Progress in Biophysics and Molecular Biology 164 (2021) 92–100

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

ENRICH: A fast method to improve the quality of flexible macromolecular reconstructions

Keywords:
Cryo-electron microscopy
Single-particle analysis
Image processing
Optical flow
Heterogeneous particles
Flexible macromolecular reconstructions

A B S T R A C T

Cryo-electron microscopy using single particle analysis requires the computational averaging of thousands of projection images captured from identical macromolecules. However, macromolecules usually present some degree of flexibility showing different conformations. Computational approaches are then required to classify heterogeneous single particle images into homogeneous sets corresponding to different structural states. Nonetheless, sometimes the attainable resolution of reconstructions obtained from these smaller homogeneous sets is compromised because of reduced number of particles or lack of images at certain macromolecular orientations. In these situations, the current solution to improve map resolution is returning to the electron microscope and collect more data. In this work, we present a fast approach to partially overcome this limitation for heterogeneous data sets. Our method is based on deforming and then moving particles between different conformations using an optical flow approach. Particles are then merged into a unique conformation obtaining reconstructions with improved resolution, contrast and signal-to-noise ratio. We present experimental results that show clear improvements in the quality of obtained 3D maps, however, there are also limits to this approach, i.e., the method is restricted to small deformations and cannot determine local patterns of flexibility of small elements, such as secondary structures, which we discuss in the manuscript.

1. Introduction

Cryogenic electron microscopy (cryo-EM) using single particle analysis is a well-known structural technique to achieve three-dimensional reconstructions of macromolecular complexes in their native state. This imaging modality has recently transitioned between being considered a promising structural technique to become a well-established approach, capable of providing high-resolution structural maps routinely. One of the main advantages of this technique is its ability to visualize macromolecule complexes that would be too large or flexible to be tackled by nuclear magnetic resonance or X-ray crystallography. Biological macromolecular complexes are not rigid entities and usually exist in a range of conformations that play a key role in their function. Macromolecular structures at sufficient resolution and in different physiological conformations provide essential insights into understanding their function and are key to efficiently design new drugs [Boland 2017]. However, obtaining high-resolution reconstructions from highly flexible samples remains a challenge. The reconstruction workflow is an inverse ill-posed problem. In cryo-EM of isolated macromolecules, normally referred to as Single Particle Analysis (SPA), macromolecular projection images are obtained without knowledge of the particle orientation. This lack of orientation information added to the low signal-to-noise ratio that characterizes cryo-EM images impose computationally demanding reconstruction procedures as marginalized maximum-likelihood based approaches [Scheres 2007; Scheres 2012a; Lyumkis 2013] or statistically robust techniques [Sorzano 2018]. High-resolution structure determination requires obtaining homogeneous sets of particles in the same state. Flexible regions of complexes will display lower resolutions correlated with the amplitude of movement. In case of macromolecular complexes exhibiting important flexibility or multiple conformations, an added level of complication is a drop in the accuracy of the particle orientation assignment. Thus, particles from a structurally heterogeneous data set must be divided into structurally homogeneous sets or classes. This processing step is called three-dimensional (3D) classification and is usually performed in combination with 3D reconstruction [Scheres 2016]. Computational developments in 3D classification methods have made possible the reconstruction of different conformations at high resolution from heterogeneous cryo-EM datasets [Scheres 2007; Scheres 2012a; Lyumkis 2013; Punjani 2017]. The most successful classification method so far is based on an iterative regularized empirical Bayesian strategy [Scheres 2012a]. This approach is based on determining the parameters of a model that shows the highest probability of being the true one taking into the account the observed data (images) and statistical prior knowledge. Most classification methods divide the data set into discrete groups (classes) with typically different sizes, although continuous classification algorithms exist [Sorzano 2019]. Homogeneous datasets show a linear relationship between the inverse of the obtained resolution (spatial frequency) and the logarithm of the number of

https://doi.org/10.1016/j.pbiomolbio.2021.01.001
0079-6107/

particles employed in the 3D reconstruction [Stagg 2014; Stagg 2008]. Then, a common problem after 3D classification is that the resolution obtained from the computed homogeneous classes may be limited by the number of particles belonging to these sets. Moreover, there are technical reconstruction issues that become more severe when the number of particles is small, such as unevenness of the angular distribution [Sorzano 2001] and even artefacts of the reconstruction method. Indeed, it is common that after 3D classification only one or maybe a few abundant classes (corresponding to the most stable states) can be resolved at high-resolution after refinement. In the worst possible scenario, only modest resolution reconstructions (between 9 and 20 Å) will be attainable for all obtained classes. This situation is typically found when the number of particles was low, or the number of selected classes was high. In all these cases, and even assuming that the 3D classification worked correctly, the number of images in these low populated classes represent a technical problem when aiming for high resolution. Modest resolutions between 9 and 20 Å are not sufficient to observe secondary element structures; thus, it is not possible to fully visualize the conformational changes performed by the macromolecule, complicating the molecular understanding of its function. Currently, there is no computational solution to assist users with such issues. The only option users have is returning to the electron microscope again with the hope of collecting enough additional data to resolve the different conformations at sufficient resolution. If resources are available, of course, this is the best approach, but in many practical cases, it may not be feasible. Recently, it has been proposed solutions to process flexible samples [Nakane 2018; Schilbach 2017]. In [Nakane 2018] is presented an approach that performs separated focused refinements with iteratively signal subtractions of rigid bodies defined by the user. In [Schilbach 2017] authors showed the capabilities of a new software called WarpCraft, which uses normal mode analysis to model the motions between different regions of a cryo-EM map.

In this work, we propose a fast and efficient computational approach that can increase the resolution of modestly populated classes without requiring collecting more data, so it is aimed at "making the best" from relatively limited data sets. As an example, our approach required only 2 hours and 50 minutes to process 45059 particles from the 80S ribosome presented below using 10 CPUs. The idea has been inspired by classical work in the field of two-dimensional (2D) electron crystallography, where a less-than-perfect 2D crystal is modified (distorted) so that an artificially improved crystal is created [Henderson 1986; Gil 2006]. The computational method proposed shares some conceptual similarities with the above described 2D crystallography approach. Our method focuses on artificially increasing the number of particles in a class by distorting particles from other similar classes and "moving" them to this class. In this way, as the approach artificially increases the number of images in a small class, then, many technical issues associated with the calculation of a map can be better tolerated, leading to increased resolution reconstructions. However, the big question here is how particles can be "moved" between different conformations, so that, and within specified limits, we can increase the quality of these maps in an objective and quantitative manner.

In this work, we use a regularized optical flow (OF) approach [Bouguet 2001] to move (deform) particles between different classes/conformations. We have applied this OF method in the past in the context of frame alignment in cryo-EM [Abrishami 2015], where we showed that this algorithm is useful to accurately capture local movements between movie frames. In this case, we use OF with a different goal, which is moving (distorting) particles between two (or more) different conformations (input and reference). In more concrete terms, for a given input particle with orientation Θ and belonging to the input conformation, our approach first computes the 2D OF motion field between the two conformations (input and reference) at orientation Θ. To this end, the map conformations are low-pass filtered and projected at the given orientation Θ, then, the 2D motion field is determined by optical flow approach that has been elaborated in the supplemental information, STAR Methods section. Map projections are not highly affected by noise, decreasing the difficulties for a robust estimation of the 2D motion field. The obtained motion field is then applied to the input particle which is, therefore, moved from the input conformation to the reference conformation. This process is repeated for all particles belonging to the input conformation. Here, we refer to the reference conformation as the one to be improved and to the input −or inputs− conformations as the ones from which particles are to be moved to the reference one. We envision two clear applications of the proposed approach, both aimed at mitigating map reconstruction problems in classes having a limited number of particles. One is the case when 3D classification results in one abundant and one or several additional scarce classes. Then, our aim is to push the resolution of the scarce classes moving particles from the abundant conformation. The other application is based on increasing the number of particles in conformations where all the obtained classes are not highly populated and then the resultant reconstructions are limited by the number of particles. In this case, particles from all classes are moved to one reference conformation.

2. Results

The goal of this method is to improve the quality of a reference 3D class by moving particles from an input conformation (or conformations) to that reference one; we refer to this approach as enRICH (Resolution Improvement in Conformational Heterogeneity maps). The inputs of the method are therefore the maps representing the different conformations and the particles to move. The particles have to be aligned previously, thus, the method requires the following steps: 1) 3D classification to obtain the different map conformations and the particles associated with them; 2) 3D refinement to obtain accurate particle angular assignments; 3) enRICH to move particles from one (or several) input conformations to a reference one; 4) 3D refinement of merged particles belonging to the reference conformation. This pipeline is summarized in Fig. 1. In this figure, as an example, our goal is to improve the quality of Class 3 using data from Class 2. The input of enRICH is then the map and particles from Class 2 and the map of Class 3. Therefore, Class 3 is the reference conformation and Class 2 is the input conformation. Steps 1) and 2) from the list above ends with the refinement of particles of each Class. enRICH then starts in Step 3), where we take as input maps from Class 2 and 3 and particles from Class 2. EnRICH produces as output a new set of Class 2 optical flow-deformed particles that we refer to in Fig. 1 as "Particles Class2 to Class3". The workflow continues in Step 4) by merging and refining together corrected particles from Class 2 and particles from Class 3 against Class 3 map, then, obtaining an improved quality map (within certain limits). Observe that enRICH can be used between any classes of Fig. 1, but that in this example we have illustrated its use with Class 2 and Class 3 only. Our workflow uses 3D classification and refinement methods from Relion 2.0 electron microscopy package (Kimanius et al., 2015) however, any method for 3D classification and refinement can be used as well. In supplementary material we provide detailed information about how particles are moved between different conformations by optical-flow approach, strategies followed to avoid model bias and discussion about limitations of the method.

We have used the proposed method with data coming from the

Plasmodium falciparum 80S ribosome, EMPIAR ID 10028 [Wong 2014], and from the type-1 ryanodine receptor (RyR1) [DesGeorges 2016].

2.1. Case with moderate flexibility: 80S ribosome data set

Ribosomal particles were obtained with a FEI Polara 300 microscope equipped with a Falcon II camera. The number of projection images deposited (EMPIAR-10028) is 105,145. We first classified the data set in three classes using Relion through Scipion framework [DelaRosa 2016]. Class 1 (25,323 particles) was composed by particles without 40S subunit, so these particles were not taken into consideration in subsequent analyses. The rest of particles (79,822 particles), distributed in Class 2 (45,059 particles) and Class 3 (34,763 particles), were further refined independently obtaining reconstructions of 4.4 Å and 4.6 Å, which were improved to 3.4 Å and 3.6 Å respectively after postprocessing (masking and B-factor correction). We then used enRICH to improve first Class 2 and then Class 3 reconstructions. The processing times were 2 hours and 50 minutes and 6 hours to deform the sets of 34,763 and 45,059 particles using 10 and 5 CPUs respectively. After refinement of merged particles, we obtained resolutions of 4 Å and 4 Å (without postprocessing), and of 3.2 Å and 3.2 Å (with postprocessing) for Class 2 and 3 respectively. The local resolution analysis of these results obtained by MonoRes [Vilas 2018] are shown in Fig. S5 (Class 2) and S6 (Class 3). These figures clearly show that enRICH improves the local resolution for Class 2 and Class 3 when compared with the reconstructions obtained without any correction. The corresponding FSCs of these results are shown in Fig. 2(a) and (b). As can be seen from these figures, there is a substantial improvement in the resolution for both Class 2 and Class 3 after running enRICH. In all cases, the obtained FSCs using enRICH surpass the ones attained without using enRICH. In Fig. 2(c–f) we show map slides of reconstructions obtained for Class 2 with (c) and without using enRICH (d) and for Class 3 with (e) and without using enRICH (f). Fig. 2(c–f) are shown without any postprocessing. Fig. 2(g) shows 3D maps of the "head" of the smaller ribosomal subunit obtained with enRICH (left) and with the consensus reconstruction using the complete set of 79,822 particles but without any correction (right). This region corresponds to the more flexible part of the structure and is marked with a red rectangle in Fig. 2(c). These maps were masked by the same soft-mask and are shown at the same density threshold value. As can be seen from Fig. 2(g) the map obtained with enRICH shows better structural details and the densities are not broken as happen in the other case. We have also compared the FSC curves obtained from reconstructions computed using all particles (79,822 particles) without applying any 3D classification or correction (consensus reconstruction) with the reconstruction computed from all particles after correction with enRICH, where we used Class 2 as the reference map. These results are shown in Fig. S7, which shows an improvement in the FSC obtained after correction with enRICH. These results show that enRICH is able, in this case, to correct the movements of the small ribosomal subunit improving, therefore, the resolution of the 3D reconstruction. Moreover, in Fig. S8, we show zoomed density structures obtained from the reconstruction computed using all particles after correction with enRICH (left) and from the consensus reconstruction computed using all particles without applying any 3D classification or correction (right). In all cases, the corresponding atomic model fitted to the density map is also shown. In Fig. S8 (A-B), we show a loop and a β-sheet located in the small and flexible 40S ribosomal subunit, while in C) an α-helix and a loop located in the large 60S subunit are rendered. Black arrows and rectangles are used to show the regions that present improvements after correction with enRICH. Note that in Fig. S8 (C) we had to set the

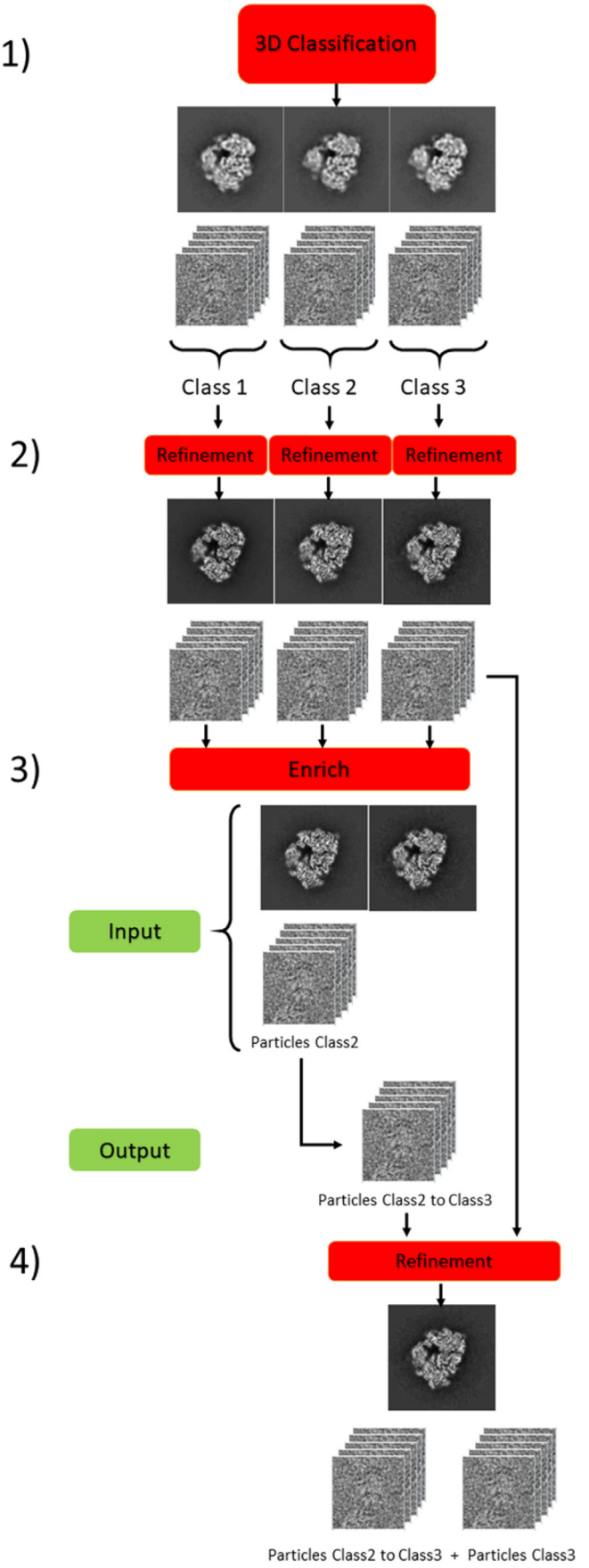

Fig. 1. Diagram of the workflow used to improve reconstructions of heterogeneous samples with enRICH.

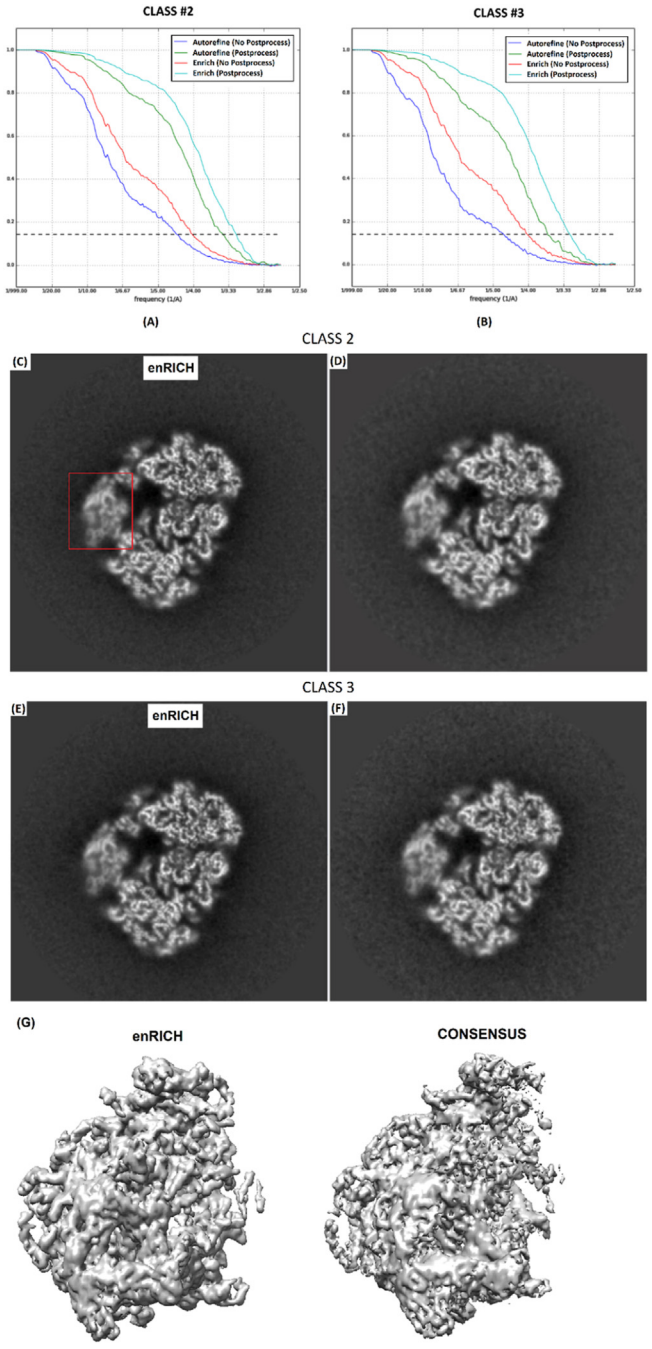

Fig. 2. (a-b) obtained FSCs for Class 2 (a) and Class 3 (b) after refinement for particles belonging to these classes (blue and green lines) and for the complete data set after correction by enRICH (red and cyan). Blue and green curves show results without applying any postprocessing, while red and cyan curves present results after masking and performing B-factor correction to reconstructions. (c-f) respective map slides of reconstructions obtained for Class 2 with (c) and without using enRICH (d) and for Class 3 with (e) and without using enRICH (f). The red mark corresponds to the "head" region of the smaller subunit. In (g) 3D reconstructions of the "head" region are shown after correction with enRICH (left) and without any correction by enRICH and using the complete data set (consensus reconstruction). In both cases the number of particles is 79,822.

threshold to a low value where the noise started to corrupt the consensus map (right) to completely render the α-helix in the case of the consensus reconstruction (right). In general, the consensus reconstruction is noisier that the enRICHed one.

We could not find differences between the maps obtained with enRICH using and without using the gold-standard approach. In this context, enRICH gold-standard means that every particle is deformed using the 2D motion field calculated from input and reference maps having the same half_id value than this particle, while non using gold-standard in enRICH just means that all particles were processed together using only one map as reference and another one as input. In other words, we used the same concept of the gold-standard in calculating the density maps resolution (see Avoiding model bias in enRICH in supplementary information). Therefore, particles with half_id = 1 (or 2) are deformed by the OF 2D motion field using map projections from maps with half_id = 1 (or 2) exclusively. This result shows that enRICH is not affected by overfitting. In Fig. 3 we compare FSCs of reconstructions obtained with enRICH using and not using a gold-standard approach for Classes 2 and 3. Fig. 3 shows no essential differences between reconstructions in both cases. As can be seen from Fig. 2(a–b), the FSCs obtained after Relion auto-refine without postprocessing (blue lines) drops off to ~0.9-0.8 in the rage between 10 and 20 Å. These small differences between maps can be propagated to the slightly differences in the FSCs shown in Fig. 3. The different 0.143-FSC resolution results are summarized in Table S1.

We compared also the obtained maps with the corresponding PDBs (PDB codes: 3j79/3j7a), since it was critical to show that at this relatively high-resolution reconstruction, the improved maps still conserved coherent structural information. PDB 3j79 corresponds to the bigger 60S subunit, while PDB 3j7a refers to the smaller 40S one. Fig. 4 shows the obtained results for Class 2 and 3 using (solid line) and not using (dash line) enRICH approach. The obtained FSC resolutions are presented in Table S2. As can be seen from Fig. 4, enRICH improves the different FSC curves for both classes at almost all frequencies.

Moreover, we compared Class 2 and 3 maps with the corresponding PDBs, using EMRinger, before and after applying enRICH, to show the improvement in enRICHed maps. EMRinger is a tool that assesses the precise fitting of an atomic model into the map and is useful for monitoring progress in resolving and modeling high-resolution features in cryo-EM [Barad 2015]. EMRinger scores can quantify improvements in the resolvability of atomic features owing to improvements obtained by different ways such as motion correction algorithms, and new data collection procedures that balance dose and radiation damages. The higher EMRinger score reports the better-refined structures and higher resolution maps. For the maps in the 3 to 4 Å range, this value is normally above 1.0. To calculate the EMRinger score for our maps, we used the implemented protocol in Scipion in the name of Phenix-EMRinger [DelaRosa 2016]. In the first step, we separated two subunit of each map. Thus, for each map, we have a bigger 60S subunit corresponding and aligned to PDB model 3j79, and a smaller 40S subunit aligned with PDB model 3j7a. Table 1 shows the results of applying EMRinger for each sub-map and corresponding PDB model.

As Table 1 shows, all maps obtained with applying our method (enRICH) have the higher EMRinger score that shows improvements in term of consistency between maps and model. Specifically, smaller 40S subunit that corresponds to the more flexible part of the structure shows better improvement in both enRICHed class 2 and 3.

For the sake of clarity and completeness, we have repeated this experiment comparing the result obtained from enRICH for Class 2 with the results computed using all particles (79,822 particles) without applying any 3D classification or correction (consensus reconstructions) when atomic models are used as references. These results are shown in Fig. S9 (bigger 60S subunit) and S10 (smaller

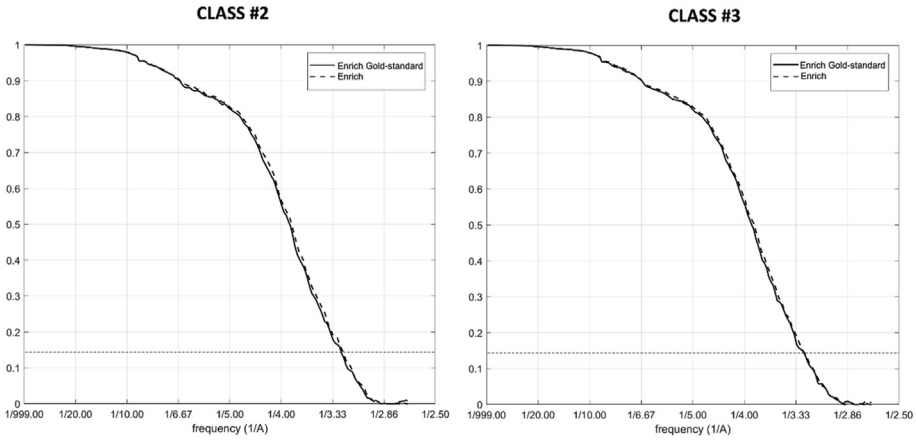

Fig. 3. FSCs of reconstructions obtained after refinement and particle correction with enRICH following and not following a gold-standard approach for Class 2 and Class 3.

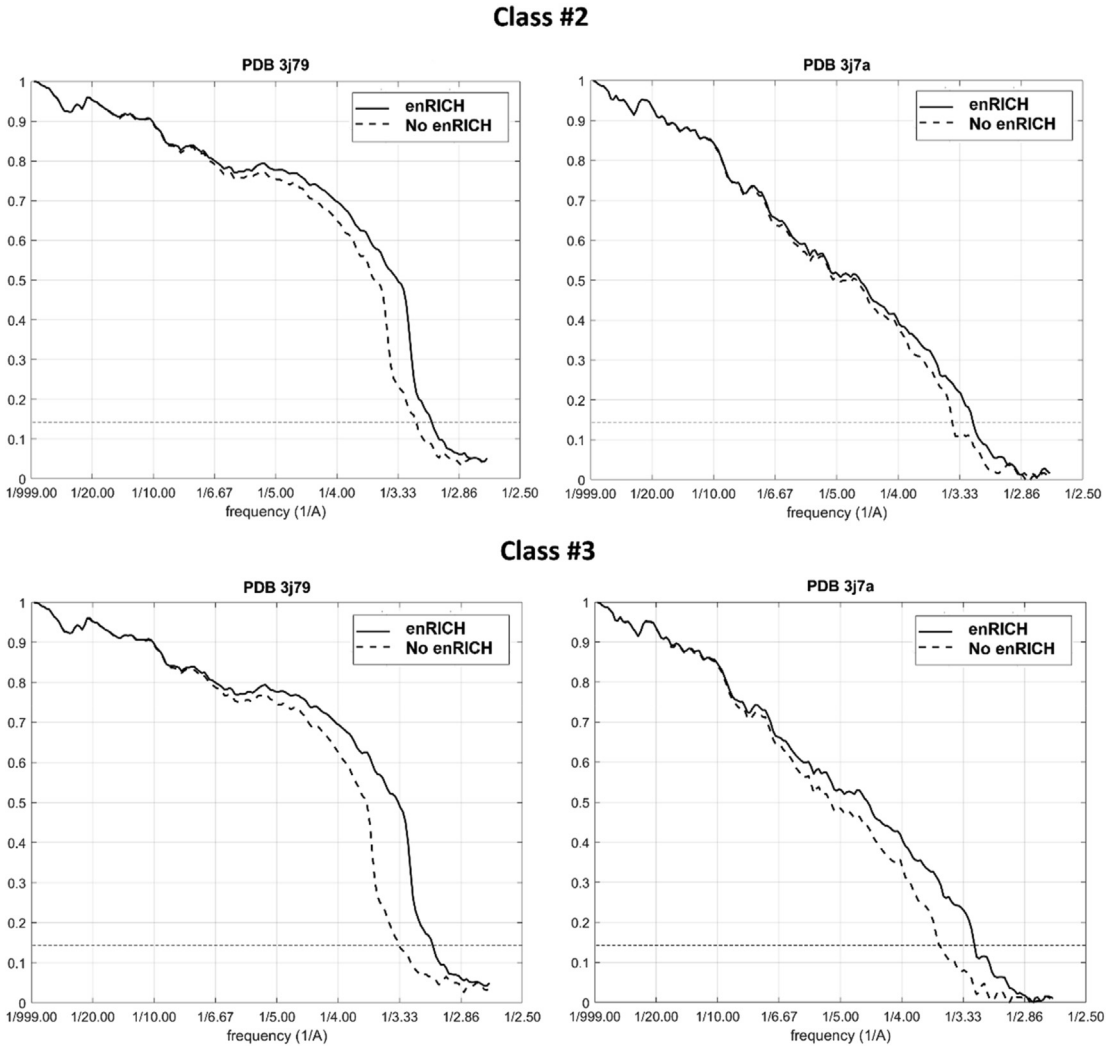

Fig. 4. FSC curves obtained confronting corresponding PDBs with maps obtained when only particles belonging to Class 2 and 3 (dashed curve) were used, and when we employed all the particles after correction by enRICH (solid line).

40S subunit). As can be seen from these figures, enRICH improves the FSCs, especially for the small moving 40S subunit when compared with the results provided by the consensus reconstruction. In Table 1 we show also the EMRinger scores provided by reconstructions obtained from all particles when compared with both the 40S and 60S atomic models. Table 1 shows higher values from enRICHed maps in comparison with results computed from the consensus reconstruction without any 3D classification or

Table 1
EMRinger scores obtained from maps reconstructed using Class 2 and Class 3 particles (with and without using enRICH), and all particles with and without correction with enRICH.

	EMRinger Score
Class2 40S subunit	0.83
Class2 40S subunit (enRICHed)	1.24
Class2 60S subunit	1.5
Class2 60S subunit (enRICHed)	1.55
Class3 40S subunit	0.93
Class3 40S subunit (enRICHed)	1.23
Class3 60S subunit	1.44
Class2+Class3 60S subunit (enRICHed)	1.59
Class2+Class3 40S subunit	1.11
Class2+Class3 40S subunit (enRICHed)	1.48
Class2+Class3 60S subunit	1.47
Class2+Class3 60S subunit (enRICHed)	1.53

correction, which approves the application of our developed method.

Finally, to show the capacity of the approach for classes with very low number of particles, we further classified the dataset composed by merged particles (without any correction) from Classes 2 and 3, into 4 classes. The population of each class was 25,087, 23,702, 5744 and 24,289 respectively. These classes refined to resolutions of 5.3, 5.7, 15.1 and 5.8 Å without any applied mask or B-factor correction. We then applied enRICH to move particles from Class1, 2 and 4 to Class 3, which is the least populated class. Fig. 5 shows the gold-standard FSCs obtained after 3D refinement by Relion auto-refine, without masking or performing B-factor correction for Classes 1, 2, 3 and 4 and after merging all particles to Class 3 using enRICH (blue curve). As can be seen from Fig. 5, the resolution is substantially improved after using enRICH.

In Fig. S11, we have compared the FSC curves of the consensus reconstruction using all particles without applying any 3D classification or correction, with the reconstruction computed from all particles after correcting and merging particles from the four 3D classes with enRICH, where we used Class 3 as reference map. Fig. S11 clearly shows that using enRICH to turn the particles to the class with lowest resolution, although improves the resolution of that class even when compared with the consensus reconstruction using all particles, it produces a reconstruction with lower resolution than the enRICHed one shown in Fig. S7. The main reason is that the reference map in this case has fewer number of particles (~5k). Thus, this reference map is of lower quality causing that the deformation maps computed from enRICH are not accurate enough to provide a better enRICHed reconstruction.

2.2. Case with large flexibility: RyR1 data set

This data set was composed by 277,044 single particles of RyR1 bound to Ca2+, ATP and caffeine. We first classified this data set into two sets composed by 129,873 (Class 1) and 147,171 (Class 2) particles respectively. Class 1 and Class 2 refined until 4.8 Å and 4.9 Å (0.143-FSC gold standard) using Relion. In this case, we found that the large conformational changes shown by the RyR1 prevented our method to improve these results. Indeed, after moving particles with enRICH from Class 2 to Class 1 and refine the merged set, the resolution achieved was 4.9 Å (0.143-FSC gold standard). As mentioned in supplemental information (section limitations of the method), complexes with large conformation movements, as the RyR1 specimen, prevent our method to obtain high-resolution reconstructions. In these situations, our approach should be used to improve maps at intermediate resolutions only. Therefore, in this case it is not realistic the use of enRICH for improving map resolutions better than 5 Å. However, although we could not improve these results, we employed this data set to show a different use of our approach, where enRICH was able to improve intermediate resolution reconstructions coming from low populated classes.

In the second experiment, we decided to further classify Class 2 (147,171 particles) into four classes. The particle distribution between classes and the obtained resolution after refinement were Class 2.1: 38,543 particles, resolution 7.2 Å; Class 2.2: 39,239 particles, resolution 7.2 Å; Class 2.3: 35,174 particles, resolution 7.3 Å; Class 2.4: 34,215 particles, resolution 7.3 Å. We then moved particles from Class 2.2 to Class 2.1, particles from Class 2.2 and from Class 2.3 to Class 2.1, and particles from Class 2.2, Class 2.3 and Class 2.4 to Class 2.1. The processing time to deform ~40,000 particles of 400 × 400 pixels using 44 CPUs was 1 hour and 35 minutes. In each case, we refined the merged sets obtaining respective FSCs from Relion that are shown in Fig. 6(a). As can be seen from Fig. 6(a), the resolution of Class 2.1 was improved from 7.2 (38,543 particles) to 5.3 Å when the dataset was composed by 147,171 particles. Additionally, in Fig. 6(b–c) we show respective map slides (without any postprocessing) obtained after merging all particles by enRICH and

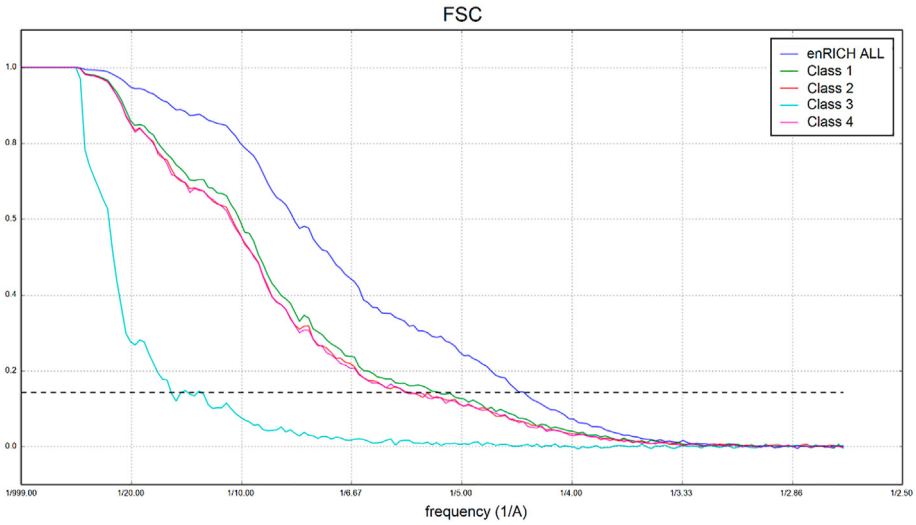

Fig. 5. FSCs obtained for Class 1, 2, 3 and 4 reconstructions (green, red, cyan and pink) and after merging all particles to Class 3 using enRICH (blue line).

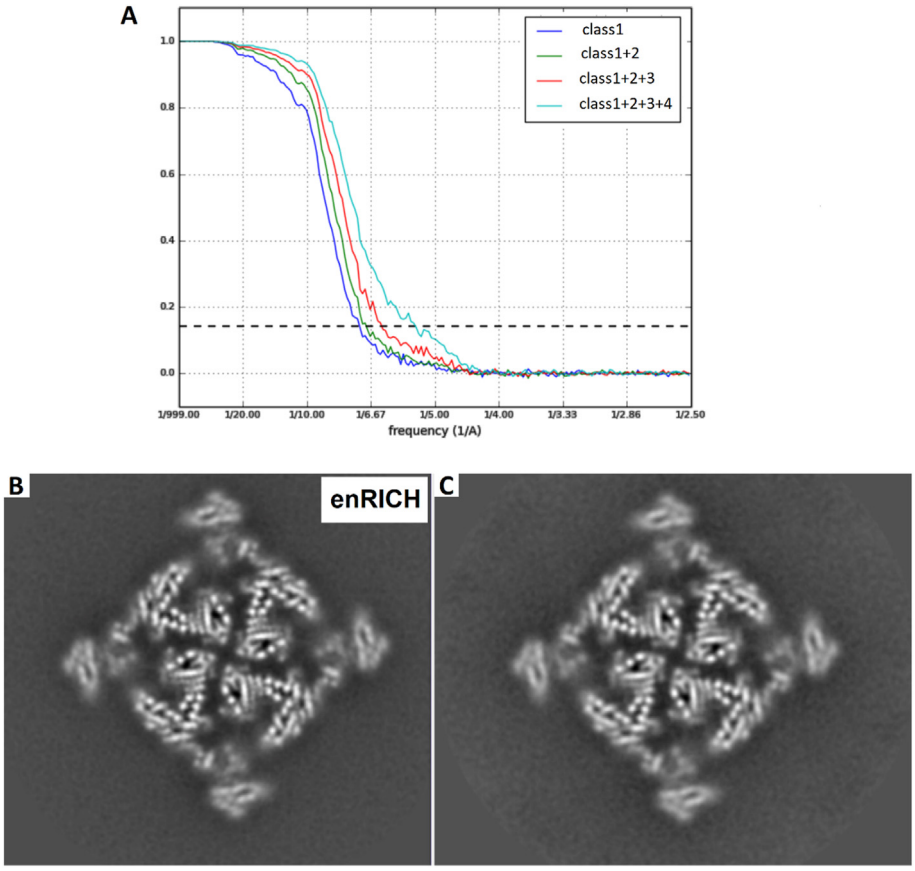

Fig. 6. (a) Gold-standard FSCs obtained from Class 2.1 particles (blue curve), Class 2.1 + Class 2.2 (corrected) (green curve), Class 2.1 + Class 2.2 (corrected) + Class 2.3 (corrected) (red curve), Class 2.1 + Class 2.2 (corrected) + Class 2.3 (corrected) + Class 2.4 (corrected) (cyan curve). (b-c) respective slides of maps obtained after merging all particles by enRICH and then refining with Relion (b), and after refining Class 2.1 data set (38,543 particles) alone (c).

refining with Relion (Fig. 6(b)), and after refining the data set of 38,543 particles alone (Fig. 6(c)). As can be seen from these figures, the map slide obtained after merging all particles with enRICH has better resolution.

For in this experiment, we have also compared the FSC curves of reconstructions obtained from all particles with and without using enRICH in Fig. S12. In this case, using enRICH to turn the particles to the class with lowest resolution (Class2.2), although improves the resolution of that class substantially so it possibilities to visualize that conformation better, also provides a reconstruction showing lower global resolution than the consensus reconstruction using all particles without any correction and 3D classification. Note that this result still is interesting as allow us to visualize the different conformations after 3D classification with improved resolution after correction with enRICH.

3. Discussion

Macromolecules are not static entities and typically exhibit different conformations when performing their function. Cryo-EM is a high-resolution structure determination technique that can capture these different states providing essential information about the macromolecule's function. Hence, 3D classification is a routine and crucial task when processing cryo-EM data. However, classifying data requires dividing the collected data into sets of homogeneous particles, a process that usually limits the attainable resolution for the different conformations. This situation is more dramatic in complexes showing a large (or infinite) number of different possible conformations [Sorzano 2016]. In this work, we propose a method to merge particles from different conformations. This approach can be used in cases where the resulting number of particles after 3D classification limits the attainable resolution. Then, this method can be used to improve (within limits) the quality of reconstructions coming from scarce classes, for example. The approach is based on obtaining the 2D motion or deformation field by a regularized optical flow approach. As shown from our results, the approach is fast and computationally efficient. The OF is computed between projections of different conformations at certain orientations given by the experimental particles to process. These motion fields are then applied to corresponding particle images to move them between conformations. The conceptual idea of the proposed method is close to crystal unbending implemented in 2D electron crystallography [Gil et al., 2006; Henderson et al., 1986] that enabled a dramatic improvement in resolution. The main goal of crystal unbending is to correct the crystal distortions, that is, its deviation from regular lattice repetition. This technique first obtains the 2D motion field of unit cells with respect to their perfect regular positions and then correct these shifts by interpolation obtaining a corrected crystal image. In both 2D crystallography and single particle analysis underlying physical limitations are crystal local deformations in the first case and large-scale low frequencies motions in the second.

We have taken special attention in avoiding problems related with model-bias when using our approach. First, we use a robust

optical flow approach which imposes strong regularization in the obtained 2D motion fields. Typical integration kernels and imposed resolutions to map projections used in the estimation of the motion fields by OF are of 25–50 Å and 15–20 Å, respectively. All the results shown in this work have been obtained using an OF integration kernel of 50 Å and the maps have been filtered to 20 Å resolution. Additionally, enRICH can be used in a gold-standard fashion. In this case the method uses half maps of each conformation (input and reference) to determine motion fields. Particles are corrected according to their corresponding halves.

It is important to mention that our method imposes a deformation model. This model assumes that the 3D motion field between conformations is small, smooth, continuous and then obtainable from map projection images. This assumption is only fulfilled when the 3D movements between conformations is small and, better still, when there is an intrinsic element of continuous flexibility in the macromolecule under study. That is the difference between enRICH and Relion multibody refinement, as if the domains or regions are small, due to masking in Relion multibody approach, the accurate alignment is challenging and most of the times is not possible. However, our focus is on a different problem which is small continuous movement without defining any mask. Also, our approach is addressed to capture and correct movements of large regions as subunits, but not to determine small local changes of smaller elements such as secondary structures. An important advance of enRICH is that it can deal with samples affected by flexibility. In these cases, the data set can be 3D classified into a large number of classes, each affected by low resolution because the limited number of particles per class. Then enrich can merge all particles from these 3D classes into one corrected class with improved resolution.

We have tested our approach on two different classes of ryanodine receptor (EMD-2751, and EMD-2752), and ribosome (EMD-6647, and EMD-6653) by simulating the particle images as has been discussed in supplemental information (section Simulations).

In addition, we have applied our approach to two different heterogeneous samples: 80S ribosome and RyR1. In the case of the 80S ribosome, we clearly improved resolution results obtained using pure classes after 3D classification. This result shows that the deformation pattern followed by these ribosomal particles aligns well with our model. On the other hand, we also showed a case where enRICH could not improve resolutions of homogeneous sets after classification. In this situation our deformation model may not describe well the large movements exhibited by the RyR1 macromolecules at high-resolution, limiting the attainable resolution to ~5 Å. However, we employed this dataset to show other practical uses of our approach. Interesting, our defined control checks (please see 'Theoretical Analysis of enRICH' in supplementary material) to evaluate the capacity of enRICH to process heterogeneous data were 0.26 and 1.17 for the ribosome and RyR1 respectively. These scores agree with the different results obtained.

Author contributions

J.V. had the main idea and wrote the manuscript, J.V and M.K developed the method, A.G provided the RyR1 data. All authors reviewed the manuscript, supervised the experiments and discussed the results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to Rouslan Efremov for helpful discussions and critical reading of the manuscript and to Josue Gomez-Blanco to help us with Chimera and in the evaluation of the maps. We are also thankful for HPC resources provided by Bio21 Institute to perform part of this analysis. This work was supported by McGill start-up funds, the Spanish Ministry of Science and Innovation through the call 2019 Proyectos de I+D+i - RTI Tipo A (PID2019-108850RA-I00), Comunidad de Madrid (S2017/BMD-3817) and the MINECO – Spain (BIO2016-76400-R (AEI/FEDER, UE)). J.V. acknowledges economical support from the Ramón y Cajal 2018 program (RYC2018-024087-I). M.K. acknowledges Isabelle Rouiller for her support and providing some computational resources.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pbiomolbio.2021.01.001.

References

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2015. Alignment of direct detection device micrographs using a robust Optical Flow approach. J. Struct. Biol. 189 (3), 163–176.
Barad, B.A., Echols, N., Wang, R., Cheng, Y., DiMaio, F., Adams, P.D., Fraser, J.S., 2015. EMRinger: side chain-directed model and map validation for #D cryo-electron microscopy. Nat. Methods 12 (10), 943–948.
Boland, A., Chang, L., Barford, D., 2017. The potential of cryo-electron microscopy for structure-based drug design. Essays Biochem. 61 (5), 543–560.
Bouguet, J.Y., 2001. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm, vol. 2. Intel Corporation.
Dela Rosa, J.M., Quintana, A., Del Cano, L., Zaldivar, A., Foche, I., Gutierrez, J., Gomez Blanco, J., Burguet, J., Cuenca Alba, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Vilas, J.L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2016. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. 195 (1), 93–99.
Des Georges, A., Clarke, O.B., Zalk, R., Yuan, Q., Condon, K.J., Grassucci, R.A., Hendrickson, W.A., Marks, A.Q., Frank, J., 2016. Structural basis for gating and activation of RyR1. Cell 167 (22), 145–157.
Gil, D., Carazo, J.M., Marabini, R., 2006, 3. In: On the nature of 2D crystal unbending, 156, pp. 546–555.
Henderson, R., Baldwin, J.M., Downing, K.H., Lepault, J., Zemlin, F., 1986. Structure of purple membrane from Halobacterium Halobium: recording, measurement and evaluation of electron micrographs at 3.5 Á resolution. Ultramicroscopy 19, 147–178.
Kimanius, D., Forsberg, B.O., Scheres, S.H.W., Lindahl, E., 2015. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. ELife 15 (5), e18722.
Lyumkis, D., Brilot, A.F., Theobald, D.L., Grigorieff, N., 2013. Likelihood-based classification of cryo-EM images using FREALIGN. J. Struct. Biol. 183 (3), 377–388.
Nakane, T., Kimanius, D., Lindahl, E., Scheres, S.H.W., 2018. Characterisation of molecular motions in cryo-EM single article data by multo-body refinement in Relion. eLife 7.
Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14 (3), 290–296.
Scheres, S.H.W., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.
Scheres, S.H.W., 2016. Chapter six - processing of structurally heterogeneous cryo-EM data in RELION. Methods Enzymol. 579, 125–157.
Scheres, S.H.W., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P.B., Frank, J., Carazo, J.M., 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods 4, 27–29.
Schilbach, S., Hantsche, M., Tegunov, D., Wigge, C., Urlaub, H., Cramer, P., 2017. Structures of transcription pre-initiation complex with TFIIHand Mediator. Nature 551, 204–209.
Sorzano, C.O.S., Marabini, R., Boisset, N., Rietzel, E., Schröder, R., Herman, G.T., Carazo, J.M., 2001. The effect of overabundant projection directions on 3D reconstruction algorithms. J. Struct. Biol. 133 (2,3), 108–118.
Sorzano, C.O.S., Alvarez-Cabrera, A.L., Kazemi, M., Carazo, J.M., Jonic, S., 2016. StructMap: elastic distance analysis of electron microscopy maps for studying conformational changes. Biophys. J. 110, 1753–1765.
Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Jiménez-Moreno, A., Melero, R., Martínez, M., Ramírez-Aportela, E., Conesa, P., Vilas, J.L., Marabini, R., Carazo, J.M., 2018. A new algorithm for high-resolution reconstruction of Single Particles by Electron Microscopy. J. Struct. Biol. 204, 329–337.

Sorzano, C.O.S., Jiménez, A., Mota, J., Vilas, J.L., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Majtner, T., Segura, J., Sánchez-García, R., Rancel, Y., del Caño, L., Conesa, P., Jonic, S., Vargas, J., Cazals, F., Freyberg, Z., Krieger, J., Bahar, I., Marabini, R., Carazo, J.M., 2019. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. Acta Crystallogr. F 75, 19–32.

Stagg, S.M., Lander, G.C., Quispe, J., Voss, N.R., Cheng, A., Bradlow, H., Bradlow, S., Carragher, B., Potter, C.S., 2008. A test-bed for optimizing high-resolution single particle reconstructions. J. Struct. Biol. 163 (1), 29–39.

Stagg, S.M., Noble, A.J., Spilman, M., Chapman, M.S., 2014. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. J. Struct. Biol. 185 (3), 418–426.

Vilas, J.L., Gómez-Blanco, J., Conesa, P., Melero, R., Miguel de la Rosa-Trevín, J., Otón, J., Cuenca, J., Marabini, R., Carazo, J.M., Vargas, J., Sorzano, C.O.S., 2018. MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. Structure 26 (2), 337–344.

Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J., Scheres, S.H.W., 2014. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. eLife 3, e03080.

M. Kazemi*
Dep. of Biochemistry and Pharmacology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC, 3010, Australia

C.O.S. Sorzano, J.M. Carazo
Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/ Darwin 3, 28049, Cantoblanco, Madrid, Spain

A. des Georges
Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY, 10031, USA

Dept. of Chemistry & Biochemistry, City College of New York, New York, NY, 10031, USA

Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, New York, NY, 10016, USA

V. Abrishami
Laboratory of Structural Biology, Helsinki Institute of Life Science HiLIFE, Finland

J. Vargas**
Departamento de Optica, Universidad Complutense de Madrid, Avda. Computense s/n, Ciudad Universitaria, 28040, Madrid, Spain

Department of Anatomy and Cell Biology, McGill University, 3640, Rue University, Montréal, QC, H3A 0C7, Canada

* Corresponding author.

** Corresponding author. Departamento de Optica, Universidad Complutense de Madrid, Avda. Computense s/n, Ciudad Universitaria, 28040, Madrid, Spain.
E-mail address: mohsen.kazemi@unimelb.edu.au (M. Kazemi).
E-mail address: jvargas@fis.ucm.es (J. Vargas).

16 July 2020
Available online 12 January 2021

Supplemental information

ENRICH: a fast method to improve the quality of flexible macromolecular reconstructions

M. Kazemi[*], C. O. S. Sorzano, A. des Georges, V. Abrishami, J. M. Carazo, J. Vargas[*]

STAR Methods

Movement of particles between different conformations

In order to move one particle with orientation Θ from conformation *A* (input) to conformation *B* (reference), we need to estimate a 2D motion field that maps projected densities from *A* to *B* at orientation Θ. The projected densities can be directly obtained projecting aligned maps from conformations *A* and *B* at the particle orientation. The 2D motion field can be then determined by optical flow (OF) approach between these two map projections. Here, we use a pyramidal implementation of the Lucas–Kanade OF algorithm with iterative refinement. In the following, we give a summary explanation of this method. For a more detailed description the interested reader is referred to [Abrishami2015; Bouguet2001]. In Figure S1 we show an illustration of the conceptual idea used in Enrich to move particles between conformations.

Optical flow is a family of algorithms originally introduced in Computer Vision to analyse the movement of objects with respect a common reference frame fixed to the camera. To this end, two consecutive images (*I*) are obtained at times *t* and *t*+ Δ*t* at similar imaging conditions. The intensities of objects within the images at different times do not change and it is their positions that are only modified. Then, the following equation, termed as brightness constancy equation holds

$$I(x,y,t) \cong I(x+\Delta x, y+\Delta y, t+\Delta t) \approx I(x,y,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \quad \text{(S1)}$$

Note that $\Delta x = \Delta x(x,y)$ and $\Delta y = \Delta y(x,y)$ refer to the apparent movement of objects between consecutive images. In Eq. (S1), we have assumed that the object movements between images are

small justifying a first order Taylor expansion. Defining the local shift components (u, v) of the 2D motion field as $u = \Delta x/\Delta t$ and $v = \Delta y/\Delta t$ we obtain the gradient constrain equation as

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \qquad \text{(S2)}$$

This expression provides one constrain at every pixel, but our goal is to determine two magnitudes (u, v), thus, we have an underdetermined system of equations and is not possible to determine univocally (u, v) at every pixel. Lucas-Kanade approach solves this problem imposing the shifts to be similar between close points in the image. Therefore, this approach defines a kernel of size $w \times w$ around each pixel and imposes the movement to be the same for all pixels inside the kernel, obtaining an overdetermined system of equations. One limitation of the approach outlined above is that it is restricted to cases with small shifts between images that justify the first order Taylor expansion employed in Eq. (S1). In cases where this first order approximation is not accurate, it is necessary to iterate multiple times on this scheme. Therefore, after the k^{th} iteration, the brightness constancy equation corresponds to

$$I(x, y, t) \cong I\Big((x + \Delta x^k) + \Delta x^{k+1}, (y + \Delta y^k) + \Delta y^{k+1}, t + \Delta t\Big) \qquad \text{(S3)}$$

The final displacement vectors obtained after K iterations are as

$$u = \sum_{k=1}^{K} \Delta x^k/\Delta t, \;\; v = \sum_{k=1}^{K} \Delta y^k/\Delta t \qquad \text{(S4)}$$

A key parameter of the LK algorithm is the kernel size. This kernel introduces a trade-off between accuracy and robustness of the approach. Large kernels give more robustness to noise and outliers, but less accuracy. Additionally, large kernels are required to handle large motions between images [Bouguet2001]. To solve these limitations, in [Bouguet2001] it was introduced a pyramidal implementation of the iterative LK algorithm presented above. This method is based on performing the recursive LK optical flow approach outlined above over different resolution representations of the input images, called pyramids representations, obtained by down sampling of the input images. The approach performs then by applying optical flow starting with the images with lowest resolution and proceeding through the ones with highest resolution. This process provides coarse estimation of the

motion vectors from low resolution pyramid representations that are refined in the next optical flow estimations using higher resolution pyramidal representation images.

Theoretical analysis of enRICH

In this section, we provide a mathematical analysis of enRICH to show in which cases the three-dimensional warping of a macromolecule can be corrected from its unwarped 2D projections. We will denote our map representing the reconstructed macromolecule as V($\mathbf{r}$), where $\mathbf{r}$=[x, y, z]^T. The map representing the macromolecule in a different conformation can be modelled by V($\mathbf{r}$+$\boldsymbol{\xi}$($\mathbf{r}$)), where $\boldsymbol{\xi}(\mathbf{r})$=[$\xi_x(\mathbf{r})$, $\xi_y(\mathbf{r})$, $\xi_z(\mathbf{r})$] represents the mapping between the two different conformations. If $\boldsymbol{\xi}(\mathbf{r})$ is small, we can approximate V($\mathbf{r}$+$\boldsymbol{\xi}$($\mathbf{r}$)) by a Taylor decomposition as

$$V\big(\mathbf{r}+\boldsymbol{\xi}(\mathbf{r})\big) \cong V(\mathbf{r}) + \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \boldsymbol{\xi}(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1} \boldsymbol{\xi}(\mathbf{r}) \qquad \text{(S5)}$$

Additionally, the relation between particle images at orientation Θ in the two different conformation can be modelled as

$$\mathrm{I}\left(\mathbf{r}+\boldsymbol{\chi}^{\Theta}(\mathbf{r})\right) \cong \mathrm{I}(\mathbf{r}) + \frac{\partial I(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \boldsymbol{\chi}^{\Theta}(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1} \boldsymbol{\chi}^{\Theta}(\mathbf{r}) \qquad \text{(S6)}$$

where here $\mathbf{r}$=[x, y]^T and $\boldsymbol{\chi}^{\Theta}(\mathbf{r})$ represents the deformation between projections at orientation Θ in two different conformations, which and is assumed small. Note from Eq. (S6) that approximately is possible to transform 2D projections between different conformations linearly if $\boldsymbol{\chi}^{\Theta}(\mathbf{r})$ is small. We can project all 3D maps in Eq. (S5) at orientation Θ using the projector operator obtaining

$$P^{\Theta}\{V\big(\mathbf{r}+\boldsymbol{\xi}(\mathbf{r})\big)\} \cong P^{\Theta}\{V(\mathbf{r})\} + P^{\Theta}\left\{\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \boldsymbol{\xi}(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1} \boldsymbol{\xi}(\mathbf{r})\right\} \qquad \text{(S7)}$$

This expression shows the link between the 3D ($\boldsymbol{\xi}(\mathbf{r})$) and 2D ($\boldsymbol{\chi}^{\Theta}(\mathbf{r})$) deformations at orientation Θ. Note from our analysis that the second terms in the left side of Expressions (S5) and (S6) should be approximately equal. Then, according to the Central Slice Theorem, we have

$$FT^{2D}\left\{\frac{\partial I(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \boldsymbol{\chi}^{\Theta}(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1} \boldsymbol{\chi}^{\Theta}(\mathbf{r})\right\} = S^{\Theta}\left\{FT^{3D}\left\{\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \boldsymbol{\xi}(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1} \boldsymbol{\xi}(\mathbf{r})\right\}\right\} \qquad \text{(S8)}$$

In Eq. (S8), $FT^{2D}\{\ \}, FT^{3D}\{\ \}$ and $S^{\phi}\{\ \}$ represents the 2D Fourier Transform, 3D Fourier Transform and the Slice operator at orientation Θ. As can be seen from Expression (S8), the Fourier

Transform of the deformation term in the 2D projection images at a given orientation Θ corresponds to the slice at this orientation of the Fourier Transform of the 3D deformation term. This equality shows that if we unwarp a 2D projection image using Eq. (S5), the 2D unwrapping term relates to the 3D unwrapping term through the Central Slice Theorem, which is indeed generally used in Single Particle Analysis to obtain 3D reconstructions of macromolecules. Moreover, relating Expressions (S6) and (S7), and assuming that $\boldsymbol{\xi}(\mathbf{r})$ represents low varying and smooth functions, which is reasonable as $\boldsymbol{\xi}(\mathbf{r})$ should be small for all voxels in the macromolecule domain, we can approximate Eq. (S7) as

$$P^{\Theta}\{V(\mathbf{r}+\boldsymbol{\xi}(\mathbf{r}))\} \cong P^{\Theta}\{V(\mathbf{r})\} + P^{\Theta}\left\{\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\left[\frac{\partial \xi(\mathbf{r})}{\partial \mathbf{r}}\right]^{-1}\right\} P^{\Theta}\{\boldsymbol{\xi}(\mathbf{r})\} \qquad \text{(S9)}$$

Comparing Equations (S6) and (S9) shows

$$\boldsymbol{\chi}^{\Theta}(\mathbf{r}) \cong P^{\Theta}\{\boldsymbol{\xi}(\mathbf{r})\} \qquad \text{(S10)}$$

Note that Eq. (S10) shows that, under our working approximations, it is possible to reconstruct the 3D deformation map between conformations from 2D deformations using for example the Central Slide Theorem.

The analysis provided before assumes that $\boldsymbol{\xi}(\mathbf{r})$ is small. To verify our working hypothesis, we propose to analyse the values of $|\boldsymbol{\chi}^{\Theta}(\mathbf{r})|$ for every orientation Θ, which can be computed directly from enRICH. Note the smaller the values of $|\boldsymbol{\chi}(\mathbf{r})|$ the more accurate will be the results provided by enRICH. As a rule of thumb, we propose as a control check that the average of $|\boldsymbol{\chi}^{\Theta}(\mathbf{r})|$ along all projections should be < 1.

Avoiding model bias in enRICH

The main risk of using OF to move particles between different conformations is the possibility to produce model bias. To avoid this undesired situation, we perform this task in a controlled way. First, the 2D motion fields are calculated by OF without "seeing" the experimental particles to modify. These fields are obtained from filtered map projections at the orientations given by the single particles. The maps are filtered at 15-20 Å resolutions before projection so only large conformational

changes will be detected by OF. Additionally, we also use large integrating kernels in OF calculations, of typically between 20×20 or 50×50 Å that give strong robustness against noise and to prevent overfitting in the determination of the 2D motion fields.

In addition to these two strategies to evade model bias, we have also implemented enRICH following a Gold-standard approach [Scheres2012b]. In this case, the approach uses as inputs half maps at each conformation (input and reference), so that we have a "Half 1" and "Half 2" data sets (maps and particles) in each conformation (input and reference) and the two halves are treated independently.

Limitations of the method

The working hypothesis of the approach is that large regions of the macromolecule (approximately of ~15-20Å) move collectively, imposing a strong spatial coherence in the motion field between conformations. Therefore, the present method can be used to capture movements of large regions as subunits, but not to determine local patterns of flexibility of smaller elements, such as secondary structures; additionally, it may introduce geometrical distortions not compatible with good stereochemistry parameters. Consequently, this approach is meant for those cases of collective movements where the 2D motion field between conformations is smooth, continuous and obtainable from map projection images. In other words, when conformational changes are not very large and when there is an element of continuous flexibility underlying structural variability. These assumptions require that 3D movements between conformations be small enough and typically limit the attainable resolution to 4-5 Å. Another issue that limits the applicability of the proposed approach to cases where 3D movements between conformations are not large is the possible distortion of the CTF information for each single particle. Note that in cases where the 3D movements are small, smooth and continuous the CTF information should not change significantly after Enrich correction. However, in the results section, we also show an example of a structure that is improved beyond 4 Å, corresponding to a case in which the hypothesis of smooth, continuous structural variations holds particularly well. However, in general, it is very difficult to give a clear answer about the range of

movements that are within reach of this approach, but we have always found that incorrect application of this approach to structures with large conformational changes can be easily detected *a posteriori*. Indeed, these cases show a decrease in the obtained resolution after particle merging when this value is compared to the resolution obtained before particle merging.

Simulations

Our first experiments were using simulated projections from some EMD density maps. Precisely, we did two different synthetic tests using two different classes of ryanodine receptor (EMD-2751, and EMD-2752), and ribosome (EMD-6647, and EMD-6653). After downloading those maps from the Protein Data Bank (PDB), and using random Euler angles, we project a number of 7770 ideal projection from each density map and simulated the effect of the microscope with adding Gaussian noise to them (SNR=0.5) as implemented in XMIPP. Following sub-sections present results and discussion on each group of synthetic datasets.

Ryanodine receptor

The first test was with two class of ryanodine receptor, EMD-2751 and EMD-2752 and reported resolution of 6.1 Å, and 8.5 Å, respectively. Since the size of the two density maps was different, before projecting them to create simulated images, we resized and resampled EMD-2751, from 384×384×384 (1.42 Å/px) to 224×224×224 (2.43 Å/px) using XMIPP to make its size and sampling rate equal to the EMD-2752. Then we projected both map with random Euler angles and then simulated the effect of the microscope with adding noise (SNR=0.5). Figure S2 shows some of the projections associated to the EMD-2752 before and after adding noise.

Using these two classes as the first test, we chose EMD-2752 (class with lower resolution) as the reference and EMD-2751 as the input volume. Using enRICH, we deformed simulated images of EMD-2751 density map to EMD-2752. Therefore, the corrected images could be used to reconstruct a new density map using EMD-2752 as a reference model.

We reconstructed several density maps using EMD-2752 as the reference map. These reconstructed volumes were computed from: a) simulated projections from EMD-2752 (7,220 particles), b) merging of simulated projections from EMD-2752 with projections of EMD-2751 without any correction (14,440 particles in total), and c) merging of simulated projections from EMD-2752 with corrected (deformed) projections of EMD-2751 using enRICH (14,440 particles in total). Figure S3 shows respective Fourier Shell Correlation (FSC) curves of each reconstructed map in comparison with the PDB model corresponding to the reference map (PDB-ID: 4uwe).

As clearly shown in Figure S3, the resolution of the reconstructed density map using only projections associated with the EMD-2752 map has been improved from ~9 Å, to ~8 Å (FSC=0.5) by adding corrected projections from EMD-2751 map. Moreover, the comparison between the green and red curves in Figure S3 shows that this improvement of resolution does not come from just increasing the number of projections, since the reconstructed volume computed by merged projections from EMD-2752 and EMD-2751 maps without any correction (green curve) shows a resolution of ~9 Å (FSC=0.5).

Ribosome

The other synthetic experiment was using two classes of ribosome, EMD-6647 and EMD-6653 (both maps 256×256×256, 1.64 Å/px), with reported resolution of 4 Å, and 3.8 Å, respectively. For this test, we selected EMD-6647 as the reference and we low-pass filtered it at 10 Å. This filtration was due to comparing the obtained map using our developed algorithm (enRICH) with the original map, so we could see how similar is our obtained map using enRICH to EMD-6647.

Therefore, we followed the same procedure on filtered EMD-6647 (as the reference) and EMD-6653 (as the input map) as we did on density maps at Ryanodine receptor. It means that, we project both reference and input maps with random Euler angles, and then simulated the effects of microscope using XMIPP. So, we had two set of simulated images composed by 7,220 projections in each set with SNR equals to 0.5.

This test was aimed to deform input images associated to the EMD-6653 to turn them to the projections of filtered EMD-6647 (as the reference). So, corrected projections from EMD-6653 were then merged with projections of the filtered EMD-6647 map.

We obtained maps from 3 different set of projections: a) simulated projections of filtered EMD-6647 (number of particles: 7,220), b) corrected particles from EMD-6653 using enRICH (number of particles: 7,220), and c) merged images from projections of filtered EMD-6647 with corrected particles from EMD-6653 using enRICH (number of particles: 14,440). Figure S4 shows corresponding FSC curves obtained comparing these maps with EMD-6647 map.

As Figure S4 shows, the new density map using joined simulated images of filtered EMD-6647 and deformed imaged associated to the EMD-6653 (output of the enRICH) has clearly better resolution, ~6.5 Å (FSC=0.5).

Supplementary Figures

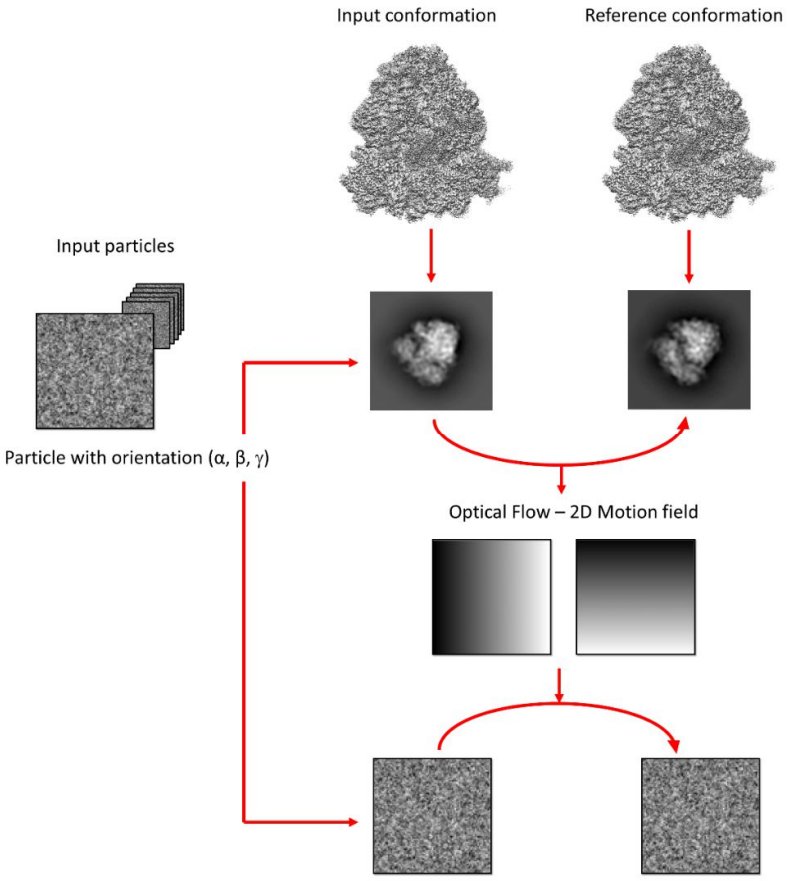

Figure S1 Conceptual idea used in enRICH. The 2D motion (deformation) field between input and reference conformations is obtained by optical flow using map projections at the orientation given by the experimental particle (in this example (α, β, γ)). This particle is then warped from the input conformation to the reference one using this 2D motion field.

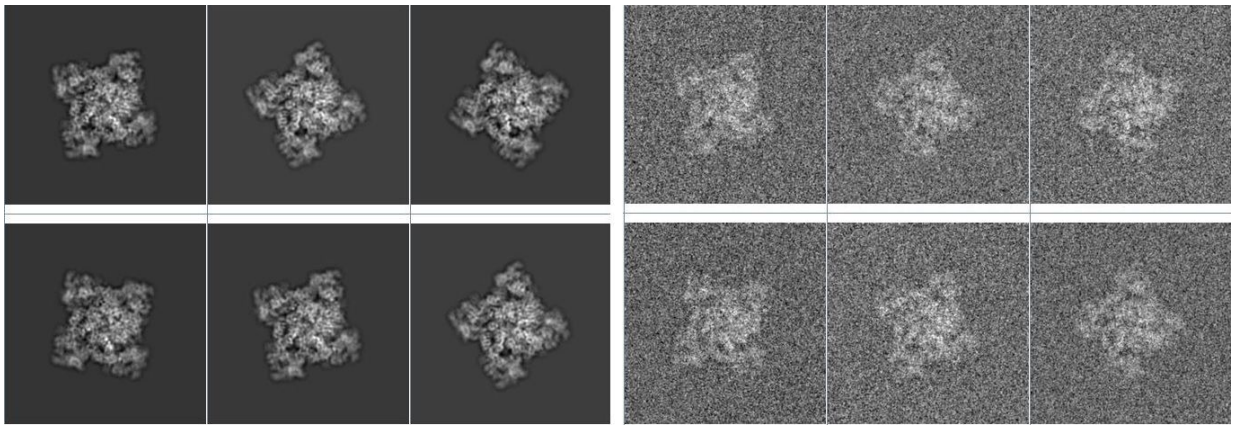

Figure S2 Some of the EMD-2752 projections creating in random Euler angles before (left images) and after (right images) simulating the effect of the microscope.

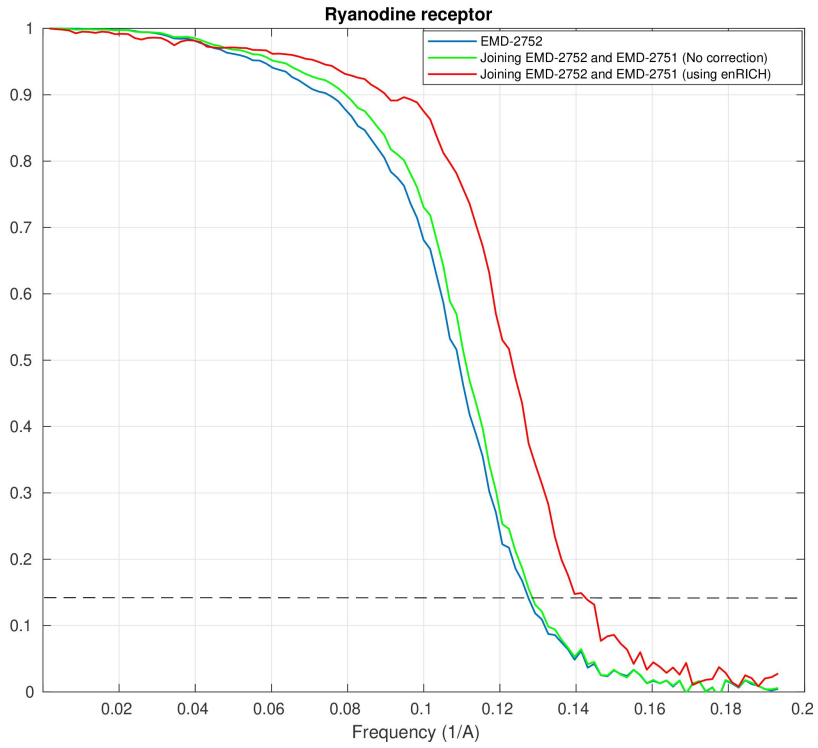

Figure S3 Fourier shell correlation (FSC) curve related to each reconstruction using different set of images and EMD-2752 as the initial model. For each FSC curve we compared each related reconstructed map with the PDB model corresponding to the reference map.

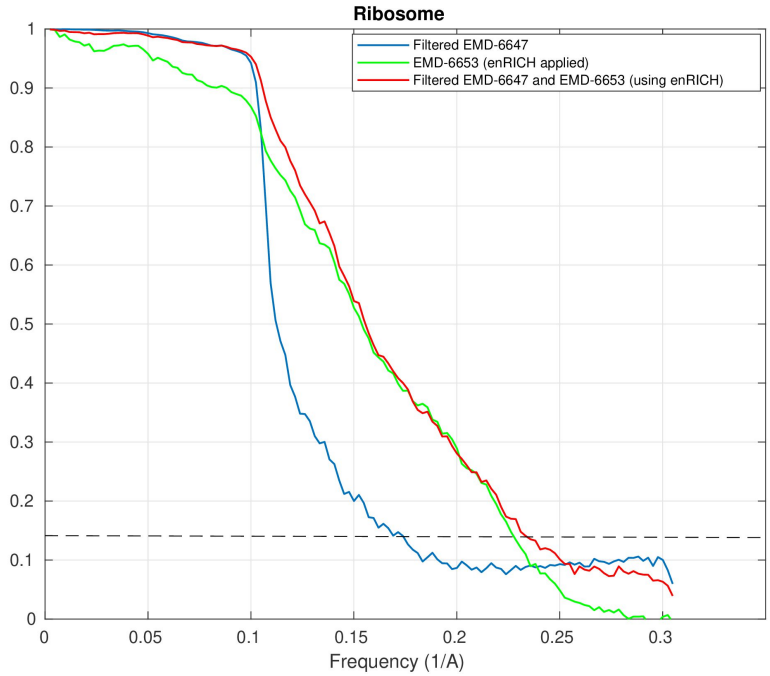

Figure S4 Fourier shell correlation (FSC) curve related to each reconstruction using different set of images and filtered EMD-6647 as the initial model. For each FSC curve we compared each related reconstructed map with the downloaded EMD-6647 map from PDB.

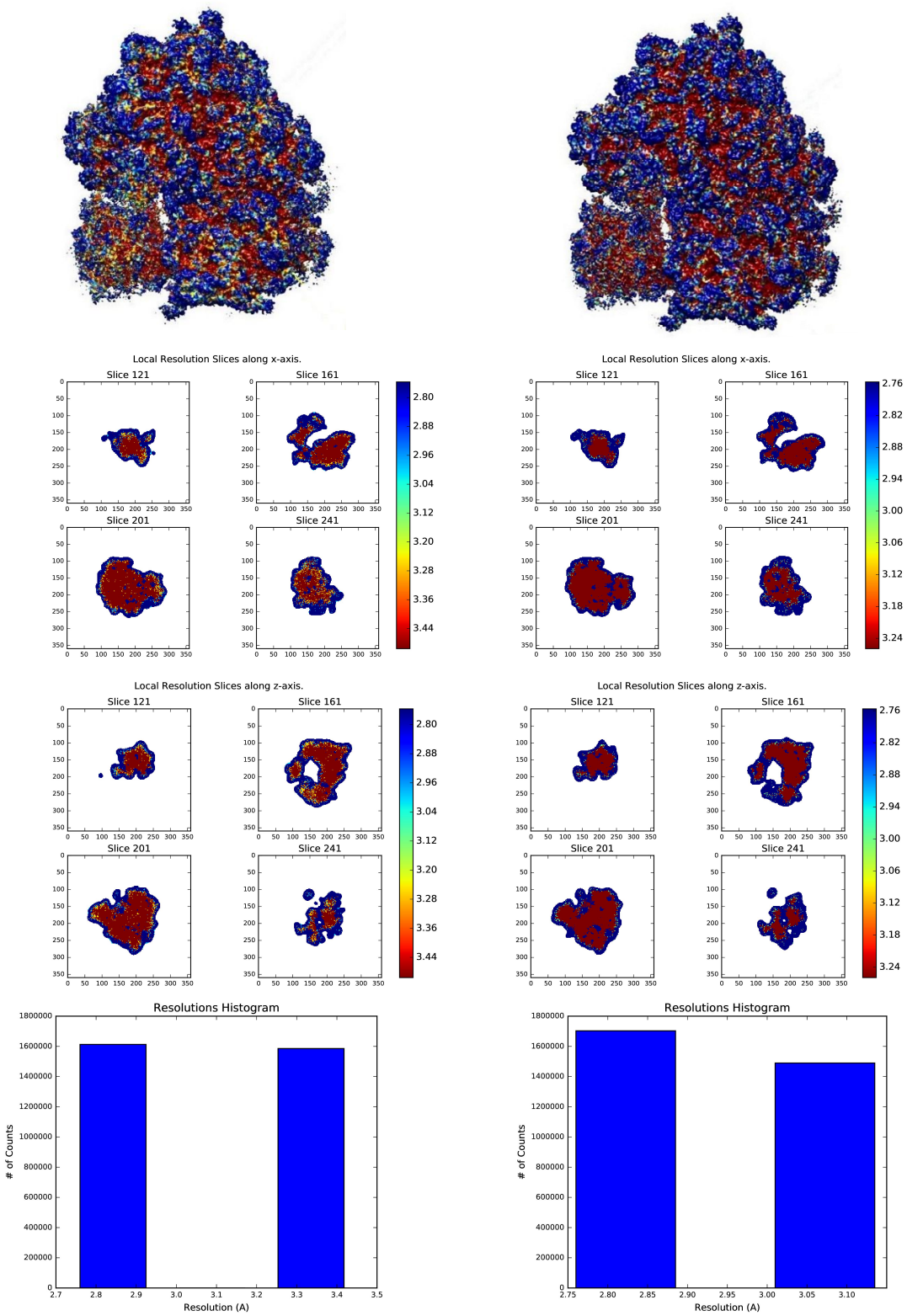

Figure S5 Local resolution information of class 2 of the 80S ribosome data set obtained by MonoRes. Left: local resolution map, local resolution slices along x-axis, local resolution slices along z-axis and resolution histogram before any correction. Right: local resolution map, local resolution slices along x-axis, local resolution slices along z-axis and resolution histogram after using enRICH.

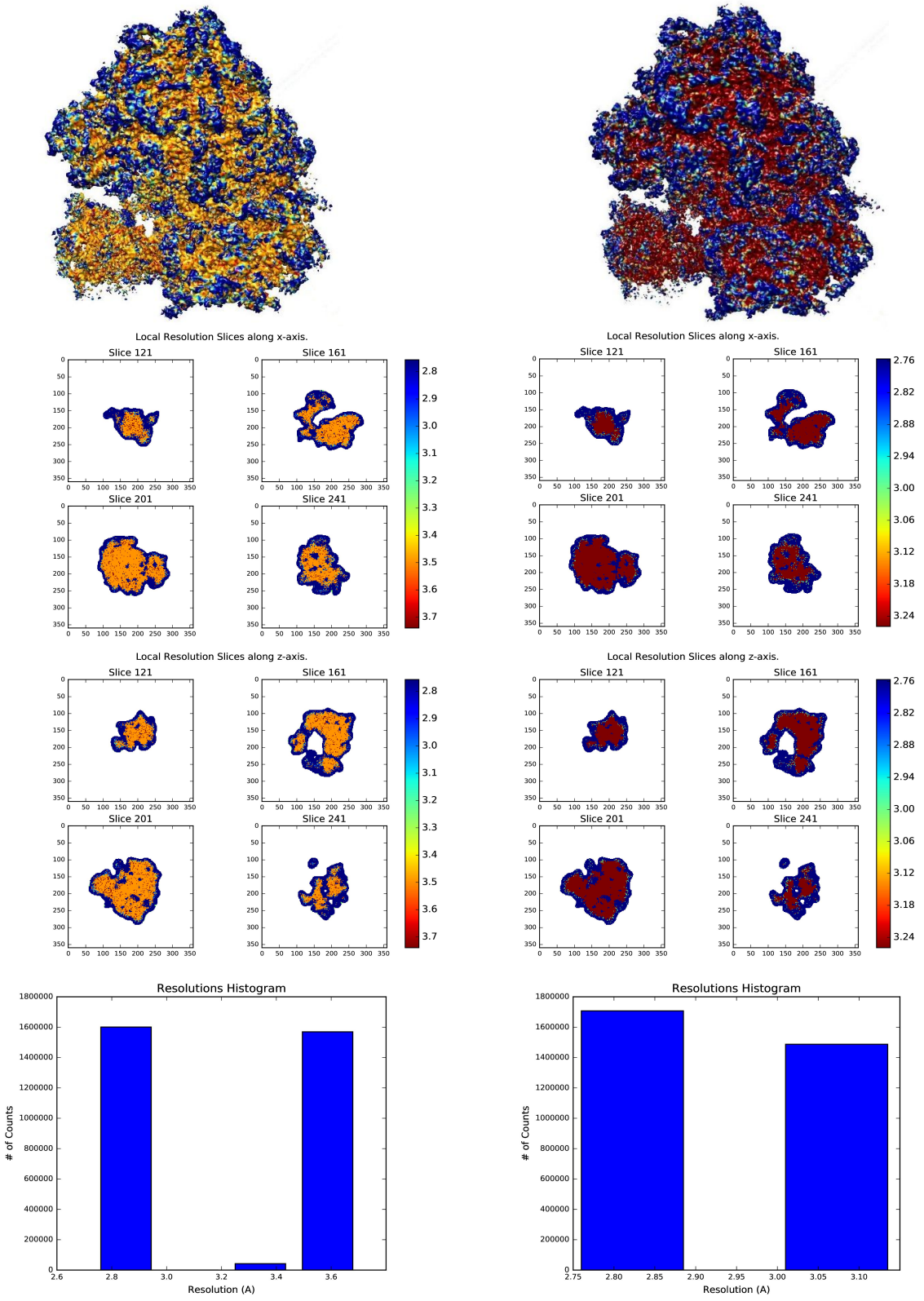

Figure S6 Local resolution information of class 3 of the 80S ribosome data set obtained by MonoRes. Left: local resolution map, local resolution slices along x-axis, local resolution slices along z-axis and resolution histogram before any correction. Right: local resolution map, local resolution slices along x-axis, local resolution slices along z-axis and resolution histogram after using enRICH.

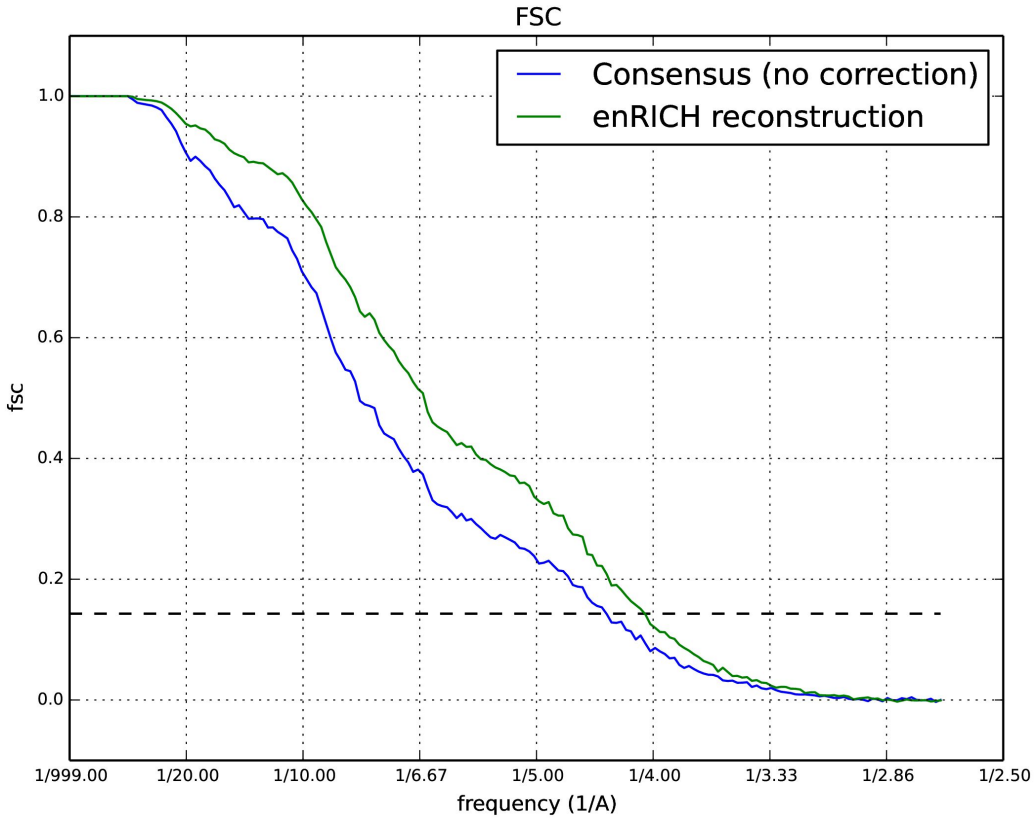

Figure S7 Comparison between FSC curves obtained from the consensus reconstruction (blue curve), computed with all ribosomal original particles (without any 3D classification or correction), and the reconstruction obtained with all particles, that have been corrected using enRICH with Class 2 as the reference (green curve)

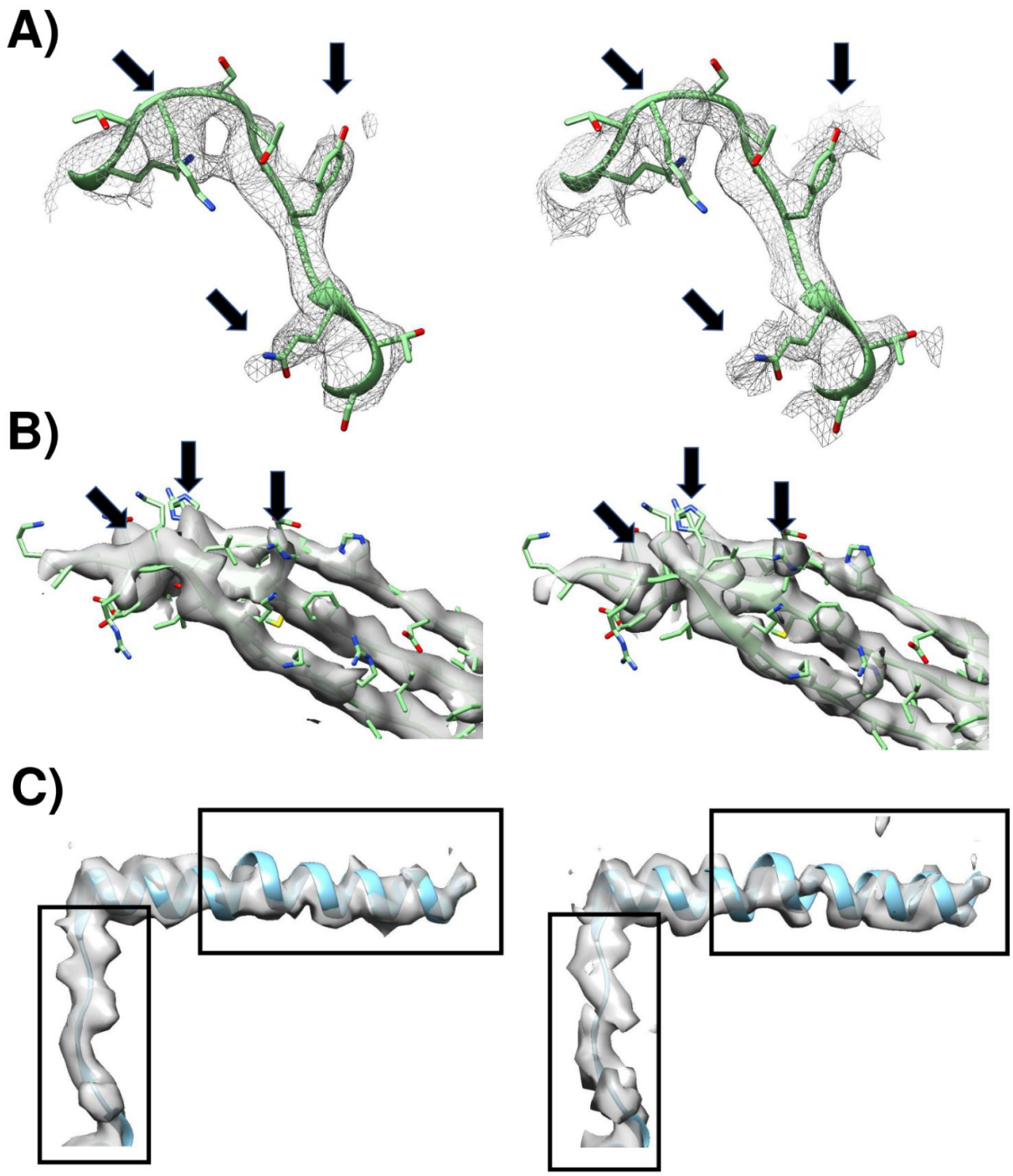

Figure S8 Left: results obtained from the ribosome dataset using enRICH; Right: results obtained from all particles without applying any 3D classification or correction (consensus reconstructions). In both cases, the corresponding PDB model is fitted to the density map; A) corresponding loops located in the small and flexible 40S subunit with black arrows pointing to the main differences of the maps; B) B-sheet located in the small 40S ribosome subunit; C) alpha helix and loop located in the large 60S ribosomal subunit with two black rectangles indicating main differences between the maps.

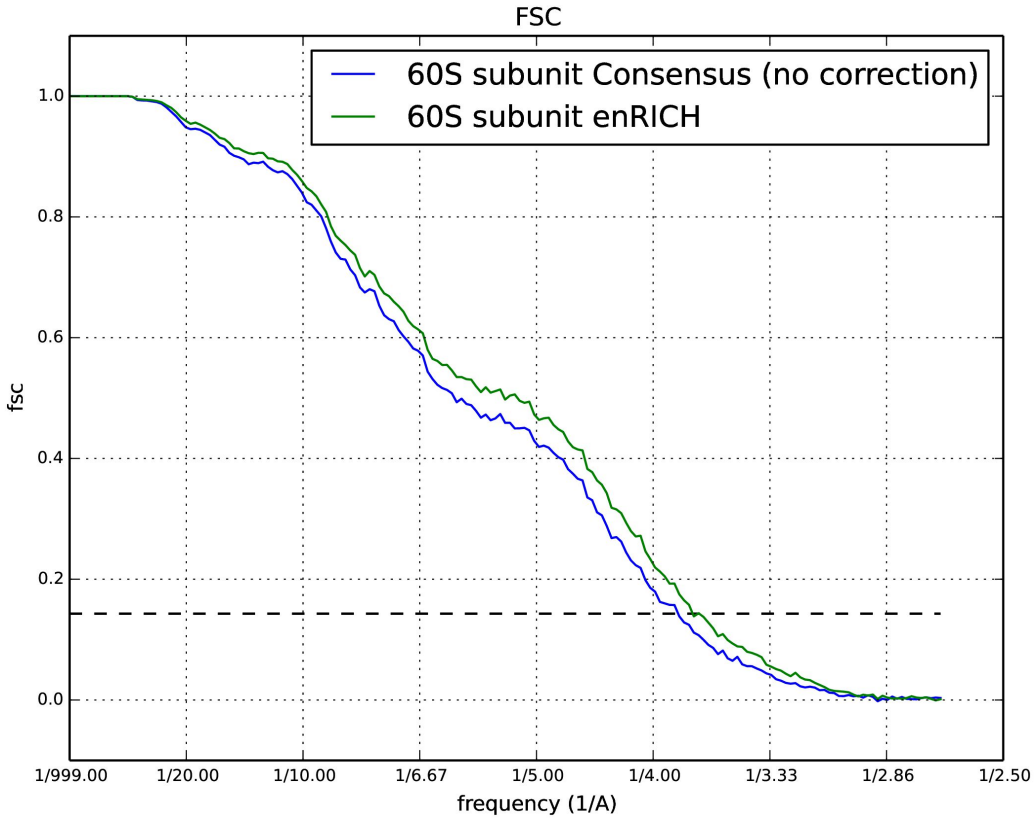

Figure S9 Comparison between FSC curves of the ribosome dataset obtained from enRICH for Class 2 (green curve) with the results computed using all particles without applying any 3D classification or correction (consensus reconstructions, blue curve) when atomic model are used as references (PDB 3j79 corresponds to the bigger 60S subunit).

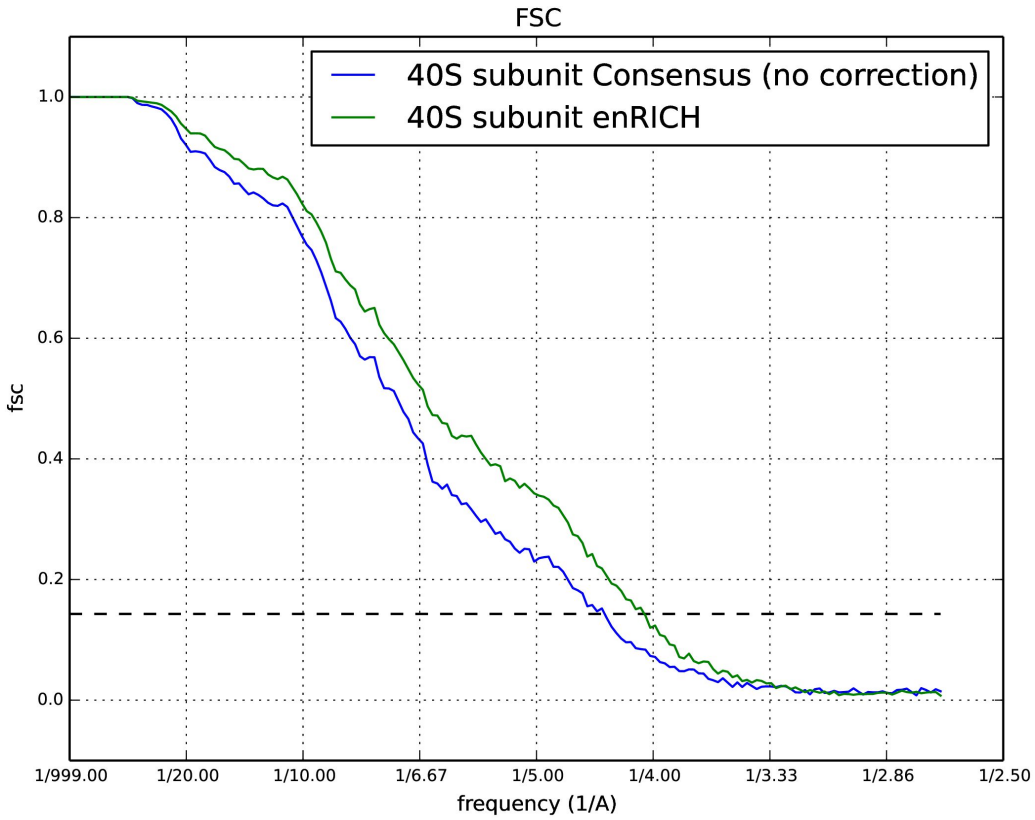

Figure S10 Comparison between FSC curves of ribosome obtained from enRICH for Class 2 (green curve) with the results computed using all particles without applying any 3D classification or correction (consensus reconstructions, blue curve) when atomic model are used as references (PDB 3j7a corresponds to the smaller 40S subunit).

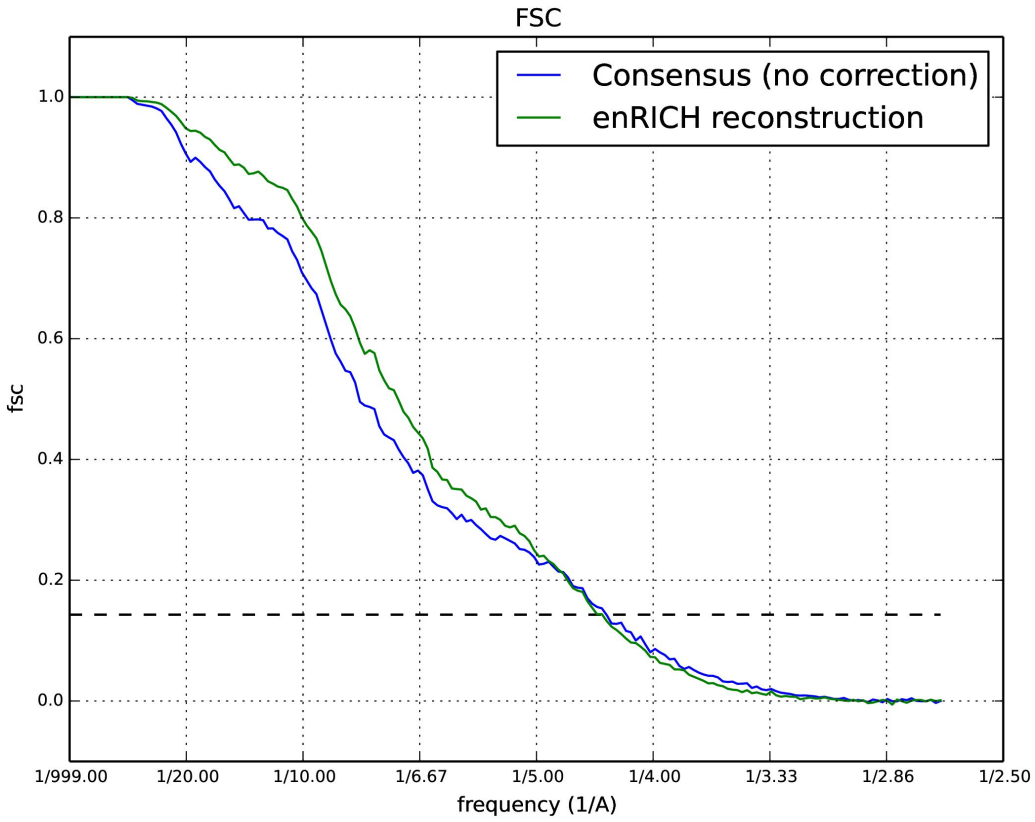

Figure S11 FSC curves of the consensus reconstruction of the ribosome dataset obtained from the original particles without any 3D classification or correction (blue curve), and the reconstruction computed with particles belonging to Class 1, 2 and 4 that have been corrected and merged to Class 3 (green curve) using enRICH.

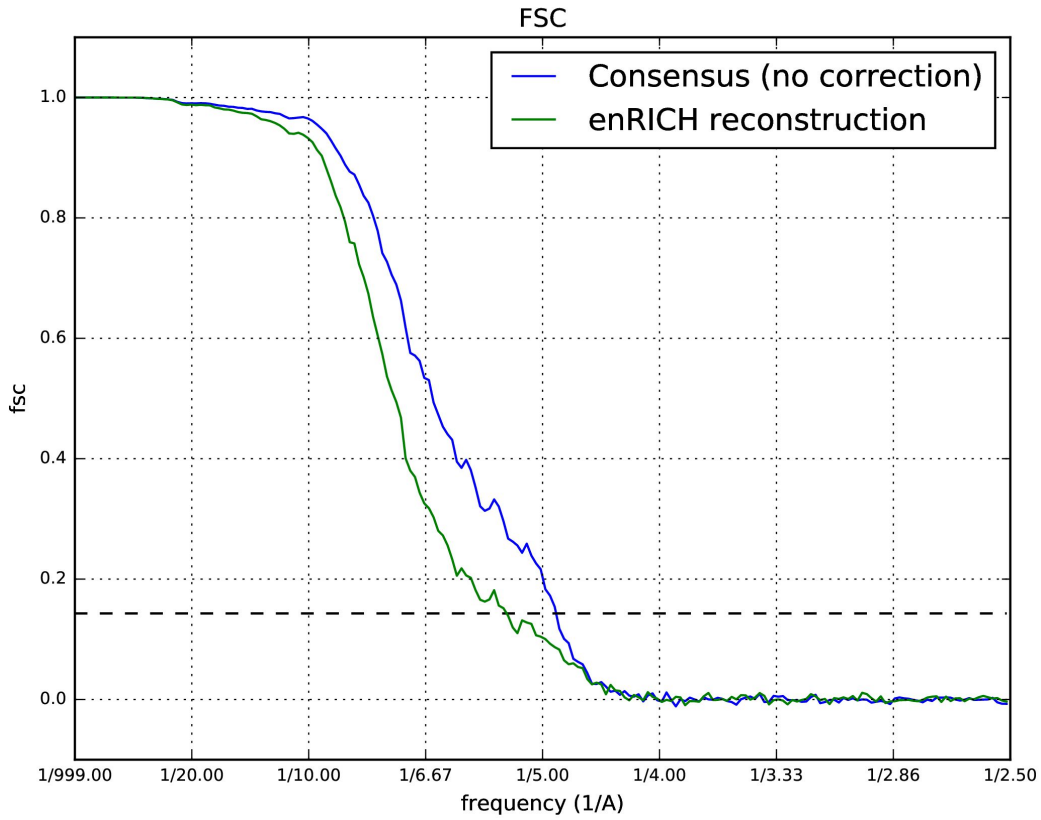

Figure S12 FSC curves of the consensus reconstruction of the RyR1 dataset obtained from the original particles without any 3D classification or correction (blue curve), and the reconstruction computed with particles belonging to Class 2.1, 2.2 and 2.4 that have been corrected and merged to Class 2.1 (green curve) using enRICH.

Supplementary Tables

	Class #2	Class #3
No enRICH + No Postprocessing	4.43	4.68
No enRICH + Postprocessing	3.45	3.63
enRICH + No Postprocessing	4.05	4.05
enRICH + Postprocessing	3.24	3.26
enRICH (Gold Standard) + No Postprocessing	4.06	4.07
enRICH (Gold Standard) + Postprocessing	3.26	3.26

Table S1 Summary of 0.143-FSC resolutions obtained for Class 2 and Class 3 reconstructions and shown in Figures 4 and 5.

	Class #2	Class #3
No enRICH – 3j7a	3.40	3.56
No enRICH – 3j79	3.17	3.34
enRICH – 3j7a	3.22	3.23
enRICH – 3j79	3.05	3.06

Table S2 0.143-FSC resolution values obtained for Class 2 and 3 using and not using enRICH approach when reconstructed maps were compared with corresponding PDBs. In all cases, the maps were masked by the same mask and b-factor corrected using similar parameters.