Journal of Structural Biology 204 (2018) 329–337

Contents lists available at ScienceDirect

Journal of Structural Biology

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

A new algorithm for high-resolution reconstruction of single particles by electron microscopy

C.O.S. Sorzano[a,b,*], J. Vargas[c], J.M. de la Rosa-Trevín[d], A. Jiménez[a], D. Maluenda[a], R. Melero[a], M. Martínez[a], E. Ramírez-Aportela[a], P. Conesa[a], J.L. Vilas[a], R. Marabini[e], J.M. Carazo[a]

[a] Centro Nac. Biotecnología (CSIC), 28049 Cantoblanco, Madrid, Spain
[b] Univ. San Pablo – CEU, Campus Urb. Montepríncipe, 28668 Boadilla del Monte, Madrid, Spain
[c] Dept. Anatomy and Cell Biology, McGill Univ., Montreal, Canada
[d] SciLife Lab, Stockholm, Sweden
[e] Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

Keywords:
Image processing
3D reconstruction
Angular assignment
Volume restoration

ABSTRACT

The Map Challenge organized by the Electron Microscopy Data Bank has prompted the development of an Xmipp high resolution reconstruction protocol (which we will refer to as *highres*) that is integrated in the software platform `Scipion`. In this work we describe the details of the image angular alignment and map reconstruction steps in our new method. This algorithm is similar to the standard projection matching approach with some important modifications, especially in the area of detecting significant features in the reconstructed volume. We show that the new method is able to produce higher resolution maps than the current *de facto* standard as measured by the Fourier Shell Correlation, the Monogenic Local Resolution and EMRinger.

1. Introduction

Single Particle Analysis of macromolecular structures by Electron Microscopy (EM) has become in the last few years one of the most successful techniques for Structural Biology (Nogales, 2016) due to its ability to achieve near-atomic resolution and to explore conformational flexibility, using low amounts of sample material. The automation of the image acquisition process at the microscope and the introduction of Direct Electron Detectors have allowed the recording of vast amounts of data whose analysis results in three-dimensional maps of the macromolecule under study from which structural models can be derived. However, the images acquired at the microscope are extremely noisy (Signal-to-Noise Ratios between 0.1 and 0.01). Such noisy measurements require robust data analysis methods.

The prerequisites to achieve near-atomic resolution include 1) a structurally homogeneous population of projection images obtained using the crucial 3D classification algorithms (Scheres, 2012), 2) a sufficiently good angular coverage to measure every region in Fourier space, and 3) a sufficiently good frequency coverage to measure every frequency and to preserve the microscope structural information at as high resolution as possible. To accomplish this last requirement, images are acquired at different defoci, especially at low defocus to preserve high frequency information. The 3D reconstruction process alternates between angular assignment and three-dimensional reconstruction to extract the maximum of structural information present at the micrographs. Overall, the whole problem can be seen as a regression in which the projection images are the data to be fitted, and the volume and the alignment parameters constitute the model.

All processing steps must be robust to high levels of noise and need to avoid overfitting (reconstruction artifacts that satisfy data constraints but that are either dominated by noise or that are far away from the best possible solution). Structural knowledge can also be incorporated in the analysis workflow, and most common ways are either using a Bayesian prior (as Relion did Scheres, 2012) or by regularization (for a review of regularization in 3D reconstruction, see Sorzano et al. (2017)).

Relion (Scheres, 2012), at the moment the most common method to refine maps in the field of Single Particle Analysis by EM, integrates the whole regression problem in a single functional that is optimized in a greedy fashion starting from an initial estimate of the volume to be reconstructed. This functional includes a Bayesian prior about the statistical distribution of the objects to be reconstructed (coefficients in Fourier space are independently distributed, with independent real and imaginary parts, Gaussianly distributed with zero-mean and a variance that is estimated from the data itself). Although in general, the prior is not accurate for macromolecular structures (Sorzano et al., 2015), it has

* Corresponding author.
E-mail address: coss@cnb.csic.es (C.O.S. Sorzano).

https://doi.org/10.1016/j.jsb.2018.08.002
Received 4 March 2018; Received in revised form 19 July 2018; Accepted 4 August 2018
Available online 24 August 2018

the advantage that it is mathematically tractable and results in a low pass filter of the reconstructed volume similar to a Wiener filter (Scheres, 2012).

This Bayesian approach has currently dominated the cryoEM 3D reconstruction field, specially for its generalization to the very important task of multiple maps reconstruction as part of a 3D classification process. However, this approach is not the only valid strategy for map reconstruction. Actually, before the introduction in cryoEM of the Maximum Likelihood (Scheres et al., 2005) and Maximum *a posteriori* methods (Scheres, 2012), the standard approach to 3D angular alignment and reconstruction was the so-called "Projection Matching" (Penczek et al., 1992; Penczek et al., 1994). Recently, novel implementations of Maximum Likelihood based on GPU processing and stochastic gradient descent have significantly reduced the processing time (Punjani et al., 2017).

Let us refer to the parameters defining the angular alignment of the whole dataset as $\boldsymbol{\Theta}$ and to the reconstructed volume as $\mathbf{V}$. In an extremely simplified manner, we may think of the Maximum Likelihood method as an algorithm that minimizes

$$\boldsymbol{\Theta}^*, \mathbf{V}^* = \arg\min_{\boldsymbol{\Theta},\mathbf{V}} ||\widetilde{\mathbf{I}} - P_{\boldsymbol{\Theta}}\mathbf{V}||^2_W \quad (P1)$$

where $\widetilde{\mathbf{I}}$ is the set of pixels from the acquired images (if the underlying algorithm allows an image to be at multiple angular orientations with different probabilities, then $\widetilde{\mathbf{I}}$ will contain multiple copies of the measured data and Θ will have several components devoted to the same image), $P_{\boldsymbol{\Theta}}$ is a projection operator that calculates the projections of the volume $\mathbf{V}$ along the directions and shifts specified by $\boldsymbol{\Theta}$ (depending on the specific implementation, this projection operator may include or not the aberrations caused by the electron microscope), and $||\cdot||_W$ is a weighted norm in which different pixels may be weighted differently according to some scheme adopted by the algorithm (the statistical distributions assumed for the noise and the alignment parameters automatically determine the form of this norm; this generic algorithmic framework may be adopted in real or Fourier space) (Sorzano et al., 2017). This problem is simply a data fidelity term (the reconstructed object has to be compatible with the acquired projections). The Bayesian approach adds *a priori* knowledge about the statistical distribution of the volumes being reconstructed that, in its turn, is translated into the minimization problem as an extra term that penalizes unlikely reconstructions

$$\boldsymbol{\Theta}^*, \mathbf{V}^* = \arg\min_{\boldsymbol{\Theta},\mathbf{V}} ||\widetilde{\mathbf{I}} - P_{\boldsymbol{\Theta}}\mathbf{V}||^2_W + f(\mathbf{V}) \quad (P2)$$

The traditional approach in cryoEM, projection matching, decomposes the minimization in $P1$ in two subproblems that are minimized separately and iteratively (k denotes the iteration number)

$$\boldsymbol{\Theta}^*_k = \arg\min_{\boldsymbol{\Theta}} ||\widetilde{\mathbf{I}} - P_{\boldsymbol{\Theta}}\mathbf{V}^*_{k-1}||^2_{W_1} \quad (P1a)$$

$$\mathbf{V}^*_k = \arg\min_{\mathbf{V}} ||\widetilde{\mathbf{I}} - P_{\boldsymbol{\Theta}^*_k}\mathbf{V}||^2_{W_2} \quad (P1b)$$

In its most traditional approach, $\widetilde{\mathbf{I}}$ is restricted to have a single copy of the experimental measurements $\mathbf{I}$, that is, only one alignment parameter set is estimated per projection. The first subproblem ($P1a$) is called angular assignment (each experimental image is assigned a set of parameters that encodes its projection direction and in-plane alignment with respect to the current guess of the macromolecular structure). P1a strongly depends on the initial map used as reference. The second subproblem uses the assigned angles to update the 3D map of the macromolecule. In practice, the Maximum Likelihood (ML) and Maximum A Posteriori (MAP) problems are solved through a numerical technique called Expectation Maximization that boils down to an iterative scheme similar to angular assignment and 3D reconstruction iterations (Scheres et al., 2005; Scheres, 2012) (the prior in MAP affects the specific form of the iterative step in ($P1b$); the fact that a Gaussian distribution is the conjugate prior of the distribution chosen for the likelihood term helps to keep the mathematical complexity of the EM iteration tractable). In the approach with subproblems ($P1a$ and $P1b$), it is still possible to introduce *a priori* knowledge about the class of volumes being reconstructed through the so-called image restoration methods (Sorzano et al., 2017). An example of map restoration is to restrict maps to be members of certain subsets defining properties a good map should have (e.g., non-negative maps, maps with compact space support, etc.) (Sorzano et al., 2008). Projection onto Convex Sets (Carazo, 1992), which were early introduced in the EM field, was a form of incorporating this *a priori* information. However, statistical properties of the volumes (like a Bayesian prior Scheres, 2012) or any other known feature of the volume being reconstructed could also be used in a restoration scheme.

The ML approach introduced an important concept in the EM community: an experimental image may occupy more than one projection direction and in-plane alignment, but with different probability. This probability gives its weighting factor during the reconstruction process. The rationale behind this idea is that images are so noisy that the maximum correlation peak calculated during the alignment is prone to errors and allowing the image to sit at different angles gives it more opportunities to find its correct localization (ideally, the likelihood distribution for a single experimental image should converge to a delta, although this is not always the case in practice for all images). This idea of more than one location was further exploited for the blind construction of an initial model (Sorzano et al., 2015). On the other hand, the fact that all experimental images are, in principle, allowed to occupy all projection directions (with different probabilities) may cause that some projection directions with intrinsically more Signal-to-Noise Ratio are over-represented (Vargas et al., 2016; Vargas et al., 2017), what we refer to as the attraction problem (Sorzano et al., 2010).

In this work, instead of using the image likelihood as weight, we used its significance (which is the probability that a random image taken from the set has a cross-correlation smaller than the correlation of this experimental image), and we promote, as in projection matching, an angular assignment in which each image receives a single angular orientation. Despite the appeal of letting an image to occupy several angular orientations (due to the uncertainty introduced by noise), in reality an image is known to come from a single (although unknown) orientation. Angular assignment algorithms will very likely commit assignment errors. For the sake of argument, let us consider that the error rate of the angular assignment is 30% (that is, 30% of the particles are assigned an incorrect orientation). If we now allow for a second angular assignment, with different weights, we know that for the 70% of particles that were correctly assigned, this second assignment will be incorrect. And for the remaining 30%, only about 70% will be correctly assigned. From the whole set of assignments (that now is twice the size of the dataset because of the double angular assignment), only 45.5% ($=0.7{\cdot}0.5 + 0.3{\cdot}0.7{\cdot}0.5$) of it has a correct angular assignment. That is, by increasing the number of positions in which a particle might sit (as a measure to fight noise), we have decreased the accuracy of our angular assignment, possibly resulting in a low pass filtering of the reconstructed volume due to the incoherent averaging in the Fourier space. In principle, this situation should be alleviated by the fact that the different projection directions have different weights, and that this weight distribution would ideally be very spiky around the true angular assignment. However, this ideal situation is not always the case. In Fig. 1 we show the weight profile for an experimental projection of a ribosome when it is compared to the whole gallery of projections of the final reconstructed volume sampled every 5 degrees and with a maximum shift of 24 Å. More than 35 million combinations of orientation and positions were explored. Relion and *highres* both agreed on the angular assignment and shift of this particle, with a precision of less than 1° in the angles and 1 pixel in the shifts. Relion is based on the l_2 norm of the difference between the volume reprojection and the experimental projection with a frequency weight given by the noise variance, while *highres* is based on the correlation of the volume

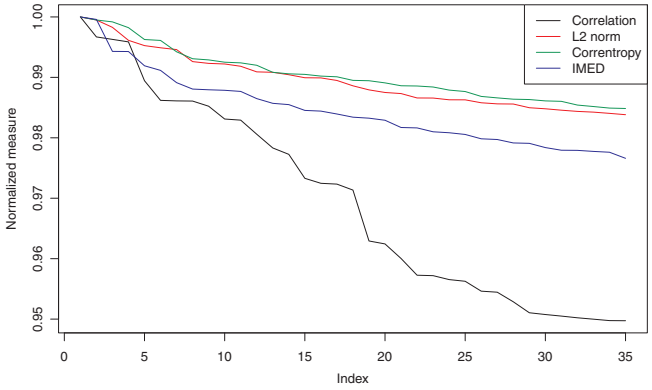

Fig. 1. Decay of the weight of an experimental particle under different image similarity measures. For comparison purposes all similarities observed in the projection gallery have been normalized to be between 0 and 1, being 1 the best matching projection. The algorithm presented in this work is using Correlation and IMED.

reprojection and the experimental projection and the IMED (Sorzano et al., 2015). For comparison purposes we have also included the correntropy (Sorzano et al., 2010), which should be dissimilar from the l_2 norm only for noise distributions deviated from the Gaussian. We only plot the similarity measures of the 99.9999% top coefficients. For representation purposes we have scaled the similarity/distance measures between 0 and 1, so that the best matching reprojection gets similarity 1. As can be seen from this figure, the weight distribution is far from being very spiky, the weights gradually drop from 1 and some of the measures drop faster than others. As the noise distribution is Gaussian, the l_2-norm and the correntropy have similar decays, as expected. IMED has a slightly better decay, especially at the beginning of the curve (for the first 1–5 reprojections), and the linear correlation is the most discriminant function (its slope at the origin is larger than any other function). For both, Relion and *highres*, the top 99.9999% coefficients span an angular coverage of ± 8°, ±9.5° and ± 5° in the 1st, 2nd and 3rd Euler angles, respectively, around the best matching angle. The shifts had an uncertainty of ± 6 Å. As suggested in the rest of the text, it is recommended to use *higres* with a single angular and shift assignment, so that the final result is not affected by these uncertainties, although this strategy does not preclude alignment errors. The fact that the correlation and IMED have larger slopes at the best matching position help them to better identify the position of each experimental projection in the presence of noise.

Multiple angular assignments may still be useful when trying to find the correct structure from an incorrect initial volume. Slow convergence from very low resolution (typically, initial volumes are filtered at about 60 Å) to high resolution, as is done in Relion, is a useful strategy to get away from local minima. However, we propose that this strategy of multiple angular assignments is dropped as high resolution is achieved.

In this paper we introduce a new algorithm for solving the problem of reconstructing the 3D density map of a macromolecule from an homogeneous population of images. It is based on the following subproblem decomposition: 1) Angular assignment in which an experimental image can occupy one or multiple orientations, although this latter option is strongly discouraged; 2) 3D Reconstruction with weights; 3) Volume restoration. The algorithm is normally used in the so-called "gold-standard" paradigm (Scheres and Chen, 2012), that is, the original dataset is split in two halves that never see each other. However, this is not a requirement of the algorithm. We now describe the specific choices of our algorithm. The algorithm is available through `Scipion` (de la Rosa-Trevín et al., 2016) from Release 1.1 under the name *highres*.

Regarding the results of the challenge we should note that the algorithm was not fully implemented by the time of the first deadline to submit the results. We submitted some results that were evaluated by the challenge assessors, and that are the ones reported in the other companion papers of this special issue. Then, the challenge organizers opened a second round of submissions in which we submitted our final results that, to the best of our knowledge, have not been reevaluated by all assessors. In this paper, we present the algorithm along with comparisons of its results with results from Relion run by us for the same dataset similarly preprocessed.

2. Methods

In this section we give the details of the algorithm: 1) We introduce the multiresolution approach as a way to prevent overfitting and removing the effect of unnecessary amounts of noise; 2) Then, we explain the details of the global and local angular assignments (including local refinement of the defocus, magnification and gray normalization); 3) It follows the explanation of the 3D reconstruction step which includes a particular weighing scheme and the correction of the CTF envelope before reconstructing; 4) We then present the dampening of all features that are not significantly different from noise; 5) Finally, we combine the two halves from the gold-standard into the reconstruction valid for this iteration using a superresolution method.

As usual, the whole algorithm is iterative and once a new reference is constructed, it can be used to refine the geometrical alignment of the particles in the next iteration.

2.1. Multiresolution

We have chosen to solve the 3D alignment and reconstruction problem in a multiresolution framework: the pixel size of the images is progressively decreased so that early iterations are performed aiming at low resolution reconstructions and, as iterations proceed, higher resolutions are sought. The advantages of this approach are not only its higher computational speed, but it also helps to smooth the landscape of solutions of the minimization problem increasing the probability of not getting trapped in local minima (Sorzano et al., 2005). Additionally, the fact that the sampling rate changes from one iteration to the next helps to avoid creating a fixed pattern in the projection gallery used for global angular assignment (fixed projection patterns may induce alignment artifacts and, ultimately, overfitting).

Too many degrees of freedom at the early iterations may produce overfitting. To avoid this overfitting we limit the target resolution at a given iteration. At each iteration the user may specify the target resolution in Å ($R_{target,k}$). Assume the resolution achieved by the algorithm (as measured by the cross of the Fourier Shell Correlation, FSC, with the 0.143 threshold) at the previous iteration was R_{k-1}. This user supplied target resolution is limited to

$$R_{target,k}{}' = \max\left(\frac{R_{k-1}}{2}, R_{target,k}\right)$$

if the alignment at this iteration is global, and to

$$R_{target,k}{}' = \max\left(\frac{4R_{k-1}}{5}, R_{target,k}\right)$$

if the alignment at this iteration is local. These target resolutions have been heuristically chosen, and their only aim is to slow the pace at which the multiresolution progresses so that the algorithm does not early fall into a local minimum.

Then, the sampling rate (Å/pixel) at this iteration is calculated as

$$T_k = \min\left(T_s, \frac{R_{target,k}{}'}{3}\right)$$

being T_s the original sampling rate. This sampling rate guarantees that, if the original sampling rate is high enough, the target resolution at this

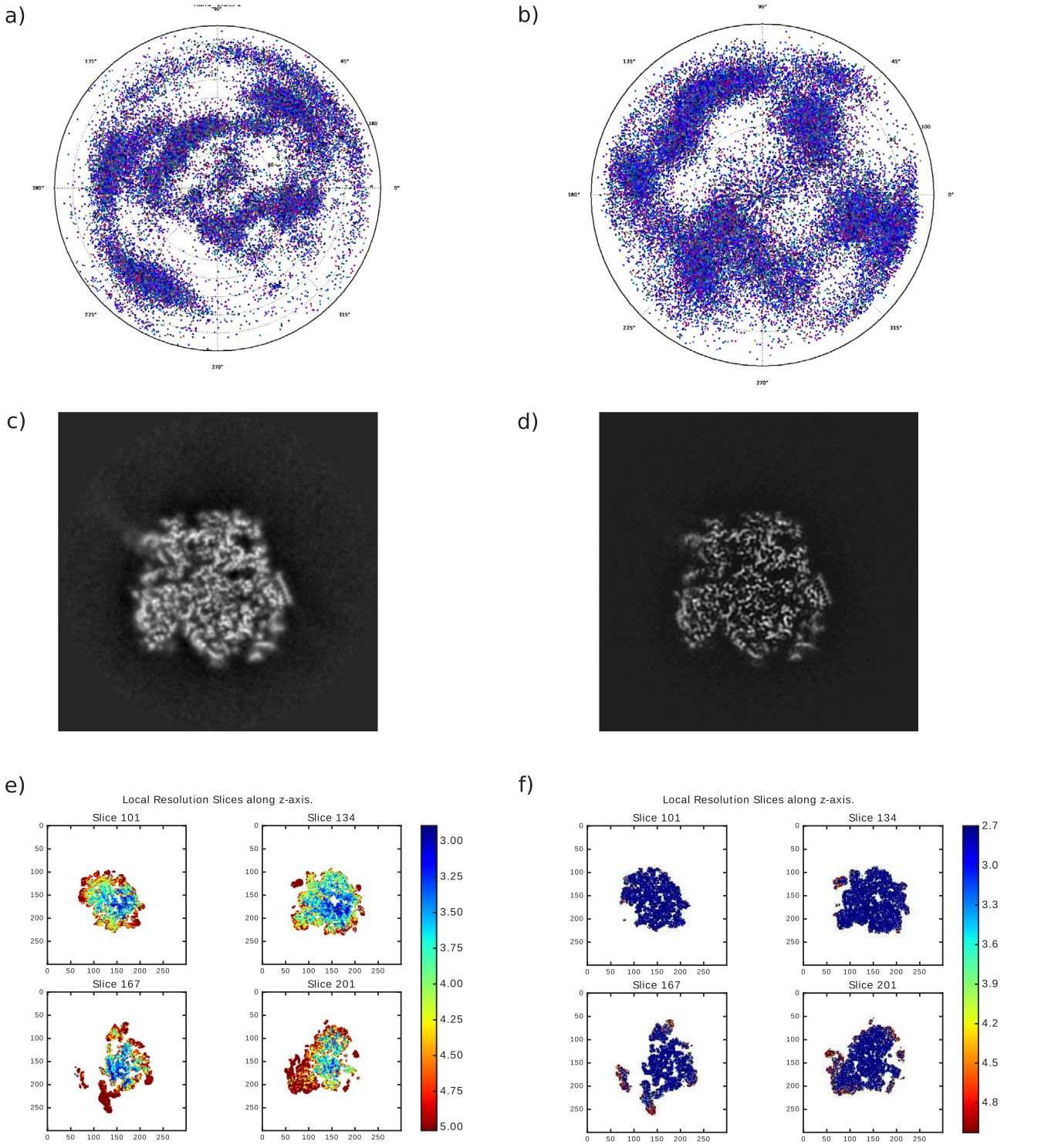

Fig. 2. Left column: Results for Relion. Right column: Results for *highres*. a) and b): Angular distribution of the dataset (each dot represents an input image on the projection sphere). c) and d) A sample slice of the 3D reconstructed volume. e) and f) Local resolution estimate at four different slices.

iteration occupies 2/3 of the Fourier space.

Given the size of the images at the current sampling rate, X_{dim}, the angular step for global angular assignment is calculated as

$$\Delta\theta = \max\left(3^\circ, \tan^{-1}\frac{R_{target,k}}{X_{dim}\,T_k}\right)$$

This limit on the angular sampling rate aims at gaining speed and, again, avoiding too many degrees of freedom at early iterations.

2.2. Angular assignment

A short summary of our approach would be: align globally until the alignment stabilizes, then refine locally. The map used as reference is a non-negative, low-pass filtered and masked map constructed from the reconstruction of the previous iteration. The process of the construction of the next reference is highly configurable in the algorithm implementation in `Scipion`, but the goal is to remove all small details that could serve as anchors for noise overfitting, in Fourier as well as in

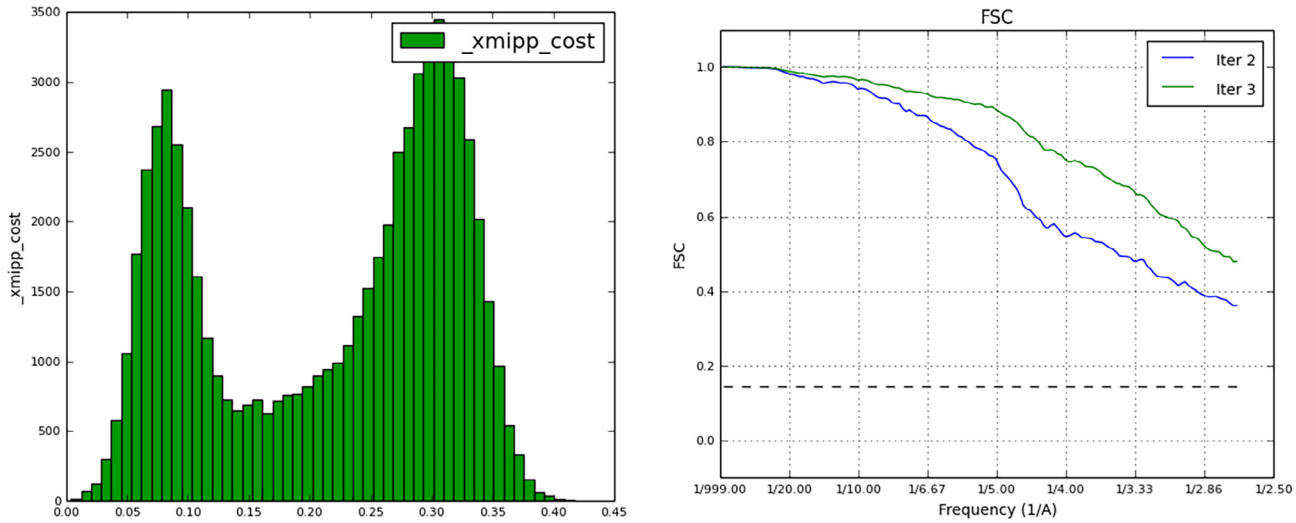

Fig. 3. Left: Histogram of the correlation within a circular mask as calculated during the continuous angular assignment. The two modes reveal the existence of two populations. Right: Difference in FSC with the full dataset (Iteration 2) and with the dataset after removing those projections whose cross correlation was below 0.14 (Iteration 3).

real space.

- *Construction of the next reference*. By default, the current reconstruction is filtered by the Fourier Shell Correlation at the current iteration as a way to avoid overfitting (Sindelar and Grigorieff, 2012). Additionally, the user may choose to apply a low-pass filter with a frequency defined by the current resolution, computed adding or subtracting a user-defined offset. The user may also choose among applying a spherical mask, a non-negative mask, with all negative voxels set to 0, or a user-defined mask. This mask should be wide enough to hold the structure features but avoiding background overfitting. Finally, the user may give a shell command or batch that filters/masks the reference volume in any desired way.
- *Global alignment*. For the global alignment we use an algorithm that takes into account the significance of a volume reprojection to the experimental image and viceversa (`xmipp_reconstruct_significant` (Sorzano et al., 2015)). The similarity between two images is measured by correlation and IMED. Based on these significances the alignment algorithm computes a weight that is later used for reconstruction (Sorzano et al., 2015 Eq. (7)). Note that this weight may be 0, implying that the image is not considered for 3D reconstruction at this iteration. The significance in this application is set very high such that each experimental image participates in very few (we recommend 1, or at most 2) projection directions, being this number specified by the user. The angular sampling at each iteration is different (see the previous section). This change of angular sampling also helps to avoid overfitting between iterations. Due to the dependence between correlation and microscope defocus, this global angular assignment is performed by defocus groups. From one iteration to the next, the implementation of the algorithm allows inspecting the set of images and filtering out images with low significance weight, low correlation, large shifts, etc. This step is optional and its need depends on the quality of the input dataset.
- *Local alignment*. The stabilization of the global angular assignment can be measured by the stabilization of the Fourier Shell Correlation curve (as it is normally done in the field), or, better still, by evaluating the angular changes for all experimental images from one iteration to the next. Typically, 3 global iterations are enough if we start from a low resolution model, and 1 global iteration if we start from a high resolution model. The user may decide at this moment to go for a local refinement. While global assignments are based on a discrete angular sampling of the projection sphere, local assignments are performed in a continuous angular space (Grigorieff, 2007; Jonić et al., 2005). We also allow for anisotropic scale (Grigorieff, 2007), particle-wise defocus, and gray value optimization (the experimental images are linearly transformed so that the distance between them and the assigned projections is minimized; this optimization aims at correcting gray normalization errors performed at the beginning of the processing). These latter options increase the number of parameters to optimize and we recommend to progressively increase the number of optimization parameters (e.g., first, angular assignment; second, angular assignment, scale and defocus; finally, all parameters). The idea is to perform the optimization in a hierarchical way, starting from the most important parameters to avoid the local minima created by the presence of many variables. The alignment parameters are optimized using Powell's method (Powell, 1964). The objective function is the cross-correlation between the reprojected volume taking into account the Contrast Transfer Function (CTF) and the experimental image within a circular mask (except when the gray value transformation must be estimated, in which the objective function is the l_1 norm of the difference between the experimental image and the gray transformed projection).

2.3. 3D Reconstruction with weights and Wiener amplitude correction

For the 3D reconstruction we use a weighted, direct Fourier inversion algorithm (Abrishami et al., 2015). This algorithm proved to produce very accurate results up to very high resolution and it is tightly related to a robust interpolation problem in Fourier space. Weights are given directly from the previous step through the significance weights calculated by the global or local alignment and scaled between w_{min} (typically, $w_{min} = 0.1$, although this value can be selected by the user) and 1. This scaling is performed by defocus groups to avoid the dependence of the correlation with the defocus. The images used for the 3D reconstruction are previously Wiener corrected in order to compensate for the Fourier envelope and oscillations caused by the microscope aberrations (CTF). However, it is important to perform a CTF correction per particle, since the envelope of the CTF depends on the specific aberrations of each micrograph as well as the defocus of each particle (Sorzano et al., 2007). Otherwise, the Fourier interpolation problem is fed with systematically inconsistent data (systematic as opposed to the random nature of the inconsistency introduced by noise). Symmetry is handled by replicating the input images at different

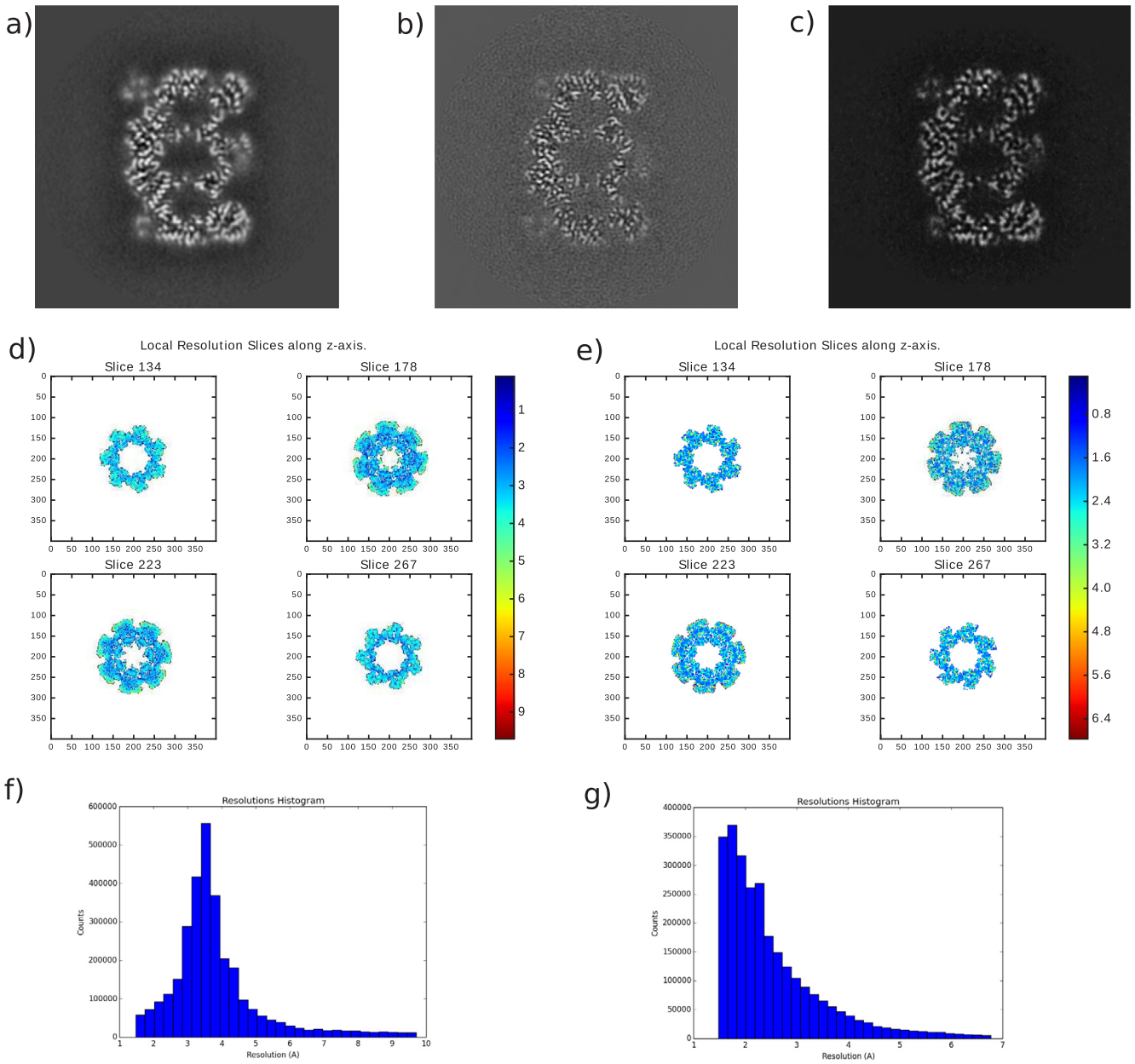

Fig. 4. Left column: Results for Relion. Right column: Results for *highres*. a), b) and c) A sample slice of the 3D reconstructed volume by Relion Autorefine, Relion Autorefine + Postprocessing, and *highres*. d) and e) Local resolution estimate at four different slices. f) and g) Histogram of the local resolution.

angular positions according to the specific symmetry, and using these symmetrized projections during the 3D reconstruction process.

2.4. Signal detection and voxel significance

Single Particle Analysis is normally performed in a context in which there is a region of the volume for which it is known that there should not be any structural detail. This area is called the background and it is normally defined by a mask. We analyze the statistical behavior of noise in this area by estimating its cumulative probability distribution, $F_N(n)$. Then, every voxel in the volume with gray value v, either in the background or foreground area, is multiplied by $F_N(v)$. That is, voxels whose value is significantly larger than noise are multiplied by 1, while those voxels not significantly larger than the noise will be dampened. The overall effect of the filter is a denoising of the input reconstructed volume.

This idea can be extended to any transformation of the input volume. For instance, if we band-pass filter the volume, we may apply the non-significant dampening above to each one of the band-pass filtered versions of the input volume. Then, we may simply combine all these denoised volumes

$$V_{\text{filterbank}} = \frac{\sum_\omega F_{N^2,\omega}(V_\omega^2)V_\omega}{\sum_\omega F_{N^2,\omega}(V_\omega^2)}$$

where ω defines the central frequency of the band-pass filter, V_ω is the band-pass filtered version of the reconstructed volume, and $F_{N^2,\omega}$ is the cumulative probability function of the instantaneous energy of the band-pass filtered noise. Note that the noise distribution of any of the volume transformations in this section is always measured in the background area.

Finally, we may further use this distinction between the noise and signal behavior to other signal operators. In particular, we may analyze the noise response to the operator $T = I + \Delta$, being I the identity operator, and Δ the Laplacian operator. This operator returns the volume

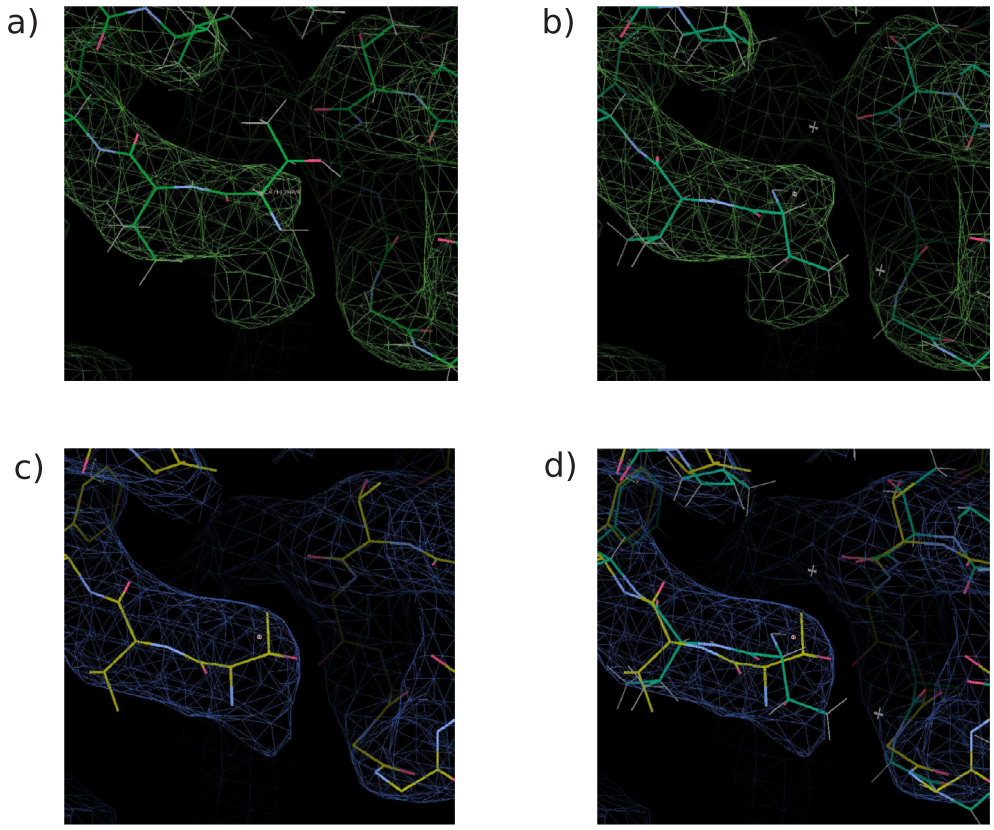

Fig. 5. a: Fitting of the challenge atomic model to the Relion map. b: Fitting of the refined atomic model to the Relion map. c: Fitting of the refined atomic model to the *highres* map. d: Both atomic models (green: refined to Relion, yellow: refined to *highres*) displayed in the *highres* map.

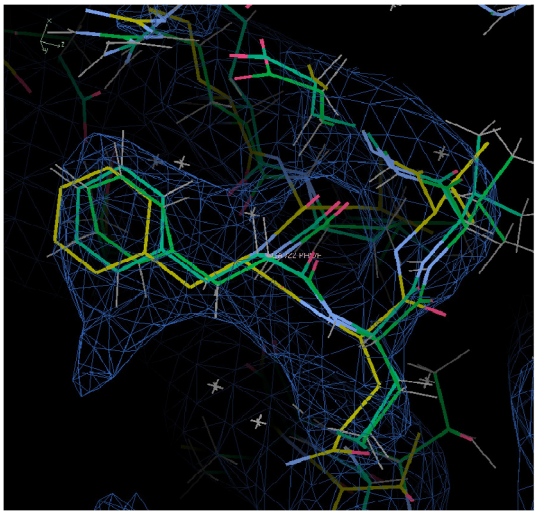

Fig. 6. Detail of the differences between the challenge atomic model (yellow), the atomic model fitted to Relion (dark green) and the *highres* (green) maps displayed in the *highres* map. Although the overall difference is only 0.79 Å in RMSD, there are regions where the orientation of side chains, like the phenylalanine in this figure, significantly change.

with enhanced borders. The rationale for involving the Laplacian of the input signal is that this operator boosts high frequencies with respect to low frequencies. In this way we may analyze which is the behavior of noise high frequencies (in the noise area) and suppress similar behaviors in the signal area. The idea is to find a denoised volume $\widetilde{V}$, such that

$$T(\widetilde{V}) = F_{N,|T(V)|}(|T(V)|)T(V)$$

where $|\cdot|$ is the absolute value, and $F_{N,|T(V)|}$ is the cumulative distribution function of the noise energy under the transformation $|T(V)|$. That is, the operator applied to the denoised volume must be equal to the significance weighted version of the same operator applied to the input volume. Without the weighting, the obvious solution is $\widetilde{V} = V$, so that the solution of the previous equation is not a volume with enhanced borders, the main effect on $\widetilde{V}$ is denoising, not sharpening. This equation is tightly related to the retinex image model and its solution is also found in Fourier space as it was done in Limare et al. (2011) (see this reference of the implementation of the solution to this equation).

Note that all these transformations aim at keeping the components of the volume that behave, under various transformations, significantly different from noise. This is exactly a task of signal detection in noise. The noise suppression operations are performed in the whole volume, not only the background, so that the overall effect is not simply masking the volume, but inside the molecule it is expected that these operators reduce the noise in the same way as outside it. All these transformations are currently available in the implementation of *highres* within `Scipion` and the user may decide to activate them or not.

2.5. Volume restoration

The two maps coming from the two dataset halves in the gold-standard may be combined using image superresolution (Park et al., 2003) rather than a simple average (note that the two half maps remain to be independent in next iterations, only their combination, which is not propagated to the following iterations, sees both volumes at the same time).

Given the two independently reconstructed volumes, V_1 and V_2, the traditional approach in EM combines them by a simple average. We here propose to use a superresolution combination. The idea is to find a higher resolution volume, V^*, and two parameters σ_1^* and σ_2^* such that

$$V^*, \sigma_1^*, \sigma_2^* = \arg\min_{V,\sigma_1,\sigma_2} ||V \star K_{\sigma_1} - V_1||^2 + ||V \star K_{\sigma_2} - V_2||^2$$

where $\star$ is the convolution operator and K_σ is a Gaussian kernel whose standard deviation is σ. Note that this volume does not participate in the calculation of the resolution between the two halves, and it is used only as the output at each iteration.

3. Results

We illustrate the performance of the proposed algorithm with two different datasets: the *Plasmodium falciparum* 80S ribosome bound to the anti-protozoan drug emetine (EMPIAR Entry: 10028 Wong et al., 2014, an asymmetric structure), and the T20S Proteasome (EMPIAR Entry: 10025 Campbell et al., 2015, a D7 symmetric structure). In both examples we use *highres* as a refinement algorithm working on the final result from a previous 3D reconstruction algorithm (in the examples below, Relion). This is not a hard constraint of the method, and results starting from the output of initial volume algorithms also converge to the same structures as the ones reported in this section. However, they involve at least three global assignment iterations to reach the high resolution regime at which the local refinement is performed. Since Relion is currently so fast on GPUs, we can easily avoid these extra global assignment iterations by directly starting at the output of Relion.

3.1. Plasmodium falciparum 80S ribosome

68,265 projection images were extracted from 600 micrographs (see the experimental details of the acquisition at Wong et al. (2014); these particles come after a particle picking performed in `Scipion` with the Xmipp automatic picking algorithm (Abrishami et al., 2013) and cleaning of the particles using CL2D stable cores (Sorzano et al., 2010; Sorzano et al., 2014)). CTFfind4 was used to estimate the CTF required by Relion and Xmipp CTF, which additionally estimates the CTF envelope, for our method. We used Relion autorefine, which resulted in a structure with a resolution of 3.9 Å (after Relion postprocessing, the resolution goes up to 3.4 Å, however, as is shown in the next section, this increase in resolution severely degrades in relation to an atomic model, while before postprocessing it has a much better agreement). We then continued with *highres*. In this example we allowed for only one angular position for each experimental image. We performed a single global angular assignment and 4 iterations of local refinement following the progressive strategy in parameters described in Section 2.2. The global alignment took 10 days in a machine with 48 CPU processors and 16 GB of RAM, while the local alignment took 13 h in the same machine. Fig. 2 shows the results for both approaches, the two maps are in the same orientation so that the projection direction diagrams are comparable. As can be seen from Figs. 2a and b, the angular coverage achieved by our method is larger than the one from Relion. Only 31.8% of the images were assigned angles by Relion and *highres* with an angular difference less than 15°, and more than 48.9% of the particles moved further than 15° between the global and the local refinement inside *highres*. The fact that only a minority of the experimental projections have a similar angular assignment between Relion and *highres* indicates that it is not only the refinement of parameters like the defocus, magnification or gray value normalization what is making a difference between Relion and the new algorithm, but the whole angular assignment procedure. Figs. 2c and d show a representative slice by Relion and *highres*. Note that the slice of our reconstruction has not been masked, but because of the signal detection operations described in Section 2.4, the noise in the background and foreground has been strongly attenuated.

The FSC resolution reported by our method was 2.7 Å (the same masks were used in Relion and our method; the original publication reported 3.2 Å (Wong et al., 2014)). The higher resolution of our method is confirmed by the local resolution analysis performed with MonoRes (Vilas et al., 2018) (Figs. 2e and f). The median local resolution for Relion was 4.9 Å, while for our method was 2.96 Å (note that MonoRes local resolution estimates are invariant to isotropic transformations like the B-factor correction of Relion postprocessing). The median local resolution for the volume deposited at EMDB (entry ID: 2660) was 6.88 Å (but this volume has a severe misalignment of the 40S subunit).

Our method allows discriminating particles by different criteria at each iteration. After the first local iteration, we inspected the cross-correlation between reprojections of our volume and the experimental projections within a circle enclosed in the projection image (this correlation is referred to as *cost*). We discovered the presence of two modes indicating the presence of two subpopulations in the dataset (see Fig. 3). After thresholding on this criterion (all images with *cost* < 0.14 were eliminated), we see an important improvement of the gold standard FSC (see Fig. 3).

3.2. T20S proteasome

22,884 projection images were extracted from 196 micrographs (see the experimental details of the acquisition at Campbell et al. (2015)). We followed the standard image processing pipeline used with Relion, including postprocessing (although this latter option resulted in a much worse agreement with the atomic model). Then, we continued the processing with *highres* making two global iterations with only one angular position per experimental image, and three local iterations progressively increasing the set of parameters to optimize, as recommended in the Methods Section. The global alignment took 16 h in the same machine as in the previous structure, and the local alignment 14 h. Fig. 4 shows the results for both approaches. The FSC reported by Relion was 3.4 Å, while our method reported 2.5 Å. The median of the local resolution analysis (Vilas et al., 2018) for Relion was 3.5 Å and for our method was 2.2 Å (the originally reported resolution was 2.8 Å (Campbell et al., 2015)).

We constructed an atomic model of the T20S proteasome based on the atomic model deposited at the Map Challenge web page. We followed the same procedure for the Relion and *highres* map. The challenge atomic model did not fit in some of the places (see Fig. 5 a). For this reason we refined the atomic model using Coot (Emsley et al., 2004) real space refinement in both maps, Relion and *highres* (see Fig. 5 b, c and d). The refined models are not largely different between each other (the RMSD between both is 0.79 Å, see Figs. 5d and 6). We evaluated the quality of the fitting using EMRinger (Barad et al., 2015). The EMRinger score for the Relion map was 2.61 and for the *highres* map 2.69; the rotamer ratio increased from 0.886 in Relion to 0.931 in *highres*; and the max Zscore increased from 4.085 to 4.208. It has become customary to perform a Relion postprocess after Relion autorefine. In this case, the EMRinger score drops to 2.01, the rotamer ratio drops to 0.729, and the max Zscore drops to 3.138. The reason is that B-factor correction flattens a middle range of the Fourier spectrum of the map, while macromolecules are not flat in this region.

4. Conclusions

In this article we have introduced a new algorithm for aligning and reconstructing EM projection images in Single Particle Analysis. The algorithm is based on a subproblem decomposition of the 3D alignment and reconstruction problem, as was traditionally done in the field before ML and MAP approaches were introduced. In our method the concept of significance is deeply involved (at the level of angular alignment, 3D reconstruction and volume denoising). The algorithm is based on a few ideas: 1) Two halves of the dataset are independently refined; 2) An image can occupy multiple positions with a merit function calculated from the significance of the angular alignment, however, we strongly recommend to use only 1 position; 3) The angular alignment is performed globally until it stabilizes at the level of angular

changes and FSC; 4) Local refinement includes anisotropic magnification, defocus and gray value optimization beside the more classical angular alignment parameters; 5) The CTF amplitude is corrected by Wiener filtering before performing the 3D reconstruction; 6) The reconstruction is performed in Fourier space with weights calculated from the significance of the merits of the alignment by defocus groups; 7) The reconstructed volumes are denoised analyzing the significance of the coefficients in the signal area under various transformations; 8) The two independent halves are combined into a single volume using a superresolution algorithm.

We have shown that the combination of these steps extract information from the electron micrographs to high resolution. The algorithm is still open for new weighting schema taking into account the Signal-to-Noise Ratio of the input images as well as their angular stability, for example.

Acknowledgements

The authors would like to acknowledge economical support from: the Comunidad de Madrid through grant CAM(S2010/BMD-2305), the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638 and FEDER BIO2016-76400-R, and the Fundación General CSIC (Programa ComFuturo).

References

Abrishami, V., Bilbao-Castro, J.R., Vargas, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., Oct 2015. A fast iterative convolution weighting approach for gridding-based direct fourier three-dimensional reconstruction with correction for the contrast transfer function. Ultramicroscopy 157, 79–87.

Abrishami, V., Zaldívar-Peraza, A., de la Rosa-Trevín, J.M., Vargas, J., Otón, J., Marabini, R., Shkolnisky, Y., Carazo, J.M., Sorzano, C.O.S., Oct 2013. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics 29 (19), 2460–2468.

Barad, B.A., Echols, N., Wang, R.Y.-R., Cheng, Y., DiMaio, F., Adams, P.D., Fraser, J.S., Oct 2015. Emringer: side chain-directed model and map validation for 3d cryo-electron microscopy. Nat. Methods 12 (10), 943–946.

Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S., Carragher, B., 2015. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. Elife 4.

Carazo, J.M., 1992. The fidelity of 3D reconstructions from incomplete data and the use of restoration methods. In: Frank, J. (Ed.), Electron Tomography. Three-Dimensional Imaging with the Transmission Electron Microscope. Plenum Press, New York, pp. 117–166.

de la Rosa-Trevín, J.M., Quintana, A., Del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, 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, 93–99.

Emsley, P., Cowtan, K., Dec. 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Sect. D, Biol. Crystallography 60, 2126–2132.

Grigorieff, N., 2007. Frealign: high-resolution refinement of single particle structures. J. Struct. Biol. 157, 117–125.

Jonić, S., Sorzano, C.O.S., Thévenaz, P., El-Bez, C., De Carlo, S., Unser, M., 2005. Spline-based image-to-volume registration for three-dimensional electron microscopy. Ultramicroscopy 103 (104), 303–317.

Limare, N., Petro, A.B., Sbert, C., Morel, J.M., 2011. Retinex poisson equation: a model for color perception. Image Process. Line 1, 39–50.

Nogales, E., 2016. The development of cryo-EM into a mainstream structural biology technique. Nat. Methods 13 (1), 24–27.

Park, S.C., Park, M.K., Kang, M.G., 2003. Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. 20 (3), 21–36.

Penczek, P., Radermacher, M., Frank, J., 1992. Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy 40, 33–53.

Penczek, P.A., Grasucci, R.A., Frank, J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53, 251–270.

Powell, M.J.D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer J. 7, 155–162.

Punjani, A., Rubinstein, J., Fleet, D.J., Brubaker, M.A., Feb. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296.

Scheres, S.H.W., 2012. Relion: implementation of a bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180 (3), 519–530.

Scheres, S.H.W., Chen, S., 2012. Prevention of overfitting in cryo-em structure determination. Nat. Methods 9 (9), 853–854.

Scheres, S.H.W., Valle, M., Núñez, R., Sorzano, C.O.S., Marabini, R., Herman, G.T., Carazo, J.M., 2005. Maximum-likelihood multi-reference refinement for electron microscopy images. J. Mol. Biol. 348, 139–149.

Sindelar, C.V., Grigorieff, N., 2012. Optimal noise reduction in 3d reconstructions of single particles using a volume-normalized filter. J. Struct. Biol. 180, 26–38.

Sorzano, C.O.S., Bilbao-Castro, J.R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernández, G., Li, M., Xu, G., Marabini, R., Carazo, J.M., 2010. A clustering approach to multireference alignment of single-particle projections in electron microscopy. J. Struct. Biol. 171, 197–206.

Sorzano, C.O.S., Jonic, S., Núñez Ramírez, R., Boisset, N., Carazo, J.M., 2007. Fast, robust and accurate determination of transmission electron microscopy contrast transfer function. J. Struct. Biol. 160, 249–262.

Sorzano, C.O.S., Thévenaz, P., Unser, M., 2005. Elastic registration of biological images using vector-spline regularization. IEEE Trans. Biomed. Eng. 52, 652–663.

Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Otón, J., Álvarez-Cabrera, A.L., Abrishami, V., Sesmero, E., Marabini, R., Carazo, J.M., 2015. A statistical approach to the initial volume problem in single particle analysis by electron microscopy. J. Struct. Biol. 189 (3), 213–219.

Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Zaldívar-Peraza, A., Otón, J., Abrishami, V., Foche, I., Marabini, R., Caffarena, G., Carazo, J.M., 2014. Outlier detection for single particle analysis in electron microscopy. In: Proc. Intl. Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO, pp. 950.

Sorzano, C.O.S., Vargas, J., Otón, J., Abrishami, V., de la Rosa Trevín, J.M., del Riego, S., Fernández-Alderete, A., Martínez-Rey, C., Marabini, R., Carazo, J.M., 2015. Fast and accurate conversion of atomic models into electron density maps. AIMS Biophys. 2, 8–20.

Sorzano, C.O.S., Vargas, J., Otón, J., Vilas, J.L., Kazemi, M., Melero, R., del Caño, L., Cuenca, J., Conesa, P., Gómez-Blanco, J., Marabini, R., Carazo, J.M., 2017. A survey of the use of iterative reconstruction algorithms in electron microscopy. BioMed Res. Intl. 2017, 6482567.

Sorzano, C.O.S., Velázquez-Muriel, J.A., Marabini, R., Herman, G.T., Carazo, J.M., 2008. Volumetric restrictions in 3DEM reconstruction. Pattern Recogn. 41, 616–626.

Vargas, J., Melero, R., Gómez-Blanco, J., Carazo, J.M., Sorzano, C.O.S., 2017. Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis. Sci. Rep. 7, 6307.

Vargas, J., Otón, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2016. Particle alignment reliability in single particle electron cryomicroscopy: a general approach. Sci. Rep. 6, 21626.

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

Supplementary Information: High-resolution reconstruction of Single Particles by Electron Microscopy

---

Supplementary Information

Mathematical/Implementation details

Multiresolution. At each iteration the user may specify the target resolution in Å($R_{target,k}$). Assume the resolution at the previous iteration was R_{k-1}. This user supplied target resolution is limited to

$$R_{target,k} = \max\left(\frac{R_{k-1}}{2}, R_{target,k}\right)$$

if the alignment at this iteration is global and to

$$R_{target,k} = \max\left(\frac{4R_{k-1}}{5}, R_{target,k}\right)$$

if the alignment at this iteration is local.

The sampling rate (Å/pixel) at this iteration is calculated as

$$T_k = \min\left(T_s, \frac{R_{target,k}}{3}\right)$$

being T_s the original sampling rate. This sampling rate guarantees that, if the original sampling rate is high enough, the target resolution at this iteration occupies 2/3 of the Fourier space. The reason for limiting the target resolution at a given iteration is avoiding overfitting by having two many degrees of freedom at the early iterations.

Given the size of the images at the current sampling rate, X_{dim}, the angular step for global angular assignment is calculated as

$$\Delta\theta = \max\left(3^\circ, \tan^{-1}\frac{R_{target,k}}{X_{dim}T_k}\right)$$

This limit on the angular sampling rate aims at gaining speed and, again, not having too many degrees of freedom at the level of global alignments.

Angular assignment.

- Construction of the next reference. By default, the current reconstruction is filtered by the Fourier Shell Correlation (FSC) at the current iteration as a way to avoid overfitting[4]. Additionally, the user may choose to apply a lowpass filter with a frequency defined by the current resolution (the current resolution plus or minus a user-defined offset), to apply a user-defined mask (normally wide enough as to hold the structure features but not background overfitting), a spherical mask, and a non-negative mask (all negative voxels are set to 0). Additionally, the user may give a shell command or batch that filters/masks the reference volume in any desired way.

- Global alignment. For the global alignment we use an algorithm that takes into account the significance of a volume reprojection to the experimental image and viceversa (`xmipp_reconstruct_significant` [5]). Based on these significances the alignment algorithm computes a weight that is later used for reconstruction (Sorzano *et al.*[5] Eq. 7). The significance in this application is set very high such that each experimental image participates at most at a few (typically, 1 or 2) projection directions, being this number specified by the user.

- Local alignment. The alignment parameters (angular orientation and in-plane shift, anisotropic scale changes, defocus refinement and gray value optimization) are optimized using Powell's method[3]. The objective function is the cross-correlation between the reprojected volume taking into account the CTF and the experimental image within a circular mask (except when the gray value transformation must be estimated, in which the objective function is the L1 norm of the difference between the experimental image and the gray transformed projection).

3D Reconstruction with weights. Our 3D reconstruction algorithm in Fourier space admits the use of weights[1]. We divide the experimental images into defocus groups. Within each defocus group, we use as weights the percentile occupied by the objective function of each experimental image within this group mapped between $w_{\min}$ (this value is defined by the user and it is typically 0.1) and 1. In this way, good images (those whose objective function is well ranked with respect to the rest of the group) have higher weights than worse images. This is a way of pushing the significance of the angular assignment of each image into the 3D reconstruction.

The images used for reconstruction are the Wiener filtered versions of the experimental images. In this way, we compensate for the effects of the CTF envelope at different frequencies. In our view, it is important to introduce amplitude compensated Fourier components in the Fourier interpolation process implied by the 3D Fourier reconstruction. Otherwise, the 3D reconstruction would incorrectly estimate the Fourier coefficients of the reconstructed volume.

Signal detection and voxel significance. Single Particle Analysis is normally performed in a context in which there is a region of the volume for which it is known for sure that there should not be any structural detail. This area is

called the background and it is normally defined as a mask (in the absence of user-defined mask the largest sphere fitting inside the box containing the macromolecule serves as a suitable mask defining the background). We analyze the statistical behavior of instantaneous energy of noise in this area by estimating its cumulative probability distribution, $F_N(n)$. Then, every voxel in the volume with gray value v, either in the background or foreground area, is multiplied by $F_N(v)$. That is, voxels whose value is significantly larger than noise are multiplied by 1 or a number close to 1, while those voxels not significantly larger than the noise will be dampened. The overall effect of the filter is a denoising of the input reconstructed volume.

This idea can be extended to any transformation of the input volume. For instance, if we bandpass filter the volume, we may apply the non-significant dampening above to each one of the bandpass filtered versions of the input volume. Then, we may simply combine all these denoised volumes

$$V_{\text{filterbank}} = \frac{\sum_{\omega} F_{N^2,\omega}(V_\omega^2) V_\omega}{\sum_{\omega} F_{N^2,\omega}(V_\omega^2)}$$

where ω defines the central frequency of the bandpass filter, V_ω is the bandpass filtered version of the reconstructed volume, and $F_{N^2,\omega}$ is the cumulative probability function of the energy of the bandpass filtered noise.

Finally, we may further use this distinction between the noise and signal behavior to other signal operators. In particular, we may analyze the noise response to the operator $T = I + \Delta$, being I the identity operator, and Δ the Laplacian operator. The idea is to find a denoised volume $\tilde{V}$, such that

$$T(\tilde{V}) = F_{N,|T(V)|}(|T(V)|)T(V)$$

where $|\cdot|$ is the absolute value. That is, the operator applied to the denoised volume must be equal to the significance weighted version of the same operator applied to the input volume. This equation is tightly related to the retinex image model and its solution is also found in Fourier space as it was done in Limare *et al*[2]. The rationale for involving the Laplacian of the input signal is that this operator boosts high frequencies with respect to low frequencies. In this way we may analyze which is the behavior of noise high frequencies (in the noise area) and suppress similar behaviors in the signal area.

Note that all these transformations aim at keeping the components of the volume that behave, under various transformations, significantly different from noise.

Post-processing. The volumes coming out from the previous step may be masked, symmetrized or a part of it symmetrized. These are options offered to the user and none of them is compulsory.

Image superresolution. Given the two independently reconstructed volumes, V_1 and V_2, the traditional approach in EM combines them by a simple average. We here propose to use a superresolution combination. The idea is to find a

higher resolution volume, V^*, and two parameters σ_1^* and σ_2^* such that

$$V^*, \sigma_1^*, \sigma_2^* = \underset{V,\sigma_1,\sigma_2}{\arg\min} \|V \star K_{\sigma_1} - V_1\|^2 + \|V \star K_{\sigma_2} - V_2\|^2$$

where $\star$ is the convolution operator and K_σ is a Gaussian kernel whose standard deviation is σ. Note that this volume does not participate in the calculation of the resolution between the two halves, and it is used only as the output at each iteration.

Particle selection. Our implementation allows executing iterations one by one. At the end of each iteration (after reconstruction), each particle in the input dataset is assigned its angular orientation and is qualified with the angular assignment objective function, the percentile this objective function occupies within its defocus group, the weight assigned by the `reconstruct_significant` algorithm, the in-plane shift and change of scale, etc. The user may look at the distribution of these parameters and decide whether particles at the extremes are of sufficient quality to participate in the next iteration. This option has not been used in this article in order to perform a fair comparison with existing image processing algorithms in the field. However, we find this possibility very useful in our daily data processing.

[1] Abrishami, V., Bilbao-Castro, J. R., Vargas, J., Marabini, R., Carazo, J. M., Sorzano, C. O. S., Oct 2015. A fast iterative convolution weighting approach for gridding-based direct fourier three-dimensional reconstruction with correction for the contrast transfer function. Ultramicroscopy 157, 79–87.

[2] Limare, N., Petro, A. B., Sbert, C., Morel, J. M., 2011. Retinex poisson equation: a model for color perception. Image Processing On Line 1, 39–50.

[3] Powell, M. J. D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer J. 7, 155–162.

[4] Sindelar, C. V., Grigorieff, N., 2012. Optimal noise reduction in 3d reconstructions of single particles using a volume-normalized filter. J. Structural Biology 180, 26–38.

[5] Sorzano, C. O. S., Vargas, J., de la Rosa-Trevín, J. M., Otón, J., Álvarez-Cabrera, A. L., Abrishami, V., Sesmero, E., Marabini, R., Carazo, J. M., 2015. A statistical approach to the initial volume problem in single particle analysis by electron microscopy. J. Structural Biology 189 (3), 213–219.