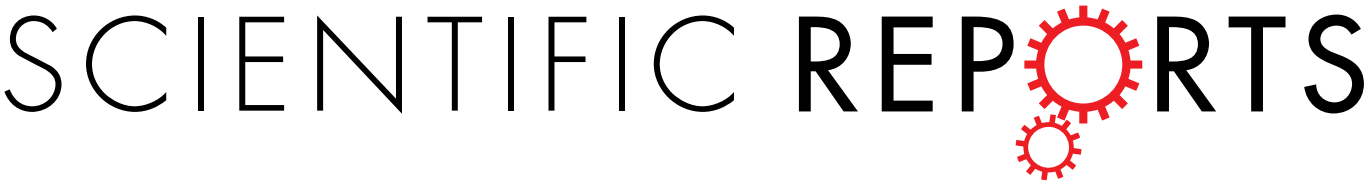

OPEN

Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis

Received: 28 February 2017
Accepted: 14 June 2017
Published online: 24 July 2017

J. Vargas[1], R. Melero[1], J. Gómez-Blanco[1], J. M. Carazo[1] & C. O. S. Sorzano[1,2]

Single Particle Analysis using cryo-electron microscopy is a structural biology technique aimed at capturing the three-dimensional (3D) conformation of biological macromolecules. Projection images used to construct the 3D density map are characterized by a very low signal-to-noise ratio to minimize radiation damage in the samples. As a consequence, the 3D image alignment process is a challenging and error prone task which usually determines the success or failure of obtaining a high quality map. In this work, we present an approach able to quantify the alignment precision and accuracy of the 3D alignment process, which is then being used to help the reconstruction process in a number of ways, such as: (1) Providing quality indicators of the macromolecular map for soft validation, (2) Assessing the degree of homogeneity of the sample and, (3), Selecting subsets of representative images. We present experimental results in which the quality of the finally obtained 3D maps is clearly improved.

Understanding how macromolecular complexes fulfill their complicated roles in the living cell is a central theme in molecular biology. Structural biology aims to deduce how such complexes function by determining the 3D arrangement of their atoms. Several techniques may be used to determine such structures. By far, the most successful technique has been X-ray crystallography. Assuming that the macromolecular complex of interest can be crystallized, this technique may yield atomic resolution and is not limited by the size of the complex. Nuclear magnetic resonance may provide unique information about dynamics and interactions, but atomic structure determination is restricted to small complexes; that is, those with molecular weights below 90 kDa. Both techniques typically require large amounts of relatively pure sample.

Single Particle Analysis (SPA) is a form of cryo-electron microscopy, which allows obtaining three-dimensional information of both large and small macromolecules close to their native state, captured in the process of performing their work, and using sample preparations with low concentrations of samples. The idea of data collection in single particle analysis is to assume that the macromolecule or particle occurs in multiple copies with (essentially) identical structure, and that its 3D orientation samples the entire angular range without leaving major gaps. Thus, instead of having to tilt the grid on which the sample is spread into multiple angles (as in electron tomography), it is then possible to take snapshots of multiple molecular views and, after suitable alignment and image processing, combine all projections into a density map depicting the molecule in three dimensions. To perform this image processing step, there are different software packages as Spider[1], Xmipp[2], Eman[3] or Relion[4], among others.

At present, one of the main problems of this technique is that there are no extensively used and accepted validation methods in SPA to assess the validity and quality of a reconstructed density map. A recent controversy highlighted this issue[5–9], imposing high priority to the development of novel methodologies to address this limitation[10]. Currently, the only way to perform a quantitative rigorous validation of a 3DEM map (especially at moderate resolution) requires the analysis of pairs of single particle images recorded at different tilt angles (tilt-pairs)[11]. The tilt-pairs validation method works comparing the discrepancy between the calculated orientations among non-tilted and tilted particles, with respect to the known tilting angle. This discrepancy is a good indicator of the 3D map quality; nevertheless collecting high-resolution and high-quality tilt-pairs is itself a relatively challenging

[1]Biocomputing Unit, Centro Nacional de Biotecnología-CSIC, C/Darwin 3, 28049, Cantoblanco (Madrid), Spain. [2]Univ. San Pablo CEU, Campus Urb. Montepríncipe s/n, 28668 Boadilla del Monte (Madrid), Spain. Correspondence and requests for materials should be addressed to J.V. (email: jvargas@cnb.csic.es)

process as often drift and/or charging occur in the tilted images[12]. Furthermore, for small cryogenic frozen samples the determination of particle correspondences between the untilted and tilted micrographs can be a difficult and error-prone process. In many occasions, the map evaluation is performed retrospectively and then tilt pairs may not be available. Additionally, this approach requires increasing the amount of data to collect and process. Finally, beam-induced movement may introduce an extra uncontrolled source of inconsistency or dispersion in the tilt-pair plot. These issues explain that although this test is currently the only fully accepted map validation methodology, it has not been widely adopted by electron microscopy practitioners. Recently, new methods to determine map quality scores have been proposed[13–15]. In ref. 13 an approach is proposed to detect "the phantom in the noise problem" meaning reconstructed density maps computed from pure noise images aligned to that map. In this case, alignment of a limited number of particle and noise images against the final released map—used as reference— is required. The resolution for the noise and particle maps is obtained through FSC, and further analysis of the resolutions achieved from particle and pure noise maps is done. In ref. 14 it an empirical quality metric is proposed for cryo-EM reconstructions by a ResLog plot. This representation shows the inverse of the resolution achieved versus the logarithm of the number of particles used, and provides heuristic information about the consistency between the obtained map and the particle dataset. Finally, in ref. 15 an approach is proposed to determine the alignment precision of a set of particles used in the map reconstruction process. This information is useful to determine the particle alignment reliability with respect to the reconstructed structure and can be used to define a map quality score. The main limitations of the latter approach are: (1) this method can determine the alignment precision for each particle but not its alignment accuracy, hence it cannot be considered a proper alignment validation approach. Note that precision refers to the reproducibility and repeatability of a measure, while accuracy is related to the degree of closeness of the measure with respect to the actual value; (2) there may exist projection images with more capacity to be aligned with precision than others. The alignment precision of a given projection highly depends on how the particle signal is distributed in the image, where highly concentrated particle signals align better that disperses ones. As a consequence, particles may not be ranked according to this score; (3) the particle alignment precision score is determined from only one reference. This reference quantifies the alignment precision when random angular orientations are given to projection images, which may not be a sufficiently demanding requirement.

In this work, we extend the previously referred method by proposing an approach that provides both the alignment precision and alignment accuracy for each particle used in the reconstruction. These scores can then be used for soft-alignment validation of a 3DEM map with respect to the set of particles used in the reconstruction process. The term soft-alignment validation refers to necessary conditions that a correct 3DEM map may verify, but it does not stand for sufficient conditions. This means that if a 3DEM reconstruction provides a low soft-alignment validation score, this map should not be accepted. However, maps with high scores can still represent incorrect structures showing a high degree of consistency with the data (false positives). The percentage of particles that aligns with both precision and accuracy is an indicator of the goodness of a reconstruction; trustworthy reconstructions should provide high scores. In addition, these indicators provide information about the homogeneity of the data. Non-homogeneous data will not be able to provide high-resolution reconstructions. In order to improve data homogeneity, we propose to use these alignment scores to prune particles.

Results

Outline of the method. The goal of this method is to provide objective information about the alignment precision and accuracy for each experimental particle image used in the reconstruction. The input of the method is a set of experimental particles previously aligned by any method and the corresponding reconstructed map. The approach is based on three steps that should be run for each particle: alignment precision estimation, alignment accuracy estimation, and determination of the percentage of reliable particles (Q value).

Particle precision estimation. Determination of the alignment precision for each particle is achieved analyzing the orientation distribution of the most similar map projections in the unit projection sphere and within the asymmetric unit. The similarity score used to compare experimental particles and map projections is normalized cross correlation. This score is always computed after aligning experimental and map projection images. The alignment consists in obtaining the two dimensional shift and orientation angle providing highest normalized cross correlation between these images.

Ideally, the set of most similar map projections for a given experimental projection image is characterized by a clustered angular spread. In this situation, the orientation determination of the projection image can be reliable. At the opposite case, a projection image may provide a very scattered distribution for the orientation of the most similar map projections. In that case, we would be unable to assign a reliable orientation to that projection image. The map projection images are obtained projecting the input reconstructed map to be evaluated into a regular grid of orientations, with a typical angular sampling rate of 5°. The more clustered is this orientation distribution, the higher the angular precision for the particle is. The clusterability of this distribution of orientations may be quantified using the Hopkins clustering tendency parameter[16], as was previously done in ref. 15. However, this latter descriptor is insensitive to cases where the orientation distribution forms two or more different clusters. In this case, we have quantified the orientation clusterability of particle m by the following weighted functional:

$$\alpha_m = \frac{\sum_{k=1}^{\xi}\sum_{\substack{n=1\\ n\neq k}}^{\xi} w_{m,n}w_{m,k}\,\arccos(\mathbf{p}_{m,n}\cdot\mathbf{p}_{m,k})}{\sum_{k=1}^{\xi}\sum_{\substack{n=1\\ n\neq k}}^{\xi} w_{m,n}w_{m,k}} \tag{1}$$

with ξ the number of most similar map projections (usually $\xi = 7$, which reflects that in an icosahedral grid (usually used to sample the projection sphere) each node has 6 closest neighbours), '$\cdot$' refers to the inner product, $p_{m,n}$ the projection direction coordinates (within the projection sphere) of the nth most similar map projection for experimental particle m and $w_{m,n}$ is the similarity score (normalized cross correlation) between experimental particle image m and map projection n. Intuitively, α_m represents the weighted sum of distances over the projection sphere (geodesic distances[15]) between orientations of the most similar map projections for a given experimental particle denoted as m. This weighted sum of distances over the projection sphere (α_m) confronts each orientation with the rest of all orientations from the set of most similar map projections. The Hopkins clustering tendency only confronts each orientation with its closest one from the set of most similar map projections. Therefore α_m is sensitive to cases where the orientation distribution forms two or more clusters, while the Hopkins parameter is not as different clusters will have always more than one element. The rationale for weighting this sum is to give more importance to highly similar map projections with respect to an experimental image m. Highly clustered cases will provide small α_m values. Observe that α_m is sensitive to cases where two or more different clusters are formed. This cluster tendency parameter (α) provides a quantitative estimation of the alignment precision for each particle. In Fig. 1 we show a comprehensive diagram to clarify the meaning of particle alignment precision. In Fig. 1(a) an experimental particle labeled as m is selected (marked in the figure with a red square). This image is confronted with the map projection set. The seven most similar map projections are then selected according to their normalized cross correlation score and marked with red squares in Fig. 1(a). The orientation of the most similar map projections is known (directions in which the input map was previously projected) and can be plotted within the projection sphere. In Fig. 1(b) and (c), red points represent examples of map projection direction coordinates of most similar map projection images for experimental particle m. Figure 1(b) and (c) represent examples of tight and spread orientation clusters respectively. We quantify the clusterability tendency of these points by the sum of geodesic distances over the sphere between all points through Eq. (1). Additionally, in Fig. 1(b) and (c) we show by discontinuous black lines geodesic distances between map projection direction coordinates $p_{m,n}$ and $p_{m,n'}$. To determine if a clusterability tendency value (α) can be considered "good" or "bad", references are required to compare it to.

A "bad" reference models the case where we cannot assign reliable orientations to given experimental particle images. The worse situation is found when all possible orientations within the asymmetric unit have exactly the same likelihood to be correct. This situation can be mathematically modeled by a random uniform orientation distribution within the asymmetric unit. Therefore, we compute the corresponding clusterability parameter α_{NOISE} using Expression (1) from ξ random angular assignments defined in the projection sphere and within the asymmetric unit. As ξ is usually small, we repeat this experiment M times (M~500) obtaining finally an average clusterability parameter $\hat{\alpha}_{NOISE}$.

To define a "good" reference we can construct a synthetic set of "perfect" particle images totally compatible with the input volume. This set is constructed projecting the map at the same orientations and distorted by the same CTFs than the experimental projections. Therefore, for each experimental particle (denoted by m index), we compute its respective "perfect" counterpart. The clusterability parameter of this "perfect" projection set is determined following the same procedure explained before obtaining $\alpha_{m,good}$. In order to decide if the clusterability parameter α_m of an experimental particle m refers to precise alignment, we should compare this value with the two obtained references ($\alpha_{m,good}$, $\hat{\alpha}_{NOISE}$). To this end, we map the clusterability scores $\hat{\alpha}_{NOISE}$ and $\alpha_{m,good}$ to subjective quality precision alignment values of $q^p_{NOISE} = 0$ and $q^p_{m,good} = 1$, respectively. From these two points ($\hat{\alpha}_{NOISE}$, $q^p_{NOISE} = 0$) and ($\hat{\alpha}_{m,good}$, $q^p_{m,good} = 1$) we can fit a straight line passing through them and determine the quality precision alignment for experimental particle m as

$$q^p_m = \left(\frac{q^p_{m,good} - q^p_{NOISE}}{\alpha_{m,good} - \hat{\alpha}_{NOISE}} \right) (\alpha_m - \hat{\alpha}_{NOISE}) + q^p_{NOISE} \tag{2}$$

Observe that the defined quality precision alignment score is an easy to interpret parameter: the closer q^p_m is to 1 (or to 0) the better (the worse) the alignment precision is for experimental particle image m. Therefore, we can establish as criteria that an experimental particle aligns with precision if $q^p_m \geq 0.5$. Note that this threshold indicates that the experimental particle aligns with precision closer to a good reference particle (obtained projecting the map at the experimental particle orientation and shift and with the same CTF) than to a bad reference coming from pure noise orientations. In Fig. 2 we show a scheme of the alignment precision estimation process for each experimental particle.

Particle accuracy estimation. Determination of the alignment accuracy for each particle is performed comparing the previously obtained particle orientation, used in the map reconstruction to be analyzed, with the orientations of the most similar map projections to that corresponding particle image. This comparison (denoted here as χ_m) is quantified by the weighted sum of geodesic distances[15] over the projection between the orientation corresponding to a given particle (say, particle m) as used in the calculation of the input map, and orientations of the most similar map projections to that experimental particle image. As before, the similarity metric used to compare experimental particles and map projections is normalized cross correlation, and the weights come from similarity values between particles and map projections after 2D alignment. The set of most similar map projections were determined after a global alignment search using the input 3DEM map as reference. Mathematically χ_m is given by

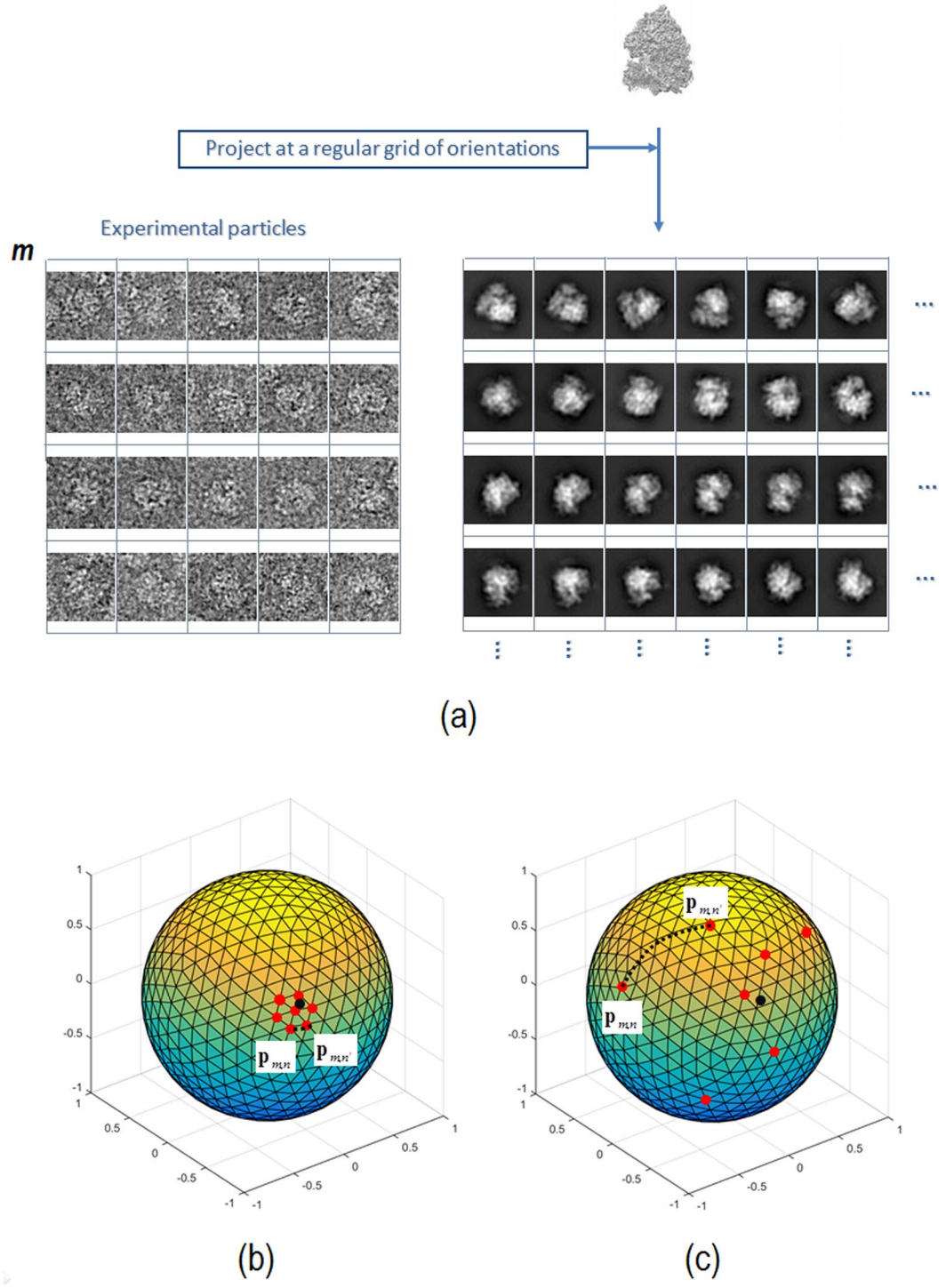

Figure 1. (**a**) Selected experimental particle labeled as *m* marked in the figure at the left part with a red square. Al the right it is shown the map projection set with the seven most similar map projections, selected according to their normalized cross correlation score, marked with red squares. In (**b**) and (**c**) are shown plots of the orientations (red points) of the most similar map projections. The plots (**b**) and (**c**) represent examples of tight and spread orientation clusters respectively.

$$\chi_m = \frac{\sum_{k=1}^{\xi} w_{m,k} \arccos(\hat{\mathbf{p}}_m \cdot \mathbf{p}_{m,k})}{\sum_{k=1}^{\xi} w_{m,k}} \tag{3}$$

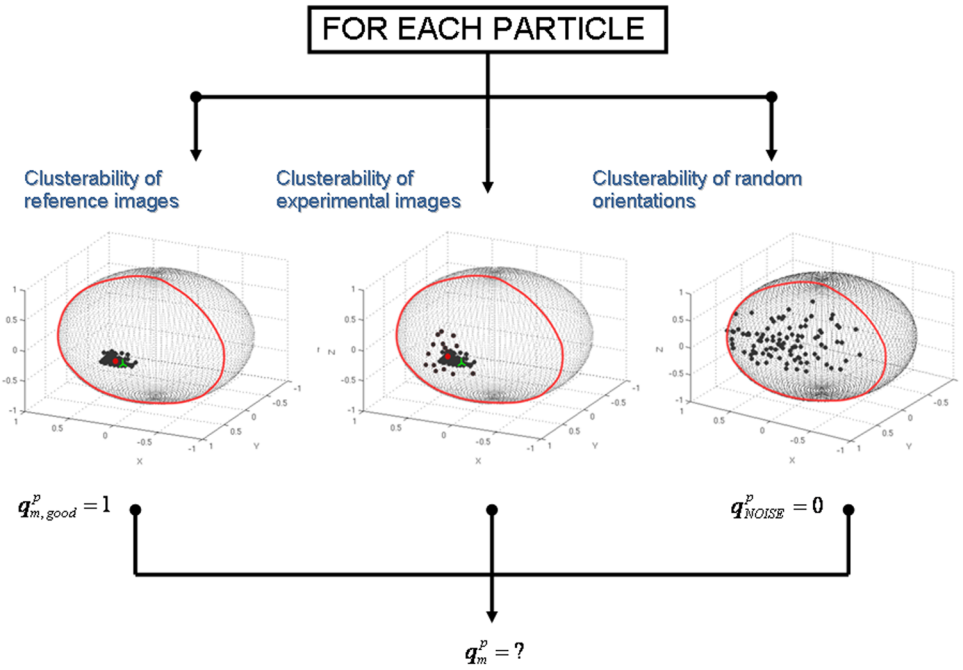

Figure 2. Diagram of the alignment precision and accuracy estimation process for each experimental particle. First, the clusterability of the orientation distribution is obtained for the experimental, perfect and pure noise particles. These values are used to give an alignment precision score to each experimental particle between 0 and 1. In addition, the alignment accuracy is computed comparing the previously obtained particle orientation (green cross) with the weighted average orientation of most similar map projections (red point).

In Eq. (3) $\hat{\mathbf{p}}_m$ is the direction projection coordinates (orientation) over the projection sphere for experimental particle m that was obtained previously in the refinement/reconstruction process. Both $\hat{\mathbf{p}}_m$ and $\hat{\mathbf{p}}_m$ correspond to Cartesian coordinates over the projection sphere obtained from Euler angles given by the refinement approach at the last iteration and by the posterior pure global alignment search, respectively. In Fig. 1(b) and (c) $\hat{\mathbf{p}}_m$ is represented by the black points over the projection sphere. Conceptually, here we are comparing the previously computed orientation after refinement for each experimental particle image, with orientations founded from a pure global alignment search using as template the final three-dimensional map. Most popular iterative reconstruction methods achieving high resolution structures perform global particle angular searches only in the first iterations of the reconstruction process. These angular explorations transform to local searches as the number of iterations increase. This means that the alignment process of reconstruction approaches is generally very dependent on previous decisions taken, with the risk of getting trapped into an alignment local minima for a given particle. Hence, if the alignment approach makes an erroneous angular assignment for one particle at any iteration, this mistake will not be amended at following iterations.

The accuracy parameter obtained for each particle (χ_m) is difficult to interpret as this geodesic distance depends on parameters as the symmetry and angular sampling, among other factors. In order to decide if a particle aligns with accuracy, we need two references to compare it to. As before, we can use the synthetic "perfect" particle set projecting the map at the same orientation and applying the same CTF than the experimental projections but without added noise. Therefore, for each experimental projection m we can determine a "good" reference accuracy parameter ($\chi_{m,good}$). On the other hand, a "bad" reference can be determined comparing the previously obtained orientation by the reconstruction method with ξ random orientations within the asymmetric unit. As before, this comparison is performed M times obtaining finally a mean accuracy parameter $\hat{\chi}_{NOISE}$. From these two references, we determine the quality accuracy alignment score for each experimental particle m as

$$q^a_m = \left(\frac{q^a_{good} - q^a_{NOISE}}{\chi_{m,good} - \hat{\chi}_{NOISE}} \right) (\chi_m - \hat{\chi}_{NOISE}) + q^a_{NOISE} \tag{4}$$

where q^a_{good} and q^a_{NOISE} equals to 1 and 0, respectively. We can establish as criteria that an experimental particle aligns with accuracy if $q^a_m \geq 0.5$. As before, this threshold indicates that the experimental particle aligns with accuracy closer to a good reference particle than to a bad reference coming from pure noise orientations. In Fig. 3 we show a scheme of the typical alignment procedure followed by three-dimensional refinement programs achieving high resolution structures as Relion[4], which starting from a low-resolution initial model performs a global angular search for each particle (a). This global search converts to a local one as the number of iterations and map resolution increases (b–d). Figure 3(b) exemplifies the case of an incorrect angular assignment can be

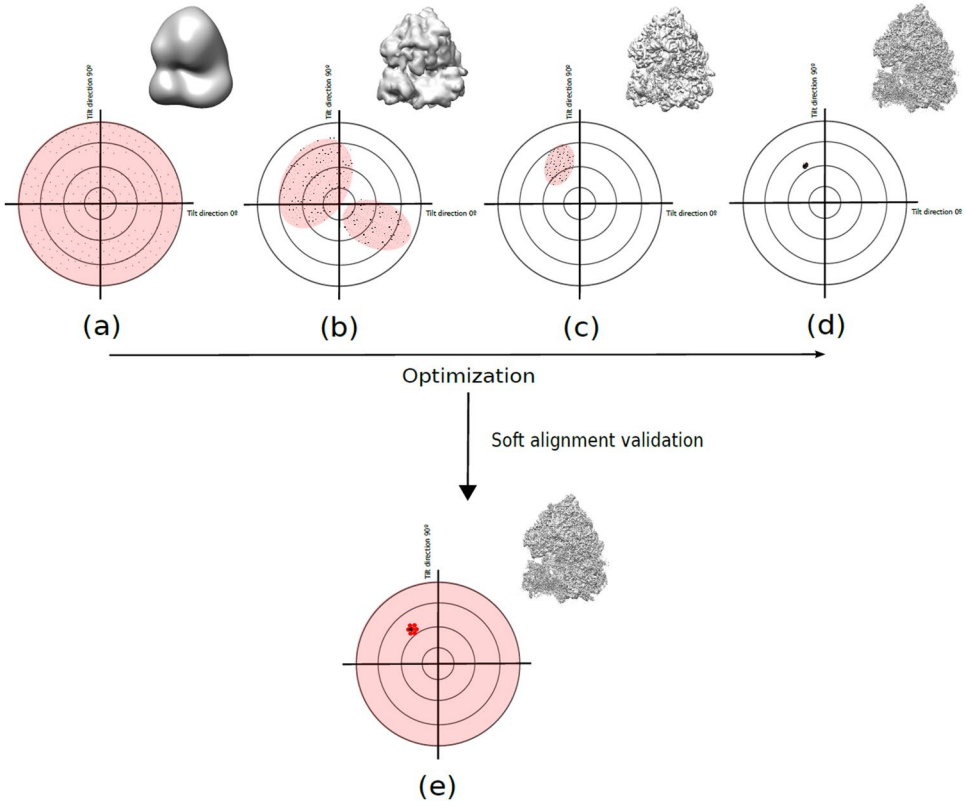

Figure 3. Typical alignment procedure performed for each particle. Starting from a low-resolution initial model the alignment method performs a global angular search with low angular sampling (**a**). This global search converts to a local one (increasing the angular sampling) as the number of iterations and the map resolution increases (**b**–**d**). Our proposed approach angular accuracy estimation is based on re-computing the particle orientation by a global search using the final reconstructed map (**e**).

made for one particle. In Fig. 3(e) we show how the particle alignment accuracy is determined by a global angular search using the final reconstructed map.

Determination of the percentage of reliable particles. After we have obtained for each experimental particle the quality precision and accuracy parameters q_m^p and q_m^a, we can determine a global alignment parameter denoted as Q, which determines the percentage of particles that aligns with both precision and accuracy

$$Q = \frac{\sum_{m=1}^{N}[(q_m^a > 0.5) \text{ AND } (q_m^p > 0.5)]}{N} \times 100 \tag{5}$$

with N the number of particles in the dataset. Obviously Q gives information about the consistency between the 3DEM map and the experimental particles. The larger Q is, the better the consistency between the particles and the 3DEM map.

Results

We have used the proposed approach in different cases, including soft-validation analysis of high-quality (β-galactosidase) and low quality (HIV-1 Env trimer) maps and application to a heterogeneous dataset (Ribosome).

First Experiment: Application to high-resolution data. We have applied our proposed approach to high-quality data corresponding to the β-galactosidase complex. We have used the data of the 2015 Map Challenge, with EMPIAR codes EMPIAR-10012 and EMPIAR-10013 that were obtained with a FEI Titan KRIOS microscope and using a Gatan K2 detector. We used the initial particle coordinates given for the Map Challenge. A gold-standard approach was performed dividing the data in two independent halves, each composed by 11,412 particles. The final reconstructed 3DEM map, obtained by Relion[4] through Scipion framework[17], has a 0.143-FSC resolution of 3.25 Angstroms, and secondary structure elements can be clearly seen. This map and the resultant FSC are shown in Fig. 4(a) and (b), respectively. In addition, we have also computed the FSC curve between the reconstructed map and the corresponding deposited PDB structure (PDB code: 3j7h). The resultant 0.143-FSC resolution is of 3.02 Angstroms. Using this data, we run our proposed soft-alignment validation approach to each half obtaining the results shown in Fig. 5. In this figure each red point represents the alignment precision

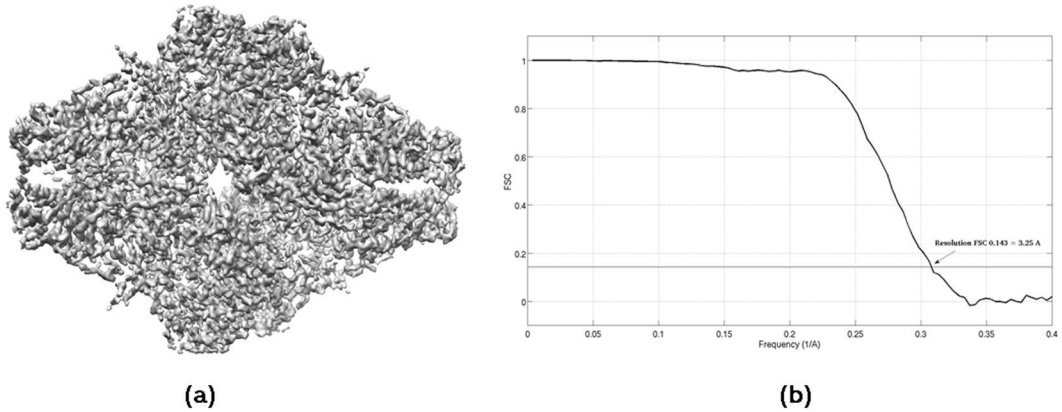

Figure 4. Reconstructed 3DEM map for the β-galactosidase complex of the 2015 Map Challenge with EMPIAR codes EMPIAR-10012 and EMPIAR-10013 (**a**) and obtained FSC curve based on Gold standard approach. The resultant 0.143-FSC resolution is of 3.25 Å.

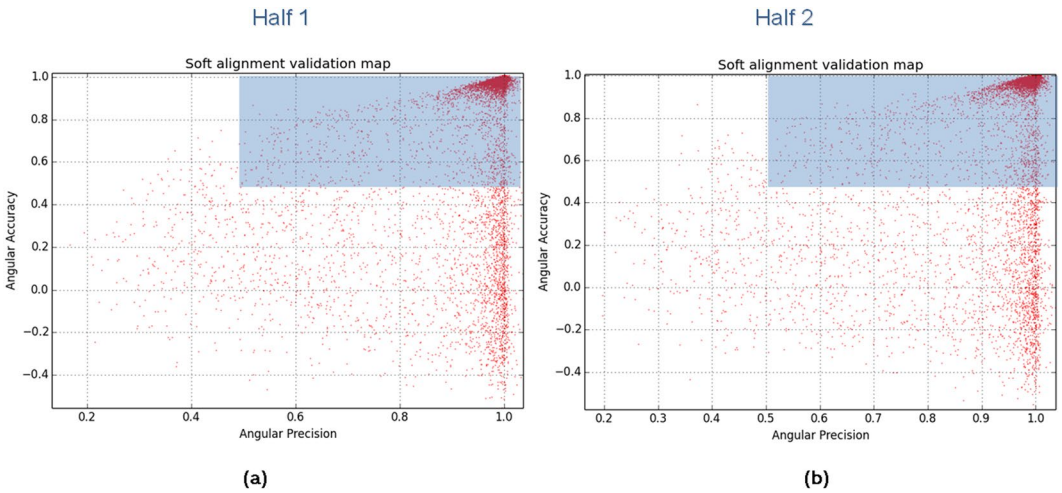

Figure 5. Soft-alignment validation maps for each half of the β-galactosidase complex and composed by 11,412 particles. The blue partially transparent rectangle indicates the particles which align with both accuracy and precision.

	Half 1	Half 2
Precision (%)	97	97
Accuracy (%)	79	79
Precision & Accuracy (%)	79	79

Table 1. Percentage of particles that align with precision, accuracy and both precision and accuracy for each half of the β-galactosidase complex.

(*x*-axis) and accuracy (*y*-axis) for one experimental particle. Observe that these soft-alignment validation maps show a clear cluster around point (1, 1) and that there are few points showing alignment precision and/or accuracy beyond 1, however this is not necessarily because the presence of image artifacts. Note that the set of most similar map projections is obtained using a defined angular sampling rate (usually 5°). Also, take into account that the used number of most similar map projections is usually small ($\xi = 7$). As the typical angular sampling rate is not high, situations where one of the obtained most similar map projections is not close to the actual one may exist, even for reference particles. Additionally, as ξ is small, we do not have a high statistically significance and therefore some unusual situations can appear if only one orientation from the set of obtained most similar map projections is not accurately determined. Note that the global alignment searches for the experimental and reference particles are totally independent. The percentage of particles that align with precision, accuracy and both precision and accuracy are given in Table 1. As can be seen from Table 1, most of the particles align with precision and accuracy, being this an indicator of the map high quality. Additionally, observe that there is a

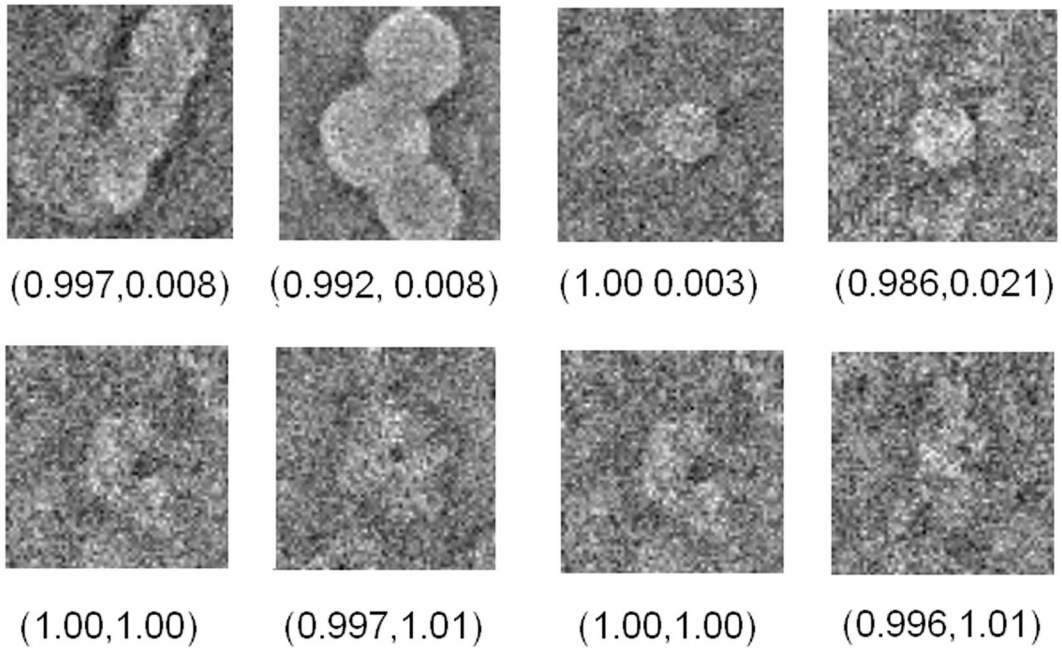

Figure 6. Exemplifying set of downsampled β-galactosidase particles showing high precision but low accuracy in the first row and both high precision and accuracy in the second row.

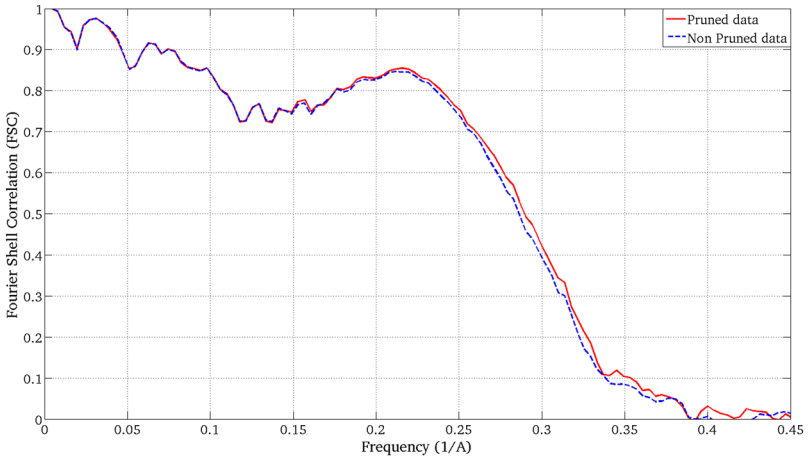

Figure 7. FSC curves obtained when the pruned (red solid curve) and non-pruned (blue dashed curve) 3DEM maps obtained for the β-galactosidase complex are confronted with the PDB map.

significant percentage of particles which aligns with precision but not with accuracy. This indicates the presence of compositional, conformational or artefact-based heterogeneity, as bright spots for example in the dataset. The β-galactosidase complex is well known to not present conformational heterogeneity. We have visually checked that a significant amount of particles are of low quality and affected by artefacts. Visually, it is confirmed that most of the particles with low alignment accuracy or both accuracy and precision are affected by artefacts. As example, in Fig. 6, we show a set of downsampled particles which show high precision but low accuracy in the first row and both high precision and accuracy in the second row. Taking into consideration that with the proposed method we can rank the particles according to its alignment quality, we have performed a particle pruning process. We have rejected particles showing values of alignment precision or accuracy lower than 0.5. As result, after pruning we have reduced the amount of particles approximately 21% and we have rejected 4,875 images from the set of 22,824 particles. We have reconstructed the density map from the selected set of 17,949 particles. The obtained 0.143 resolution using gold-standard approach is 3.23 Angstroms. Moreover, the resulting 0.143-FSC resolution is of 3.00 Angstroms when this map is confronted with the corresponding PDB. In Fig. 7, we show the resultant FSC curves obtained when the PDB is confronted with the pruned (red solid curve) and not pruned 3DEM maps (blue dashed curve). As can be seen from these curves the pruned map presents better quality than the other over all range of frequencies and using approximately 21% less data.

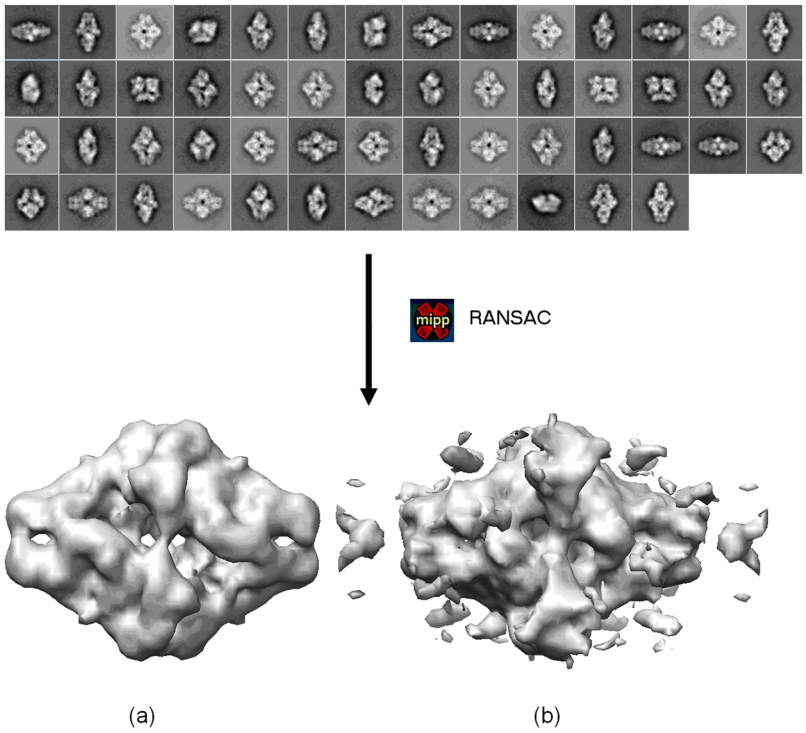

Figure 8. Experimental class averages of β-galactosidase particles and obtained "good" (**a**) and "bad" quality (**b**) *ab initio* initial volumes.

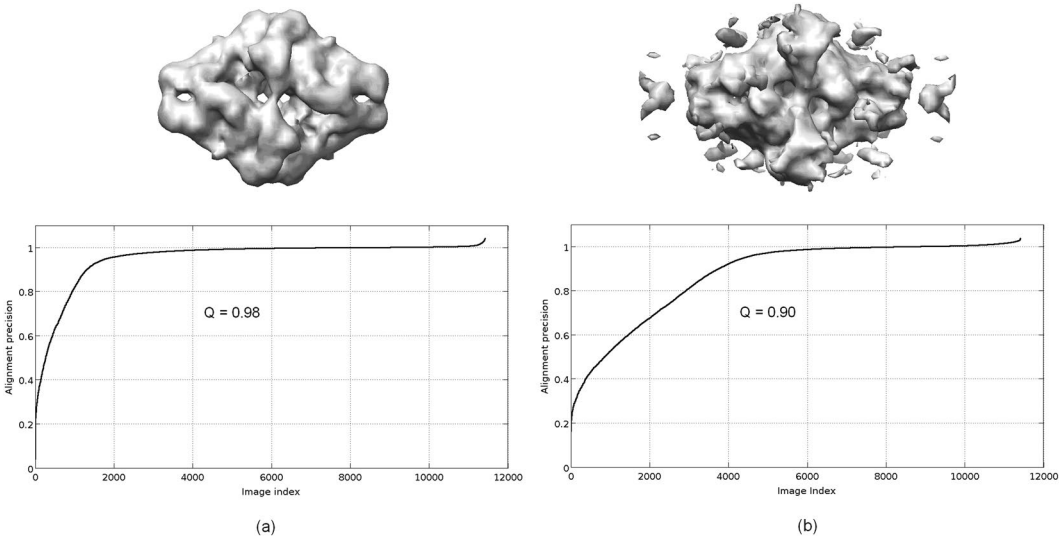

Figure 9. Alignment precision curves computed from the "good" (**a**) and "bad" quality (**b**) *ab initio* β-galactosidase initial volumes.

Second Experiment: Application to ranking of *ab initio* initial maps. We have applied the soft-alignment validation approach to the task of ranking different *ab initio* initial maps. We have used as before the β-galactosidase complex. 2D classes were obtained from picked particles using CL2D method[18]. From these averages we run RANSAC *ab initio* initial volume estimation method[19], obtaining ten different low resolution maps. From this set, we selected two maps, one clearly correct and another of lower quality. In Fig. 8 we show the used 2D classes and the two initial volumes selected. The obtained results for the alignment precision when these two maps are confronted with 11,412 particles, composing each of the two halves of the full dataset, are 0.98 and 0.90 for maps shown in Fig. 8(a) and (b), respectively. These values indicate that in both cases we can align the particles with precision but (a) map gives better results. Note that in this experiment we cannot determine the alignment accuracy as at the moment of obtaining the initial map, we have not yet aligned the particles. In Fig. 9 we show the respective alignment precision plots obtained when the maps shown in Fig. 8(a) and (b) are

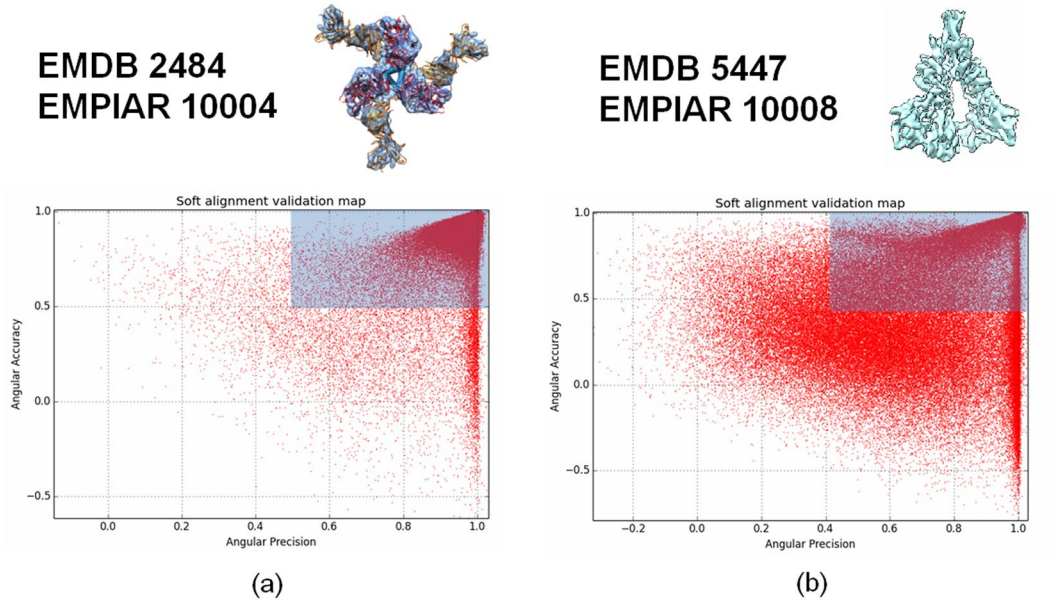

Figure 10. Resultant soft-alignment validation plots for HIV-1 trimmer reconstructions with EMDB codes 2484 (**a**) and 5447 (**b**) and using the deposited particles in EMPIAR with codes 10004 (88,125 particles) and 10008 (124,478 particles), respectively.

	EMDB 5447	EMDB 2484
Precision (%)	75	91
Accuracy (%)	45	98
Precision & Accuracy (%)	38	90

Table 2. Percentage of particles that align with precision, accuracy and both precision and accuracy for the HIV-1 trimmer with EMDB codes 5447 (a) and 2484 (b).

confronted with the particles. As can be seen from this figure the "correct" map presents higher alignment precision than the other one.

Third Experiment: Application to detect incorrect maps. We have applied our method to maps that have been subjected to recent controversy in the field, corresponding to reconstructions of the HIV-1 trimer with EMDB codes 5447 and 2484. We have used the experimental images previously used by the authors and deposited in EMPIAR with codes 10008 and 10004. The number of images for EMPIAR 5447 and 2484 corresponds to 124,478 and 88,125, respectively. We run our proposed soft-alignment validation approach confronting all the available particles with their respective EMDB maps. In Fig. 10 we show the resulting soft-alignment validation plots for both cases. It can be visually seen that for EMPIAR 10008 (Fig. 10(b)) the resultant soft-alignment plot is of low quality and most of the particles do not align with precision and accuracy. On the other hand for, EMPIAR 10004 most of the particles aligns with precision and accuracy showing a clear cluster at point (1,1). Table 2 shows the percentage of particles which aligns with precision, accuracy and both precision and accuracy in both cases. This table clearly shows that EMDB 5447 map cannot be a valid map in terms of its alignment quality.

Fourth Experiment: Application to heterogeneous data. We have also used our soft-alignment validation approach to data coming from the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine[20]. Ribosomes are well-known examples of flexible samples presenting conformational heterogeneity. However, in this dataset ribosomes are bounded to a drug, which partially inhibits its movement. From our analysis of the data, we have found compositional heterogeneity resulting on particles of the 80S ribosome bound to the drug and of the 60S subunit alone. Additionally, we have found continuous conformational heterogeneity in a small region of the 40S subunit. We used the data deposited in the 2015 Map Challenge with EMPIAR ID of 10028. This data was obtained with a FEI Polara 300 microscope equipped with a Falcon II camera. The number of projection images deposited was 105,247; from this data we reconstructed a 3DEM map using Relion[4] through Scipion framework[17]. The resultant map has a 0.143-FSC resolution of 3.28 Å when it is compared to the PDB (PDB codes: 3j79/3j7a) and 3.24 Å using gold-standard from Relion.The resulting soft-alignment validation plot is given in Fig. 11. The percentage of particles which aligns with precision, accuracy and precision and accuracy corresponds to 0.92, 0.75 and 0.72, respectively. In this case the plot shows the highly variability of the data because sample heterogeneity. Using this information we pruned particles with low alignment score so that 34% of the particles were removed. We reconstructed an additional map using the remaining 69,402 particles

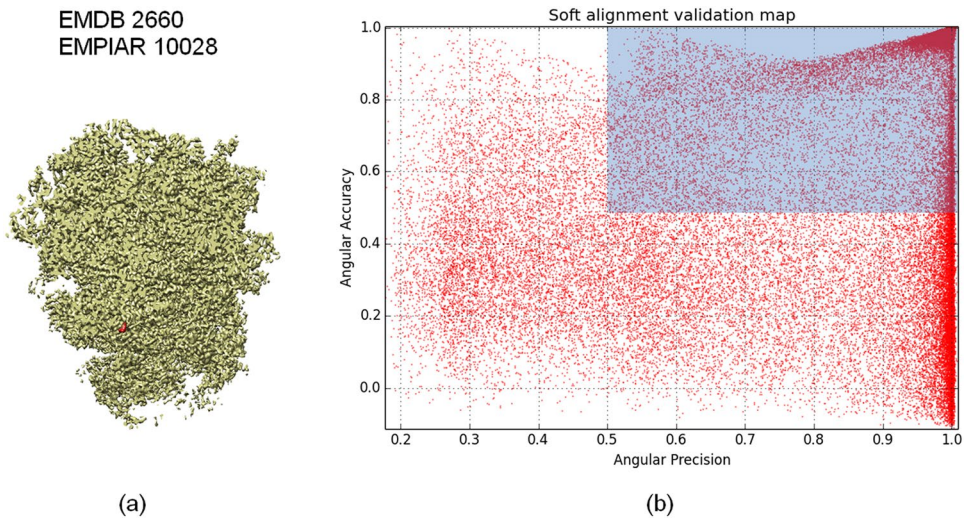

Figure 11. 80S ribosome with EMDB code 2660 and used in the 2015 Map Challenge (**a**) and resultant soft-alignment validation plot (**b**) obtained when 69,402 particles deposited in EMPIAR 10028 were used.

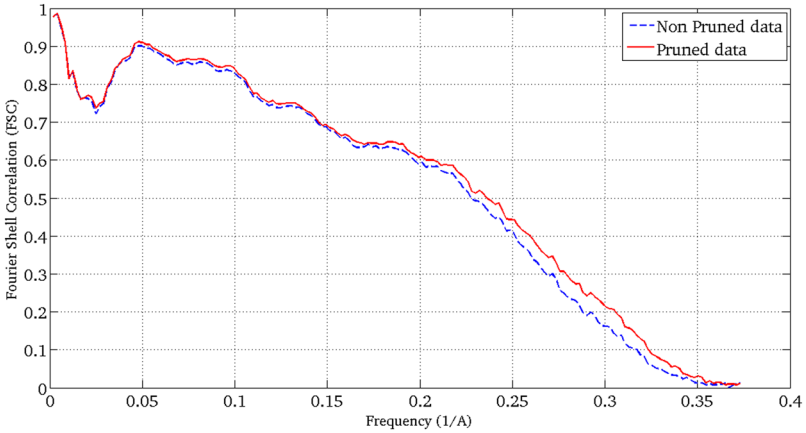

Figure 12. FSC curves when the pruned (red dashed curve) and non-pruned (blue solid curve) 3DEM structures were confronted with the PDB.

	Class1	Class2	Class3
Precision (%)	0.94	0.94	0.93
Accuracy (%)	0.92	0.93	0.49
Precision & Accuracy (%)	0.92	0.91	0.45

Table 3. Percentage of particles that align with precision, accuracy and both precision and accuracy for an intermediate resolution map.

obtaining a 0.143-FSC resolution of 3.16 Å when it was compared with the PDB (PDB codes: 3j79/3j7a) and 3.23 Å using gold-standard from Relion. In Fig. 12 we show the corresponding FSC curves when the pruned and non-pruned structures were confronted with the PDB. As can be seen from Fig. 12 the FSC curve obtained from the pruned data presents higher FSC values for all the frequencies, clearly indicating the better quality of this data in terms of particle homogeneity.

We have also used more than one 3D reference for soft-alignment validation and particle pruning. We have classified the set of 105,247 particles in three classes composed by 45,059 (Class1), 34,763 (Class2) and 25,323 (Class3) using Relion. Each class has been finally refined through Relion auto-refine method, obtaining structures at resolutions of 3.45 Å, 3.63 Å and 5.61 Å after amplitude correction and using gold-standard approach. The soft-alignment validation results when classified particles were confronted with their respective 3D maps are shown in Table 3. These results show high consistency between particles and respective 3D classes for Class 1 and Class2. The number of resultant particles which aligns with both precision and accuracy is of 41,415, 31,474

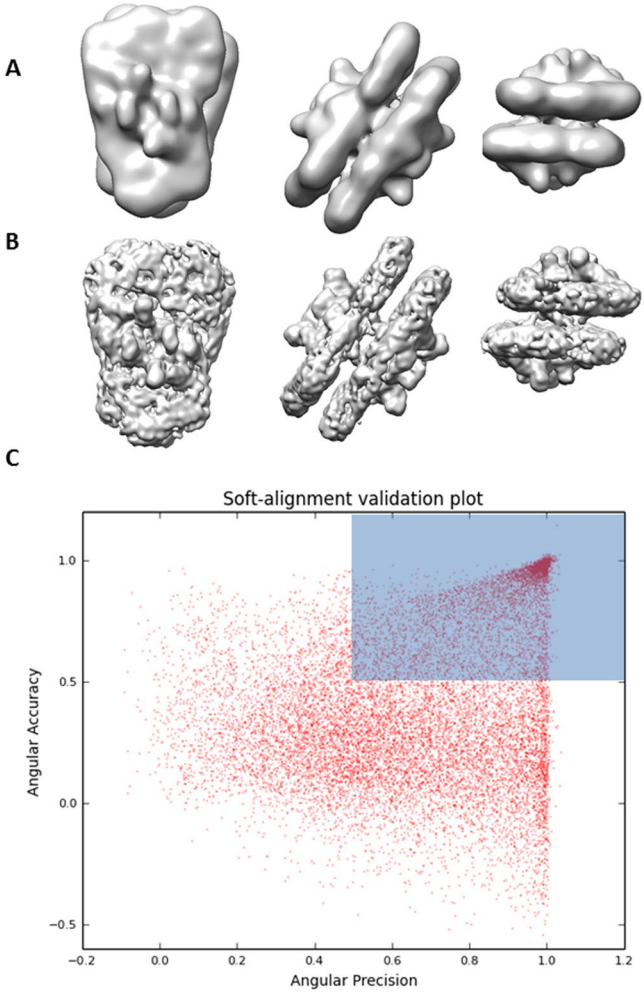

Figure 13. Views at different orientations of the map without (**A**) and with (**B**) pruning process and respective soft-alignment validation plot (**C**).

and 11,614 for Class1, Class2 and Class3, respectively. The resultant 3D maps after refining pruning particles have resolutions of 3.45 Å, 3.57 Å and 5.82 Å using gold-standard. These results indicate that the pruning method can only improve the map resolution for Class 2. For Classes 1 and 3 likely the most important resolution limiting factor is the number of particles.

Fifth Experiment: Application to improve intermediate resolution maps. We have also used our approach to significantly improve the resolution of an intermediate-resolution map, removing bad quality data from the reconstruction process. We used cryo-EM data of a plant protein complex with molecular weight higher than 2 MDa, which presents high degree of compositional heterogeneity (unpublished data). We followed the typical Relion workflow to obtain a 3D map at ~30 Å resolution (0.5 FSC criteria) using 15,976 particles after particle screening through 2D and 3D classification from a set of 33,729. Figure 13(A) shows this obtained map at different views. In order to improve the map resolution, we run our soft-alignment validation approach to give a second chance to particles previously removed by particle screening using Relion. We used the previously obtained map from the set of 15,976 particles and the complete set of extracted particles (without any particle screening process) composed by 33,729 particles. The corresponding soft-alignment plot is given in Fig. 13(C), which shows that only 16% of particles align with both precision and accuracy resulting on a "clean" set of 5,308 particles used to reconstruct a map at ~14 Å shown in Fig. 13(B).

Discussion

Currently, the field of cryo-EM is living the so-called 'resolution revolution' thanks to a series of technical advantages, and is starting to provide high-resolution reconstructions routinely. However, cryo-EM has not been without concerns. At present, this technique is not a 'push-button' method and a lot of experience and knowledge has to be added to the reconstruction process. As a consequence, errors can been committed. In this work, we propose a novel method to perform quantitative analysis of the 3D alignment quality, which can be used to: (1) soft-validation of cryo-EM determined structures, (2) particle ranking and pruning and (3) data homogeneity analysis. The approach is based on determining the alignment precision and accuracy of each particle that

participated in the reconstruction of a 3DEM map. Proper aligned particles should align at the same time with precision and accuracy. Usually, particles with low alignment precision score represent noise images that cannot be aligned with reliability. In addition, particles with low alignment accuracy depict projections affected by artefacts or particles representing different conformations than the reconstructed 3DEM map. Pruning particles with both low alignment accuracy and precision improves the data homogeneity and therefore the quality of the reconstructed map.

The conceptual principle of the method is based on the idea that for a given map most of the experimental particles should present a cluster distribution for their most likely map orientations when a global alignment process is performed. This cluster distribution is a consequence of the spatial coherence that a correct map should show. In addition, the previously computed particle orientations, used in the map reconstruction, should be compatible with the weighted average of the most similar map orientations after the global angular search soft-alignment validation score. In order to decide if a particle shows good or bad alignment precision and accuracy scores we use two references. One comes from a pure uniform random orientation distribution within the asymmetric unit. This reference exemplifies a bad alignment, where all the possible orientations have the same probability of being true. Opposed to this, we define a good reference projecting the input 3DEM map at the same orientations and in the same conditions in terms of CTF than the experimental particles but without added noise. Each experimental particle has a "perfect" counterpart which will align with precision and alignment. With these two references it is possible to define numbers between 1 and 0 describing the alignment precision and accuracy of each experimental particle. These alignment scores can be used to determine the percentage of particles that align with both accuracy and precision. We have defined as criteria that a particle aligns with precision and accuracy if these scores are equal and higher than 0.5. These are the default thresholds, which indicate that the experimental particles align better than noise. However, depending on the user goals, it may be a good idea to be more restrictive defining higher thresholds such as 0.8–0.9 for example. A good application case may be the improvement of map resolution through particle pruning in cases not limited by the experimental number of particles. Using this information we can determine the percentage of particles which aligns with both precision and accuracy (Q value) as an indicator of the reconstruction alignment quality. Nonetheless, indicating a direct relation between the Q value and the map validity is difficult. It is clear that maps with low Q scores are affected by high heterogeneity and/or a high number of pure noise particles, among other problems, which can compromise the alignment refinement process. However, it is not possible to define a unique Q threshold to determine the validity of a given 3DEM map. For example, a reconstruction with a Q value of 0.5 can be correct if bad quality particles are just pure noise images and enough good particles are available. However, this Q value can be problematic if the projection images are affected by artefacts. It is clear that low values of Q (between 0–0.5) clearly indicate that the quality of the data is low and likely the reconstruction is not correct, and additional test should be added to show the validity of the reconstruction. However, it should be understood that this method does not validate the map, as this would imply validation of all the computational steps leading to the three-dimensional reconstruction, but the angles assignments are at least soft-validated. Note that soft-validation refers to necessary conditions that a correct 3DEM map may verify, but it does not stand for sufficient conditions. We have used our proposed method in different situations such as high resolution data (β-galactosidase complex), ranking of *ab initio* initial maps (β-galactosidase complex), controversial maps (HIV-1 trimmer), heterogeneous data (80S ribosome) and intermediate resolution data. In all these cases we have computed the soft-alignment validation map which gives information about the goodness of the particle alignment process and of the particle homogeneity. This information was used to correctly rank initial maps in terms of the quality and clearly discard the HIV-1 trimmer map in terms of its alignment quality. In addition, these soft-alignment validation plots were used to improve the quality of the data in terms of its homogeneity by a pruning process. After rejecting particles with low alignment scored we clearly improved the resolution of the reconstructed maps. In the case of the β-galactosidase and 80S ribosome macromolecules we improved from 3.02 and 3.28 to 3.00 and 3.16 Å, respectively.

References

1. Shaikh, T. *et al*. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**(12), 1941–1974 (2008).
2. de la Rosa-Trevín, J. M. *et al*. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**(2), 321–328 (2013).
3. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single particle reconstructions. *J. Struct. Biol.* **128**(1), 82–97 (1999).
4. Scheres, S. H. W. Relion: implementation of a Bayesian approach to cryo-em structure determination. *J. Struct. Biol.* **180**(3), 519–530 (2012).
5. Mao, Y. *et al*. Molecular architecture of the uncleaved hiv-1 envelope glycoprotein trimer. *Proc. Natl. Acad. Sci. USA* **110**, 12438–12443 (2013).
6. Mao, Y., Castillo-Menendez, L. R. & Sodroski, J. G. Reply to subramaniam, van heel, and henderson: Validity of the cryo-electron microscopy structures of the hiv-1 envelope glycoprotein complex. *Proc. Natl. Acad. Sci. USA* **110**, E4178–E4182 (2013).
7. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. USA* **110**, 18037–18041 (2013).
8. Subramaniam, S. Structure of trimeric hiv-1 envelope glycoproteins. *Proc. Natl. Acad. Sci. USA* **110**, E4172–E4174 (2013).
9. van Heel, M. Finding trimeric hiv-1 envelope glycoproteins in random noise. *Proc. Natl. Acad. Sci. USA* **110**, E4175–E4177 (2013).
10. Henderson, R. *et al*. Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).
11. Henderson, R. *et al*. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J. Mol. Biol.* **413**, 1028–1046 (2011).
12. Murray, S. C. *et al*. Validation of cryo-em structure of {IP3R1} channel. *Structure.* **21**, 900–909 (2013).
13. Heymann, J. B. The Phantom in the Noise and Validation of 3D EM Reconstructions. *Microsc. Microanal.* **20**, 792–793 (2014).
14. Stagg, S. C. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *J. Struct. Biol.* **185**(3), 418–426 (2014).
15. Vargas, J., Oton, J., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. Particle alignment reliability in single particle electron cryomicroscopy: a general approach. *Sci. Rep.* **6**, 21626 (2016).

16. Banerjee, A. & Dave, R. Validating clusters using the hopkins statistic. In Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE04), vol. 1, 149–153, Budapest (Hungary) (2004).
17. de la Rosa-Trevín, J. M. *et al*. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**(1), 93–9 (2016).
18. Sorzano, C. O. S. *et al*. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010).
19. Vargas, J., Álvarez-Cabrera, A. L., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30**, 2891–2898 (2014).
20. Wong, W. *et al*. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e03080 (2014).

Acknowledgements

This work was supported by grants from the Comunidad de Madrid (S2010/BMD-2305), the MINECO – Spain (AIC-A-2011–0638) and the Fundación General CSIC (Programa ComFuturo to Javier Vargas). Carlos Oscar S. Sorzano is the recipient of a Ramón y Cajal fellowship.

Author Contributions

J.V. had the main idea, developed the method and wrote the manuscript. J.V., R.M., J.G., J.M.C. and C.O.S. reviewed the manuscript, supervised the experiments and discussed the results.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.