes (EMD-5262, EMD-5265, and EMD-5267) were padded from the size 125^3 voxels (voxel size $(3\ Å)^3$) to the size of 128^3 voxels (the image size of 128^2 pixels and the volume size of 128^3 voxels were used throughout this paper regarding the experiments with the ribosome). We computed 300 images (128^2-pixels images with the pixel size of $(3\ Å)^2$) for each of the three EM structures, using the parameters of its orientation and position (rigid-body parameters) whose values were generated randomly, as in **Experiment 1** (two out-of-plane rotations were distributed uniformly in the ranges [0°,360°] and [0°,180°], the in-plane

rotation was 0°, and each of the two in-plane translations was distributed uniformly in the range [-5,5] pixels). The noise (SNR of 0.3 or 0.1) and the CTF (defocus of 0.5 μm, 1 μm, or 2 μm) were applied onto the ideal projections, as in **Experiment 1**. The total number of images was 900 in each of 7 analyzed datasets (one set without and six sets with noise and CTF effects).

Pseudo-atomic structure and pseudo-atomic modes
The volume with the unratcheted conformation (EMD-5262) was converted into a pseudo-atomic structure with 9411 pseudo-atoms for σ = 3 and the volume approximation error of 3.4% (**Fig. S2D**) and twenty lowest-frequency normal modes were computed for the pseudo-atomic structure with the pseudo-atom interaction cutoff distance of 22 Å using the methods of 3DEM Loupe web server (Nogales-Cadenas et al., 2013). The volume approximation error was larger when approximating this volume than the synthetic one (**Supplemental experimental procedure S2**), for the same voxel size and the same value of σ, because of a higher level of noise of the EM volume with respect to the synthetic one.

Image analysis
The images were analyzed using the reference pseudo-atomic structure and its mode reflecting the ratchet-like motion of the 30S and 50S subunits. The pseudo-atomic mode 9 (**Movie S3**) was thus used here as it overlapped the best with the atomic mode responsible for the ratchet-like motion (**Experiment 1, Movie S1**). Note that the atomic structure 3I1OP was aligned with the EMD-5262 structure before any further processing. As in **Experiment 1**, the results showed that the accuracy of the estimated rigid-body parameters decreases as the SNR decreases and the defocus increases and, in the most difficult tested case (SNR=0.1, defocus of 2 μm), the angular and shift errors were 1.3° and 0.2 pixel, respectively. Accuracy of the estimated elastic parameters could not be checked here as we had no knowledge about the ground-truth real conformational change but the analysis of the estimated deformation amplitudes showed distinct mean values for the three conformations (the average deformation amplitude over different SNR and defocus values is around -1 for the unratcheted conformation, -135 for the intermediate conformation, and -280 for the ratcheted conformation). The images were thus separated into three equal-size classes of estimated deformation-amplitude values (300 highest, 300 lowest, and 300 intermediate values).

Even for this small number of images per class (300), 3D reconstruction from the classes produced structures that were similar to the given EM structures (*e.g.*, **Figure S2E-G** shows the results in the most difficult tested case). **Figure S2F** clearly shows a rotation of the 30S subunit with respect to the 50S subunit (ratchet-like movement). **Figure S2F** also shows a movement of the L1 stalk with respect to the 30S subunit, which is an important result because the L1-stalk movement was not explicitly used to analyze images. Distinct degrees of rotation of the 30S subunit with respect to the 50S subunit and distinct positions of the L1 stalk are even better visualized in **Figure S2G** that shows an overlap between the structure reconstructed from the class with the smallest deformation amplitudes (images at the left side of the tan class-separation line in **Figure S2E**) and the structure reconstructed from the class with the largest deformation amplitudes (images at the right side of the cyan class-separation line in **Figure S2E**). Thus, this experiment with synthetic data showed that the ratchet-like mode alone can be used to classify images of the ribosome at distinct states of the conformational change combining ratchet-like and L1-stalk motions, and to reconstruct the structure at distinct states of the conformational change

Supplemental References

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

Velazquez-Muriel, J. A., Sorzano, C. O., Fernandez, J. J., and Carazo, J. M. (2003). A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment. Ultramicroscopy *96*, 17-35.

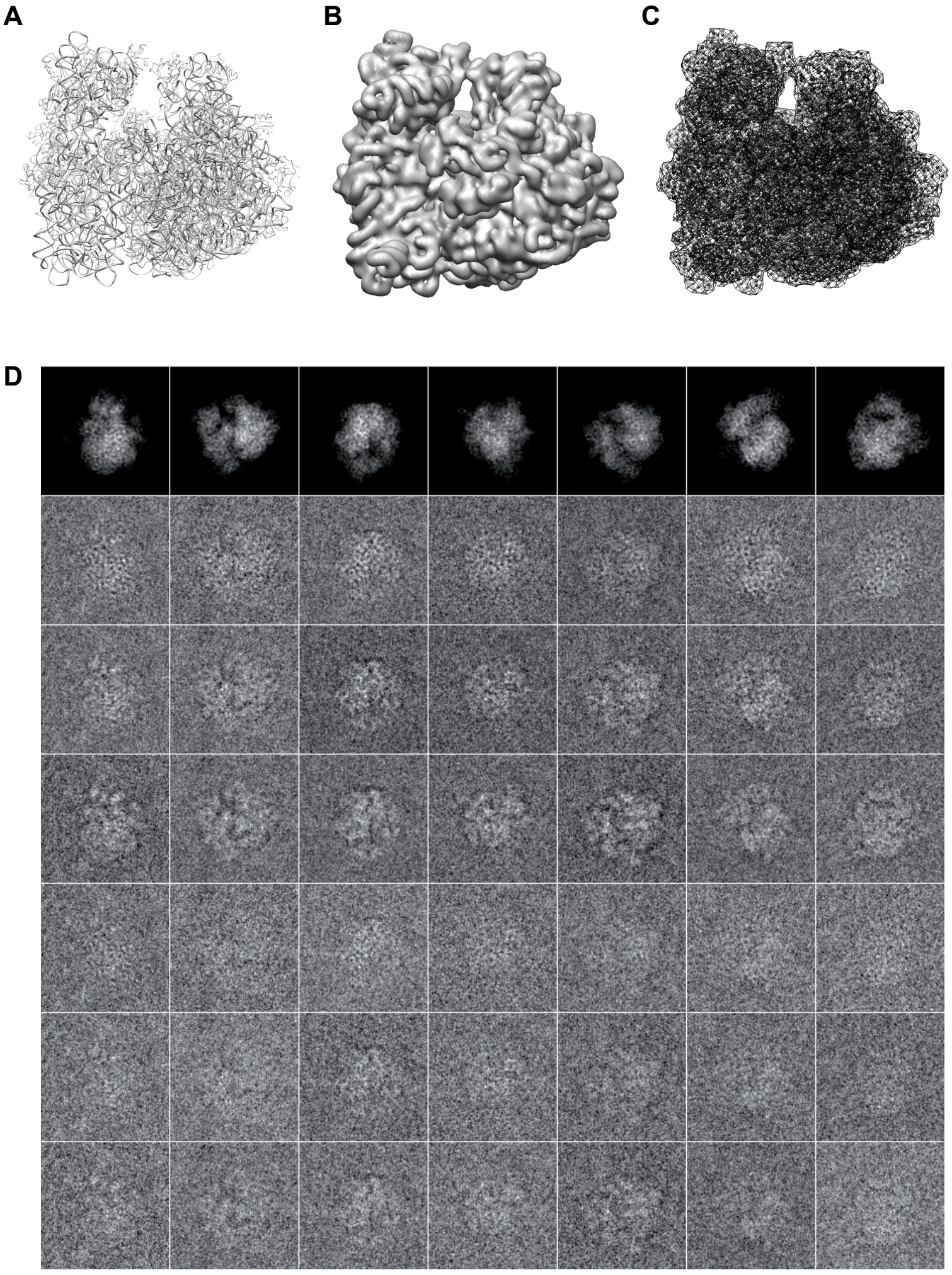

Figure S1

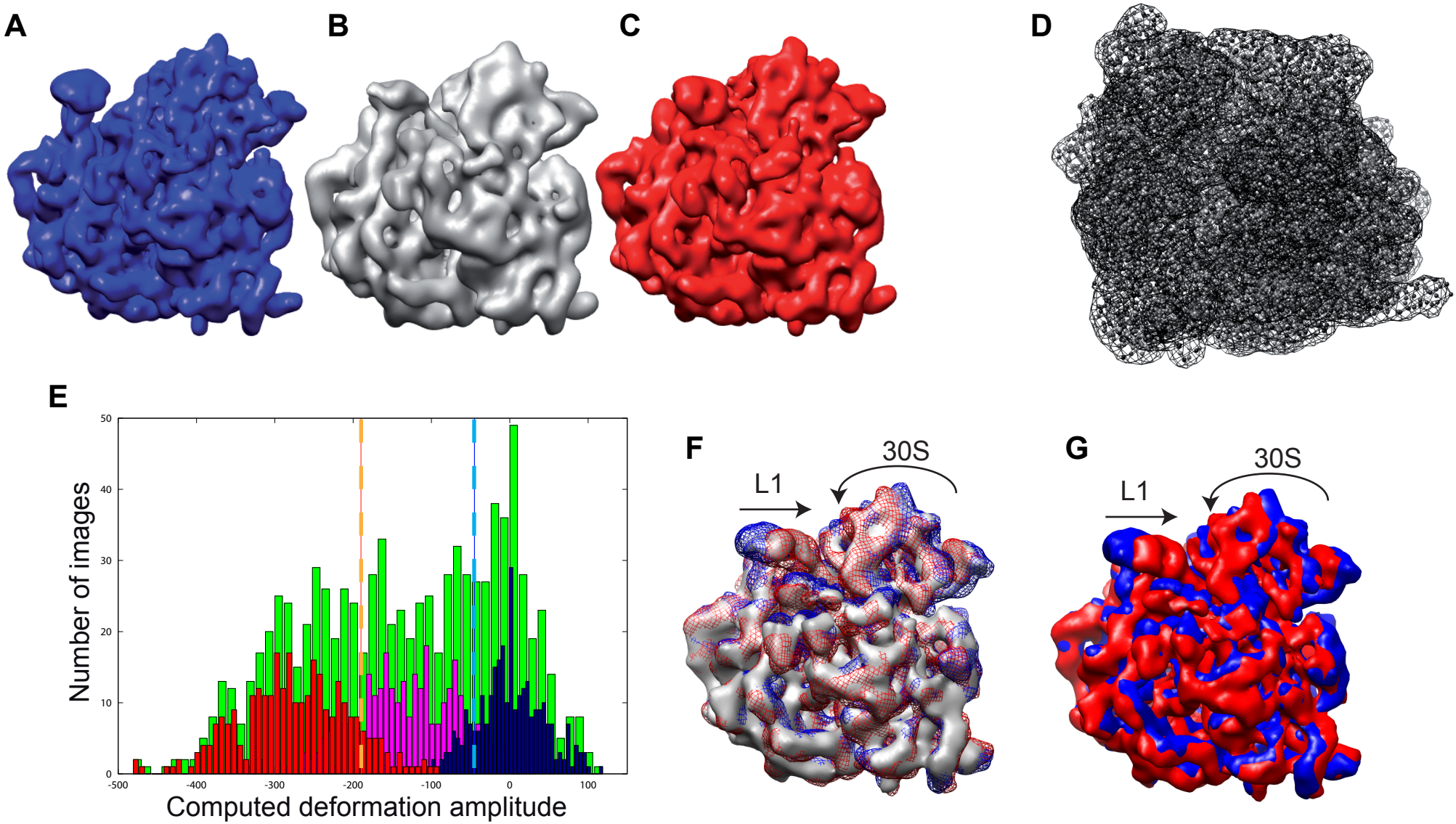
A
B
C
D
E
Number of images
Computed deformation amplitude
F
30S
L1
G
30S
L1

Figure S2

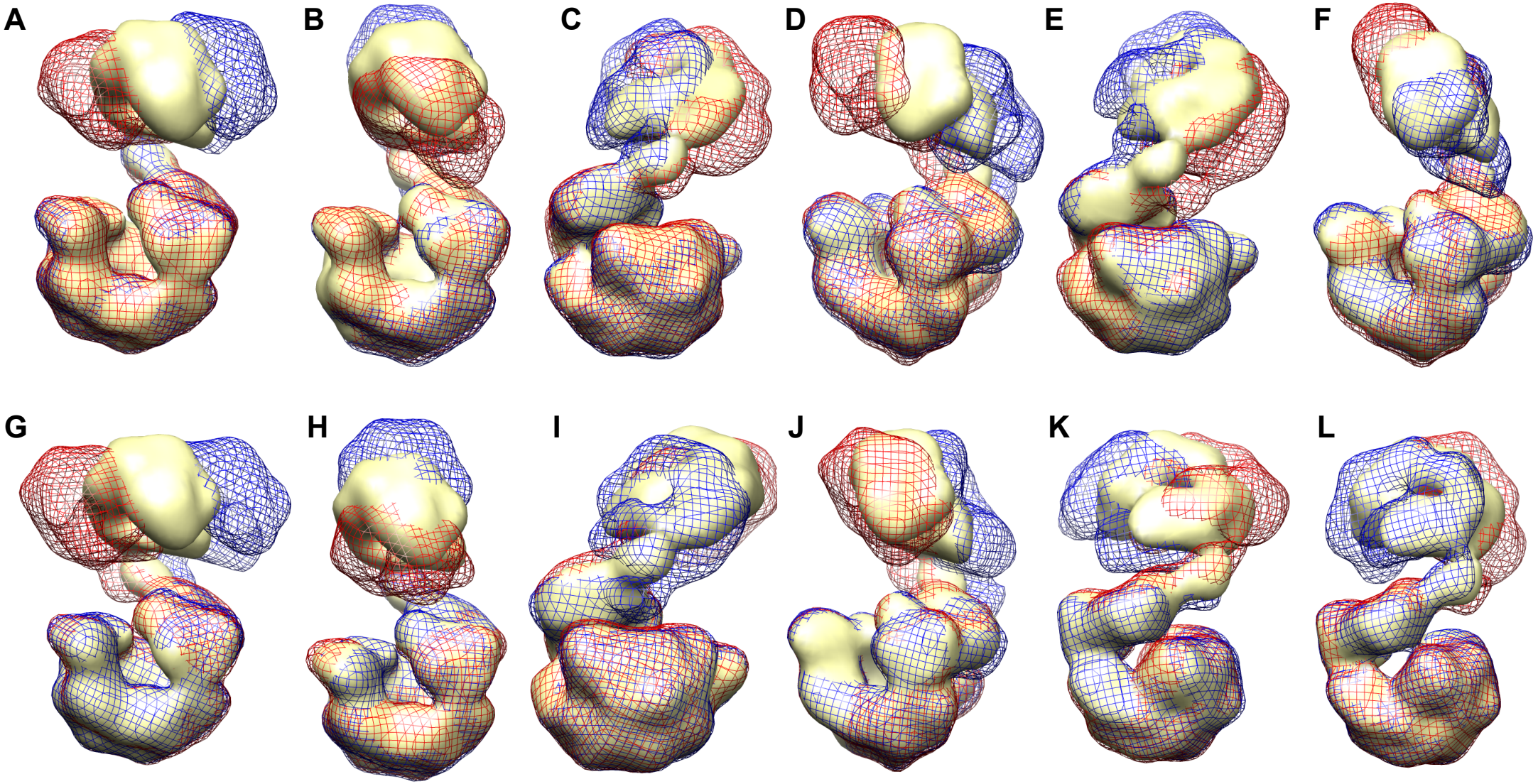

Figure S3

A

	Relative amplitude error along mode for no noise [%]		Relative amplitude error along mode for SNR=0.3 [%]		Relative amplitude error along mode for SNR=0.1 [%]	
	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms
Defocus 0.5 μm	NA	NA	3.4	6.7	4.1	8.8
Defocus 1 μm	NA	NA	3.4	7.3	3.9	8.2
Defocus 2 μm	NA	NA	3.7	8.3	4.2	8.9
No CTF	1.6	4.3	NA	NA	NA	NA

B

	Global absolute angular and shift error for no noise				Global absolute angular and shift error for SNR=0.3				Global absolute angular and shift error for SNR=0.1			
	Orientation [deg]		Position [pix]		Orientation [deg]		Position [pix]		Orientation [deg]		Position [pix]	
	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms	Atoms	Pseudo-atoms
Defocus 0.5 μm	NA	NA	NA	NA	0.30	0.37	0.07	0.08	0.45	0.54	0.11	0.12
Defocus 1 μm	NA	NA	NA	NA	0.32	0.43	0.07	0.09	0.37	0.51	0.08	0.11
Defocus 2 μm	NA	NA	NA	NA	0.38	0.48	0.09	0.10	0.50	0.59	0.12	0.12
No CTF	0.15	0.15	0.03	0.04	NA	NA	NA	NA	NA	NA	NA	NA

C

Volume-to-pseudoatoms conversion			Normal mode analysis		Image analysis (no noise case)		
Volume approximation error	Gaussian-function standard deviation	Number of pseudo-atoms	Pseudo-atoms interaction cutoff distance	30S-rotation mode	Relative deformation-amplitude error along mode [%]	Global absolute angular error [deg]	Global absolute shift error [pix]
0.05	2	5155	22	12	7.0	0.27	0.07
0.05	3	1510	22	13	6.9	0.38	0.09
0.07	2	1246	35	10	24.9	8.69	0.28
0.07	3	450	40	17	27.3	13.65	0.46

Table S1