Visualizing conformational space of functional biomolecular complexes by deep manifold learning

Zhaolong Wu[1,2], Enbo Chen[1], Shuwen Zhang[1], Yinping Ma[1], Congcong Liu[3], Chang-Cheng Yin[3] & Youdong Mao[1,2,4,5]*

[1]State Key Laboratory for Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University, Beijing, China. [2]Center for Quantitative Biology, Peking University, Beijing, China. [3]Department of Biophysics, School of Basic Medical Sciences, Peking University, Beijing 100871, China. [4]Intel Parallel Computing Center for Structural Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [5]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA.

*To whom correspondence may be addressed. E-mail: ymao@pku.edu.cn

Abstract

The cellular functions are executed by biological macromolecular complexes in nonequilibrium dynamic processes, which exhibit a vast diversity of conformational states. Solving conformational continuum of important biomolecular complexes at atomic level is essential to understand their functional mechanisms and to guide structure-based drug discovery. Here we introduce a deep learning framework, named AlphaCryo4D, which enables atomic-level cryogenic electron microscopy reconstructions of conformational continuum. AlphaCryo4D integrates 3D deep residual learning with manifold embedding of energy landscapes, which directs 3D clustering of markedly improved accuracy via an energy-based particle-voting algorithm. By applications of this approach to analyze five experimental datasets, we examine its generality in breaking resolution limit of visualizing dynamic components of functional complexes, in discovering 'invisible' lowly populated intermediates and in exploring their 'hidden' conformational space. Our approach expands the realm of structural ensemble determination to the nonequilibrium regime at atomic level, thus potentially transforming biomedical research and therapeutic development.

Essential cellular functions are executed and regulated by macromolecular complexes comprising many subunits, such as the ribosome, proteasome, replisome and spliceosome. Their structures in functional cycles are often compositionally heterogeneous and conformationally dynamic, involving many reversibly associated components in cells. Visualizing the conformational continuum of highly dynamic megadalton complexes at high resolution is crucial for understanding their functional mechanisms and for guiding structure-based therapeutic development. The current approaches in cryogenic electron microscopy (cryo-EM) structure determination allow atomic-level visualization of major conformational states of these dynamic complexes [1-12]. However, structure determination of transient, nonequilibrium intermediates connecting their major states during their functional cycles has been prohibitory at large.

The problem of structural heterogeneity in cryo-EM structure determination has been investigated with numerous computational approaches including maximum-likelihood-based classification and multivariate statistical analysis [5,6,11,13-18]. As an extension or generalization of multivariate statistical analysis, several machine-learning approaches were recently proposed to estimate a continuous conformational distribution in the latent space [11,12,19-22]. The estimation of continuous conformational distributions does not guarantee an improvement of 3D classification accuracy nor warrant identification of hidden conformational states of biological importance. These methods often trade off the reconstruction resolution for gaining a more complete representation of conformational landscape in the latent space. The insufficient resolution would preclude their potential applications in structure-based drug discovery.

To date, the most widely used method to counteract structural heterogeneity in cryo-EM for resolution gain is the hierarchical maximum-likelihood-based 3D classification due to its ease of application [1]. To improve the resolution of major conformers, low-quality classes are often manually removed during iterations of curated classification, inevitably causing an incomplete representation of the true conformational landscape [23,24]. The outcome of refined cryo-EM maps depends on user expertise of decision-making in class selection, which can be biased by user experience and subjectivity. Importantly, a considerable portion of misclassified images can limit the achievable resolution of observed conformational states and result in missing states that could be biologically important.

In addition to the approaches characterizing global structural heterogeneity, other methods were proposed to analyze local structural dynamics or to improve resolution of local regions in

cryo-EM. These includes the multi-body or focused refinement [4,25,26], normal mode analysis [27,28], flexible refinement [29,30], non-uniform refinement [31] and integrating traditional molecular dynamics (MD) simulations with or without use of deep learning [32,33]. These methods were not designed to recover a complete picture of structural dynamics hidden in cryo-EM datasets.

The energy landscape is a statistical representation of the conformational space of a macromolecular complex and is the basis of the transition-state theory of reaction dynamics [34,35]. The minimum-energy path (MEP) on the energy landscape theoretically represents the most probable trajectory of conformational transitions and can inform the activation energy for chemical reactions [35,36]. Previous studies have demonstrated the benefit of energy landscape estimation in characterizing conformational variation of macromolecules from cryo-EM data [37-43]. The combination of linear dimensionality reduction via 3D principal component analysis (PCA) and energy landscape estimation has been shown to be potentially effective in capturing the overall conformational landscape [38-40]. The benefit of using nonlinear dimensionality reduction and manifold embedding, such as diffusion map [44], to replace PCA for estimating energy landscapes has also been investigated at limited resolution, with the assumption that the conformational changes can be discerned through a narrow angular aperture [37,41,42]. This assumption restricts its potential applications to more complicated conformational dynamics.

Breaking the resolution limit in visualizing the conformational space of highly dynamic complexes represent a major objective of technical advances for leveraging cryo-EM applications in both basic biomedical science and drug discovery. This would require substantial improvement in 3D classification techniques that are optimized for resolution gain. We hypothesize that a high-quality estimation of energy landscape could be potentially used to improve 3D classification of cryo-EM data. If different conformers or continuous conformational changes can be sufficiently mapped and differentiated on the energy landscape, the energetic visualization of conformational continuum could then be used to discover previously missing conformers and achieve higher resolution for lowly populated, transient or nonequilibrium states. To explore these ideas, we developed a novel machine learning framework named AlphaCryo4D that can break the existing limitations and enable 4D cryo-EM reconstruction of highly dynamic, lowly populated intermediates or transient states at atomic level. We examined the general applicability of AlphaCryo4D in analyzing three large synthetic datasets over a wide range of single-to-noise ratios (SNRs) and five experimental cryo-EM datasets of diverse sample

behaviors in complex dynamics, including the mammalian calcium-conducting channel complex RyR1 [45], human 26S proteasome [23], malaria parasite 80S ribosome [46], pre-catalytic spliceosome [47] and bacterial ribosomal assembly intermediates [48]. Our approach pushes the envelope of visualizing conformational space of macromolecular complexes beyond the previously achieved scope toward the atomic level.

Results

Design of deep manifold learning

The conceptual framework of AlphaCryo4D integrates unsupervised deep learning with manifold embedding to learn a free-energy landscape, which directs cryo-EM reconstructions of conformational continuum or transient states via an energy-based particle-voting algorithm. AlphaCryo4D consists of four major steps (Fig. 1A). First, hundreds to thousands of 3D volumes are bootstrapped with *M*-fold particle shuffling and Bayesian clustering. To this end, all curated particles are aligned in a common frame of reference through a consensus 3D refinement. They are then randomly divided into *M*+1 groups of equal data sizes. In the step of '*M*-fold particle shuffling', one group of particles is taken out of the dataset to form a shuffled dataset. This procedure is repeated for $M + 1$ times, each time with a different particle group being left out, resulting in $M + 1$ shuffled datasets (Fig. S1). Each shuffled dataset is subject to 3D volume bootstrapping separately and is clustered into tens to hundreds of 3D reconstructions through Bayesian clustering in RELION [3,7]. In total, hundreds to thousands of volumes from all shuffled datasets are expected to be bootstrapped through these steps.

Next, all bootstrapped volumes are learned by a 3D autoencoder made of a deep residual convolutional neural network in an unsupervised fashion [49,50] (Fig. 1B, Table S1). A 3D feature map is extracted for each volume by the deep residual autoencoder and is juxtaposed with the volume data for nonlinear dimensionality reduction by manifold embedding with the t-distributed stochastic neighbor embedding (t-SNE) algorithm [51] (Fig. 1D, E).

Third, the learned manifold is used to estimate an energy landscape via the Boltzmann relation using the particle density on the manifold [37]. A string method can be optionally used to search the MEP on the energy landscape [52,53]. The local energy minima or transition states connecting adjacent minimum-energy states can be defined as the centers of 3D clustering, with a circular range defined as the cluster boundary for subsequent particle voting (Fig. 1F).

Last, because each particle is used *M* times during volume bootstrapping, it is mapped to *M* locations on the energy landscape. The mapping of each copy of the particle is called a 'vote'. By counting the number (*K*) of votes of the same particle casted within the same cluster boundary on the energy landscape, the reproducibility of deep manifold learning can be evaluated at single-particle level. Each particle is classified to the 3D cluster that receives more than *M*/2 votes of this particle within its voting boundary (Fig. 1F, Fig. S1B). If none of the 3D clusters on the energy landscape receives more than *M*/2 votes of a given particle, the corresponding particle is voted out and excluded for further 3D reconstruction. The resulting 3D classes are expected to be conformationally homogeneous enough for high-resolution cryo-EM refinement.

Since many protein complexes exhibit profound conformational changes in different local regions, we also implemented a focused classification strategy of AlphaCryo4D that applies a local 3D mask [6] throughout the entire procedure, often being executed as an iterative step after initial 3D classification by AlphaCryo4D in the absence of any 3D mask (Fig. 1A, Methods).

Solving conformational continuum at atomic level

To assess the numerical performance of AlphaCryo4D, we generated three large synthetic heterogeneous cryo-EM datasets with signal-to-noise ratios (SNRs) of 0.05, 0.01 and 0.005. Each dataset includes 2 million randomly oriented single particles computationally simulated from 20 hypothetical conformer models of the ~130-kDa NLRP3 inflammasome protein [54]. These conformers imitate conformational continuum of the NACHT domain rotating around the LRR domain over an angular range of 90° during inflammasome activation [54] (Fig. 1C). We conducted blind assessments on 3D classification and heterogeneous reconstructions by AlphaCryo4D, without providing any information of particle orientations, translations and conformational identities (Fig. 1D-F, Fig. S2, Methods). The 3D classification precision of a retrieved conformer was computed as the ratio of the particle number of correct class assignment (based on the ground truth) versus the total particle number in the class. The results were then compared with several alternative methods, including conventional maximum-likelihood-based 3D (ML3D) classification in RELION [3,4,6], 3D variability analysis (3DVA) in cryoSPARC [2,11], and deep generative model-based cryoDRGN [12]. In all blind tests by presetting the class number to 20, AlphaCryo4D retrieved all 20 conformers and markedly outperformed the alternative

methods, with an average of 3D classification precision at 0.83, 0.82 and 0.65 for datasets with SNRs of 0.05, 0.01 and 0.005, respectively (Fig. 2A, Fig. S3A, S4). By contrast, all alternative methods missed two to seven conformers entirely and exhibited 3D classification precisions in the range of 0.2-0.5 in general (Fig. 2B-D, Figs. S3B-J). By increasing the preset class numbers to 25 or 30 that is more than the number of ground-truth conformers, all tested methods appear to be improved in the precision of 3D classification marginally but also reduced in the recall of classification (defined as the percentage of correctly assigned particles versus the number of ground-truth particles for a conformer) (Fig. S5). In these cases, AlphaCryo4D still outperformed all alternative methods considerably, with the highest average classification precision reaching 0.91 at the SNR of 0.05.

The 3D classification precision appears to be strongly correlated with the map quality and resolution homogeneity across the density map (Fig. 2, Fig. S2H). All density maps from AlphaCryo4D consistently show homogeneous local resolutions (at 2.6-2.9 Å for SNR of 0.01) between the NACHT and LRR domains (Fig. 2E, I, Figs. S2H, 3K, 4A). By contrast, all density maps by the alternative methods show lower average resolutions and notably heterogeneous local resolutions, with the NACHT domain exhibiting substantially lower resolution than that of the LRR domain, causing blurred features, broken loops and invisible sidechains in NACHT (Fig. 2F-H, J-L, Figs. S3L-N, 4B-D). Thus, the significantly improved 3D classification accuracy by AlphaCryo4D enables 4D reconstructions of conformational continuum at the atomic level.

To understand how the 3D classification accuracy is improved, we further studied the algorithmic mechanism by removing or replacing certain components of AlphaCryo4D in 24 control tests (Figs. S2D-G, 6, see Methods for details). By removing the component of deep manifold learning from AlphaCryo4D, the statistical distributions of high-precision 3D classes after particle voting are reduced by ~60% (Fig. S6D-F). Similarly, classifying particles without particle voting or in the absence of deep residual autoencoder reduces the high-precision 3D classes by ~15-30% at a lower SNR (Fig. S6M-R). Further, replacing t-SNE with other four dimensionality reduction techniques resulted in the loss of at least half of ground-truth conformers in 3D classification, including PCA, multidimensional scaling (MDS) [55], isometric mapping [56] and locally linear embedding [57] (Figs. S2D-G). Throughout AlphaCryo4D, the steps of particle shuffling, defining voting boundaries on energy landscapes and energy-based particle voting contribute to ~16%, ~20% and ~40% improvements of high-precision 3D classes,

respectively (Fig. S6G-L, S-U). Taken together, these results indicate that all components of AlphaCryo4D contribute to the improved accuracy of 3D classification.

Breaking the resolution limit of the dynamic domains of RyR1

Having conducted the proof-of-principle study of AlphaCryo4D using the simulated datasets, we then turned to five experimental datasets. We first applied AlphaCryo4D to solve the dynamic domains of the 2.2-MDa ryanodine receptor (RyR), a tetrameric calcium-activated channel protein complex that mediates the release of calcium ions from intracellular stores. The mammalian RyR1 has been determined at near-atomic resolution in its central core in multiple conformations [58,59], which was recently further improved to sub-3 Å resolution [45]. However, approximately half of the entire RyR1 structure located at peripheral domains remain at moderate to low resolution due to their complex dynamics that precludes atomic modeling with reliable amino acids assignment. Extensive previous efforts have failed in further improving the resolution of the peripheral domains by the conventional methods [45,58,59]. Thus, we applied AlphaCryo4D to analyze the RyR1 dataset [45] to test if our approach can overcome the conformational heterogeneity and break the resolution barrier. We computed a focused energy landscape using 310 bootstrapped density maps to characterize the conformational distribution of the peripheral domains (see Methods). The energy landscape reveals three new sub-states from the original dataset corresponding to state 1 conformation [45] (Fig. 3A). The peripheral domains of the three states vibrate vertically against the central domains with a very small amplitude (Fig. 3C), which is sufficient to blur the local density when the entire particles are used for alignment.

Using the AlphaCryo4D-classified particles for refinement, the local resolution of the peripheral domains was refined to 3.3 Å, representing an unprecedented improvement over the previous resolution of 5-8 Å for nearly half of RyR1. The resulting density map showed high-resolution features including helical pitches and side-chain densities (Fig. 3D, E). However, the small P2/RY3&4 domain abutting the helical C-Sol domain via flexible links still has low resolution in the refined map. To improve the visualization of P2, we computed the zoomed-in energy landscape by focusing on the P2 domain. This energy landscape also exhibited three conformational clusters (Fig. 3B), from which we found the P2 domain oscillating relative to the helical C-Sol domain. Further refinement of the AlphaCry4D-classified particles improved the clarity of the secondary structure in the density map, but it did not reach the atomic level likely

due to the very small size of this domain limiting the alignment accuracy (Fig. S7). In summary, the application to RyR1 demonstrates that AlphaCryo4D is capable of breaking the resolution barrier of the dynamic domain by resolving hierarchically entangled conformational dynamics of the RyR1 peripheral domains.

Visualizing hidden dynamics of the 26S proteasome

The 26S proteasome is one of the most complex, dynamic, heterogeneous holoenzyme machineries that regulates nearly all aspects of cell physiology in eukaryotes [23,60-63]. Visualizing atomic structures of nonequilibrium intermediates connecting the major states of the functional proteasome in the act of substrate degradation has been unsuccessful [4,6,7,23,60]. To test this possibility, we applied AlphaCryo4D to analyze a large experimental dataset of substrate-engaged human 26S proteasome, which comprises more than 3 million particles. AlphaCryo4D was able to significantly enrich the number of visualized proteasome conformers to 64 using this dataset, which was previously used to determine atomic structures of seven proteasome states [23]. As the majority of the states were refined to near-atomic resolution, they allowed us to obtain a more complete picture of the functional dynamics of the proteasome, thus providing unprecedented insights into the molecular mechanism of proteasome autoregulation at the atomic level. Several important biological findings and the corresponding analysis are described in detail elsewhere in a companion paper [24].

To demonstrate that AlpahCryo4D can simultaneously improve 3D classification and resolution of key dynamic features, we highlight a typical case among numerous improvements of solving the proteasome dynamics by AlphaCryo4D. Here we focus on state E_{A2} of substrate-engaged 26S proteasome, which is the initial state that a substrate-conjugated ubiquitin moiety binds the deubiquitylating enzyme RPN11 to prepare for substrate engagement and deubiquitylation. In this case, we expected to use AlphaCryo4D to detect if an intermediate substrate binding to RPN11 proceeds the deubiquitylation step of proteasome [23]. Previous work on state E_{A2} has showed a moderate resolution of RPN11-bound ubiquitin and no substrate was observed on RPN11 [23]. On the energy landscape of sub-dataset corresponding to state E_{A2}, we examined two conformational clusters, designated $E_{A2.0}$ and $E_{A2.1}$, which were refined to 3.5 and 3.2 Å, respectively (Fig. 4A). Comparing these two density maps reveals subtle conformation changes of the lid subcomplex against the base, although both show no substrate binding in the

central channel of hexameric AAA-ATPase ring (Fig. 4B). Markedly, state $E_{A2.1}$ exhibits a segment of substrate bound to the cryptic hydrophobic groove at the interface between RPN11 and RPT4 near the active site of RPN11, whereas no substrate is observed at the same site in state $E_{A2.0}$ (Fig. 4C). This polypeptide substrate segment has a length of about 4 amino acids, which is comparable in size to many small-molecule compounds, suggesting that AlphaCryo4D may be capable of classifying cryo-EM data for very small features relevant to drug discovery.

The conformation of state $E_{A2.0}$ appears identical to that of the previously published state E_{A2}. The RPN11-bound, 8.5-kDa ubiquitin density in state E_{A2} is around 6-9 Å, although the overall resolution of state E_{A2} was measured to 3.3 Å. Only half particles previously classified to E_{A2} were also classified to $E_{A2.1}$, indicating the mixture of different conformers that caused the low-resolution feature of ubiquitin. Notably, this RPN11-bound ubiquitin in $E_{A2.1}$ clearly shows higher resolution features consistent with 3.8 Å, with sufficiently separated β-strands (Fig. 4D). It indicates that the RPT5 N-loop, C-terminal strand of ubiquitin and insert-1 hairpin already form a four-stranded β-sheet prior to substrate insertion into the proteasomal AAA-ATPase channel. These features were only sufficiently observed in state E_B previously, where the substrate has completed its insertion into the AAA-ATPase ring [23]. These high-resolution features allow us to confidently interpret state $E_{A2.1}$ as a transient intermediate between previously published states E_{A2} and E_B.

Revealing residual conformational heterogeneity in the *Pf*80S ribosome

We further evaluated the capability of AlphaCryo4D in detecting residual heterogeneity in a mostly homogeneous dataset of the Plasmodium falciparum 80S (*Pf*80S) ribosome bound to anti-protozoan drug emetine (EMPIAR-10028) [46]. The original cryo-EM analysis of the *Pf*80S ribosome dataset (105,247 particles) determined a major conformational states and detected flexibility in the head region of the 40S small subunit, which was originally unsolved [46]. Application of AlphaCryo4D to analyze this dataset reveals notable rotation of 40S in the *Pf*80S ribosome. By reconstitution of the approximately continuous energy landscape using 79 density maps, we simultaneously identified two 40S-rotated states (R3 and R4) and another state (R2) with missing 40S head in addition to the original major conformational state (R1) located in the deepest energy well (Fig. 5A, B). Notably, AlphaCryo4D was able to maintain the original resolution of the major state at 3.2 Å, while allowing the three new states R2, R3 and R4 to be

refined to 3.6, 4.1 and 4.6 Å, respectively, without using more data (Fig. S8).

On the energy landscape, the distance between states R4 and R1 is longer than that between states R3 and R1, indicating that a larger conformational change occurs in state R4. Indeed, superimposing the density maps of states R3 and R1 together reveals that the rotation of 40S in state R3 is localized within its head region with concerted movement in the central protuberance of 60S, while the rest of the 40S exhibits no obvious movement (Fig. 5F). By contrast, the entire 40S rotates above the 60S in state R4 relative to R1, and this coincides with the intrinsic structural rearrangement within 60S near the central protuberance and L1 stalk, which appears to also drive the 40S translation relative to the 60S (Fig. 5G), indicating a different mode of intersubunit rotation. Moreover, the reconstructions of R3 and R4 also identified differential rearrangements of small structural elements, including the stepwise disappearance of an rRNA helix and of the intersubunit bridge formed by eL8 C-terminal helix, in line with previous studies on the *Pf*80S dynamics during translation [64] (Fig. 5C-E). Our analysis suggests that the superflically simple rotation of the 40S subunit is a manifestation of complicated dynamics involving the entire *Pf*80S ribosome. Given that the *Pf*80S ribosome was bound to the anti-protozoan drug, the newly observed conformational states at high resolution open the door to investigate the drug-modulated allosteric effect relevant to therapeutic improvement [46].

Visualizing conformational continuum of the pre-catalytic spliceosome

We next evaluated the applicability of AlphaCryo4D in simultaneously visualizing continuous motion and discrete compositional heterogeneity in the pre-catalytic spliceosome (EMPIAR-10180) [47]. Previous study using focused classification has provided a multistate-averaged map of the pre-catalytic spliceosome and suggested that the SF3b subcomplex may sample a continuum of conformations. To characterize the pre-catalytic spliceosome dynamics, we bootstrapped 160 density maps using 327,490 particles to reconstitute its energy landscape (Fig. 6A), which allowed us to determine many intermediate states (S1-S9) representative of a continuous motion of the SF3b and helicase subcomplexes as well as several compositionally distinct states (S10-S14). On the energy landscape, the nine states S1-S9 representing the continuous motions are intuitively located in adjacent clusters, forming a major MEP traversing the long axis of the energy landscape (Fig. 6C). These states show both independent movements of the SF3b and helicase, as well as concerted motion of the two subcomplexes. For instance, comparison

between states S3 and S4 exhibits that the SF3b and helicase move away from each other, whereas states S5 and S6 show co-directional motion of both the SF3b and helicase. Interestingly, the Cus1 subunit in the SF3b subcomplex appears to switch its contact between the N-terminal helicase cassettes of RecA-1 and RecA-2, implicating that part of conformational rearrangement is mediated by the transient interface B between the SF3b and helicase.

In addition to mapping the continuous inter-subunit motion, AlphaCryo4D also reveals potential allosteric effect of local structural variation relevant to spliceosome activation. Notably, comparison between states S4 and S13 exhibits that the U2 snRNP and helicase motions among these states appear to be allosterically coupled to the disappearance of the Spp381, Prp38 and Snu23 densities in state S13 (Fig. 6B). As the proteins Spp381, Prp38 and Snu23 are required for spliceosome activation [47], these coexisting conformers may be important to understand the intermediate steps of spliceosome activation. These biologically important dynamic features have not been previously reported from the same dataset. Taken together, our analysis demonstrates that AlpahCryo4D is capable of characterizing continuous conformational changes on a dataset with a moderate size.

Discovery of hidden assembly intermediates of bacterial ribosome

In the last test, we applied AlphaCryo4D to analyze a highly heterogenous dataset of the *Escherichia coli* 50S large ribosomal subunit (LSU) (EMPIAR-10076) [48]. This dataset (131,899 particles) is known to be both compositionally heterogeneous and conformationally dynamic in that the 50S LSU undergoes bL17-depleted intermediate assembly. We first computed the overall energy landscape of the ribosomal assembly using 119 density maps (Fig. 7B). On this energy landscape, the originally reported five major states A-E were straightforwardly reproduced by energy-based particle voting. Obviously, state A containing both 50S and 30S subcomplexes is a rare state distinct from other conformers (Fig. 7A). The density map of state A reconstructed by only 766 particles still retains enough details, benefiting from the strategy of energy-based particle voting.

As the remaining particle numbers of states B, C, D and E are all adequate for deeper classification, we then computed their zoomed-in energy landscapes separately. This allowed us to discover eight new conformational states (designated B1, B2, C5, D5, E6, E7, E8) that have evaded all other methods [11,12,19,48] (Fig. 7D-K), as well as all 13 previously reported sub-class

conformers classified from this dataset [12,48] (Fig. S10). On the zoomed-in energy landscape of state B, state B1 exhibits a low occupancy of domain III and is perhaps a new least mature LSU intermediate on the assembly pathway (Fig. 7E). The rRNA helix 101 exhibits small rotation between states B1 and B2 (Fig. 7E and Fig. S10B). The differential occupancy of rRNA helix 63 and the movement of uL1 stalk are observed in states C5 and E1-E8 (Fig. 7G, I, K). Among the sub-states derived from state E, the new state (E7) with uL1 missing was reconstructed by only 357 particles (Fig. 7K), which was another rare state discovered via AlphaCryo4D that is only 0.3% of the entire dataset. These results suggest that AlphaCryo4D is highly efficient in finding transient states of extremely low populations and in exploring the 'dark' conformers in the conformational space of highly heterogeneous systems.

Discussion

One primary objective of the AlphaCryo4D development is to simultaneously improve reconstruction resolution and 3D classification without assumption on the behavior of sample dynamics and heterogeneity. A key lesson from this study is that the improved accuracy of 3D classification is often positively correlated with the resolution gain in general, as misclassification is presumably a key factor limiting achievable resolution. Because of extremely low SNR, most objective functions in image similarity measurement and machine learning algorithms tend to poorly perform or entirely fail on cryo-EM data. To date, no methods have included any implicit metrics or strategies for validating 3D classification accuracy in cryo-EM. We address this issue by introducing *M*-fold particle shuffling method and the energy-based particle voting algorithm in AlphaCryo4D that automatically checks the reproducibility and robustness of the 3D classification. This allows users to objectively reject non-reproducible particles and permits a maximal number of particles to be assessed and classified in an integrative procedure based on uniform, objective criteria. Our extensive tests suggest that the particle-voting algorithm synergistically enhances the performance of deep manifold learning. Moreover, it can potentially rescue certain particles that are prone to be misclassified if processed only once by deep manifold learning, thus boosting the efficiency of particle usage without necessarily sacrificing quality and homogeneity of selected particles. This mechanism simultaneously optimizes the usage of all available particles and their 3D classification accuracy for achieving higher resolution.

Our particle-shuffling and Bayesian bootstrap strategy is conceptually related to the stochastic bootstrap method previously proposed for estimating 3D variance map [65] or performing 3D PCA [13] to detect conformational variability. One essential difference lies in that our bootstrap requires that only particles of similar or identical conformations via likelihood-based similarity estimation are grouped together instead of being combined or resampled in a stochastic manner [1,18]. Another difference is that the goal of *M*-fold particle shuffling is to improve the quality of reconstituted energy landscape and to enable energy-based particle voting for intrinsic cross-validation of 3D classification.

Using energy landscape to represent the conformational variation provides several advantages. First, it has theoretical roots in physical chemistry and protein dynamics, in which extensive research tools being built around the energy landscape and transition-state theory allows correlation with other complementary approaches, such as single-molecule florescence microscopy, NMR and molecular dynamics simulation. Second, the energy landscape allows users to intuitively examine kinetic relationships between the adjacent conformers and to potentially discover new conformations. Our approach to use many bootstrapped density maps to estimate the ideologically continuous energy landscape is statistically analogous to other methods proposed to map single particles to continuous distributions in the latent space [11,12,19]. The latter may suffer from the lack of energetic information that can restrict the interpretation of the resulting reconstruction. Third, while our method of mapping 3D volumes to manifolds increases the sampling grain in approximating the continuity of conformational distribution as compared to other methods directly mapping single particles to the latent space [11,12,19], AlphaCryo4D trades off the fineness of sampling grain in the continuity approximation for resolution gain. In nearly all cases we examined, AlphaCryo4D apparently still oversampled the conformational space at large in terms of finding those conformational states that can potentially achieve high resolution.

AlphaCryo4D does not assume any prior knowledge regarding the conformational landscape, its continuity and topology, as well as chemical composition of the macromolecular system. Our approach allows multiple consensus models being used for optimizing the particle alignment prior to the machine-learning-based reconstitution of energy landscape and 3D classification. This can theoretically deal with more complicated conformational dynamics when no stable core structure is available to guide the consensus alignment over a single model.

The limitation of AlphaCryo4D lies in that it requires generally larger datasets and more computational costs to fully exploit its advantages and potentials. The data size must proportionally scale with and match the degree of conformational heterogeneity in order to visualize a full spectrum of conformational states at the atomic level. Moreover, its outcomes depend on the success of initial consensus alignment of all particles during particle shuffling and volume bootstrapping. Certain conformational dynamics or heterogeneity can interfere with image alignment, in which case the performance of AlpahCryo4D will be restricted by the alignment errors in during consensus refinement. This is a common problem for all existing methods developed so far. Our tentative solution is to use multiple consensus models for improving image alignment. Failure in obtaining accurate alignment parameters can lead to futile, erroneous estimation of conformational heterogeneity in the subsequent steps no matter which methods are used.

The use of t-SNE to reconstitute the energy landscape poses both potential advantages and caveats. The t-SNE algorithm can well preserve the local topology but is not warranted to preserve global topology, which is nonetheless a common limitation for existing manifold learning techniques. This issue is partly alleviated by the particle-voting algorithm that is performed at the local areas of energy minima on the energy landscape. Dimensionality reduction to 2D by t-SNE, as well as MEP searching on a 2D energy landscape might also limit its capability in disentangling complex dynamics. Further investigations are required to understand how to map individual particles to a continuous energy landscape without trading off reconstruction resolution, to estimate the energy landscape at higher dimensions, to preserve its global topology with high fidelity and to fully automatize MEP solution.

References

1 Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**, 406-418, doi:10.1016/j.jmb.2011.11.010 (2012).

2 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

3 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).

4 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, doi:10.7554/eLife.36861 (2018).

5 Wu, J. *et al.* Massively parallel unsupervised single-particle cryo-EM data clustering via

statistical manifold learning. *PLoS One* **12**, e0182130, doi:10.1371/journal.pone.0182130 (2017).
6 Scheres, S. H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157, doi:10.1016/bs.mie.2016.04.012 (2016).
7 Scheres, S. H. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat Methods* **4**, 27-29, doi:10.1038/nmeth992 (2007).
8 Galves, M., Rathi, R., Prag, G. & Ashkenazi, A. Ubiquitin Signaling and Degradation of Aggregate-Prone Proteins. *Trends Biochem Sci*, doi:10.1016/j.tibs.2019.04.007 (2019).
9 Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152-156, doi:10.1038/s41586-020-2829-0 (2020).
10 Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157-161, doi:10.1038/s41586-020-2833-4 (2020).
11 Punjani, A. & Fleet, D. J. 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*, 107702, doi:10.1016/j.jsb.2021.107702 (2021).
12 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185, doi:10.1038/s41592-020-01049-4 (2021).
13 Penczek, P. A., Kimmel, M. & Spahn, C. M. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure* **19**, 1582-1590, doi:10.1016/j.str.2011.10.003 (2011).
14 Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H. & Rao, M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *J Struct Biol* **191**, 245-262, doi:10.1016/j.jsb.2015.05.007 (2015).
15 Sanchez Sorzano, C. O., Alvarez-Cabrera, A. L., Kazemi, M., Carazo, J. M. & Jonic, S. StructMap: Elastic Distance Analysis of Electron Microscopy Maps for Studying Conformational Changes. *Biophys J* **110**, 1753-1765, doi:10.1016/j.bpj.2016.03.019 (2016).
16 Katsevich, E., Katsevich, A. & Singer, A. Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. *SIAM J Imaging Sci* **8**, 126-185, doi:10.1137/130935434 (2015).
17 Anden, J., Katsevich, E. & Singer, A. Covariance Estimation Using Conjugate Gradient for 3d Classification in Cryo-Em. *Proc IEEE Int Symp Biomed Imaging* **2015**, 200-204, doi:10.1109/ISBI.2015.7163849 (2015).
18 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* **183**, 377-388, doi:10.1016/j.jsb.2013.07.005 (2013).
19 Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat Methods* **18**, 930-936, doi:10.1038/s41592-021-01220-5 (2021).
20 Gupta, H., Phan, T. H., Yoo, J. & Unser, M. 429-444 (Springer International Publishing).
21 Moscovich, A., Halevi, A., Anden, J. & Singer, A. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Probl* **36**, doi:10.1088/1361-6420/ab4f55 (2020).

22 Lederman, R. R., Andén, J. & Singer, A. Hyper-molecules: on the representation and recovery of dynamical structures for applications in flexible macro-molecules in cryo-EM. *Inverse Problems* **36**, 044005, doi:10.1088/1361-6420/ab5ede (2020).
23 Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49-55, doi:10.1038/s41586-018-0736-4 (2019).
24 Wu, Z. *et al.* Deep manifold learning reveals hidden dynamics of proteasome autoregulation. *bioRxiv*, 2020.2012.2022.423932, doi:10.1101/2020.12.22.423932 (2021).
25 Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, doi:10.7554/eLife.11182 (2015).
26 Ilca, S. L. *et al.* Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat Commun* **6**, 8843, doi:10.1038/ncomms9843 (2015).
27 Jonic, S. & Sorzano, C. O. Versatility of Approximating Single-Particle Electron Microscopy Density Maps Using Pseudoatoms and Approximation-Accuracy Control. *Biomed Res Int* **2016**, 7060348, doi:10.1155/2016/7060348 (2016).
28 Sorzano, C. O. S. *et al.* Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallogr F Struct Biol Commun* **75**, 19-32, doi:10.1107/S2053230X18015108 (2019).
29 Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209, doi:10.1038/nature24282 (2017).
30 Punjani, A. & Fleet, D. J. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *bioRxiv*, 2021.2004.2022.440893, doi:10.1101/2021.04.22.440893 (2021).
31 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods* **17**, 1214-1221, doi:10.1038/s41592-020-00990-8 (2020).
32 Bonomi, M., Pellarin, R. & Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy. *Biophys J* **114**, 1604-1613, doi:10.1016/j.bpj.2018.02.028 (2018).
33 Matsumoto, S. *et al.* Extraction of protein dynamics information from cryo-EM maps using deep learning. *Nature Machine Intelligence* **3**, 153-160, doi:10.1038/s42256-020-00290-y (2021).
34 Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545-600, doi:10.1146/annurev.physchem.48.1.545 (1997).
35 Henriksen, N. E. & Hansen, F. Y. *Theories of Molecular Reaction Dynamics: The Microscopic Foundation of Chemical Kinetics*. 2nd edn, (Oxford University Press, Oxford, 2019).
36 Atkins, P., de Paula, J. & Keeler, J. *Physical Chemistry*. 11th edn, (Oxford University Press, Oxford, 2018).
37 Dashti, A. *et al.* Trajectories of the ribosome as a Brownian nanomachine. *Proc Natl Acad Sci U S A* **111**, 17492-17497, doi:10.1073/pnas.1419276111 (2014).
38 Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat Commun* **8**, 15578,

doi:10.1038/ncomms15578 (2017).
39 Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329-333, doi:10.1038/nature09206 (2010).
40 Haselbach, D. *et al.* Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. *Cell* **172**, 454-464 e411, doi:10.1016/j.cell.2018.01.010 (2018).
41 Dashti, A. *et al.* Retrieving functional pathways of biomolecules from single-particle snapshots. *Nat Commun* **11**, 4734, doi:10.1038/s41467-020-18403-x (2020).
42 Frank, J. & Ourmazd, A. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61-67, doi:10.1016/j.ymeth.2016.02.007 (2016).
43 Maji, S. *et al.* Propagation of Conformational Coordinates Across Angular Space in Mapping the Continuum of States from Cryo-EM Data by Manifold Embedding. *J Chem Inf Model* **60**, 2484-2491, doi:10.1021/acs.jcim.9b01115 (2020).
44 Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci U S A* **102**, 7426-7431, doi:10.1073/pnas.0500334102 (2005).
45 Wu, Z. *et al.* Chemistry of cation hydration and conduction in a skeletal muscle ryanodine receptor. *bioRxiv*, 732172, doi:10.1101/732172 (2019).
46 Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **3**, doi:10.7554/eLife.03080 (2014).
47 Plaschka, C., Lin, P. C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617-621, doi:10.1038/nature22799 (2017).
48 Davis, J. H. *et al.* Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell* **167**, 1610-1622 e1615, doi:10.1016/j.cell.2016.11.020 (2016).
49 He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 770-778, doi:10.1109/CVPR.2016.90 (2016).
50 Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507, doi:10.1126/science.1127647 (2006).
51 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).
52 E, W., Ren, W. & Vanden-Eijnden, E. String method for the study of rare events. *Phys Rev B* **66**, 052301 (2002).
53 E, W., Ren, W. & Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J Chem Phys* **126**, 164103, doi:10.1063/1.2720838 (2007).
54 Sharif, H. *et al.* Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome. *Nature* **570**, 338-343, doi:10.1038/s41586-019-1295-z (2019).
55 Cox, T. & Cox, M. *Multidimensional Scaling*. (Chapman & Hall, London, 1994).
56 Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319-2323, doi:10.1126/science.290.5500.2319 (2000).
57 Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323-2326, doi:10.1126/science.290.5500.2323 (2000).
58 Yan, Z. *et al.* Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution.

Nature **517**, 50-55, doi:10.1038/nature14063 (2015).
59 des Georges, A. *et al.* Structural Basis for Gating and Activation of RyR1. *Cell* **167**, 145-157 e117, doi:10.1016/j.cell.2016.08.075 (2016).
60 Mao, Y. Structure, Dynamics and Function of the 26S Proteasome. In: Harris J.R., Marles-Wright J. (eds) Macromolecular Protein Complexes III: Structure and Function. *Subcell Biochem* **96**, 1-151, doi:10.1007/978-3-030-58971-4_1 (2021).
61 Zhang, S. & Mao, Y. AAA+ ATPases in Protein Degradation: Structures, Functions and Mechanisms. *Biomolecules* **10**, doi:10.3390/biom10040629 (2020).
62 Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc Natl Acad Sci U S A* **113**, 12991-12996, doi:10.1073/pnas.1614614113 (2016).
63 Zhu, Y. *et al.* Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nat Commun* **9**, 1360, doi:10.1038/s41467-018-03785-w (2018).
64 Sun, M. *et al.* Dynamical features of the Plasmodium falciparum ribosome during translation. *Nucleic Acids Res* **43**, 10515-10524, doi:10.1093/nar/gkv991 (2015).
65 Penczek, P. A., Yang, C., Frank, J. & Spahn, C. M. Estimation of variance in single-particle reconstruction using the bootstrap technique. *J Struct Biol* **154**, 168-183, doi:10.1016/j.jsb.2006.01.003 (2006).
66 Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics* **18**, 348, doi:10.1186/s12859-017-1757-y (2017).
67 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).
68 Ioffe, S. & Szegedy, C. in *Proceedings of the 32nd International Conference on Machine Learning* Vol. 37 (eds Bach Francis & Blei David) 448--456 (PMLR, Proceedings of Machine Learning Research, 2015).
69 Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Netw* **12**, 145-151, doi:10.1016/s0893-6080(98)00116-6 (1999).
70 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).

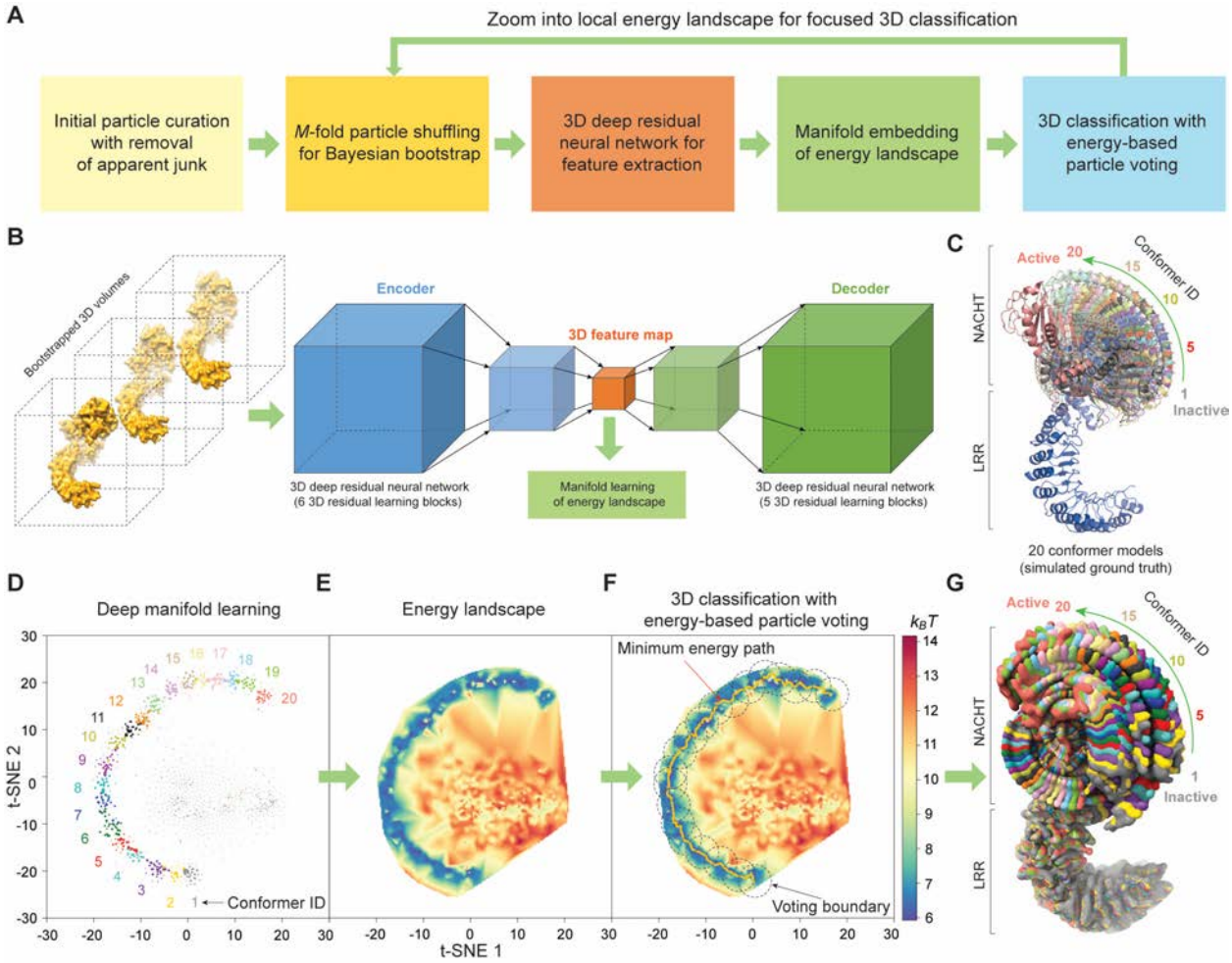

Figure 1. Conceptual framework of AlphaCryo4D for 4D cryo-EM reconstruction. (**A**) Schematic showing the major conceptual steps of single-particle cryo-EM data processing in AlphaCryo4D. (**B**) Illustration of deep residual learning of 3D feature maps by an autoencoder conjugated to a decoder for unsupervised training. (**C**) The 20 atomic models of hypothetical conformers of NLRP3 in cartoon representations simulate a 90° rotation of the NACHT domain relative to the LRR domain based on the structure of NLRP3 (PDB ID: 6NPY). (**D**) Manifold learning of the bootstrapped 3D volumes and their feature maps learnt by the autoencoder. Each data point corresponds to a 3D volume. The color labels the conformer identity of ground truth for the purpose of verification. **(E**) Energy landscape computed from the manifold shown in (**D**) using the Boltzmann distribution. (**F**) Minimum energy path (orange line) calculated by the string method is used to find the approximate cluster centers of 20 conformers. The cluster boundaries for energy-based particle voting are shown as dashed circles. (**G**) The 20 reconstructions resulting from the blind assessment of AlphaCryo4D using simulated dataset with SNR of 0.01 are superimposed together, shown in surface representation, and aligned against its LRR domain.

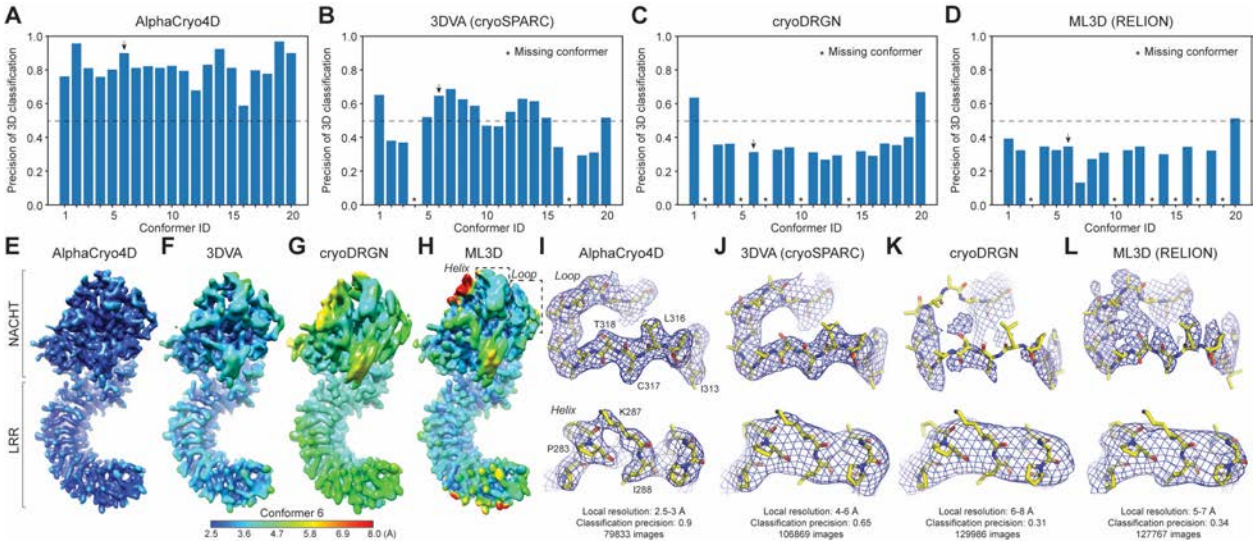

Figure 2. Performance evaluation of AlphaCryo4D for reconstructions of conformational continuum at the atomic level. (**A-C**) Plots of 3D classification precisions of the 20 NLRP3 conformers from blind assessments on the simulated NLRP3 dataset with SNR of 0.01, using AlphaCyo4D (**A**), 3DVA algorithm in cryoSPARC[11] (**B**), cryoDRGN[12] (**C**), and conventional ML3D in RELION[3] (**D**). All 3D conformers in panel (A) were reconstructed to 2.6-2.9 Å resolution (Figure S3M). Asterisks mark the missing conformers that were completely lost due to misclassification by 3DVA, cryoDRGN and ML3D. (**E-H**) Typical side-by-side comparison of density map quality and local resolution of the same conformer (ID 6) reconstructed by AlphaCryo4D (**E**), 3DVA (**F**), cryoDRGN (**G**) and ML3D (**H**). The maps are colored by their local resolutions calculated by Bsoft blocres program. (**I-L**) Closeup side-by-side comparison of the same two secondary structures, including a loop (upper row) and a helix (lower row), in the NACHT domain illustrates considerable improvements in local density quality and resolution by AlphaCryo4D (**I**) as opposed to 3DVA/cryoSPARC (**J**), cryoDRGN (**K**) and ML3D/RELION (**L**). The locations of the loop and helix in the NLRP3 structure are marked by dashed boxes in panel (**H**). The same ground-truth atomic model of Conformer 6 shown in stick representations is superimposed with the density maps shown in blue mesh representations, all from the same perspective. The atomic model was not further refined against each map for visual validation of the map accuracy.

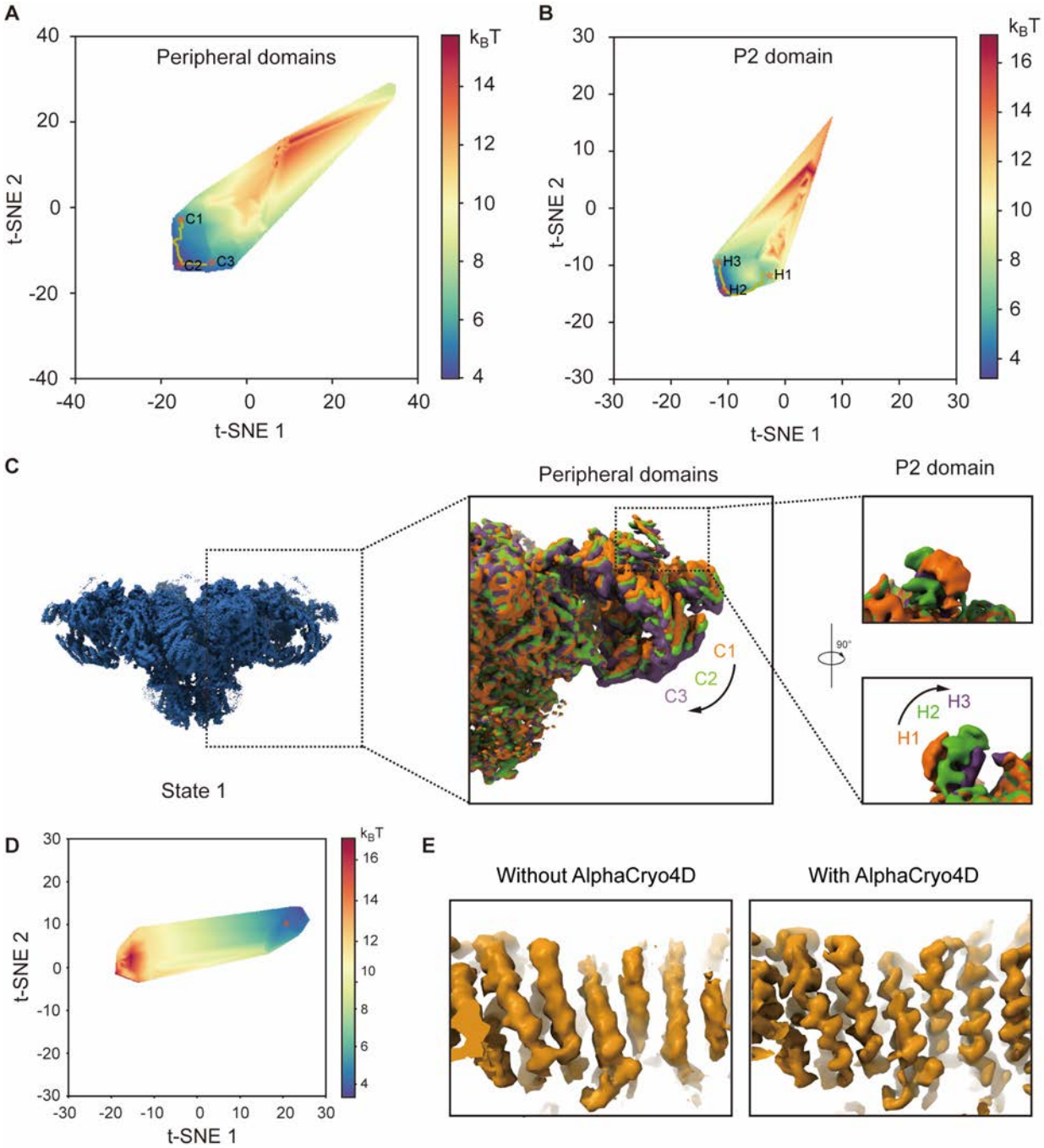

Figure 3. Improving the resolution of the dynamic domains of RyR1 to the atomic level by AlphaCryo4D. (**A**) Energy landscape of the peripheral domains of the RyR1. Clusters C1, C2 and C3 were selected for particle voting by this energy landscape. (**B**) Focused energy landscape of the P2 domain of the RyR1. Clusters H1, H2 and H3 were uncovered on this energy landscape. (**C**) Cryo-EM density maps of clusters C1-C3 and H1-H3 classified by AlphaCryo4D. The peripheral domains exhibited hierarchical dynamics in its P2 domain, which was not well observed in the original density map of state 1. (**D**) Energy landscape estimated with the alignment of the peripheral domains. This energy landscape is aimed to improve the quality of

the final density map. (**E**) Comparison of the helical domain density map after AlphaCryo4D with the original state 1. The process described here improved the resolution of the density.

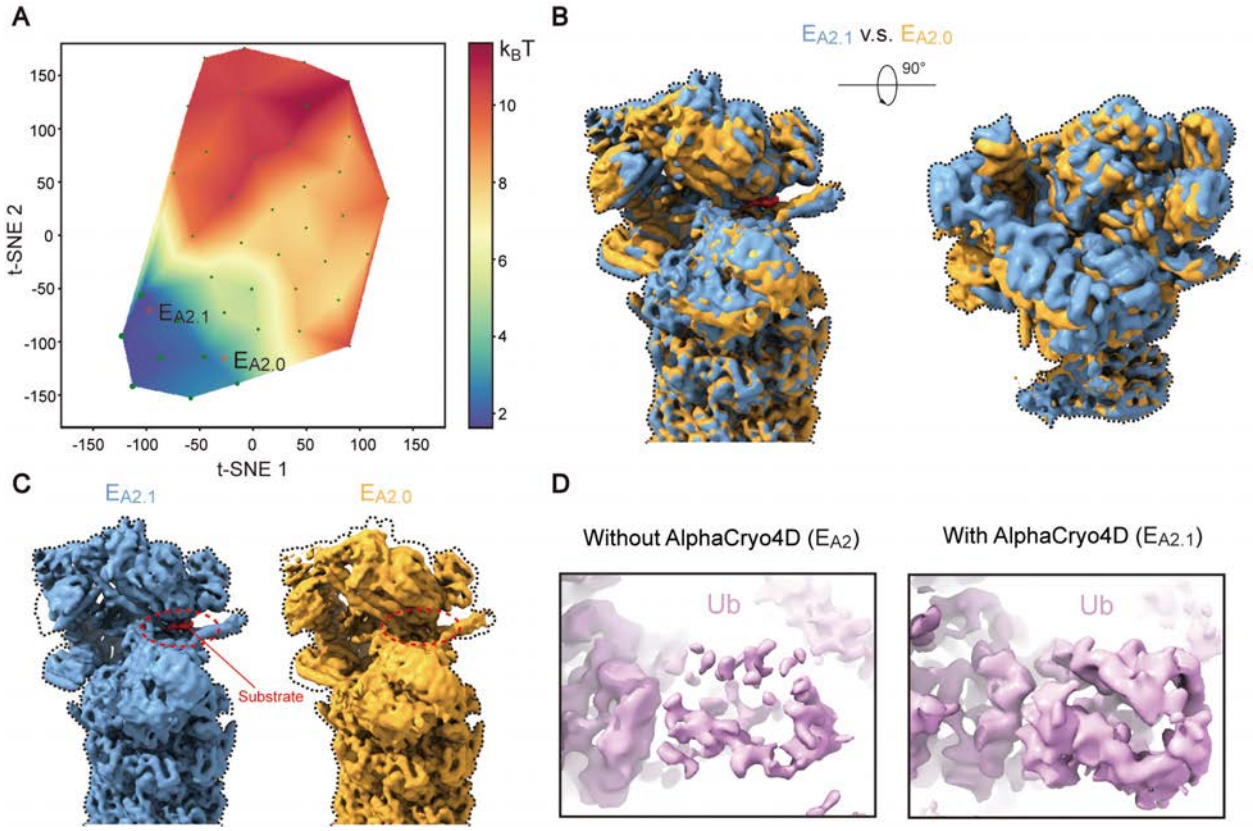

Figure 4. Discovery of novel intermediate of 26S proteasome with improved local resolution of 8.5-kDa ubiquitin by AlphaCryo4D. (**A**) Energy landscape calculated on the regulatory particles of 26S proteasome state E_{A2}. Clusters $E_{A2.0}$ and $E_{A2.1}$ were determined on this energy landscape, revealing the states of substrate-unbound and substrate-bound RPN11, respectively. (**B**) Cryo-EM density map of cluster $E_{A2.0}$ superimposed with the density map of cluster $E_{A2.1}$ from two orthogonal viewing angles. The lid subcomplex undergoes conformational changes during substrate binding. (**C**) Side-by-side comparison of the two states. The red region denoted the substrate in the density map of 26S proteasome. (**D**) Density map quality improvement of the region around the RPN11-bound ubiquitin compared to the previously published structure [23]. Ub, ubiquitin.

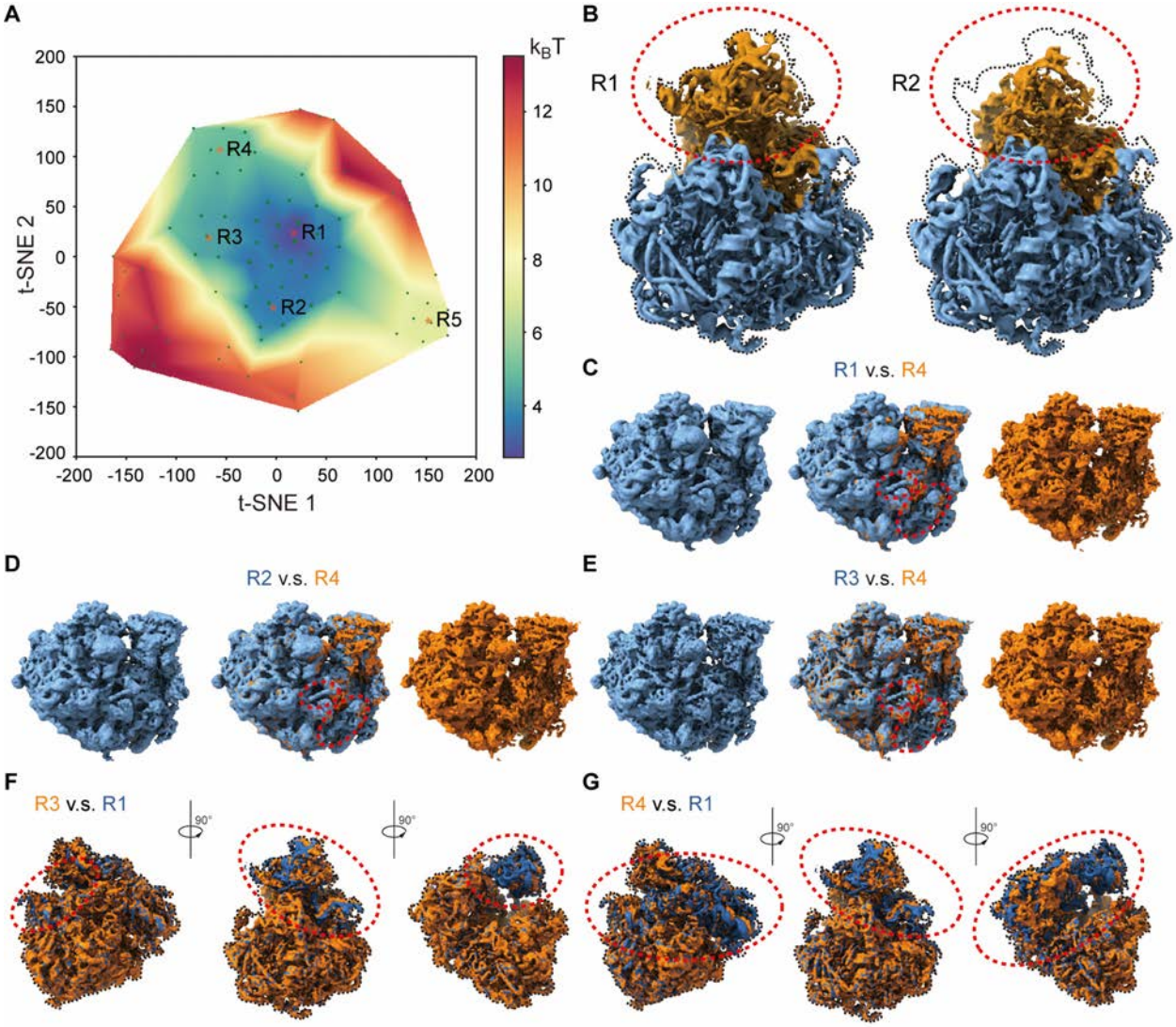

Figure 5. Conformational dynamics of the *Pf*80S ribosome bound to anti-protozoan drug. (**A**) Clusters R1-R5 on the energy landscape by AlphaCryo4D. Energy-based particle voting were conducted within each cluster to keep particles of high reproducibility. (**B**) Comparison of the density maps of cluster R1 and R2. Cluster R1 is the major state in the *Pf*80S ribosome dataset, while cluster R2 accounts for the state missing the head of 40S (red circles). The dynamic region of the density map is colored yellow and the stable region is colored blue. (**C-E**) Comparisons of states R1 and R4 (C), states R2 and R4 (D), states R3 and R4 (E) reveal the differential appearance of the C-terminal helix of eL8 and an rRNA helix. Note that the rRNA helix is only partially presented in state R3. (**F**) State R3 superimposed with state R1 from three different viewing angles. The density of the 40S head region (red circles) rotated relative to the 60S. (**G**) State R4 superimposed with state R1 from three different viewing angles. The density of the whole 40S (red circles) is rotated relative to the 60S.

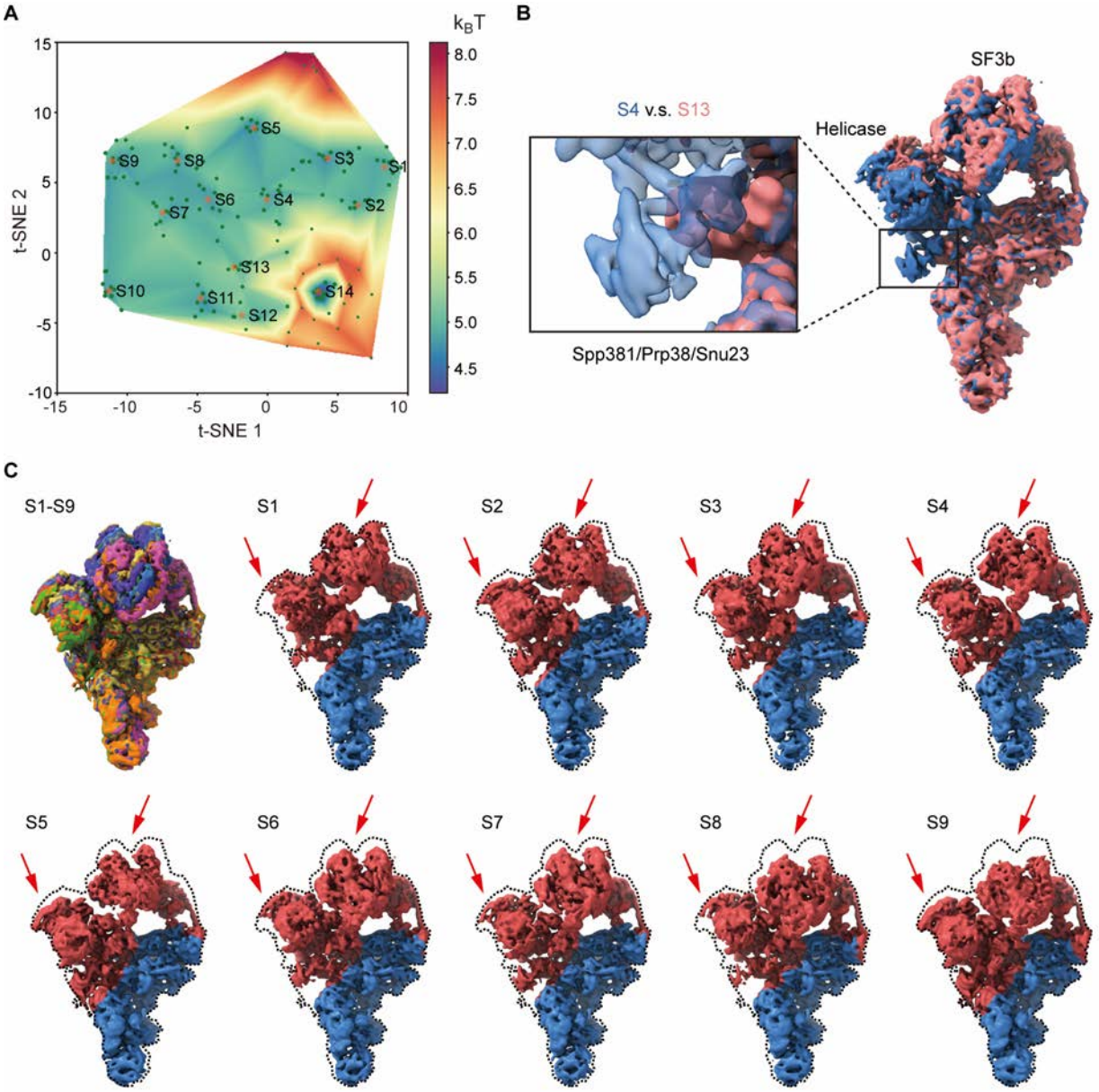

Figure 6. Multi-scale conformational dynamics of the pre-catalytic spliceosome. (**A**) Energy landscape by AlphaCryo4D uncovering 14 clusters designated as S1-S14. Clusters S1-S9 were related to the continuous rotational motion of the SF3b and helicase subcomplexes of pre-catalytic spliceosome, and the density maps of clusters S10-S14 were shown in Figure S9B. (**B**) Density map of the Spp381/Prp38/Snu23 missing state S13 superimposed with the state S4. Closeup comparison showed the density difference between the state S13 and S4. (**C**) Cryo-EM density maps of Clusters S1-S9. The dynamic region of the density map was colored red and the rigid region was colored blue. The red arrows suggested the rotational motion of the SF3b and helicase subcomplexes within the dashed outline.

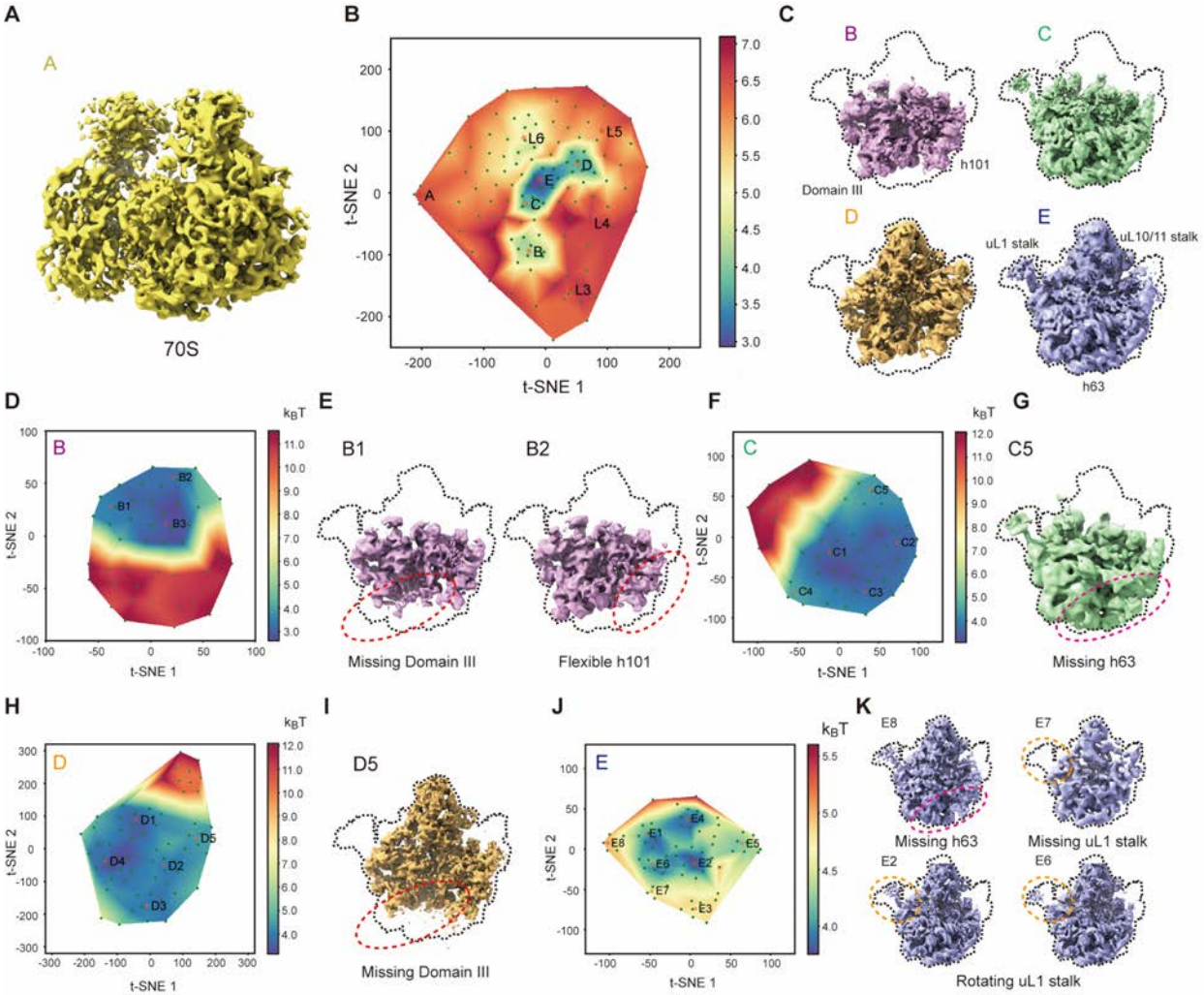

Figure 7. Conformational space of the bacterial ribosomal assembly intermediates. (**A**) Cryo-EM density map of state A determined by AlphaCryo4D. This state was located far from other states on the energy landscape (B). (**B**) Global energy landscape of bacterial ribosomal assembly intermediates. Clusters A-E and L3-L6 were reconstructed for further analysis. (**C**) Density maps of the major states B, C, D and E reconstructed by AlphaCryo4D. The density maps of these four states were refined using the particles of clusters B, C, D and E respectively from the energy landscape. (**D**) Zoomed-in energy landscape of the major state B. (**E**) New minor classes B1 and B2 of B. Comparing to the original state B (Fig. S10B), the states missing Domain III (B1) and with flexible rRNA helix 101 (B2) were discovered by AlphaCryo4D. (**F**) Zoomed-in energy landscape of the major state C generating clusters C1-C5. (**G**) A new minor state (C5) of C determined by AlphaCryo4D. The rRNA helix 63 was not observed in the density map of C5. (**H**) Clusters D1-D5 from the zoom-in energy landscape of the major state D. (**I**) A newly discovered minor state D5. Like the minor state B1 of the major state B, Domain 3 was not seen in the density map of D5. (**J**) Clusters E1-E8 on the zoom-in energy landscape of the major state E by AlphaCryo4D. (**K**) Discovery of new classes in the major state E. Like C3, the class E8 lacks the rRNA helix 63. Comparing to the state E3 (Fig. S10E), the uL1 stalk is not

found in the class E7. And two subclasses of E2 with different conformational changes of the uL1 stalk were discovered by AlphaCryo4D.

Methods

AlphaCryo4D framework

***M*-fold particle shuffling and Bayesian bootstrapping**. A key philosophy of the AlphaCryo4D design is to avoid subjective judgement on the particle quality and usability as long as it is not apparent false particles like ice contaminants, carbon edges and other obvious impurities. Deep-learning-based particle picking in DeepEM [66], Torpaz [67] or other similarly performed software is favored for data preprocessing prior to AlphaCryo4D. To prepare particle datasets for AlphaCryo4D, an initial unsupervised 2D image classification and particle selection, preferentially conducted by the statistical manifold-learning-based algorithm in ROME [5], is necessary to ensure that no apparent false particles are selected for further analysis and that the data have been collected under an optimal microscope alignment condition, such as optimized coma-free alignment. No particles should be discarded based on their structural appearance during this step if they are not apparent false particles. Any additional 3D classification should be avoided to pre-maturely reject particles prior to particle shuffling and volume bootstrapping in the first step of AlphaCryo4D processing. Pre-maturely rejecting true particles via any 2D and 3D classification is expected to introduce subjective bias and to impair the native conformational continuity and statistical integrity intrinsically existing in the dataset.

In raw cryo-EM data, 2D transmission images of biological macromolecules suffer from extremely heavy background noise, due to the use of low electron dose to avoid radiation damage. To tackle the conformational heterogeneity of the macromolecule sample of interest in the presence of heavy image noise, the particle shuffling and volume bootstrapping procedure was devised to incorporate the Bayesian or maximum-likelihood-based 3D clustering in RELION [3]. To enable the particle-voting algorithm in the late stage of AlphaCryo4D, each particle is reused *M* times during particle shuffling to bootstrap a large number of 3D volumes (Fig. S1A). First, all particle images are aligned to the same frame of reference in a consensus 3D reconstruction and refinement in RELION [3] or ROME [5] to obtain the initial alignment parameters of three Euler angles and two translational shifts. Optimization for alignment accuracy should be pursued to avoid the error propagation to the subsequent steps in AlphaCryo4D. For a dataset with both compositional and conformational heterogeneity, coarsely classifying the dataset to a few 3D classes during initial image alignment, limiting the 3D alignment to a moderate resolution like 10 Å or 15 Å during global orientational search,

progressing to small enough angular steps in the final stage of consensus refinement, may be optionally practiced in order to optimize the initial 3D alignment. Failure of initial alignment of particles would lead to failure of all subsequent AlphaCryo4D analysis.

Next, based on the results of consensus alignment, in the particle-shuffling step, all particles were divided into $M + 1$ groups, and M was set to an odd number of at least 3. Then the whole dataset was shuffled $M + 1$ times by removing a different group out of the dataset each time. Each shuffled dataset is classified into a large number (B) of 3D volumes, often tens to hundreds, by the maximum-likelihood 3D classification algorithm without further image alignment in RELION (with 'skip-align' option turned on). This is necessary because the alignment accuracy often degrades when the sizes of 3D classes decrease considerably. This step is repeated $M + 1$ times, each time on a shuffled dataset missing a different group among the $M + 1$ groups. Due to the effect of M-fold particle shuffling, the outcome of this entire process is expected to bootstrap up to thousands of 3D volumes in total (i.e., $B(M+1) > 1000$). Each particle is used and contributed to the 3D reconstructions of M volumes, which prepare it for the particle-voting algorithm to evaluate the robustness and reproducibility of each particle with respect to the eventual 3D classification.

For processing a large dataset including millions of single-particle images, it becomes infeasible even for a modern high-performance computing system to do the consensus alignment by including all particles once in a single run due to limitation of supercomputer memory and the scalability of the alignment software. To tackle this issue, the whole dataset is randomly split into a number (D) of sub-datasets for batch processing, with each sub-dataset including about one to two hundred thousand particles, depending on the scale of available supercomputing system. In this case, the initial reference should be used for the consensus alignment of different sub-datasets to minimize volume alignment errors in the later step. The total number of resulting bootstrapped volumes becomes $BD(M+1) > 1000$. This strategy can substantially reduce the supercomputer memory pressure and requirement. In each sub-dataset, all particles were divided into $M + 1$ groups and subject to the particle shuffling and volume bootstrapping procedure as described above.

Deep residual autoencoder for 3D feature extraction. The bootstrapped volumes may still suffer from reconstruction noises and errors due to variation of particle number, misclassification

of conformers and limited alignment accuracy. Thus, we hypothesize that unsupervised deep learning could help capture the key features of structural variations in the bootstrapped volumes and potentially enhance the quality of subsequently reconstituted energy landscapes for improved 3D classification. To this end, a 3D autoencoder was constructed using a deep Fully Convolutional Network (FCN) composed of residual learning blocks [49,50]. The architecture of the 3D autoencoder consists of the encoder and the decoder, which are denoted as $\mathcal{E}$ and $\mathcal{D}$, respectively (Fig. S1B). The relation between the output $\boldsymbol{y}$ and the input $\boldsymbol{x}$ of the network can be expressed as $\boldsymbol{y} = \mathcal{D}\big(\mathcal{E}(\boldsymbol{x})\big)$, in which $\boldsymbol{x}$ is the input 3D density volume with the size of N^3, where N is the box size of the density map in pixel units. For reconstruction of the 3D volumes and further optimization, the decoding output $\boldsymbol{y}$ should be in the same size and range with the input data $\boldsymbol{x}$. In this way, the framework of FCN is established to restore the input volume, using the sigmoid function $S(x) = \frac{1}{1+\exp(-x)}$ as the activation function of the decoding layer to normalize the value of $\boldsymbol{y}$ into the range (0, 1). Meanwhile, all 3D density maps $\boldsymbol{x}$ are normalized with the following equation before input to the deep neural network:

$$x_{ijk} := \frac{x_{ijk} - x_{min}}{x_{max} - x_{min}}, \quad i, j, k = 1,2, \dots, N.$$

where x_{min} and x_{max} are, respectively, the minimum and minimax value in all x_{ijk}.

The distance between the decoded maps and the input volumes can be used for constructing the loss function to train the 3D kernels and bias of the networks. The value distribution of the encoded 3D feature maps $\boldsymbol{z} = \mathcal{E}(\boldsymbol{x})$ is expected to be an abstract, numerical representation of the underlying structures in the volume data, which may not necessarily have any intuitive real-space physical meanings. The neural network is capable of extracting such abstract information in the prediction step, with no restriction on the feature maps $\boldsymbol{z}$ in the expression of training loss. The loss function is then formulated as:

$$L(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N^3} \sum_{i,j,k=1}^{N} \left\| x_{ijk} - y_{ijk} \right\|^2 + \lambda \|\boldsymbol{\theta}\|^2$$

where $\boldsymbol{\theta}$ denotes the weights and bias of the network, and λ is L2 norm regularization coefficient. As the feature of the complex structure is difficult to be learned from the 3D volume data, the value of λ is set to 0 in default to focus on the first term of the expression. In the applications to several experimental datasets, it was set to 0.001 to prevent overfitting (see

below).

To improve the learning capacity of the 3D autoencoder, residual learning blocks containing 3D convolutional and transposed convolutional layers are employed in the encoder and decoder, respectively. In each residual block, a convolutional layer followed by a Batch Normalization [68] (BN) layer and activation layer appears twice as a basic mapping, which is added with the input to generate the output. The mathematical expression of the lth block can be shown as:

$$\begin{cases} \boldsymbol{y_l} = \mathcal{F}(\boldsymbol{x_l}; \boldsymbol{\theta}_l) + \boldsymbol{x_l} \\ \mathcal{F}(\boldsymbol{x_l}; \boldsymbol{\theta}_l) = \mathcal{C}\big(\mathcal{C}(\boldsymbol{x_l})\big) \\ \mathcal{C}(\boldsymbol{x_l}) = a(b(c(\boldsymbol{x_l}))) \end{cases}$$

where $\boldsymbol{x_l}$ and $\boldsymbol{y_l}$ represent the input and the output of the block, respectively. $\mathcal{F}(x_l; \boldsymbol{\theta}_l)$ denotes the basic mapping of the lth block parameterized by $\boldsymbol{\theta}_l$, and $\mathcal{C}(\boldsymbol{x_l})$ is the sequential operation of convolution or transposed convolution c, BN b and activation a. The rectified linear unit (ReLU) function is used in all the activation layers but the last one. In addition, the output of mapping function $\mathcal{F}(\boldsymbol{x_l}; \boldsymbol{\theta}_l)$ must have the same dimension as the input $\boldsymbol{x_l}$. If this is not the case, the input $\boldsymbol{x_l}$ must be rescaled along with $\mathcal{F}(\boldsymbol{x_l}; \boldsymbol{\theta}_l)$ using a convolutional or transposed convolutional transformation $c'(\boldsymbol{x_l})$ with an appropriate stride value, the parameters of which can be updated in the training step.

Autoencoder training. To analyze a large number of volumes, we carefully tuned the training of the 3D deep residual autoencoder to obtain suitable kernels and bias. First, parallel computation with multiple GPUs has been implemented to reduce the training time. Then the parameters of the network are optimized by the stochastic gradient descent Adam (Adaptive moment estimation) algorithm, in which the gradients of the objective function $L(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{y})$ with respect to the parameters $\boldsymbol{\theta}$ can be calculated by the chain rule. Moreover, the learning rate is reduced to one tenth when the loss function does not decrease in three epochs based on the initial value of 0.01. After trained about 50 epochs, the best model is picked to execute the task of structural feature extraction. Using the unsupervised 3D autoencoder, the feature maps $\boldsymbol{z} = \mathcal{E}(\boldsymbol{x})$ encoding the structural discrepancy among the 3D volume data can be extracted automatically without any human intervention.

Autoencoder hyperparameters. The recommended hyperparameters for the deep residual

autoencoder architecture are provided in Table S1. Although we expand the residual network (ResNet) into 3D, we keep the original design rules of ResNet for constructing the 3D autoencoder [49,50]. The first and last convolutional layers have 5 × 5 × 5 filters (kernels). The remaining convolutional layers have 3 × 3 × 3 filters. Because the cubic filter in convolutional layer is computationally expensive, only one 3D filter is used in each of the last three convolutional layers in the encoder and of the first two and last transposed convolutional layers in the decoder. To accommodate a large volume input, two filters are used in the first three convolutional layers and in the third and fourth transposed convolutional layers. Downsampling is directly performed by convolutional layers with a stride of 2. The output dimension of encoding layers is set to be a quarter of the input dimension, in order to avoid over-compressing the feature maps. We employed six 3D convolutional layers in the encoder and five transposed convolutional layers in the decoder based on the expected tradeoff between the learning accuracy and training cost, which is roughly comparable to a 2D ResNet with more than 200 layers with respect to the training cost. Lastly, one should randomly inspect some feature maps to ensure that all hyperparameter setting works as expected. If a feature map clearly shows no structural correspondence to its input volume and exhibits strong noise or abnormal features, it might signal the existence of overfitting. In this case, a non-zero value of λ in the L2 norm regularization should be empirically examined to avoid overfitting.

Manifold embedding of energy landscape. To prepare for the energy landscape reconstitution, each bootstrapped 3D volume was juxtaposed with its 3D feature map learned by the 3D autoencoder to form an expanded higher-dimensional data point. Each expanded data point is either low-pass filtered at a given resolution (5 Å in default setting) or standardized by the z-score method prior to t-SNE processing:

$$\begin{cases} x_{ijk} := \dfrac{x_{ijk} - \mu_x}{\sigma_x}, \quad i,j,k = 1,2,\dots,N. \\ z_{ijk} := \dfrac{z_{ijk} - \mu_z}{\sigma_z}, \quad i,j,k = 1,2,\dots,N'. \end{cases}$$

(4) in which μ_x and σ_x denote the mean and standard deviation of the 3D volume voxels x_{ijk} with the box size of N, respectively. Similarly, μ_z and σ_z are the mean and standard deviation of the 3D feature map voxels z_{ijk} with the box size of N', respectively. All the input data points were then embedded onto a low-dimensional manifold via the t-SNE algorithm by preserving the

geodesic relationships among all high-dimensional data [51]. During manifold embedding, it is assumed that the pairwise similarities in the high dimensional data space and low dimensional latent space follow a Gaussian distribution and a Student's t-distribution, respectively. To find the low-dimensional latent variable of each data point, the relative entropy, also called the Kullback-Leibler (KL) divergence, is used as an objective function measuring the distance between the similarity distribution p_{ij} in the data space and q_{ij} in the latent space:

$$KL(P \parallel Q) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

which is minimized by the gradient descent algorithm with the momentum method [69]. The hyperparameter of t-SNE may be tuned to better preserve both local and global geodesic distances (https://distill.pub/2016/misread-tsne/).

The idea of using juxtaposed data composed of both volumes and their feature maps for manifold embedding is similar to the design philosophy of deep residual learning, in which the input and feature output of a residual learning block are added together to be used as input of the next residual learning block. Such a design has been demonstrated to improve learning accuracy and reduce the performance degradation issues when the neural network goes much deeper. [49,50]. Although the 3D feature maps or the bootstrapped volumes alone can be used for manifold embedding, both appear to be inferior in 3D classification accuracy (Fig. S4). The juxtaposed format of input data for manifold learning is potentially beneficial to the applications in those challenging scenarios, such as visualizing a reversibly bound small protein like ubiquitin (~8.6 kDa) of low occupancy by the focused AlphaCryo4D classification (see below).

After the manifold embedding by t-SNR, each 3D volume is mapped to a low-dimensional data point in the learned manifold. The coordinate system, in which the low-dimensional representation of the manifold is embedded, is used for reconstructing energy landscape. The difference in the Gibbs free energy ΔG between two states with classified particle numbers of N_i and N_j is defined by the Boltzmann relation $N_i/N_j = \exp(-\Delta G/k_B T)$. Thus, the free energy of each volume can be estimated using its corresponding particle number:

$$\Delta\Delta G_i = -k_B T \ln \frac{N_i}{\sum_k N_k},$$

where $\Delta\Delta G_i$ denotes the free energy difference of the data point with the classified particle number of N_i against a common reference energy level, k_B is the Boltzmann constant and T is

the temperature in Kelvin. The energy landscape was plotted by interpolation of the free energy difference in areas with sparse data. We suggest that linear interpolation is used for the energy landscape with loosely sampled areas to avoid overfitting by polynomial or quadratic interpolation. For densely sampled energy landscape, polynomial interpolation could give rise to smoother energy landscape that is easier to be tackled by the string method for MEP solution (see below).

The coordinate system of the embedded manifold output by t-SNE is inherited by the reconstitution of energy landscape as reaction coordinates. They do not have intuitive physical meaning of length scale in the real space. However, they can be viewed as transformed, rescaled, reprojected coordinates from the real-space reaction coordinates along which the most prominent structural changes can be observed. Alternatively, they can be intuitively understood as being similar to transformed, reprojected principal components (PCs) in principle component analysis (PCA). In the case of the 26S proteasome, the two most prominent motions are the rotation of the lid relative to the base and the intrinsic motion of AAA-ATPase motor [23], which can be approximately projected to two reaction coordinates. In real space, both motions are notably complex and in fact are characterized by a considerable number of degrees of freedom, which are partially defined by the solved conformers.

String method for finding minimum energy path (MEP). The string method is an effective algorithm to find the MEP on the potential energy surface [52]. To extract the dynamic information implicated in the experimental energy landscape, an improved and simplified version of the string method has been previously developed [53]. Along the MEP on the energy landscape, the local minima of interest could be defined as 3D clustering centers to guide the particle-voting algorithm for 3D classification to generate high-resolution cryo-EM density maps (Fig. S1D). The objective of the MEP identification in energy barrier-crossing events lies in finding a curve γ having the same tangent direction as the gradient of energy surface ∇G. It can be expressed as:

$$(\nabla G)^{\perp}(\gamma) = 0,$$

where $(\nabla G)^{\perp}$ denotes the component of ∇G perpendicular to the path γ. To optimize this objective function, two computational steps, referred to as 'evolution of the transition path' and 'reparameterization of the string', are iterated until convergence within a given precision threshold.

After initialization with the starting and ending points, in the step of evolution of the transition path, the positions of interval points along the transition path were updated according to gradient of the free energy at the tth iteration:

$$\varphi_i^{*(t)} = \varphi_i^{(t-1)} - h\nabla G\left(\varphi_i^{(t-1)}\right),$$

with $\varphi_i^{(t)}$ $(i = 0,1,\dots,N)$ being the ith intermediate point along the transition path at the tth iteration ($*$ denoting the temporary values), and h the learning rate.

In the step of reparameterization of the string, the values of positions $\varphi_i^{(t)}$ $(i = 0,1,\dots,N)$ were interpolated onto a uniform mesh with the constant number of points. Prior to interpolation, the normalized length $\alpha_i^{*(t)}$ $(i = 0,1,\dots,N)$ along the path was calculated as:

$$\alpha_0^{*(t)} = 0,\ \ \alpha_i^{*(t)} = \alpha_{i-1}^{*(t)} + \frac{\left\|\varphi_i^{*(t)} - \varphi_{i-1}^{*(t)}\right\|}{\sum_{i=1}^{N}\left\|\varphi_i^{*(t)} - \varphi_{i-1}^{*(t)}\right\|},\ \ i = 1,2,\dots,N.$$

Given a set of data points $(\alpha_i^{*(t)}, \varphi_i^{*(t)})$, the linear interpolation function was next used to generate the new values of positions $\varphi_i^{(t)}$ $(i = 0,1,\dots,N)$ at the uniform grid points $\alpha_i^{(t)}$ $(i = 0,1,\dots,N)$. The iteration was terminated when the relative difference $\sum_{i=0}^{N}\left\|\varphi_i^{(t)} - \varphi_i^{(t-1)}\right\|^2/N$ became small enough.

String method parameters. The string method of searching a rational MEP on the energy landscape can only guarantee the solution of local optimum and is not capable of ensuring a solution of global optimum [53]. The outcome of the string method depends on the initialization of the starting and ending points of the MEP on the energy landscape. If there are too many local minima to sample along a long MEP, the string method could retrieve a MEP solution that partly misses some local energy minima by going off the pathway. In this case, the search of an expected MEP solution can be divided into several segments of shorter MEPs connecting one another, with each MEP being defined by a closer pair of starting and ending points, which travels through a smaller number of energy minima. Another parameter affecting the MEP solution is the step size being set to explore the MEP on energy landscapes. Although a default value of 0.1 empirically recommended may work in many cases, it may be helpful to tune the step size according to the quality of the energy landscape. The value of step size can be decreased if the computed path runs out of the energy landscape and can be increased if the path

updates too slowly during iterative searching by the string method.

Energy-based particle voting algorithm. The particle-voting algorithm was designed to conduct 3D classification, particle quality control, reproducibility test and particle selection in an integrative manner. The particle-voting algorithm mainly involves two steps (Fig. S1D). First, we count the number of votes for each particle mapped within the voting boundaries of all 3D clusters on the energy landscape. One vote is rigorously mapped to one copy of the particle used in reconstructing a 3D volume and to no more than one 3D cluster on the energy landscape where the corresponding volume is located. Thus, each particle can have M votes casted for no more than M 3D clusters. If the vote is mapped outside of any 3D cluster boundary, it becomes an 'empty vote' with no cluster label. Each non-empty vote is thus labeled for both its particle identify and corresponding cluster identity. For each pair of particle and cluster, we compute the total number (K) of votes that the cluster receives from the same particle. Each particle is then assigned and classified to the 3D cluster that receives $K > M/2$ votes from this particle (Fig. S1D). Note that after particle voting, each particle is assigned no more than once to a 3D class, with its redundant particle copies removed from this class. This strategy only retains the particles that can reproducibly vote for a 3D cluster corresponding to a homogeneous conformation, while abandoning those non-reproducible particles with divergent, inconsistent votes.

Alternative strategy of 3D clustering without particle voting. Because the particle-voting algorithm imposes strong constraints on the reproducibility of particle classification by deep manifold learning, some 3D classes might be assigned with a small number of particles that are insufficient to support high-resolution reconstruction. To remedy this limitation, an alternative, distance-based classification algorithm was devised to replace the particle-voting algorithm when there are not enough particles to gain the advantage of particle voting (Fig. S1E). In this method, the distances of all M copies of each particle to all 3D cluster centers on the energy landscape are measured and ranked. Then, the particle is classified to the 3D cluster of the shortest distance. A threshold could also be manually preset to remove particles that are too far away from any of the cluster centers. The distance-based classification method can keep more particles, but it ignores the potential issue of irreproducibility of low-quality particles. Thus, it is proven to be less accurate in 3D classification (Fig. S6M-O). In other words, it trades off the classification

accuracy and class homogeneity to gain more particles, which is expected to be potentially useful for small datasets or small classes. By contrast, the energy-based particle-voting algorithm imposes a more stringent constraint to select particles of high reproducibility during classification, resulting in higher quality and homogeneity in the classified particles, which is superior to the distance-based classification method (Fig. S6S-U).

Practical consideration of particle shuffling and voting parameters. The parameter M determines the degree of implicit cross-validation of classification reproducibility, as well as the sampling densities of energy landscape. To establish reproducibility for 3D classification, M should be no less than 3. The variation of M is not supposed to change the energy landscape because it multiplies on both denominator and enumerator in the Boltzmann relation and is cancelled in the ratio of particle densities in computing the free-energy differences. Increasing M value will allow more volumes to be bootstrapped, which then leads to a higher sampling density in computing the manifold of energy landscape and potentially enhances its reconstitution quality. The default classification threshold $M/2$ is considered to be minimal for verifying the reproducibility of 3D classification. However, a higher threshold, such as $2M/3$, will give rise to more stringent criteria in cross-validation, with a tradeoff of voting out more particles. The particle voting algorithm does not entirely eradicate misclassified particles. However, increasing either M or the classification threshold could theoretically have similar impact and improve the conformational homogeneity, because it gets less probable for a particle being misclassified to the same cluster for more times.

Several considerations may be applied to the choice of M and to set up particle shuffling and voting. First, to obtain a high-quality energy landscape, a thousand or more bootstrapped volumes are expected for datasets of more than 1-2 million particles. Second, the average particle number per volume is expected to be no less than 5000 or more to ensure that majority of volumes include sufficient particles for quality reconstructions. The expected particle number per volume must increase if the average SNR per particle is decreased. It may need to reach 10000-20000 or more for small proteins or lower SNR datasets. Third, for a dataset of a moderate size, M can be adjusted to a higher value to mitigate the lack of image data for bootstrap.

Data-processing workflow of AlphaCryo4D. The following summarizes the complete

procedure of AlphaCryo4D, with the detailed algorithm rationale explained in the subsequent subsections.

Input: Single-particle cryo-EM dataset after initial particle rejection of apparent false particles.

Output: Energy landscape, MEP, 3D class assignment of each particle.

Step 1. Bootstrap a large number of 3D volumes through particle shuffling, consensus alignment and Bayesian clustering.

1. Split the particle dataset randomly to many sub-datasets, if necessary, in case of a large dataset (>150,000 particle images), for batch processing of particle shuffling and volume bootstrapping. Otherwise, skip this step if the dataset is small enough (<150,000).
2. For each sub-dataset, conduct a consensus alignment to generate initial parameters of Euler angles and translations in RELION with one or more consensus models.
3. Divide each sub-dataset into $M + 1$ groups, shuffle the sub-dataset $M + 1$ times and each time take a different group out of the shuffled sub-dataset, giving rise to $M + 1$ shuffled sub-datasets all with different collection of particles.
4. Conduct 3D Bayesian classification on all the $M + 1$ shuffled sub-datasets to generate hundreds of 3D volumes, making each particle to contribute to M different volumes.
5. Execute steps (2) to (4) using the same initial model (low-pass filtered at 60-Å) for all sub-datasets.

Step 2. Extract deep features of all volume data with the 3D deep residual autoencoder.

6. Align all 3D volumes and adjust them to share a common frame of reference.
7. Initialize the hyper-parameters of the 3D autoencoder (Table S1).
8. Train the neural network with the 3D volume data to minimize mean square error between the decoding layer and the input by the Adam algorithm of initial learning rate 0.01.
9. Extract the 3D feature maps of all volumes from the encoding layer.

Step 3. Embed the volume data to two-dimensional manifolds through the t-SNE algorithm and compute the energy landscape and find the MEP.

10. Calculate the pairwise similarities between volumes using their feature-map-expanded volume vectors, and randomly initialize the low-dimensional points.
11. Minimize the Kullback-Leibler divergence by the Momentum algorithm to generate 2D manifold embeddings with t-SNE.

12. Compute the energy landscape from the manifold using the Boltzmann relation.
13. Initialize searching of the MEP with a straight line between given starting and ending points.
14. Find the optimal MEP solution using the string method.

Step 4. Classify all particles through the energy-based particle-voting algorithm.

15. Sample the clustering centers along the MEP and calculate the recommended clustering radius.
16. Define the local energy minima as the 3D clustering centers and their corresponding cluster boundary for particle voting.
17. For each particle, cast a labelled vote for a 3D class when a volume containing one of the *M* particle copies is located within the voting boundary.
18. Count the number of votes of each particle with respect to each 3D class and assign the particle to the 3D class that receives more than *M*/2 votes from this particle.
19. Refine each 3D density map separately to high resolution in RELION or cryoSPARC using particles classified into the same 3D classes.

Blind assessments with simulated datasets

Three simulated datasets. Three simulated large datasets with the SNRs of 0.05, 0.01 and 0.005, each including 2-million particle images, were employed to benchmark AlphaCryo4D and to compare its performance with alternative methods. For each synthetic dataset, the particles were computationally simulated by projecting the 20 3D density maps calculated from 20 hypothetical atomic models emulating continuous rotation of the NLRP3 inflammasome protein. The 20 atomic models were interpolated between the inactive NLRP3 structure and its hypothetical active state, which was generated through homology modeling using the activated NLRC4 structure [54] (Fig. S1C). The 20 atomic models represent sequential intermediate conformations during a continuous rotation in its NATCH domain against its LRR domain over an angular range of 90°. Each conformation is thus rotated 4.5° over its immediate predecessor in the conformational continuum sequence. 100,000 simulated particle images per conformational state were generated with random defocus values in the range of -0.5 to -3.0 μm, resulting in 2 million particles for each dataset of a given SNR. The pixel size of the simulated image was set to the same as the pixel size (0.84 Å) of the real experimental dataset of NLRP3-NEK7 complex

[54]. To emulate realistic circumstances in cryo-EM imaging, Gaussian noises, random Euler angles covering half a sphere and random in-plane translational shifts from -5.0 to 5.0 pixels were then applied to every particle image.

Each simulated heterogeneous NLRP3 dataset was analyzed separately by AlphaCryo4D and used to characterize the performance and robustness of AlphaCryo4D against the variation of SNRs. In the step of particle shuffling and volume bootstrapping, 2,000,000 particles in the dataset of any given SNR were divided randomly into 10 sub-datasets for batch processing. The orientation of each particle was determined in the initial 3D consensus alignment in RELION, which did not change in the subsequent 3D classification. In this step, the maximum number of iterations of the 3D alignment was set up as 30, with the initial reference low-pass filtered to 60 Å. 3-fold particle shuffling (indicated as × 3 below) was conducted on each sub-dataset for volume bootstrapping. The first round of maximum-likelihood 3D classification divided the input shuffled particle sub-dataset into 5 classes, each of these classes were then further classified into 8 classes. This procedure was separately executed on all shuffled particle sub-datasets. The particle shuffling and volume bootstrapping generated 1,372, 1,489, and 1,587 volumes by the datasets with SNRs of 0.05, 0.01 and 0.005, respectively. These volume data were used as inputs for deep residual autoencoder to compute low-dimensional manifolds and energy landscapes (Fig. 1B-D, Figs. S1D, 2A, B). After searching the MEP on the energy landscapes by the string method, 20 cluster centers along the MEP were defined by the local energy minima along the MEP by approximately equal geodesic distance between adjacent minima, which represent potentially different conformations of the molecule (Fig. S2A-C). The particle-voting algorithm was applied in every cluster to determine the final particle sets for all 3D classes. For the purpose of validation of the methodology and investigation of 3D classification improvement, we labeled each bootstrapped 3D volume with the conformational state that held the maximum proportion of particles in the class and computed its 3D classification precision as the ratio of the particle number belonging to the labelled class versus the total particle number in the volume (Fig. 1B, Figs. S2A, B, 3A).

Analysis of algorithm mechanism. To understand how the 3D classification accuracy is improved by AlphaCryo4D, we analyzed the algorithmic mechanism by a number of control experiments (Figs. S2D-G, 4). In total, we conducted 24 conditional control tests using 10

variations of algorithmic design by removing or replacing certain components of AlphaCryo4D. First, by removing the entire component of deep manifold learning, the distributions of high-precision 3D classes obtained by *M*-fold particle shuffling and voting alone drop by ~60% relative to that from complete AlphaCryo4D processing (Fig. S6D-F). In these control experiments, the particle voting was achieved through counting the votes of a particle against the same conformers classified by RELION via computing the Intersection-over-Union (IoU) metric after the step of volume bootstrapping. Similarly, by removing the 3D deep residual autoencoder but keeping the manifold learning by t-SNE for energy landscape reconstitution, the distributions of high-precision 3D classes is reduced by ~15% at the SNR of 0.005 (Fig. S6P-R). The effect of accuracy degradation is less prominent at a higher SNR (0.05 or 0.01), indicating that the unsupervised feature learning using deep residual autoencoder promotes the tolerance of the algorithm against higher noise level in the data. Together, these control tests suggest that deep manifold learning plays a crucial role in improving 3D classification accuracy with low SNR data.

Next, we replaced t-SNE with four other algorithms in manifold learning step, including two classic linear dimensionality reduction techniques, principal component analysis (PCA) [13] and multidimensional scaling (MDS) [55], and two nonlinear dimensionality reduction algorithms, isometric mapping (Isomap) [56] and locally linear embedding (LLE) [57]. We applied the four algorithms to reduce the dimensionality of the same sets of bootstrapped volume data. Although being capable of differentiating a portion of ground-truth conformers, the resulting 2D mappings by the four techniques exhibit considerable overlap between adjacent conformers and are incapable of distinguishing all 20 conformers of ground truth, thus inevitably missing many conformers (Fig. S2D-G). The PCA and Isomap approximately missed half of the ground-truth conformers, whereas MDS and LLE both missed around 70% of the ground-truth conformers. The inferior performance of these techniques is consistent with the original control experiments conducted by the t-SNE developers [51].

Further, we conducted control experiments to evaluate the impact of particle voting on the 3D classification accuracy by replacing the particle voting component with a clustering strategy that directly classifies each particle to the cluster of the nearest clustering center (Fig. 1C). In this case, the distributions of high-precision 3D classes are reduced by ~15-30% (Fig. S6M-O). The reduction is more prominent at the lower SNR. This indicates that the particle voting

considerably improves the 3D classification accuracy but is less impactful than the component of deep manifold learning.

Last, by tracking the statistics of classification precisions step by step, we evaluated how the 3D classification accuracy is improved over the intermediate steps of AlphaCryo4D (Fig. S6G-L). We found that the steps of particle shuffling, defining cluster boundaries on the energy landscapes and energy-based particle voting contribute to ~16%, ~20% and ~40% improvements in the distributions of high-precision 3D classes, respectively. Taken together, these results indicate that all components contribute to the improved performance of AlphaCryo4D in 3D classification toward the atomic level. Neither deep manifold learning nor particle voting alone is sufficient to achieve the present level of 3D classification accuracy.

Comparisons with alternative methods. Using 3D classification precision as a benchmark indicator, the performance of AlphaCryo4D was compared with several other methods: (1) the conventional maximum-likelihood-based 3D (ML3D) classification in RELION [3,4,6] with and without a hierarchical strategy, (2) the conventional heterogeneous reconstruction in cryoSPARC [2], (3) the 3DVA algorithm in cryoSPARC [11] and (4) the deep generative model-based cryoDRGN [12]. A total of 18 comparative tests by these alternative methods have been conducted blindly using the three simulated datasets, which includes 6-million images in total. In all our comparative tests on AlphaCryo4D and the alternative methods, the ground-truth information of particle orientations, in-plane translations and conformational identities were completely removed from the methods being tested and were not used for any steps of data processing or training. The ground-truth conformational identities of particles were only used as validation references to compute the 3D classification precisions of the blind testing results.

For the tests using ML3D in RELION, we classified all particles directly into 20 classes and hierarchically into 4 × 5 classes, which first divided the dataset into 4 classes, with each class further classified into five sub-classes (Fig. S3). For testing the conventional discrete ab initio heterogeneous reconstructions in cryoSPARC, each synthetic dataset was directly classified into 20 classes without providing any low-resolution reference model (Fig. S3). For comparison with the 3DVA algorithm in cryoSPARC, we first did the blind consensus alignment of the entire dataset to find the orientation of each particle. Then the alignment and the mask generated from the consensus reconstruction were used as inputs into the 3DVA calculation, with the number of

orthogonal principal modes being set to 2 or 3 in 3D classification. The 3D variability display module in the cluster mode was used to analyze the results of 3D classification. For blind tests with cryoDRGN, a default 8D latent variable model was trained for 25 epochs. The encoder and decoder architectures were 256 × 3, as recommend by the cryoDRGN authors [12]. The particle alignment parameters prior to cryoDRGN training were obtained by the same blind consensus refinement in RELION used for other parallel tests. The metadata of 3D classification precisions as well as the 3D density maps from all the algorithms applied on the three simulated datasets were collected to conduct the statistical analysis (Fig. 2 and Fig. S3).

Computational costs. Although the computational cost of AlpahCryo4D is generally higher than the conventional approach, it does not appear to increase drastically and likely falls in an affordable range, while reducing the average cost of computation per conformational state. In a nutshell, we can have a brief comparison of the computational efficiency on the simulated 2-million image dataset with an SNR of 0.01. In the step of 3D data bootstrapping, we split the dataset into 20 subsets, which contained 100,000 particles each. The 3D consensus alignment of all 2,000,000 particles cost about 75 hours using 8 Tesla V100 GPUs interconnected with the high-speed NVLink data bridge in a NVIDIA DGX-1 supercomputing system. Within each subset, the 3D Bayesian classification for one leave-one-group-out dataset cost about 2.5 hours using 320 CPU cores (Intel Xeon Gold 6142, 2.6 GHz, 16-core chip), so the total time spent in one subset was about 10 hours using 320 CPU cores in an Intel processor-based HPC cluster. In addition, it spent about 3 hours to extract feature via deep neural network using 8 Tesla V100 GPUs of the NVIDIA DGX-1 system. In contrast to about 160 hours cost in traditional classification methods, this approach cost a little more than 200 hours using 8 Tesla V100 GPUs and 320 CPU cores. Taken together, these observations suggest that the computational cost and efficiency of AlphaCryo4D are within the acceptable range considering the output of more high-resolution conformers yielded by the procedures. The benefits likely outweigh the tradeoff of computational costs, particularly for very large datasets that equally slow down any methods.

Applications to experimental cryo-EM datasets

Five experimental cryo-EM datasets were processed using AlphaCryo4D to examine its applicability in analyzing experimental cryo-EM data, as described in detail in the following.

The membrane channel protein complex RyR1. To determine the dynamics of the peripheral domains of the RyR1, density subtraction and focused energy landscape were applied in the further analysis of state 1 conformation [45]. Due to the low SNR of the particles, a large $M = 30$ was used in the particle shuffling of the AlphaCryo4D. After symmetry expansion of particles of state 1, a total of 1,557,168 particles with the core domain subtracted were utilized to compute the energy landscape with 310 volumes. Three 3D classes with the particle numbers of 196,669, 848,655 and 189,796 were obtained by the particle-voting procedure on the energy landscape. All these 3 classes were merged together for high resolution refinement with a mask of the peripheral domains using RELION. To further improve the quality of the density map, a zoom-in energy landscape with $M = 22$ was computed to result in a final particle set with the number of 1,198,946. Furthermore, as the P2 domain showed high flexibility, we subtracted the non-P2 high-resolution components from the particles and computed the P2 domain-focused energy landscape using $M = 23$. This energy landscape resulted in three classes with the particle numbers of 14,089, 1,108,282 and 57,381, which was combined to refine the final density map of the P2 domain with a specific mask using RELION.

The human 26S proteasome. The substrate-engaged human 26S proteasome dataset [23] includes 3,254,352 RP-CP particles (combined with particle images from both the DC and SC proteasomes) in total, with the super-resolution counting mode pixel size of 0.685 Å and the undecimated box size of 600 × 600 pixels. The substrates of 26S proteasome first appear in the state E_{A2}. In this focused study, 147,108 particles of E_{A2} were utilized for computing the specific energy landscape with 40 volumes. The alignment of these particles was firstly refined with a 19S mask. $M = 3$ particle shuffling was then conducted to bootstrap the 3D volumes with a soft mask of 19S. By clustering on the energy landscape and particle voting, two classes containing 99,043 and 47,389 particles were generated for high resolution refinement using RELION, which were designated state $E_{A2.1}$ and $E_{A2.0}$, respectively. The density map of state $E_{A2.0}$ exhibits a conformation identical to the previously reported state E_{A2} [23].

The *Pf*80S ribosome. The *Pf*80S ribosome dataset (EMPIAR-10028) contains 105,247 particle images. First, the alignment of all particles was refined with a global mask in RELION. To

bootstrap enough 3D volumes for the energy landscape, we set $M = 7$ in the particle shuffling step, resulting in 79 volumes in total with a global mask and 15 Å resolution limit in the expectation step. Moreover, the regularization coefficient was set to 0.001 when training the 3D residual network with this small dataset. All volumes were then low-pass filtered at 5 Å prior to manifold embedding. Five clusters were obtained from the energy landscape, which had 66,035, 21,482, 9,922, 6,424, 657 particle images respectively after voting. All these clusters were refined independently using RELION.

The pre-catalytic spliceosome. The pre-catalytic spliceosome dataset (EMPIAR-10180) with the particle number of 327,490 shows high conformational dynamics. The consensus alignment of these particles in the original dataset was used in the first step. $M = 7$ particle shuffling was utilized to bootstrap 160 volumes with a soft global mask and 15 Å resolution limit in the expectation step. The regularization coefficient of 0.001 was set to train the 3D Autoencoder with this dataset. All volumes were then low-pass filtered at 5 Å prior to manifold embedding. Based on the energy landscape, we obtained 14 classes with the particle number of 15,749, 17,703, 27,522, 13,820, 20,841, 14,989, 16,377, 22,047, 25,049, 21,327, 18,949, 6,157, 9,790 and 36,878 after voting. Each class was refined independently using RELION.

The bacterial 50S ribosomal intermediates. The 131,899 particle images of the 50S ribosomal large subunit (EMPIAR-10076) are highly heterogeneous. We refined the alignment of all particles with a global mask of 50S using RELION. In the first step, 119 bootstrapped volumes were utilized to calculate the energy landscape. $M = 7$ particle shuffling was conducted to generate these volumes without any mask. The regularization coefficient of 0.001 was set when training the deep neural network. All volumes were then low-pass filtered at 5 Å prior to manifold embedding. Then the first energy landscape resulted in nine clusters with the particle number of 766 (A), 15,129 (B), 1,236, 322, 670, 22,037, 25,445 (D), 33,976 (E) and 25,115 (C) respectively. For the clusters of states B-E, four zoom-in energy landscapes were then computed to discover more sub-states with 46, 56, 66 and 72 bootstrapped volumes respectively. In the zoom-in energy landscape calculation, $M = 7$ was set in particle shuffling for bootstrapping 3D volumes with a global mask and 20 Å resolution limit in the expectation step. Then the regularization coefficient of 0.001 was used to train the deep residual network. All volumes were

then low-pass filtered at 5 Å prior to manifold embedding. After particle voting, the energy landscape of state B resulted in three sub-clusters with the particle number of 3,494, 3,600 and 7,803. The state C energy landscape resulted in five sub-clusters with the particle number of 2,444, 7,442, 711, 7,633 and 5,557. The state D energy landscape resulted in five sub-clusters with the particle number of 6,231, 8,785, 2,123, 3,718 and 4,092. And the energy landscape of state E resulted in eight sub-clusters with the particle number of 4,783, 1,588, 5,891, 3,620, 7,352, 5,265, 357 and 2,241. All these clusters were refined using RELION.

Code availability

Source code of AlphaCryo4D is freely available at http://github.com/alphacryo4d/alphacryo4d/.

Acknowledgments. The authors thank Q. Ouyang and W. L. Wang for constructive discussions. This work was funded in part by the Beijing Natural Science Foundation (grant No. Z180016/Z18J008), the National Natural Science Foundation of China (grant No. 11774012) and a gift academic grant from Intel Corporation.

Author contributions. Y.Mao and Z.W. conceived this study and devised the methodology. Z.W. developed the prototype source code of the AlphaCryo4D system, conducted numerical studies of the system using the synthetic datasets. Z.W., E.C., and S.Z. analyzed the experimental cryo-EM datasets and refined the density maps. C.L. and C.C.Yin contributed to the sample preparation and data collection on the RyR1 dataset. Y.Ma contributed to early studies in algorithm design and testing. Y.Mao supervised this study, analyzed the data and wrote the manuscript with inputs from all authors.

Competing interests. The authors declare no competing financial interests.

Correspondence and requests for materials should be addressed to Y.M.

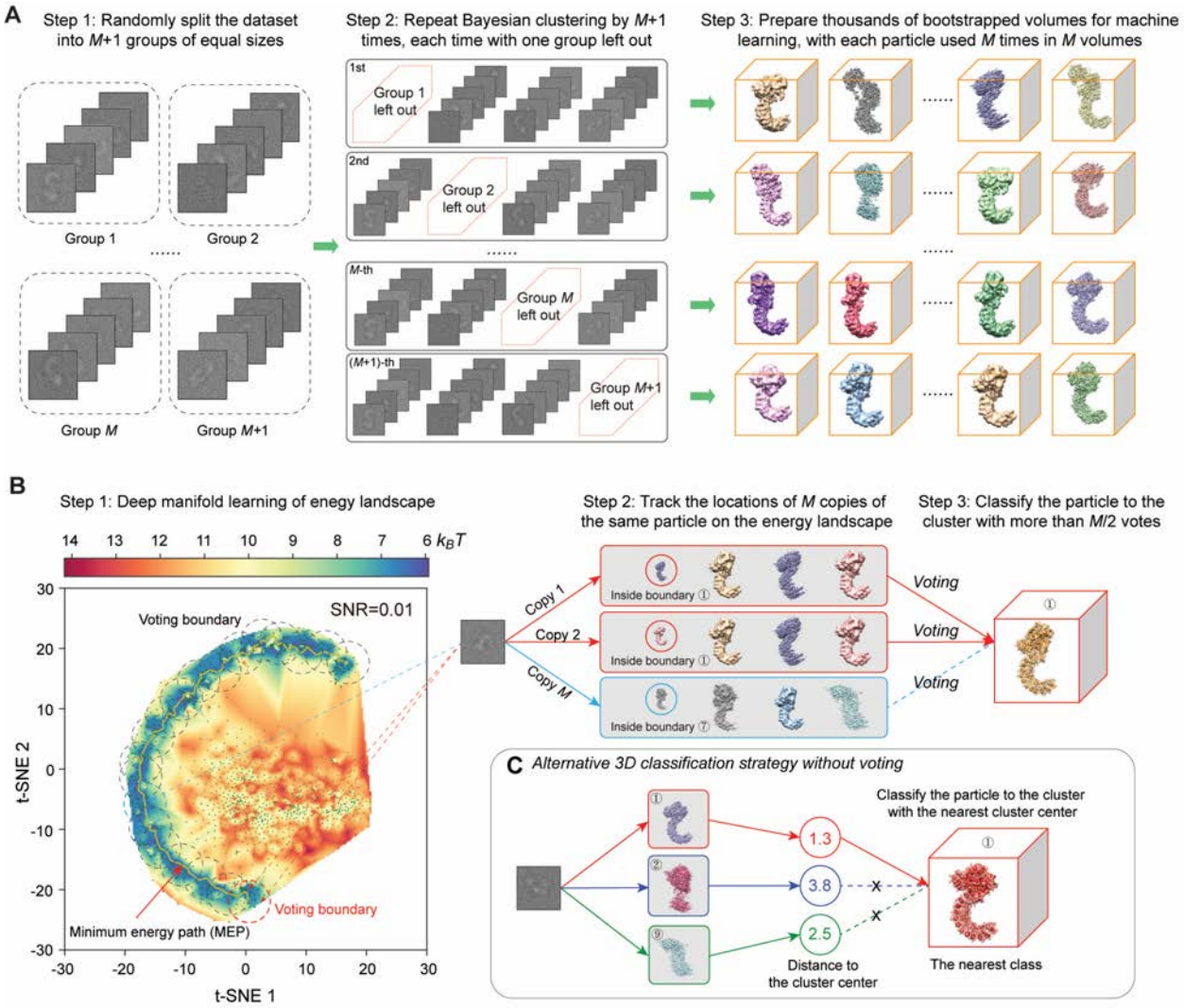

Figure S1. Detailed algorithmic design of particle shuffling and voting in AlphaCryo4D. (**A**) Schematic showing the method of *M*-fold particle shuffling and bootstrapping of 3D volumes. In the step 1, all particles are split randomly into *M* + 1 groups equally. Then the step 2 carries out the Bayesian clustering for reconstructions of 3D density maps within each of the *M* + 1 particle sets that are shuffled via the leave-one-group-out approach. After these two steps, thousands of 3D volumes are generated for the subsequent 3D deep learning, with each particle contributing to *M* volumes. (**B**) Schematic showing the algorithmic concept of energy-based particle voting for 3D classification. The left panel shows the energy landscape obtained by deep manifold learning. After clustering along the minimum energy path, all *M* locations of each particle on the energy landscape can be tracked to cast *M* votes. A vote of the particle is only counted for the cluster when it is mapped within the voting boundary of the cluster, as indicated by the circles marked on the energy landscape. Eventually, this particle is classified to the 3D cluster with over *M*/2 votes from this particle. (**C**) Alternative distance-based 3D classification method as a control in the analysis of algorithmic performance of particle voting. Instead of

particle voting, each particle is directly classified to the cluster with the nearest clustering center among the M volume data points in the distance-based 3D classification strategy.

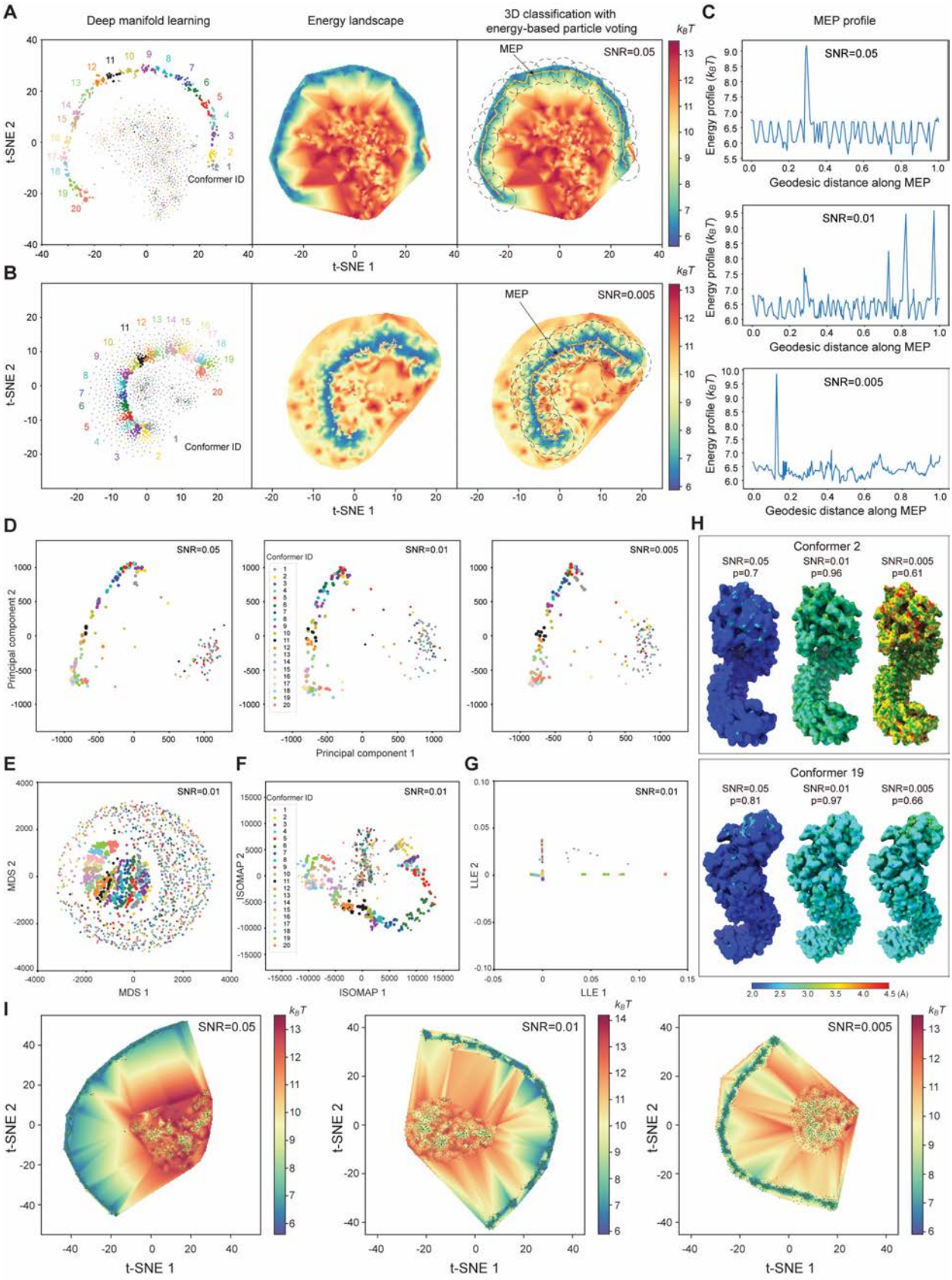
A
Deep manifold learning
Energy landscape
3D classification with energy-based particle voting
MEP
SNR=0.05
Conformer ID
t-SNE 1
t-SNE 2
B
SNR=0.005
C
MEP profile
SNR=0.05
SNR=0.01
SNR=0.005
Energy profile (k_BT)
Geodesic distance along MEP
D
Principal component 1
Principal component 2
E
MDS 1
MDS 2
F
ISOMAP 1
ISOMAP 2
G
LLE 1
LLE 2
H
Conformer 2
SNR=0.05 p=0.7
SNR=0.01 p=0.96
SNR=0.005 p=0.61
Conformer 19
SNR=0.05 p=0.81
SNR=0.01 p=0.97
SNR=0.005 p=0.66
I

Figure S2. Blind assessments of AlphaCryo4D and its comparison with the 3D PCA method using the simulated heterogeneous NLRP3 datasets of different SNRs. (**A** and **B)** Reconstruction of the energy landscape of the simulated NLRP3 datasets at SNRs of 0.05 (**A**) and 0.005 (**B**) by the t-SNE algorithm using the bootstrapped volumes and their corresponding feature maps. Colors in the left panels indicate the ground truth of 3D volume data points. (**C**) Free energy profiles along the MEP calculated by the string method in the 2D energy landscapes of the simulated NLRP3 datasets at SNRs of 0.05 (left), 0.01 (middle) and 0.005 (right). (**D**) Linear dimensionality reduction of bootstrapped 3D volumes from the simulated datasets at three distinct SNRs by the 3D PCA method. (**E-G**) Dimensionality reduction of bootstrapped 3D volumes from the simulated dataset at SNR of 0.01 by the multidimensional scaling (MDS) [55] (panel **E**), isometric mapping (Isomap) [56] (panel **F**) and locally linear embedding (LLE) [57] algorithms (panel **G**). Colors of data points indicate the ground truth of their corresponding 3D volumes. (**H**) Comparison of local resolution assessment of AlphaCryo4D-classified NLRP3 reconstructions of conformers 2 (upper inset) and 19 (lower inset) from the simulated datasets of three distinct SNRs. The local resolutions were computed by ResMap [70]. Conformers 2 and 19 have been missed 10 and 8 times in 18 control tests using several other methods (see Fig. S3B-G), which are the most and second-most frequently missed conformers, respectively. The 3D classification precision (P) is labelled above each density map. The color bar of local resolution is shown in the lower insert. (**I**) Energy landscape calculated by 3D volumes with preprocessing of 5 Å low-pass filtering instead of standardization. The gap between the majority and minority data points was widened at different SNRs.

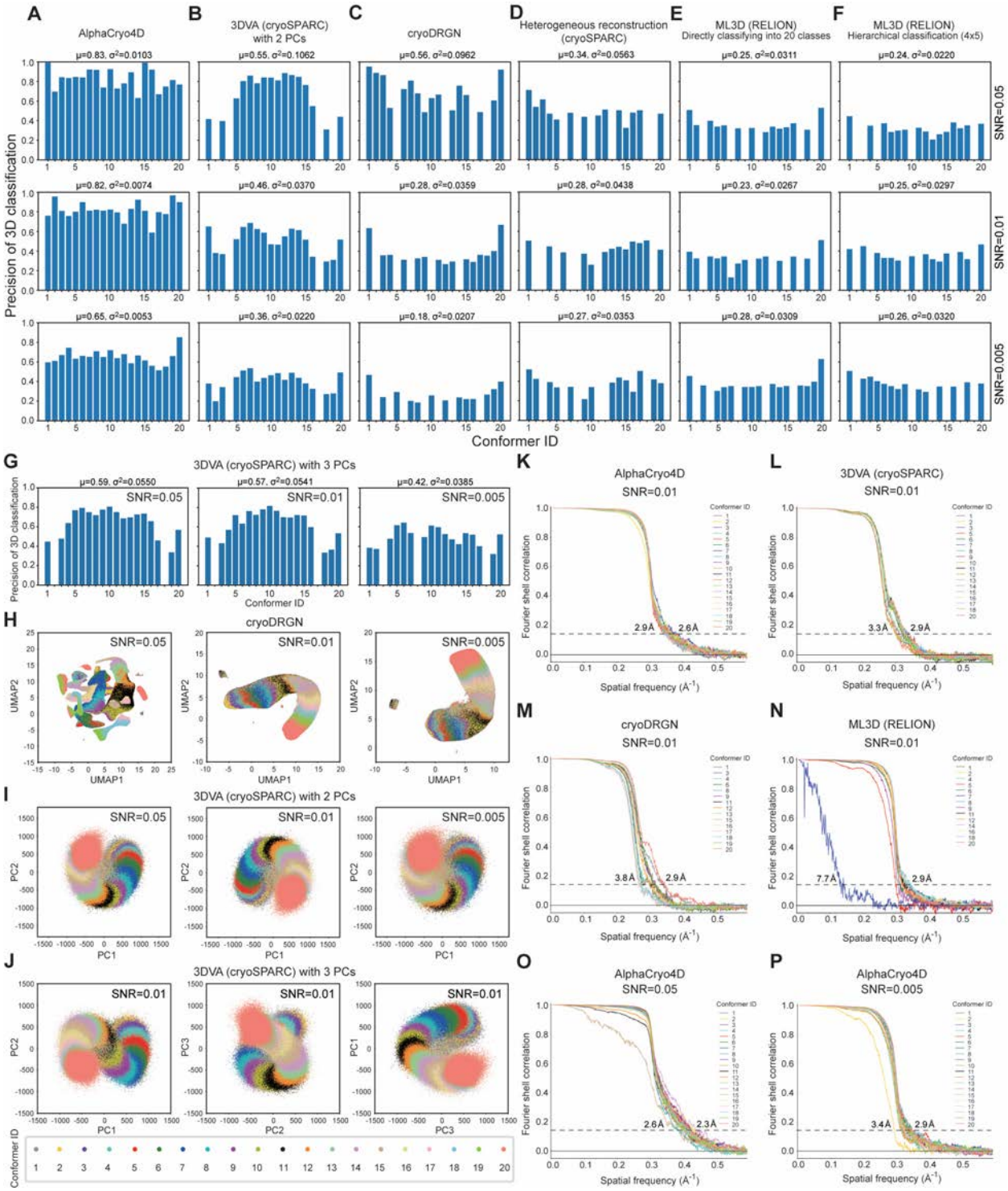

Figure S3. Performance comparison of AlphaCryo4D with alternative methods using the simulated heterogeneous NLRP3 datasets of different SNRs. (**A**-**G**) 3D classification precision of the simulated datasets by AlphaCryo4D (**A**), 3DVA with two principal components (PCs) in cryoSPARC [11] (**B**), cryoDRGN (**C**), heterogeneous reconstruction in cryoSPARC (**D**) and maximum-likelihood classification in RELION (**E**, **F**). The results of SNRs of 0.05 (the first

row), 0.01 (the second row) and 0.005 (the third row) are shown on three rows for side-by-side comparison. On the top of each panel, the symbols of μ and σ^2 denote the mean and variance of precision, respectively, with the values of missing classes treated as zeros. In the maximum-likelihood classification of RELION, both direct and hierarchical strategies are compared in the study. (**G**) 3D classification precision of the simulated datasets by 3DVA using three PCs. (**H-J**) Visualization of 3D classification by cryoDRGN in autoencoder-learned feature space (**H**), and by 3DVA with 2 PCs (**I**) and 3 PCs (**J**). (**K-N**) The gold-standard FSC plots of the 20 maps resulting from 3D classification by AlphaCryo4D (**K**), of the 18 maps resulting from the 3D classification by 3DVA in cryoSPARC (**L**), of the 15 maps resulting from cryoDRGN (**M**) and of the 14 maps resulting from the maximum-likelihood 3D classification in RELION (**N**) on the simulated data of 0.01 SNR. They correspond to the precision results presented in the second row of panels (**A**), (**B**), (**C**) and (**E**), respectively. (**O, P**) Additional gold-standard FSC plots of the refined density maps resulting from AlphaCryo4D on the simulated datasets of SNRs of 0.05 and 0.005.

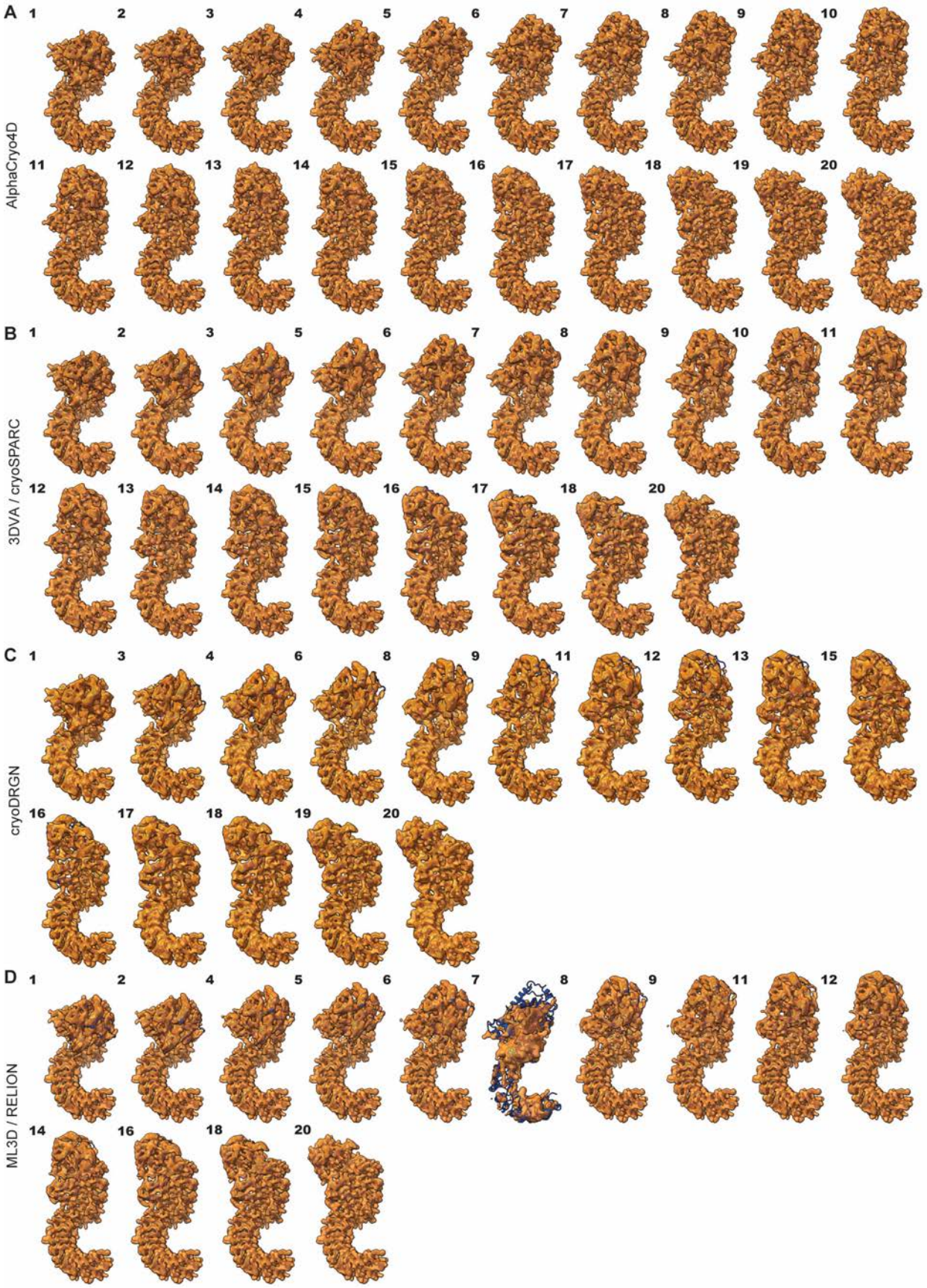
A
AlphaCryo4D
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
B
3DVA / cryoSPARC
1 2 3 5 6 7 8 9 10 11
12 13 14 15 16 17 18 20
C
cryoDRGN
1 3 4 6 8 9 11 12 13 15
16 17 18 19 20
D
ML3D / RELION
1 2 4 5 6 7 8 9 11 12
14 16 18 20

Figure S4. Map assessments of AlphaCryo4D in reconstructions of conformational continuum in comparison with conventional methods on the simulated data of 0.01 SNR. (**A**) The 20 maps of distinct NLRP3 conformers resulting from the 3D classification by AlphaCryo4D. (**B**) The 18 maps of NLRP3 resulting from 3DVA in cryoSPARC. (**C**) The 15 maps of NLRP3 resulting from the 3D classification by cryoDRGN. (**D**) The 14 maps of NLRP3 resulting from the 3D classification by ML3D in RELION. All maps are shown in transparent surface representations superimposed with their corresponding atomic models of the ground truth in cartoon representations, which are fitted to the maps as rigid bodies without further atomic modelling. The conformer ID numbers are marked on the upper left of each map panel. The results shown in panels (**A**), (**B**), (**C**) and (**D**) correspond to the FSC results shown in panels (**K**), (**L**), (**M**) and (**N**) of Fig. S3, respectively.

Figure S5. Performance evaluation of 3D classification with alternative parameters and comparison of the recall of 3D classification using the simulated datasets of three distinct SNRs. (**A**) 3D classification precision of AlphaCryo4D, 3DVA using two PCs, 3DVA using three PCs and cryoDRGN when the class number K was set to 25 or 30, which are more than the class number 20 of ground truths. (**B**) 3D classification recall of AlphaCryo4D, 3DVA using two PCs, cryoDRGN, heterogeneous reconstruction of cryoSPARC and maximum-likelihood classification of RELION at SNRs of 0.05 (the first row), 0.01 (the second row) and 0.005 (the third row). Particles recall of each state was calculated as the ratio of true positive particles in all

ground truth particles. In all methods, the class number was set to 20. (**C**) 3D classification recall of AlphaCryo4D, 3DVA using two PCs, 3DVA using three PCs and cryoDRGN when the class number was set to 25 or 30 (more than the class number 20 of ground truths). (**D**) 3D classification recall of the simulated datasets by 3DVA using three PCs.

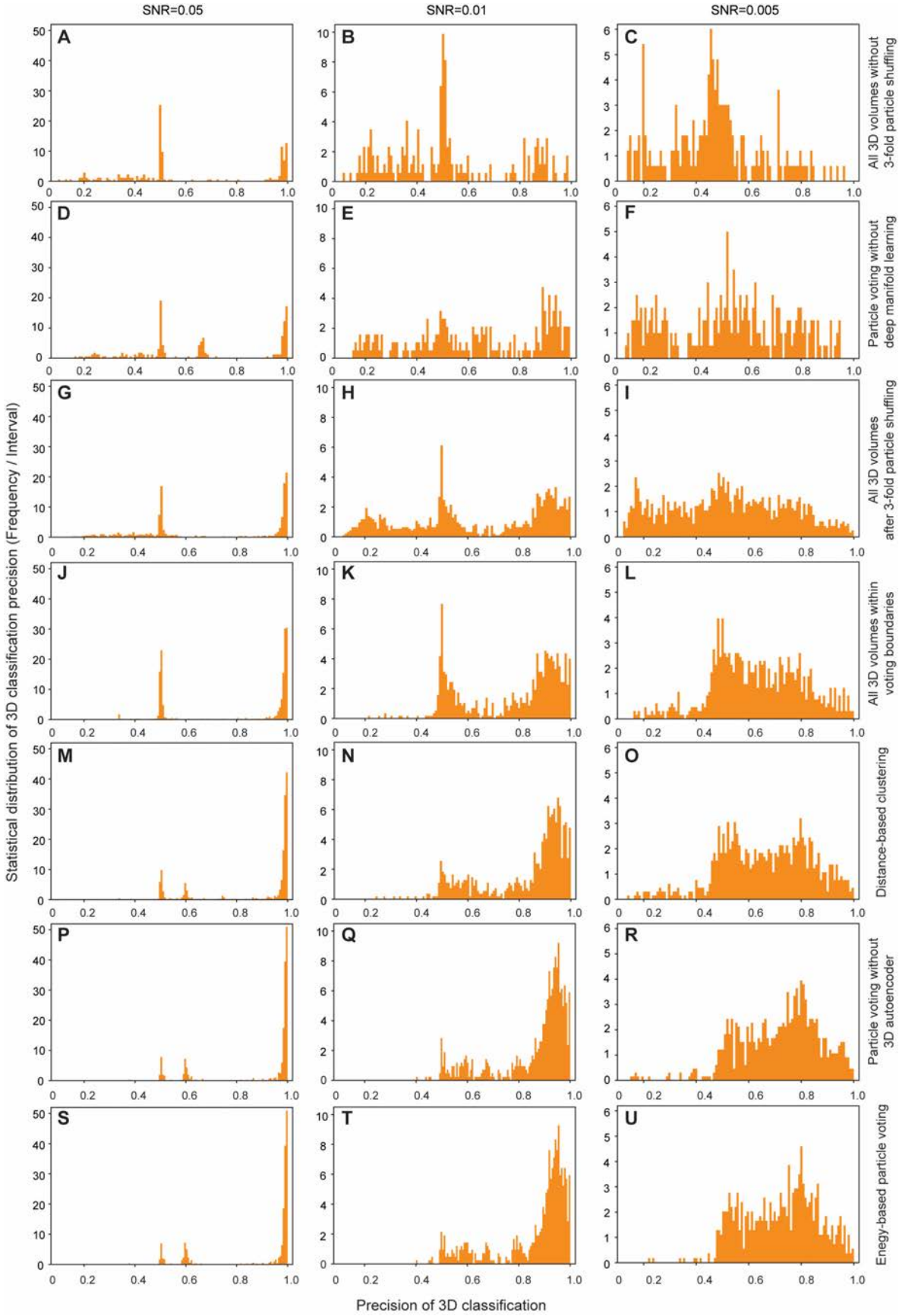
SNR=0.05
SNR=0.01
SNR=0.005
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
All 3D volumes without 3-fold particle shuffling
Particle voting without deep manifold learning
All 3D volumes after 3-fold particle shuffling
All 3D volumes within voting boundaries
Distance-based clustering
Particle voting without 3D autoencoder
Enegy-based particle voting
Statistical distribution of 3D classification precision (Frequency / Interval)
Precision of 3D classification

Figure S6. Mechanistic characterizations of the improvement of 3D classification accuracy by AlphaCryo4D using the simulated NLRP3 datasets of three typical SNRs. Left, middle and right vertical columns show the precision analysis on the simulated datasets with SNRs of 0.05, 0.01 and 0.005, respectively. In total, there are 18 conditional controls under six different algorithmic design variations being analyzed in panel (**A**)-(**R**). (**A-C**) Statistical distribution of 3D classification precision in bootstrapped 3D volumes without particle shuffling. (**D-F**) Statistical distribution of 3D classification precision by implementing particle voting directly on the bootstrapped 3D volumes without using deep manifold learning. The procedure of 3-fold particle shuffling and volume bootstrapping is identical to AlphaCryo4D. (**G-I**) Distribution of 3D classification precision in bootstrapped 3D volumes after 3-fold particle shuffling in the intermediate step of AlphaCryo4D. (**J-L**) Distribution of 3D classification precision in bootstrapped 3D volumes screened by voting boundary on the energy landscape in the intermediate step of AlphaCryo4D. (**M-O**) Distribution of 3D classification precision after distance-based 3D clustering in the absence of energy-based particle voting with all prior steps identical to AlphaCryo4D. (**P-R**) Distribution of 3D classification precision after particle voting by a modified AlphaCryo4D variation only without using deep residual autoencoder in the manifold embedding step. (**S-U**) Distribution of 3D classification precision after energy-based particle voting via a complete AlphaCryo4D procedure.

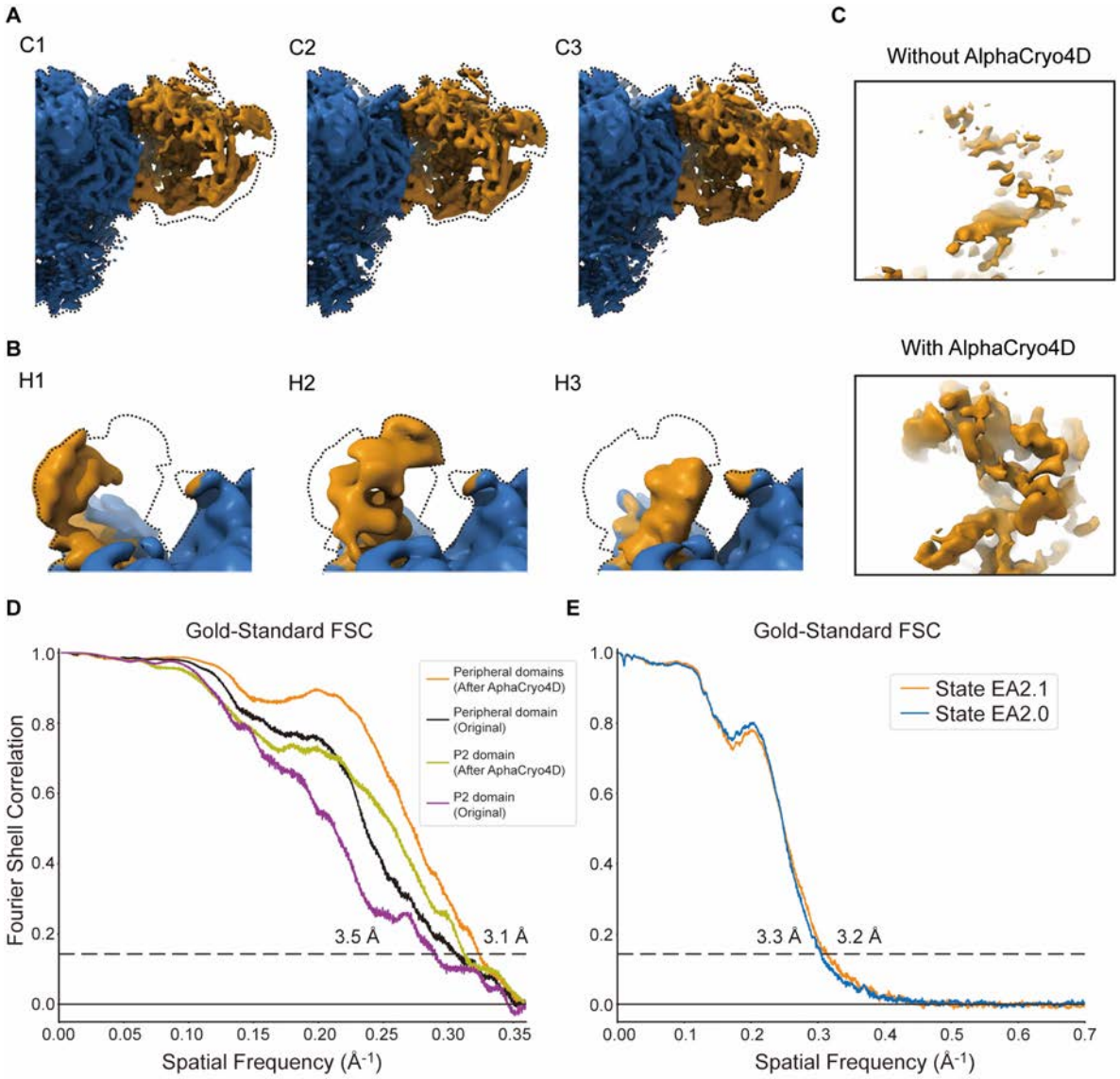

Figure S7. Different states of the RyR1 and resolution measure of 26S proteasome reconstructed by AlphaCryo4D. (**A**) Density maps of clusters C1, C2 and C3 from the energy landscape of the peripheral domain. The dashed line indicated the outline of all these density maps superimposed together. The dynamic region of the density map is colored yellow and the stable region is colored blue. (**B**) Density maps of clusters H1, H2 and H3 on the energy landscape of the P2 domain. (**C**) Comparison of the density map of the P2 domain after AlphaCryo4D with the original state 1. (**D**) Gold-standard FSC of the peripheral and P2 domains with and without AlphaCryo4D. (**E**) Gold-standard FSC of states $E_{A2.0}$ and $E_{A2.1}$ of 26S proteasome reconstructed by AlphaCryo4D.

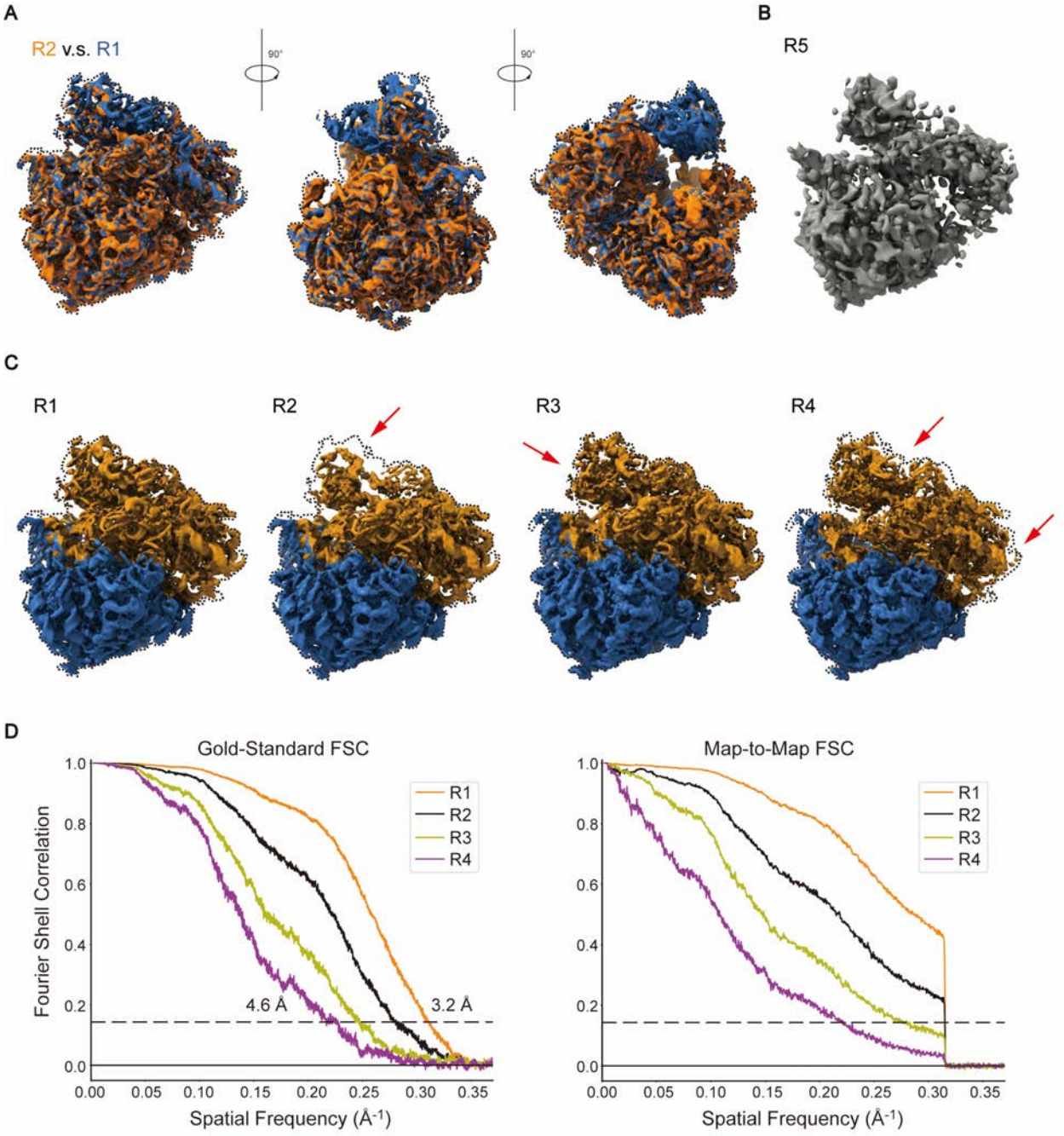

Figure S8. Heterogeneous analysis of the pf80S ribosome reconstructed by AlphaCryo4D. (**A**) Cryo-EM density map of the 40S head missing state R2 superimposed with the major state R1 from three different viewing angles. (**B**) Density map of cluster R5 on the energy landscape (Figure 5A). The particle number of this cluster limited the resolution of its final reconstructed density map. (**C**) Comparison of the density maps of clusters R1-R4. The red arrows emphasized the conformational difference of states R2-R4 with the state R1. (**D**) Gold-standard FSC of the R1-R4 states reconstructed by AlphaCryo4D. (**E**) Map-to-map FSC of the R1-R4 states with the

published cryo-EM structure of the pf80S ribosome (EMD-2660). The cliff of the FSC curve was caused by the low-pass filtering of released cryo-EM maps.

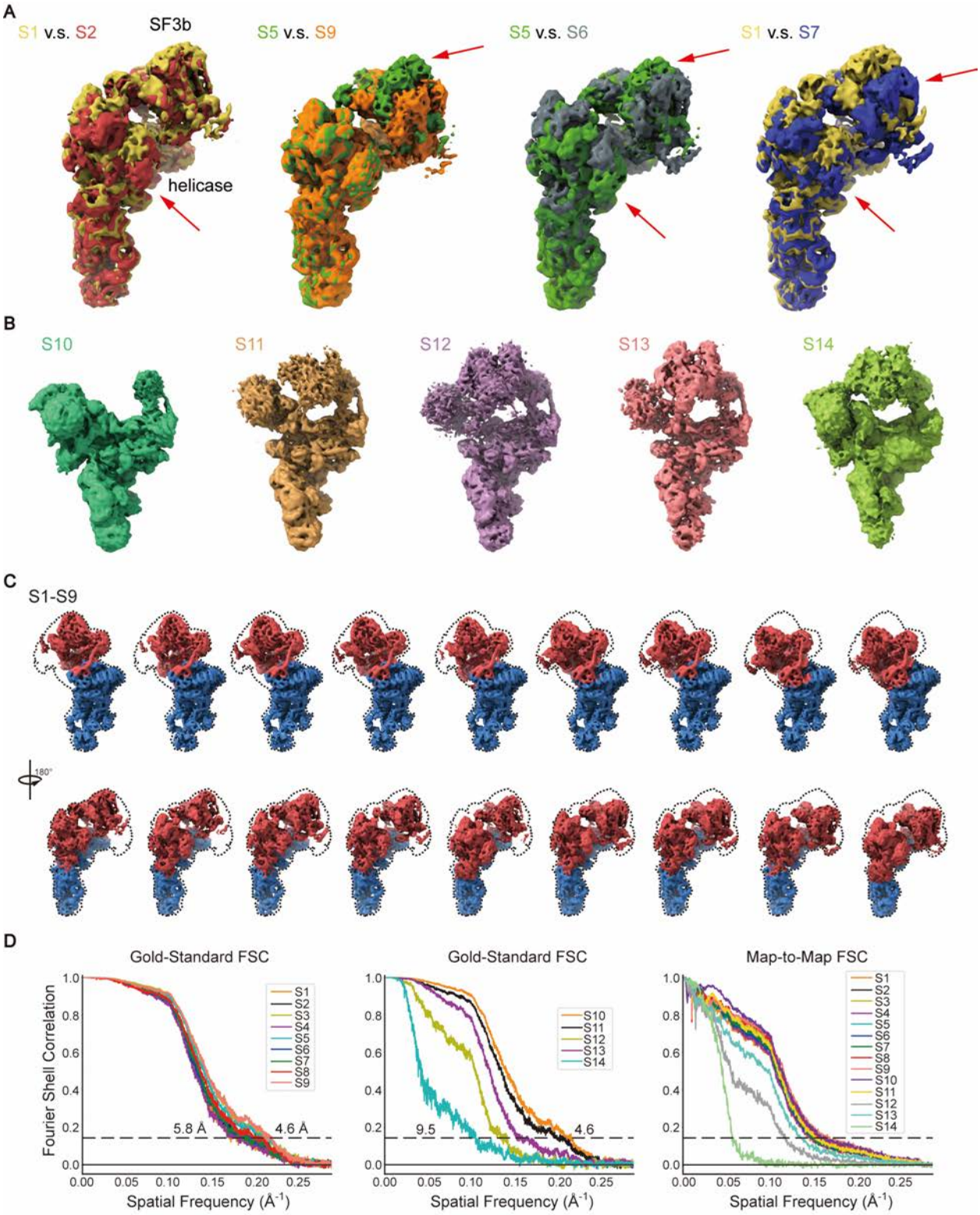

Figure S9. Multiple motion patterns of pre-catalytic spliceosome. (A) Different potential motion modes of the SF3b and helicase subcomplexes. The red arrows pointed the regions where the conformational change happened. These four comparisons showed that the continuous

motion the SF3b or helicase subcomplexes is independent. **(B)** Density maps of clusters S10-S14. Discrete dynamics occurred in these states. **(C)** Cryo-EM density maps of Clusters S1-S9 from other two viewing angles. The dynamic region of the density map is colored red and the rigid region is colored blue. Continuous dynamics was observed between these states. (**D**) Gold-standard FSC of the S1-S14 states classified by AlphaCryo4D. (**E**) Map-to-map FSC of the S1-S14 states over the published B4 cryo-EM density map of the pre-catalytic spliceosome (EMD-3685).

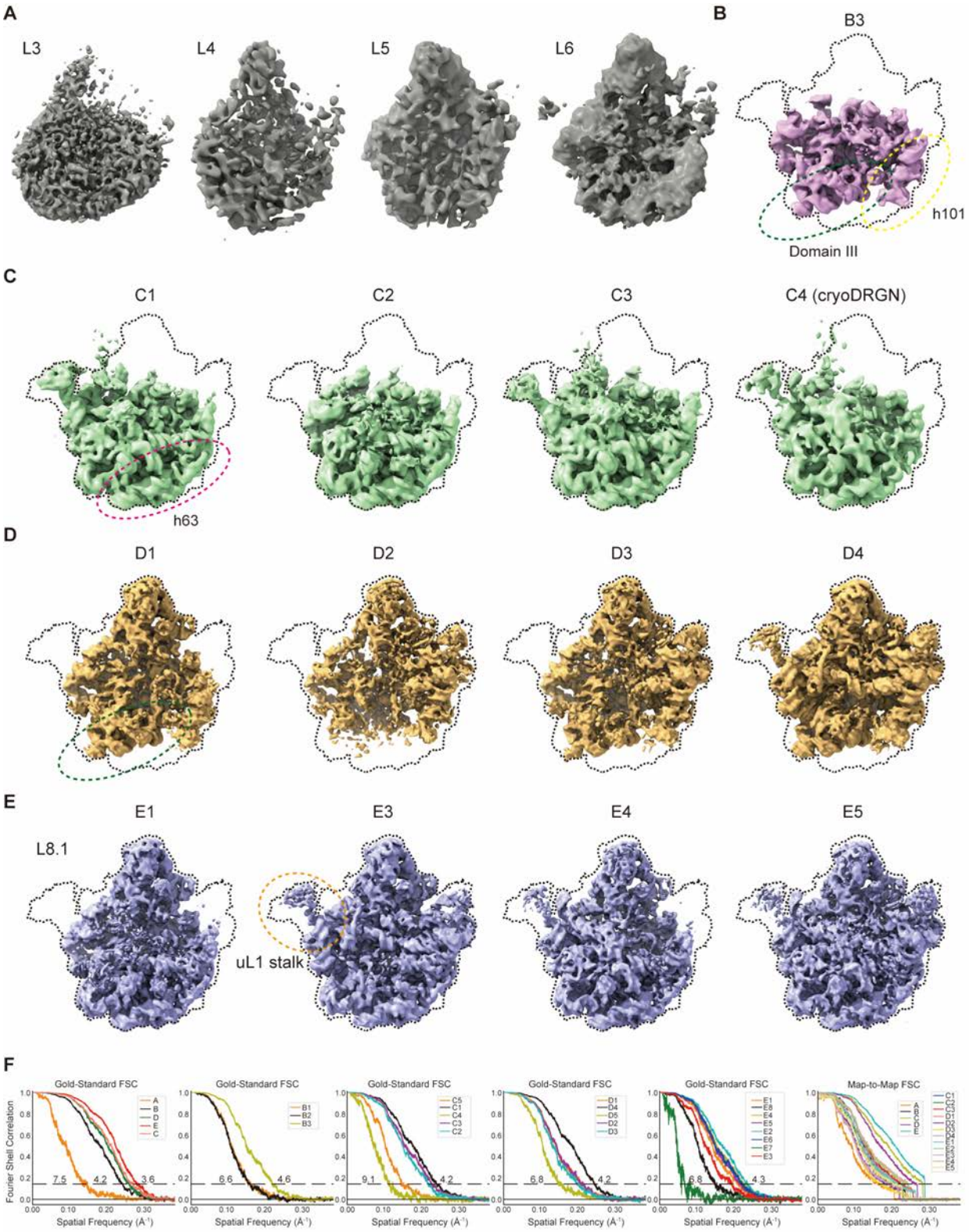

Figure S10. Reconstructions of bacterial ribosomal assembly intermediates by AlphaCryo4D. (**A**) Cryo-EM density maps of clusters L3-L6. The classes contained particles of likely poor quality. (**B**) Reconstruction of the cluster B3 from the state B energy landscape

(Figure 7D). The cluster B3 revealed the main conformation identical to state B. (**C**) Reconstructions of the state C1, C2, C3 and C4 from the state C energy landscape (Figure 7F). The state C4 contained the density of rRNA helix 68, which was previously found by cryoDRGN. (**D**) Reconstructions of the state D1, D2, D3 and D4 from the state D energy landscape (Figure 7H). (**E**) Reconstructions of the state E1, E3, E4 and E5 from the state E energy landscape (Figure 7J). (**F**) Gold-standard FSC of the 5 major states (A, B, C, D and E) and 21 minor states (B1-B3, C1-C5, D1-D5 and E1-E8) reconstructed by AlphaCryo4D. (**G**) Map-to-map FSC of the existing states over the published cryo-EM structures of the states A (EMD-8434), B (EMD-8440), C (EMD-8441), D (EMD-8445), E (EMD-8450), C1 (EMD-8442), C2 (EMD-8443), C3 (EMD-8444), D1 (EMD-8446), D2 (EMD-8447), D3 (EMD-8448), D4 (EMD-8449), E1 (EMD-8451), E2 (EMD-8452), E3 (EMD-8453), E4 (EMD-8455) and E5 (EMD-8456) of the bacterial ribosomal assembly intermediates.

Table S1. Hyperparameters of the deep residual networks in the 3D autoencoder. Optimizer: Adam. Epochs: 50. Initial learning rate: 0.01. A factor of 0.1 and patience of 3 means that the learning rate will times 0.1 (factor) if the loss function does not improve in 3 (patience) epochs. Batch size: 32 (for the three simulated datasets), 8 (for the substrate-bound proteasome dataset).

Layer name	Output size (simulation)	Output size (RyR1)	Output size (Proteasome)	Output size (pf80S Ribosome)	Output size (Spliceosome)	Output size (Bacterial Ribosomal Intermediates)	Network structure
Conv1	$200 \times 200 \times 200$	$360 \times 360 \times 360$	$300 \times 300 \times 300$	$360 \times 360 \times 360$	$320 \times 320 \times 320$	$320 \times 320 \times 320$	$5 \times 5 \times 5, 2$
Conv2_x	$100 \times 100 \times 100$	$180 \times 180 \times 180$	$150 \times 150 \times 150$	$180 \times 180 \times 180$	$160 \times 160 \times 160$	$160 \times 160 \times 160$	$3 \times 3 \times 3, 2$, stride 2 $3 \times 3 \times 3, 2$
Conv3_x							$3 \times 3 \times 3, 2$ $3 \times 3 \times 3, 2$
Conv4_x	$50 \times 50 \times 50$	$90 \times 90 \times 90$	$75 \times 75 \times 75$	$90 \times 90 \times 90$	$80 \times 80 \times 80$	$80 \times 80 \times 80$	$3 \times 3 \times 3, 1$, stride 2 $3 \times 3 \times 3, 1$
Conv5_x							$3 \times 3 \times 3, 1$ $3 \times 3 \times 3, 1$
Conv6_x (encoding)							$3 \times 3 \times 3, 1$ $3 \times 3 \times 3, 1$
TransConv7_x	$100 \times 100 \times 100$	$180 \times 180 \times 180$	$150 \times 150 \times 150$	$180 \times 180 \times 180$	$160 \times 160 \times 160$	$160 \times 160 \times 160$	$3 \times 3 \times 3, 1$, stride 2 $3 \times 3 \times 3, 1$
TransConv8_x							$3 \times 3 \times 3, 1$ $3 \times 3 \times 3, 1$
TransConv9_x	$200 \times 200 \times 200$	$360 \times 360 \times 360$	$300 \times 300 \times 300$	$360 \times 360 \times 360$	$320 \times 320 \times 320$	$320 \times 320 \times 320$	$3 \times 3 \times 3, 2$, stride 2 $3 \times 3 \times 3, 2$
TransConv10_x							$3 \times 3 \times 3, 2$ $3 \times 3 \times 3, 2$
TransConv11							$5 \times 5 \times 5, 1$