SCIENTIFIC REP⚙RTS

OPEN

Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited

C. O. S. Sorzano[1,2], M. Alcorlo[3], J. M. de la Rosa-Trevín[1], R. Melero[3], I. Foche[1], A. Zaldívar-Peraza[1], L. del Cano[1], J. Vargas[1], V. Abrishami[1], J. Otón[1], R. Marabini[4] & J. M. Carazo[1]

Cryo-Electron Microscopy (cryo-EM) of macromolecular complexes is a fundamental structural biology technique which is expanding at a very fast pace. Key to its success in elucidating the three-dimensional structure of a macromolecular complex, especially of small and non-symmetric ones, is the ability to start from a low resolution map, which is subsequently refined with the actual images collected at the microscope. There are several methods to produce this first structure. Among them, Random Conical Tilt (RCT) plays a prominent role due to its unbiased nature (it can create an initial model based on experimental measurements). In this article, we revise the fundamental mathematical expressions supporting RCT, providing new expressions handling all key geometrical parameters without the need of intermediate operations, leading to improved automation and overall reliability, essential for the success of cryo-EM when analyzing new complexes. We show that the here proposed RCT workflow based on the new formulation performs very well in practical cases, requiring very few image pairs (as low as 13 image pairs in one of our examples) to obtain relevant 3D maps.

The use of Random Conical Tilt (RCT)[1–5] as a way to produce the initial volume for a subsequent projection matching refinement has been in place for over 30 years. The technique has revealed as a very useful approach, especially when little *a priori* structural information is available. A proof of this statement is the more than 1,100 citations of the 4 references above. Normally this technique is applied to stained samples, since its application in cryo-EM setups is cumbersome. As it is currently applied, the projection images in untilted micrographs are subjected to a 2D classification. Those classes with a high internal homogeneity are selected. Finally, RCT is applied to each class individually obtaining, in this way, a set of RCT models (one model per 2D class of untilted images). These models can then be combined, at least in some cases, in order to reduce the effect of the missing cone in Fourier space, which, otherwise, results in an elongation of the resolved structure along the missing direction[6,7].

Mathematical developments related to RCT have concentrated on analysing the geometrical relationships between the Euler angles of the untilted and tilted micrographs[5]. However, a number of simplifications have traditionally been made, such as the omission of the explicit consideration of the shifts among projections introduced by inaccurate particle selection (the user cannot pick exactly at the particle origin in all projections) and the possible misalignment introduced by micrograph acquisition. Additionally, some particles may have been flipped within the holder and, in the current formulation, they have to be excluded from the 3D reconstruction because the current RCT formulation cannot manage them. However, we show that with the proper geometrical treatment, they can also be included in the reconstruction process and indeed provide useful structural information (note that the inclusion of mirrored

[1]National Center for Biotechnology (CSIC), c/Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain. [2]Universidad CEU San Pablo, Campus Urb. Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, Spain. [3]Center for Biological Research (CSIC), c/Ramiro de Maeztu, 9, 28040, Madrid, Spain. [4]Escuela Politécnica Superior, Universidad Autónoma de Madrid, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain. Correspondence and requests for materials should be addressed to C.O.S.S. (email: coss@cnb.csic.es)

images must be subjected to the consideration that partial staining may invalidate the assumption that mirrored and unmirrored images contain the same structural information, consequently, whether it is more beneficial to include them or not depend on each specific case).

In this paper we address the problems mentioned above. In this way, we propose a geometrical framework that explicitly accounts for possible shifts introduced during the picking process, angular misalignment during the micrograph scanning, as well as it allows the inclusion within the same 2D class of projections that are related by a mirror operation. Additionally, we propose to align tilted experimental projections to the class average of the corresponding untilted projections, again without intermediate steps. In this way, the 2D alignment of the tilted images becomes much more robust since the reference image to which the tilted images are aligned to has higher signal-to-noise ratio (SNR). All these changes together result in a very robust RCT algorithm capable of producing a 3D reconstruction of sufficient quality with very few image pairs (as few as 13 image pairs in one of our examples, as shown in the Results section).

Method

All mathematical details of the geometrical relationships between the untilted and tilted projections are described in Supplementary Material. Let us, at this point, summarize their main steps, going from the more general to the more particular aspects:

- Determination of the tilt axis: RCT requires the acquisition of two electron micrographs of the same field of view: one tilted and another one untilted. The location of the tilt axis needs to be determined in each of the two micrographs, and this is normally done by manual identification of a set of corresponding coordinate pairs in both micrographs. In Supplementary Material we provide a rigorous justification of the equation normally used in the field for this task (Eq. 37 of[5]), identifying its key assumptions. Namely: 1) that the location of the tilt axis relies heavily on the correctness of the hypothesis that the grid being visualized is flat; 2) that the untilted micrograph is perpendicular to the electron beam; and 3) that the centers of all selected particles are in the $z = 0$ plane of the grid. Note that this latter assumption could be avoided by an explicit refinement of the z location of each particle within the grid, which has never been attempted, probably because of the use of RCT for low resolution models only.

- Alignment between the untilted and tilted images: RCT requires the alignment between the tilted projection of a particle and its untilted projection (Eq. (12) in the Supplementary Material). Note that the parameters describing this alignment are related to the location of the tilt axis in both micrographs (tilted and untilted) as well as a shift between the two projections. The location of the tilt axis was determined in the previous step and, to the best of our knowledge, it is not the main source of errors in RCT reconstructions. Therefore, the shift between the tilted and untilted projections is the only new parameter that needs to be estimated at this step. This shift should be zero if, for each projection pair, the user (or a program) were able to pick the tilt and untilted projections so that they were mutually centered, indeed a rather difficult task considering the low SNR and the need for comparing two images that may look quite different (micrographs have been tilted by a substantial amount). The way commonly used in the field to handle this issue is by performing an intermediate step of aligning the untilted image to a centered reference. Of course, it is essential to relate the untilted to the tilted images, and this is normally done through a stretching operation, which currently requires the additional intermediate step of aligning the tilt axis with respect to the Y axis[5]. One of the novel contributions of this work is to provide a generalized mathematical formulation that avoids all these intermediate operations, explicitly taking into account the respective centers in the tilted and the untilted images, as well as generalizing the "stretching factor" to a "stretching matrix", able to account for an arbitrary orientation of the tilt axis with respect to the image coordinate axis both in the tilted and untilted micrographs (Equation 12 of Supplementary Material). It is important to note that the use of a stretching operation relies on the assumption that the particle being reconstructed is infinitely thin (as rigorously proved in our Supplemental Material). In the limit, for a globular protein, the stretching of the tilted projection to match the untilted one may do more harm than good. In our implementation in Xmipp[8,9] the user is asked whether she desires to perform this stretching or not.

- Calculation of 3D alignment parameters: If we are to produce a single 3D reconstruction out of multiple image pairs, we need to not only align the tilted projection to the untilted one, but also all the untilted projections among themselves, so that they are compatible with a single 3D structure. Implicitly, this requirement is usually handled by relying on the series of intermediate steps described in the previous section, considering that a common centered reference is used for all pairs. However, since we are aiming at avoiding all intermediate operations, we need to explicitly consider this common centering of all pairs. In Supplementary Material (Section 3.2), we show how to account for these inaccuracies in the selection of the untilted coordinates. The parameters needed for the 3D reconstruction normally involve the determination of 3 Euler angles and 2 in-plane shifts. We show how to calculate them as a function of an arbitrary tilt axis location, relating selection errors in the tilted and untilted projections. Additionally, we also consider the in-plane rotation needed to bring each untilted projection into the same orientation as the volume to be reconstructed. Finally, our derivation allows the determination of the 3D alignment parameters for an arbitrary orientation of the tilt axis, avoiding in

this way the interpolation needed by the traditional approach in which the tilt axis is assumed to be along the vertical image axis.

- Improving the estimation of in-plane alignment parameters: One of the problems of aligning tilted and untilted projections, as well as untilted projections among themselves, is the low SNR of the images. This problem can be alleviated if at least one of the images has a larger SNR. This is actually the case if instead of aligning with respect to experimental untilted projections, we align to the class average of the untilted projections. The class average has a larger SNR and makes the estimation of the alignment parameters more reliable. As a drawback, we need to account for a new geometrical transformation: the alignment between the experimental untilted projections and their corresponding class average. Section 3.3 in Supplementary Material shows how to incorporate this extra alignment into the 3D reconstruction parameters to be used with the tilted projections.
- Inclusion of mirror images: Depending on the molecule, target resolution and experimental SNR, some of the untilted projections may be classified by the 2D classification algorithm into the same class by applying a mirroring operation (that is, images assigned to a class may correspond to either mirrored or unmirrored projections). In our experience, mirror images may account for between 10% and 20% of the images in a class. The standard RCT approach does not consider the mirroring effect, so it precludes the possibility of utilizing this extra 10–20% of images. In Supplementary Material (Section 5), we show how mirrors can be explicitly modelled, so that mirrored and unmirrored images can be jointly used in a single 3D reconstruction.
- Revisited RCT procedure: All these changes result in a new RCT procedure that is summarized in Section 4 of Supplementary Material. It basically follows the same steps as in the standard RCT procedure, but it provides new formulas for the calculation of the 3D alignment parameters of the tilted projections (the main output of the RCT process) without intermediate operations, like setting the tilt axis vertically.

In summary, previously published RCT procedures had a number of geometrical simplifications that required the use of a number of intermediate steps. The consequence was the introduction of a number of degradation steps in the process, besides making the whole approach more error prone and difficult to automatize and trace. In fact, it is probably the higher chance to make errors in these intermediate steps the main drawback of traditional approaches to RCT. In our approach we have relaxed these simplifications, providing general formulas for an arbitrary orientation of the tilt axis and arbitrary large errors in picking positions. We have also included the possibility of considering mirrored images and the alignment of the tilted images using a class average (which increases the SNR of one of the images being aligned and, consequently, has a higher alignment accuracy). Note that this methodology is compatible with the posterior alignment of all reconstructed volumes (from different 2D classes and corresponding RCT reconstructions) into a single volume in which the missing cone is complemented by information from other volumes[7]. However, this alignment has to be performed with care because the different volumes may correspond to different conformations. In this case, an alignment and classification approach may be preferred[10]. The improvements brought in by our geometrical framework goes well beyond interpolation errors or slight resolution improvements. The possible accumulation of geometrical errors in the process, from which the traditional RCT theory was more susceptible to suffer, may result in 3D reconstructions which can be greatly improved, as shown by our results. Additionally, by omitting intermediate steps, we make the whole RCT workflow to be easily traceable, improving the reporting and reproducibiity of the results.

Results

To show a practical case of the use of these new geometrical relationships into a reconstruction procedure, we present here the results obtained with C3b as a test sample. This protein is one of the members of the complement system, a major component of innate immunity that comprises over thirty soluble and membrane-associated proteins having crucial roles in pathogen and apoptotic clearance, immune complexes handling, and modulation of adaptive immune responses. The central event in the complement activation cascade is the proteolytic cleavage of C3 to generate the activated fragment C3b. When C3b is generated, a reactive thioester group is exposed resulting in covalent attachment of C3b to pathogenic or apoptotic surfaces and thereby inducing several biological processes[11,12].

C3b is a small multidomain molecule (180 kDa), which poses a challenge to the accuracy of orientation determination of individual single particles, since this process depends critically on molecular mass[13]. However, we wished to present a representative case of the broad range of relatively small and probably labile complexes that can now be studied at high resolution by cryo-EM. C3b images were subjected to 2D reference-free classification to group those images coming from similar views of the protein, revealing several detailed structural features. C3b appeared on the EM support film in different orientations, but the most common view showed the macroglobulin (MG) ring clearly visible with a hole at the bottom end of the ring as well as globular densities corresponding to C345C (complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1) and TED (thioester containing domain) domains at the top and at the lower part of the ring, respectively (see Fig. 1).
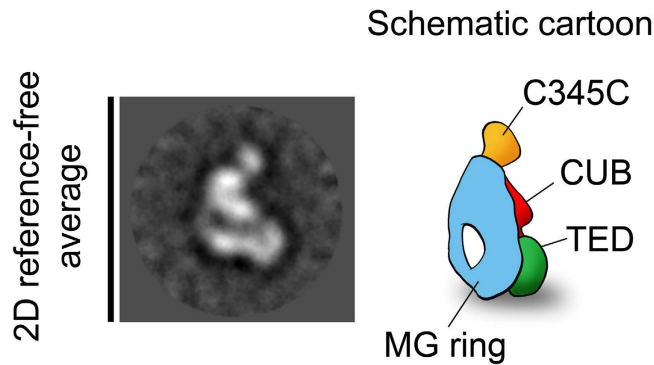
Figure 1. Selected 2D reference-free average showing a typical view of C3b. The cartoon representation highlights the location of the MG ring and C345C and TED domains.

A few μl of freshly purified complexes were adsorbed onto glow-discharged carbon-coated grids and negatively stained with 2% (mass/vol) uranyl formate (for a description of the purification process the reader is referred to[14]). Grids were visualized in a JEOL 1230 transmission electron microscope operated at 100 kV and a TemCam-F416 detector from Tietz Video and Image Processing Systems (TVIPS) using EM-TOOLS (TVIPS). Micrographs were recorded under low-dose conditions at a nominal magnification of 39,499x. Untilted (0°) and tilted (45°) micrographs were sequentially taken from a particular area in the grid following an RCT data collection scheme. Three thousands pairs of untilted and tilted projections were manually selected using Xmipp[8,9].

An independent data set was obtained to serve as control for the RCT reconstruction. The collection of non-tilted micrographs was complemented by additional micrographs recorded at 30° and 45° tilting in order to increase the number of views obtained for C3b. A total of 32,595 C3b images were manually selected using e2boxer.py[15] and subjected to 2D reference-free alignment and classification using Xmipp[16]. An initial volume was obtained using e2initialmodel.py from eman2. To investigate the presence of conformational heterogeneity, the data was refined using the 3D maximum-likelihood classification scheme implemented in RELION[17]. After convergence of the 3D classification refinement, the dataset was sorted into different homogeneous subsets (details of this analysis are given in[14] and[18]). The output volume of the most abundant conformation was further refined using the 3D-autorefine-procedure implemented in RELION and the projection matching refinement procedure implemented in Xmipp[19]. The resolution of the structure was estimated as 26Å using Fourier shell correlation of two independent reconstructions and a cutoff of 0.5. This 3D map will be referred to in the following as the control map. Note that we intentionally used a completely different image processing workflow to serve as control, since the use of RCT as starting volume for a 3D reconstruction would invalidate the comparison between the RCT reconstruction and the final result. The biological interpretation of the structure of C3b is reported in[14] and[18].

We then compared the reconstructions obtained following both the new RCT procedure (as implemented in Xmipp) and the traditional RCT workflow (as described in[20] and carefully detailed in http://spider.wadsworth.org/spider_doc/spider/docs/techs/ranconical/docs/rct.htm, which also performs a certain 2D classification), to the control map of C3b. In RCT, it is normally recommended to have small, very homogeneous 2D classes of untilted images, perform their 3D reconstruction using the RCT methodology, and then combine the different 3D structures to reduce the missing cone. We have followed the same strategy and have selected two examples of 2D classes with low and medium number of images (less than 20, and about 60–80). As can be appreciated in Fig. 2, the volumes reconstructed using the new RCT workflow are more similar to the control map than those obtained using the traditional RCT approach. The similarity is even more striking when we analyze the central slices of the aligned volumes (Supplementary Figs 1, 2, 3, 4 and 5) (it should be mentioned that this comparison can only address the gross morphology of the reconstructions, among other reasons because the sample is heterogeneous and the particular classes used in each analysis may not fully correspond to the same structure).

Discussion

The importance of RCT in the analysis of macromolecular complexes by Electron Microscopy is unquestionable. It provides a reference-free initial map that can be used for subsequent refinement using an independent set of images. To date, it has successfully performed its task, contributing to many important biomedical results. However, some simplifications underlying its mathematical derivation required a number of intermediate, otherwise unnecessary, operations, which degraded the information, increased the probability of errors and difficulted a complete tracing of the processing workflow. Considering the new capabilities of cryo-EM thanks to the direct detectors and the large range of new biomedical systems it is expected to be applied to, it is of paramount importance to ensure that RCT can be routinely applied to its fullest extent. In an effort to streamline the mathematical formulation of RCT, we provide in this
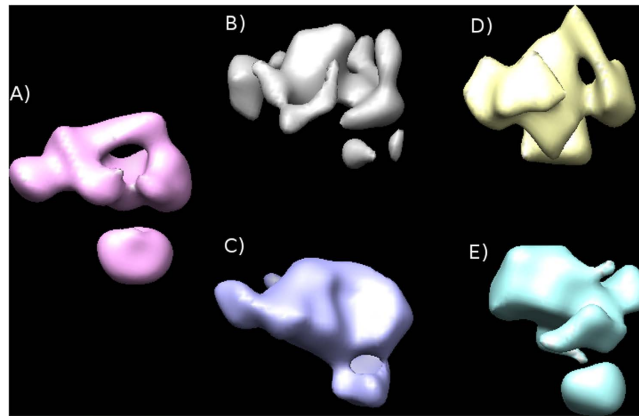
Figure 2. (**A**) Control map of C3b refined from 32,595 images. RCT applied to small 2D classes (less than 20 image pairs): (**B**) traditional RCT workflow as implemented in Spider (20 image pairs have been used); (**C**) new RCT workflow as implemented in XMIPP (13 image pairs used). RCT applied to medium size classes (about 60-80 image pairs): (**D**) traditional RCT workflow as implemented in Spider (79 image pairs used); (**E**) new RCT workflow as implemented in XMIPP (61 image pairs used).

work a generalized derivation of the relationships between the untilted and tilted projections along with the class average of the untilted projections. This generalization includes the user picking errors as part of the model and, consequently, it is capable of correcting them, resulting in a better 3D reconstruction. It also shows how to directly align the tilted images to the class average of the untilted projections. In this way, the direct estimation of the in-plane shifts becomes much more reliable and accurate since one of the images being compared (the class average) has a much larger SNR. Additionally, it allows the inclusion of mirrored projections, thus increasing the number of images that can be used during the reconstruction. Finally, our mathematical derivation no longer requires the tilt axis to be aligned with the vertical axis, avoiding in this way interpolation errors. The use of this generalized formulation results in a simpler method, with less chances of user errors and potential intermediate degradations. As a practical case, we show how the implementation of this new approach in Xmipp can indeed obtain very meaningful RCT reconstructions with a very low number of images (13 in our example), requiring less steps than traditional approaches. This possibility opens new perspectives in the treatment of RCT reconstructions as if they were subtomograms, making use of all new alignment and classification procedures that are currently being developed in the field.

References

1. Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T. S. Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy* **3,** 283–290 (1978).
2. Radermacher, M. & Hoppe, W. Properties of 3-D reconstruction from projections by conical tilting compared to single axis tilting. In *Proc. Seventh European Congress on Electron Microscopy*. **vol. I,** 132–133 (1980).
3. Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli. *J Microsc* **141,** RP1–RP2 (1986).
4. Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-Dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli. J. Microscopy* **146,** 113–136 (1987).
5. Radermacher, M. Three-Dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microscopy Technique* **9,** 359–394 (1988).
6. Sander, B., Golas, M. M., Lhrmann, R. & Stark, H. An approach for de novo structure determination of dynamic molecular assemblies by electron cryomicroscopy. *Structure* **18,** 667–676 (2010).
7. Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J. M. & Förster, F. Fast and accurate reference-free alignment of subtomograms. *J. Structural Biology* **182,** 235–245 (2013).
8. Sorzano, C. O. S. *et al.* XMIPP: A new generation of an open-source image processing package for electron microscopy. *J. Structural Biology* **148,** 194–204 (2004).
9. de la Rosa-Trevn, J. M. *et al.* Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Structural Biology* **184,** 321–328 (2013).
10. Chen, Y., Pfeffer, S., Fernández, J. J., Sorzano, C. O. S. & Förster, F. Autofocused 3D classification of cryoelectron subtomograms. *Structure* **22,** 1528–1537 (2014).
11. Gros, P., Milder, F. J. & Janssen, B. J. C. Complement driven by conformational changes. *Nat Rev Immunol* **8,** 48–58 (2008).
12. Rodríguez de Córdoba, S., Harris, C. L., Morgan, B. P. & Llorca, O. Lessons from functional and structural analyses of disease-associated genetic variants in the complement alternative pathway. *Biochim Biophys Acta* **1812,** 12–22 (2011).
13. 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).
14. Alcorlo, M. *et al.* Unique structure of iC3b resolved at a resolution of 24 Å by 3d-electron microscopy. *Proc Natl Acad Sci USA* **108,** 13236–13240 (2011).
15. Tang, G. *et al.* Eman2: an extensible image processing suite for electron microscopy. *J. Structural Biology* **157,** 38–46 (2007).
16. Sorzano, C. O. S. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Structural Biology* **171,** 197–206 (2010).

17. Scheres, S. H. W. Relion: implementation of a bayesian approach to cryo-em structure determination. *J Struct Biol* **180,** 519–530 (2012).
18. Alcorlo, M., Tortajada, A., Rodrguez de Córdoba, S. & Llorca, O. Structural basis for the stabilization of the complement alternative pathway c3 convertase by properdin. *Proc Natl Acad Sci USA* **110,** 13504–13509 (2013).
19. Scheres, S. H. W., Núñez Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using xmipp. *Nature Protocols* **3,** 977–990 (2008).
20. Huang, T. *et al.* The group II intron ribonucleoprotein precursor is a large, loosely packed structure. *Nucleic Acids Res* **39,** 2845–2854 (2011).

Acknowledgements

Author Contributions

C.O.S.S. designed and implemented the algorithm, run the experiments and wrote the manuscript. M.A., R.M., I.F. and T.S. helped with the experiments. J.M.R.T., A.Z.P., L.C., J.V. V.A., J.O., R.M. and J.M.C. helped with technical details of the implementation. J.M.C. and C.O.S.S. evaluated the methods comparison. All authors revised the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sorzano, C.O.S. *et al.* Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited. *Sci. Rep.* **5,** 14290; doi: 10.1038/srep14290 (2015).

Supplementary Material: Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited

C.O.S. Sorzano, M. Alcorlo, J.M. de la Rosa-Trevín, R. Melero,
I. Foche, A. Zaldívar-Peraza, L. del Cano, J. Vargas,
V. Abrishami, J. Otón, R. Marabini & J.M. Carazo

June 23, 2015

1 Determination of the tilt axis

Let us assume that each micrograph in the tilt pair is a projection of a large density volume containing all particles. Let us refer to this large volume as $V_{\mathrm{micrograph}}(\tilde{\mathbf{r}}) : \mathbb{R}^3 \times \{1\} \to \mathbb{R}$ (note that the volume argument has been expressed in homogeneous coordinates, i.e., $\tilde{\mathbf{r}} = (x, y, z, 1)^T$ and T represents the transpose operator). We also express images in homogeneous coordinates. For instance, $I_{\mathrm{micrograph},u}(\tilde{\mathbf{s}}) : \mathbb{R}^2 \times \{1\} \to \mathbb{R}$ represents the untilted projection image from the previous volume ($\tilde{\mathbf{s}} = (x, y, 1)^T$). The untilted micrograph can be modelled as

$$I_{\mathrm{micrograph},u}(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\mathrm{micrograph}}(\tilde{R}_Z^{-1}(\alpha_u)\tilde{H}^T\tilde{\mathbf{s}})dz, \qquad (1)$$

where

$$\tilde{H}^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z \\ 0 & 0 & 1 \end{pmatrix}, \qquad (2)$$

and $\tilde{R}_Z(\alpha_u)$ is the homogeneous rotation matrix about the Z axis by α_u degrees:
$\tilde{R}_Z(\alpha_u) = \begin{pmatrix} \cos\alpha_u & \sin\alpha_u & 0 & 0 \\ -\sin\alpha_u & \cos\alpha_u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. The rotation by α_u simply states that the tilt axis in the untilted image forms an angle α_u around the vertical axis. Similarly, we can model the tilted micrograph as

$$I_{\mathrm{micrograph},t}(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\mathrm{micrograph}}((\tilde{T}\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{\mathbf{s}})dz, \qquad (3)$$

where θ is the tilt angle, $\tilde{R}_Y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ is the rotation

matrix around Y, and $\tilde{T} = \begin{pmatrix} 1 & 0 & 0 & t_{\text{micrograph},X} \\ 0 & 1 & 0 & t_{\text{micrograph},Y} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ represents a possible

shift introduced when tilting the sample holder. In a way, the formulation above can be seen as an extension of the work in [Guckenberger(1982)].

Eqs. (1) and (3) can be used to relate coordinates in the 3D volume $V_{\text{micrograph}}$ to their projection locations. Particularly, consider the 3D homogeneous location $\tilde{\mathbf{r}}$, that projects into the untilted and tilted micrographs in locations

$$\begin{aligned} \tilde{\mathbf{s}}_u &= \tilde{H}_0 \tilde{R}_Z(\alpha_u)\tilde{\mathbf{r}} \\ \tilde{\mathbf{s}}_t &= \tilde{H}_0 \tilde{T}\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{\mathbf{r}} \end{aligned} \tag{4}$$

where $\tilde{H}_0$ is the $\tilde{H}$ matrix in Eq. (2) with $z = 0$. If we now restrict ourselves to 3D locations in the XY plane (i.e., $\tilde{\mathbf{r}} = (x, y, 0, 1)^T$), then we can recover $\tilde{\mathbf{r}}$ from its location in the untilted micrograph as $\tilde{\mathbf{r}} = \tilde{R}_Z^{-1}(\alpha_u)\tilde{H}_0^T\tilde{\mathbf{s}}_u$. Substituting in the expression for $\tilde{\mathbf{s}}_t$, we get

$$\tilde{\mathbf{s}}_t = \tilde{H}_0 \tilde{T}\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z^{-1}(\alpha_u)\tilde{H}_0^T\tilde{\mathbf{s}}_u \tag{5}$$

Similar equations without justification are presented by [Voss et al.(2009), Hauer et al.(2013)]. [Shatsky et al.(2014)] presents an evolution of this formulation in which particles are considered to be in a slightly bent grid.

If we know corresponding pairs of landmark locations in the untilted and tilted micrographs, this equation allows us to estimate the orientation of the tilt axis in both micrographs. From the knowledge of corresponding landmarks we can compute a matrix that meets in a least-squares sense that

$$\tilde{\mathbf{s}}_t = \tilde{A}\tilde{\mathbf{s}}_u = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \tilde{\mathbf{s}}_u. \tag{6}$$

a_{13} and a_{23} directly provide an estimate for $\tilde{T}$. Then, we need to find angles α_u, α_t and θ such that the 2×2 top-left submatrix of $\tilde{E} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z^{-1}(\alpha_u)$ approximates as well as possible the 2×2 top-left submatrix of A. In fact, Eq. (36) in [Radermacher(1988)] provides an excellent estimate of θ, and [Voss et al.(2009)] introduces a weighted version of this estimate. α_u and α_t can be determined by any optimization algorithm that minimizes the error between the 2×2 top-left submatrix of $\tilde{A}$ and the 2×2 top-left submatrix of $\tilde{E}$.

This derivation is totally equivalent to the one presented by Dr. Radermacher [Radermacher(1988)] and already used in the field [Hegerl et al.(1991)]. However, the formulation has been presented in a different way to highlight an

important limitation. Indeed, the determination of the tilt axis is imperfect due to the assumption that image landmarks correspond to 3D points in the XY plane (i.e., $\tilde{\mathbf{r}} = (x, y, 0, 1)^T$). This does not need to be necessarily the case, since each particle may have a different height within the sample holder. Although it is feasible to further refine the estimate of the tilt axis location to include the different height effect, this is not normally done in the field, and it will not be done in this work. For thin samples (as is normally the case in EM) and due to error cancellation effects, we do not expect this error to be of a large magnitude.

2   Image formation model without mirrors

Let $V_{\mathrm{ref}}(\tilde{\mathbf{r}}) : \mathbb{R}^3 \times \{1\} \to \mathbb{R}$ be a density volume (note that the volume argument has been expressed in homogeneous coordinates, i.e., $\mathbf{r} = (x, y, z, 1)^T$, T represents the transpose operator). This is the volume we would like to recover using RCT. The micrograph volume in the previous section, $V_{\mathrm{micrograph}}(\tilde{\mathbf{r}})$, is supposed to be composed of multiple copies of this $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$ in different orientations and spatial locations.

Picking the particles in the tilt pair amounts to identifying the locations of the different copies of $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$ in the untilted and tilted micrographs. We will refer to the extracted particles from the untilted micrograph as the untilted images. 2D classification of the untilted images amounts to sorting the images according to the different orientations of $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$.

Once a single 2D class is selected, we have chosen a particular orientation of $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$ (with no loss of generality we will assume that $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$ corresponds to the particular orientation selected by the 2D class) and its $0°$ projection should be similar to the 2D class average calculated by the 2D classification algorithm ($I_{\mathrm{ref},0°}(\tilde{\mathbf{s}}) \approx I_u^{\mathrm{2Dclass}}(\tilde{\mathbf{s}})$). Note that we do not have experimental access to $I_{\mathrm{ref},0°}$ since it is a projection calculated from a volume we do not actually have. However, we do have access to I_u^{2Dclass} since it is the result of averaging all experimental images assigned to the same 2D class by the classifier.

Each image pair corresponds to the projection of a certain copy of $V_{\mathrm{ref}}(\tilde{\mathbf{r}})$, that we will refer to as $V(\tilde{\mathbf{r}})$. The observed untilted image corresponds to a $0°$ projection of this volume

$$I_u(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V(\tilde{R}_Z^{-1}(\alpha_u) \tilde{H}^T \tilde{\mathbf{s}}) dz, \tag{7}$$

where we have explicitly accounted for the fact that the tilt axis in the untilted micrograph is not aligned with the vertical axis. The corresponding tilted image would be

$$I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V((\tilde{R}_Z(\alpha_t) \tilde{R}_Y(\theta))^{-1} \tilde{H}^T \tilde{\mathbf{s}}) dz. \tag{8}$$

However, the user seldom picks exactly at corresponding points between the untilted and tilted images and, therefore, a small shift can be expected. The

actual tilted image selected by the user would then correspond to

$$I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V((\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz, \tag{9}$$

being $\tilde{T}_t$ a 2D homogeneous matrix that reflects a shift introduced by the user picking.

3 Alignment of the tilted image

Traditionally, the alignment has been done by comparing the untilted image and the corresponding stretched tilted image once they have been both corrected to align the tilt axis with the vertical axis [Zampighi et al.(2004), Zheng et al.(2007)] (in fact, Dr. Radermacher[Radermacher(1988)] suggests to do the stretching in the direction perpendicular to the tilt axis in the tilted image, but it does not give any specific formula to do it). However, both images have very low Signal-to-Noise Ratio (SNR) making the alignment quite difficult. Alternatively, it is possible to substitute the untilted image by the class average that has a much better SNR. In the following sections, we show how to perform this alignment. We divide the derivation in three parts: first, we present the formal definition of streching in the direction perpendicular to the tilt axis and present the assumptions that justifies the use of that stretching; second, we relate the orientation of the individual copies of the reconstructed volume to the standard position defined by the 2D class average; finally, we extend the traditional in-plane alignment method so that we can use the 2D class average.

3.1 Stretching the tilted image and its assumptions

Let us define a new volume in which the angle of the tilt axis in the untilted image has been corrected:

$$V'(\tilde{\mathbf{r}}) = V(\tilde{R}_Z^{-1}(\alpha_u)\tilde{\mathbf{r}}) \leftrightarrow V(\tilde{\mathbf{r}}) = V'(\tilde{R}_Z(\alpha_u)\tilde{\mathbf{r}}). \tag{10}$$

The experimental projections in terms of the new volume become:

$$I_u(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V'(\tilde{H}^T\tilde{\mathbf{s}})dz$$

$$I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V'(\tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz = \int\limits_{-\infty}^{\infty} V'(\tilde{E}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz,$$
$$\tag{11}$$

where $\tilde{E} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z(-\alpha_u)$. If V is infinitely thin and concentrated in the XY plane (that is $V(\tilde{\mathbf{r}}) = V(\tilde{\mathbf{r}}_X, \tilde{\mathbf{r}}_Y, 0)\delta(\tilde{\mathbf{r}}_Z)$, being $\delta(x)$ the Dirac's delta distribution), then so is V' (since it is only a Z-rotated version of V). In this

4

way, we have (see Appendix for its proof)

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T \tilde{\mathbf{s}}) \\
I_t(\tilde{\mathbf{s}}) &= \int_{-\infty}^{\infty} V'(\tilde{E}^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz = V'(\tilde{H}_0^T (\tilde{H}_0 \tilde{E} \tilde{H}_0^T)^{-1} \tilde{T}_t^{-1} \tilde{\mathbf{s}}) \\
&= I_u((\tilde{H}_0 \tilde{E} \tilde{H}_0^T)^{-1} \tilde{T}_t^{-1} \tilde{\mathbf{s}})
\end{aligned}
\tag{12}
$$

In the particular case that $\alpha_u = \alpha_t = 0$ (i.e., the tilt is performed around the Y axis in both images), then the previous equation becomes

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T \tilde{\mathbf{s}}) \\
I_t(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T S \tilde{T}_t^{-1} \tilde{\mathbf{s}}) = I_u(S \tilde{T}_t^{-1} \tilde{\mathbf{s}})
\end{aligned}
\tag{13}
$$

where

$$
S = \begin{pmatrix} \frac{1}{\cos(\theta)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\tag{14}
$$

That is, the tilted image is a shifted and stretched (by a factor $\frac{1}{\cos(\theta)}$) version of the untilted one, as stated by Dr. Radermacher[Radermacher(1988)]. We, therefore, see that the traditional stretching for aligning RCT tilt pairs is a paticular case of the more general case given by Eq. (12), and more importantly, this derivation is only valid for infinitely thin objects. As soon as we depart from this assumption, the equality becomes only an approximation or may be even invalid for thick objects.

3.2   3D alignment parameters

The in-plane orientation of volume V' is not the same as the orientation of V_{ref} (in fact, this is the key of the success of Random Conical Tilt). Let us define $\tilde{M}_u$ as an affine matrix that encodes an in-plane rotation ($\tilde{R}_Z(\alpha'_u)$) followed by a translation ($\tilde{T}_u$), that is

$$
\tilde{M}_u = \tilde{T}_u \tilde{R}_Z(\alpha'_u),
\tag{15}
$$

such that the two volumes are registered to each other, i.e.,

$$
V_{\text{ref}}(\tilde{\mathbf{r}}) = V'(\tilde{M}_u^{-1} \tilde{\mathbf{r}}) \leftrightarrow V'(\tilde{\mathbf{r}}) = V_{\text{ref}}(\tilde{M}_u \tilde{\mathbf{r}}).
\tag{16}
$$

$\tilde{M}_u$ can be easily estimated from the transformation needed to align $I_u(\tilde{\mathbf{s}})$ to the 2D class average selected ($I_u^{\text{2Dclass}}(\tilde{\mathbf{s}})$) because

$$
I_u^{\text{2Dclass}}(\tilde{\mathbf{s}}) \approx I_u((\tilde{H}_0 \tilde{M}_u \tilde{H}_0^T)^{-1} \tilde{\mathbf{s}}).
\tag{17}
$$

Note that $\tilde{H}_0 \tilde{M}_u \tilde{H}_0^T$ is a compact way of selecting the top-left 2×2 matrix of $\tilde{M}_u$ and making it a 2D homogeneous matrix.

Combining all this information, we can relate the experimentally observed images to the volume we aim to reconstruct by

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{M}_u \tilde{H}^T \tilde{\mathbf{s}}) dz \\
I_t(\tilde{\mathbf{s}}) &= \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{M}_u \tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz \\
&= \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_u \tilde{R}_Z(\alpha'_u)\tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz \quad (18) \\
&= \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_u(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z(-(\alpha'_u + \alpha_u)))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz \\
&= \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_u \tilde{E}^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz,
\end{aligned}
$$

where we have defined the matrix $\tilde{E}$ as $\tilde{E} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z(-(\alpha'_u + \alpha_u))$. This is an Euler rotation matrix in the convention ZYZ whose first rotation is around Z by $-(\alpha'_u + \alpha_u)$ degrees (rotational angle), the second is arout Y by θ degrees (tilt angle), and finally arout Z by α_t degrees (in-plane rotation).

In Xmipp [Sorzano et al.(2004)], the orientation and translation of a projection with respect to a volume are defined through the following relationship

$$
I_t(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{E}^{-1} \tilde{H}^T \tilde{T}_t'^{-1} \tilde{\mathbf{s}}) dz. \quad (19)
$$

So we need to transform Eq. (18) into the form of Eq. (19). For doing so we find a new matrix $\tilde{T}_u'$ such that

$$
I_t(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_u \tilde{E}^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz = \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{E}^{-1} \tilde{T}_u' \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz \quad (20)
$$

The new $\tilde{T}_u'$ must, therefore, satisfy $\tilde{T}_u \tilde{E}^{-1} = \tilde{E}^{-1} \tilde{T}_u'$. It can be easily shown that

$$
\tilde{T}_u' = \tilde{E} \tilde{T}_u \tilde{E}^{-1} \quad (21)
$$

is the translation matrix sought. By a change of variable, it can be easily proved that

$$
I_t(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{E}^{-1} \tilde{T}_u' \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz = \int_{-\infty}^{\infty} V_{\text{ref}}(\tilde{E}^{-1} \tilde{H}^T \tilde{T}_{u,2D}' \tilde{T}_t^{-1} \tilde{\mathbf{s}}) dz,
$$
$$(22)$$

where $\tilde{T}_{u,2D}' = \tilde{H}_0 \tilde{T}_u' \tilde{H}_0^T$ (this is a compact way of turning a 3D homogeneous translation matrix into a 2D homogeneous matrix by simply dropping the translation along Z).

Finally, we find that the translation needed by Eq. (19) is

$$
\tilde{T}_t' = \tilde{T}_t \left(\tilde{T}_{u,2D}' \right)^{-1} \quad (23)
$$

6

3.3   Aligning the tilted images to the 2D class average

The derivation in the previous section allows us to align the untilted and tilted images in a tilt pair. However, these two images have very low SNR since both of them are experimental. We may substitute the untilted image by the class average produced by the classification algorithm. This class average has a larger SNR, and it is supposed to be identical to the untilted image except for some misalignment provided by Eq. (17). Let us now derive the relationship between $I_t(\tilde{\mathbf{s}})$ and $I_u^{2\text{Dclass}}(\tilde{\mathbf{s}})$ in the case of an infinitely thin sample (so that we can reuse the stretching theory above).

We reproduce here the main results of Eqs. (17) and (12)

$$
\begin{aligned}
I_t(\tilde{\mathbf{s}}) &= I_u((\tilde{E}^{2D})^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \\
I_u^{2\text{Dclass}}(\tilde{\mathbf{s}}) &\approx I_u((\tilde{M}_u^{2D})^{-1}\tilde{\mathbf{s}})
\end{aligned}
\tag{24}
$$

In which we have defined $\tilde{E}^{2D} = \tilde{H}_0 \tilde{E} \tilde{H}_0^T$ and $\tilde{M}_u^{2D} = \tilde{H}_0 \tilde{M}_u \tilde{H}_0^T$. From the first equation we can conclude that

$$
I_u(\tilde{\mathbf{s}}) = I_t(\tilde{T}_t \tilde{E}^{2D} \tilde{\mathbf{s}}).
\tag{25}
$$

Therefore,

$$
I_u^{2\text{Dclass}}(\tilde{\mathbf{s}}) \approx I_t(\tilde{T}_t \tilde{E}^{2D} (\tilde{M}_u^{2D})^{-1}\tilde{\mathbf{s}}).
\tag{26}
$$

From a computational point of view and for reasons that will become apparent now, it is more convenient to define two new auxiliary matrices:

$$
\begin{aligned}
\tilde{A}^{2\text{Dclass}} &= \tilde{M}_u^{2D}(\tilde{E}^{2D})^{-1} \\
\tilde{T}_t^{2\text{Dclass}} &= \tilde{A}^{2\text{Dclass}}\tilde{T}_t^{-1}(\tilde{A}^{2\text{Dclass}})^{-1}
\end{aligned}
\tag{27}
$$

such that

$$
I_u^{2\text{Dclass}}(\tilde{\mathbf{s}}) \approx I_t((\tilde{A}^{2\text{Dclass}})^{-1}(\tilde{T}_t^{2\text{Dclass}})^{-1}\tilde{\mathbf{s}}).
\tag{28}
$$

At this point, let us define the image

$$
I_t^{2\text{Dclass}}(\tilde{\mathbf{s}}) = I_t((\tilde{A}^{2\text{Dclass}})^{-1}\tilde{\mathbf{s}}),
\tag{29}
$$

that is nothing more than the tilted image transformed so that it becomes in the same orientation as the 2D class average of the untilted images. Now we have

$$
I_u^{2\text{Dclass}}(\tilde{\mathbf{s}}) \approx I_t^{2\text{Dclass}}((\tilde{T}_t^{2\text{Dclass}})^{-1}\tilde{\mathbf{s}}).
\tag{30}
$$

The translation encoded by $\tilde{T}_t^{2\text{Dclass}}$ can be found at this point by simple cross correlation in Fourier space, and then we recover the sought translation by

$$
\tilde{T}_t = (\tilde{A}^{2\text{Dclass}})^{-1}(\tilde{T}_t^{2\text{Dclass}})^{-1}\tilde{A}^{2\text{Dclass}}.
\tag{31}
$$

4 Revisited RCT algorithm

According to the theory explained in this document, a RCT reconstruction can be performed as follows:

For each pair of micrographs (untilted and tilted):

- Step 1: Mark a set of corresponding particle centers (set of $\tilde{s}_u$ and $\tilde{s}_t$ coordinates).

- Step 2: Estimate by Least Squares or any other method a matrix that transforms untilted coordinates into tilted coordinates (see Eq. (6)).

- Step 3: Use Eq. (36) in [Radermacher(1988)] to estimate the tilt angle, θ.

- Step 4: Use any optimization algorithm to find the angles α_u and α_t that minimize the error between the 2×2 top-left submatrix of $\tilde{A}$ and the 2×2 top-left submatrix of $\tilde{E}$ (see Eq. (6) and text around).

- Step 5: Extract the particle image pairs $I_u(\tilde{\mathbf{s}})$ and $I_t(\tilde{\mathbf{s}})$.

For all untilted particle projections ($I_u(\tilde{\mathbf{s}})$):

- Step 6: Perform a 2D classification into different classes.

- Step 7: Choose a certain class for which the 3D model is desired, $I_u^{\text{2Dclass}}(\tilde{\mathbf{s}})$

For all particle pairs ($I_u(\tilde{\mathbf{s}})$, $I_t(\tilde{\mathbf{s}})$):

- Step 8: Compute $\tilde{M}_u$ such that the 2D class and the untilted image become aligned (see Eq. (17))

- Step 9: If the particle to reconstruct is thin enough so that stretching the tilted image is not a strong violation, stretch the tilted image to look like the untilted 2D class average according to Eq. (29).

- Step 10: Find the best shift, $\tilde{T}_t^{\text{2Dclass}}$, so that the image in Step 9 is registered with the 2D class (see Eq. (30)).

- Step 11: Estimate the particle picking misalignment according to Eq. (31).

- Step 12: Compute the shift needed by the reconstruction algorithm, in the case of Xmipp use Eq. (23).

- Step 13: The ZYZ Euler angles of the tilted image are $(-(\alpha_u' + \alpha_u), \theta, \alpha_t)$ (first, second, and third rotation, respectively).

Once all particle geometry parameters have been determined:

- Step 14: Use any 3D reconstruction algorithm using the tilted images (without stretching) and the shifts and Euler angles calculated in Steps 12 and 13.

- Step 15: Angles and shifts can be further refined using any projection matching algorithm.

5 Geometrical framework with mirrors

Let us now consider the case in which there are untilted images in the 2D class related by mirroring over the Y axis (X is flipped). This is the case of some classification algorithms as CL2D [Sorzano et al.(2010)] and ML2D [Scheres et al.(2005)]. The traditional approach to handling mirrors in RCT has been to discard those images that are mirrors or to classify them into a separate class. But doing so discards useful information that can be exploited to construct a more reliable RCT model.

Eqs. (7) and (9) are still valid, only that they now refer to a volume V_{mirror}, that is

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V_{\text{mirror}}(\tilde{R}_Z^{-1}(\alpha_u)\tilde{H}^T\tilde{\mathbf{s}})dz \\
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V_{\text{mirror}}((\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz,
\end{aligned}
\tag{32}
$$

with

$$
V_{\text{mirror}}(\tilde{\mathbf{r}}) = V(\tilde{R}_Y^{-1}(\pi)\tilde{\mathbf{r}}).
\tag{33}
$$

Note that now Eqs. (15) and (16) become

$$
\begin{aligned}
\tilde{M}_u &= \tilde{R}_Y^{-1}(\pi)\tilde{T}_u\tilde{R}_Z(\alpha_u') \\
V_{\text{ref}}(\tilde{\mathbf{r}}) &= V'_{\text{mirror}}(\tilde{M}_u^{-1}\tilde{\mathbf{r}}) \leftrightarrow V'_{\text{mirror}}(\tilde{\mathbf{r}}) = V_{\text{ref}}(\tilde{M}_u\tilde{\mathbf{r}}),
\end{aligned}
\tag{34}
$$

and Eq. (17) is still valid.

The relationship between the tilted image and V_{ref} (Eq. (18)) now is

$$
\begin{aligned}
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{M}_u\tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz \\
&= \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{R}_Y^{-1}(\pi)\tilde{T}_u\tilde{R}_Z(\alpha_u')\tilde{R}_Z(\alpha_u)\tilde{R}_Y^{-1}(\theta)\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz \\
&= \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{R}_Y^{-1}(\pi)\tilde{T}_u\tilde{R}_Z(\alpha_u' + \alpha_u)\tilde{R}_Y(\theta)^{-1}\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz
\end{aligned}
\tag{35}
$$

Now, we define $\tilde{T}_{u,\text{mirror}} = \tilde{R}_Y(\theta + \pi)^{-1}\tilde{T}_u\tilde{R}_Y(\theta + \pi)$ (this simply amounts to change the sign of the shift in X). In this way, we can rewrite the previous expression as

$$
I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_{u,\text{mirror}}\tilde{R}_Y^{-1}(\pi)\tilde{R}_Z(\alpha_u' + \alpha_u)\tilde{R}_Y(\theta)^{-1}\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz
\tag{36}
$$

At this point we make use of the fact $\tilde{R}_Y^{-1}(\pi)\tilde{R}_Z(\alpha_u' + \alpha_u) = \tilde{R}_Z(-(\alpha_u' + \alpha_u))\tilde{R}_Y^{-1}(\pi)$, then

$$
\begin{aligned}
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_{u,\text{mirror}}\tilde{R}_Z(-(\alpha_u' + \alpha_u))\tilde{R}_Y(\theta + \pi)^{-1}\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz \\
&= \int\limits_{-\infty}^{\infty} V_{\text{ref}}(\tilde{T}_{u,\text{mirror}}\tilde{E}_{\text{mirror}}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz
\end{aligned}
\tag{37}
$$

9

where $\tilde{E}_{\text{mirror}} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta + \pi)\tilde{R}_Z(\alpha'_u + \alpha_u)$.

Regarding the tilt image alignment, we can rewrite Eq. (11) in terms of the mirrored volume

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'_{\text{mirror}}(\tilde{H}^T\tilde{\mathbf{s}})dz \\
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'_{\text{mirror}}(\tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz,
\end{aligned}
\tag{38}
$$

with $V'_{\text{mirror}}(\tilde{\mathbf{r}}) = V'(\tilde{R}_Y^{-1}(\pi)\tilde{\mathbf{r}})$. Substituting V'_{mirror} in the previous expression, we get

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'(\tilde{R}_Y^{-1}(\pi)\tilde{H}^T\tilde{\mathbf{s}})dz \\
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'(\tilde{R}_Y^{-1}(\pi)\tilde{R}_Z(\alpha_u)\tilde{R}_Y^{-1}(\theta)\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz \\
&= \int\limits_{-\infty}^{\infty} V'(\tilde{R}_Z(-\alpha_u)\tilde{R}_Y^{-1}(\theta + \pi)\tilde{R}_Z^{-1}(\alpha_t)\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz, \\
&= \int\limits_{-\infty}^{\infty} V'(\tilde{E}_{\text{mirror}}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz,
\end{aligned}
\tag{39}
$$

where $\tilde{E}_{\text{mirror}} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta + \pi)\tilde{R}_Z(\alpha_u)$. Again, as we did for Eq. (12), we can show (see Appendix) that

$$
I_t(\tilde{\mathbf{s}}) = I_u(\tilde{M}(\tilde{H}_0\tilde{E}_{\text{mirror}}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}),
\tag{40}
$$

where $\tilde{M} = \tilde{H}_0\tilde{R}_Y^{-1}(\pi)\tilde{H}_0^T$. Combining this equation with Eq. (17) we have

$$
I_u^{\text{2Dclass}}(\tilde{\mathbf{s}}) \approx I_t(\tilde{T}_t\tilde{E}_{\text{mirror}}^{2D}\tilde{M}(\tilde{M}_u^{2D})^{-1}\tilde{\mathbf{s}}),
\tag{41}
$$

that is the mirror version equivalent of Eq. (26). From this point on, we can follow the same derivation as in Section 3.3 except that we need to define

$$
\tilde{A}^{\text{2Dclass}} = \tilde{M}_u^{2D}\tilde{M}(\tilde{E}^{2D})^{-1}
\tag{42}
$$

Therefore, the only modifications to the previous RCT algorithm are:

- Step 8: Compute $\tilde{M}_u$ such that the 2D class and the untilted image become aligned (see Eq. (34))

- Step 9: If the particle to reconstruct is thin enough so that stretching the tilted image is not a strong violation, stretch the tilted image to look like the untilted 2D class average according to Eq. (29) (but using $\tilde{A}^{\text{2Dclass}}$ as defined in Eq. (42)).

- Step 11: Estimate the particle picking misalignment according to Eq. (31) (but using $\tilde{A}^{\text{2Dclass}}$ as defined in Eq. (42))

- Step 13: The ZYZ Euler angles of the tilted image are $(\alpha'_u + \alpha_u, \theta + \pi, \alpha_t)$ (first, second, and third rotation, respectively).

Appendix

Proof of Eq. (12)

Let us prove Eq. (12), which we reproduce here:

$$I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V'(\tilde{E}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz = V'(\tilde{H}_0^T(\tilde{H}_0\tilde{E}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \qquad (43)$$

with $\tilde{E} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z(-\alpha_u)$. Note that $\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}} = (\mathbf{s}_X - t_{t,X}, \mathbf{s}_Y - t_{t,Y}, z, 1)^T$. For notation simplicity, we define two new variables such that $\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}} = (x, y, z, 1)^T$. With this notation the argument of V' within the integral simplifies to

$$\tilde{E}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}} = \begin{pmatrix} x(\sin(\alpha_u)\sin(\alpha_t) + \cos(\alpha_u)\cos(\alpha_t)\cos(\theta)) + y(\sin(\alpha_u)\cos(\alpha_t) - \cos(\alpha_u)\sin(\alpha_t)\cos(\theta)) + z\cos(\alpha_u)\sin(\theta) \\ x(\cos(\alpha_u)\sin(\alpha_t) - \sin(\alpha_u)\cos(\alpha_t)\cos(\theta)) + y(\cos(\alpha_u)\cos(\alpha_t) + \sin(\alpha_u)\sin(\alpha_t)\cos(\theta)) - z\sin(\alpha_u)\sin(\theta) \\ -x\cos(\alpha_t)\sin(\theta) + y\sin(\alpha_t)\sin(\theta) + z\cos(\theta) \\ 1 \end{pmatrix}.$$

$$(44)$$

Since V' is a delta distribution concentrated in the XY-plane, its integral along Z coincides with the value of V' when the third coordinate is 0, that is

$$-x\cos(\alpha_t)\sin(\theta) + y\sin(\alpha_t)\sin(\theta) + z\cos(\theta) = 0,$$

or, what is the same, when

$$z = x\cos(\alpha_t)\tan(\theta) - y\sin(\alpha_t)\tan(\theta).$$

Substituting this value into Eqs. (44) and (43), we have

$$I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V'(\tilde{E}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz = V'(\tilde{H}_0^T\tilde{E}'\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \qquad (45)$$

where

$$\tilde{E}' = \begin{pmatrix} \sin(\alpha_u)\sin(\alpha_t) + \frac{\cos(\alpha_u)\cos(\alpha_t)}{\cos(\theta)} & \sin(\alpha_u)\cos(\alpha_t) - \frac{\cos(\alpha_u)\sin(\alpha_t)}{\cos(\theta)} & 0 \\ \cos(\alpha_u)\sin(\alpha_t) - \frac{\sin(\alpha_u)\cos(\alpha_t)}{\cos(\theta)} & \cos(\alpha_u)\cos(\alpha_t) + \frac{\sin(\alpha_u)\sin(\alpha_t)}{\cos(\theta)} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$(46)$$

It turns out that $\tilde{E}'$ is the homogeneous inverse matrix of the 2×2 top-left submatrix of $\tilde{E}$, i.e.

$$\tilde{E}' = (\tilde{H}_0\tilde{E}\tilde{H}_0^T)^{-1}. \qquad (47)$$

Substituting $\tilde{E}'$ by its value in Eq. (45), we obtain Eq. (43), which finishes the proof.

Proof of Eq. (40)

Let us start with Eq. (39), which we reproduce here for convenience:

$$
\begin{aligned}
I_u(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'(\tilde{R}_Y^{-1}(\pi)\tilde{H}^T\tilde{\mathbf{s}})dz \\
I_t(\tilde{\mathbf{s}}) &= \int\limits_{-\infty}^{\infty} V'(\tilde{E}_{\text{mirror}}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz,
\end{aligned}
\tag{48}
$$

with $\tilde{E}_{\text{mirror}} = \tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta+\pi)\tilde{R}_Z(\alpha_u)$. Since V' corresponds to an infinitely thin volume, we know

$$
I_u(\tilde{\mathbf{s}}) = V'(\tilde{R}_Y^{-1}(\pi)\tilde{H}_0^T\tilde{\mathbf{s}}) = V'(\tilde{H}_0^T\tilde{M}\tilde{\mathbf{s}}).
\tag{49}
$$

On the other hand, making the same change of variable as in the previous proof $(\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}} = (x,y,z,1)^T)$

$$
\tilde{E}_{\text{mirror}}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}} = \begin{pmatrix} -x(\sin(\alpha_u)\sin(\alpha_t) + \cos(\alpha_u)\cos(\alpha_t)\cos(\theta)) - y(\sin(\alpha_u)\cos(\alpha_t) - \cos(\alpha_u)\sin(\alpha_t)\cos(\theta)) - z\cos(\alpha_u)\sin(\theta) \\ x(\cos(\alpha_u)\sin(\alpha_t) - \sin(\alpha_u)\cos(\alpha_t)\cos(\theta)) + y(\cos(\alpha_t)\cos(\alpha_u) + \sin(\alpha_u)\sin(\alpha_t)\cos(\theta)) - z\sin(\alpha_u)\sin(\theta) \\ x\cos(\alpha_t)\sin(\theta) - z\cos(\theta) - y\sin(\alpha_t)\sin(\theta) \\ 1 \end{pmatrix}.
\tag{50}
$$

Since V' is a delta distribution concentrated in the XY-plane, its integral along Z coincides with the value of V' when the third coordinate is 0, that is

$$
x\cos(\alpha_t)\sin(\theta) - z\cos(\theta) - y\sin(\alpha_t)\sin(\theta) = 0,
$$

or, what is the same, when

$$
z = x\cos(\alpha_t)\tan(\theta) - y\sin(\alpha_t)\tan(\theta),
$$

which is exactly the same condition as in the previous proof. Substituting this value into Eqs. (50) and (48), it can be checked that

$$
I_t(\tilde{\mathbf{s}}) = \int\limits_{-\infty}^{\infty} V'(\tilde{E}_{\text{mirror}}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz = V'(\tilde{H}_0^T(\tilde{H}_0\tilde{E}_{\text{mirror}}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}).
\tag{51}
$$

According to Eq. (49) we have

$$
\begin{aligned}
I_u(\tilde{M}(\tilde{H}_0\tilde{E}_{\text{mirror}}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T\tilde{M}\tilde{M}(\tilde{H}_0\tilde{E}_{\text{mirror}}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \\
&= V'(\tilde{H}_0^T(\tilde{H}_0\tilde{E}_{\text{mirror}}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \\
&= I_t(\tilde{\mathbf{s}}),
\end{aligned}
\tag{52}
$$

which concludes the proof.

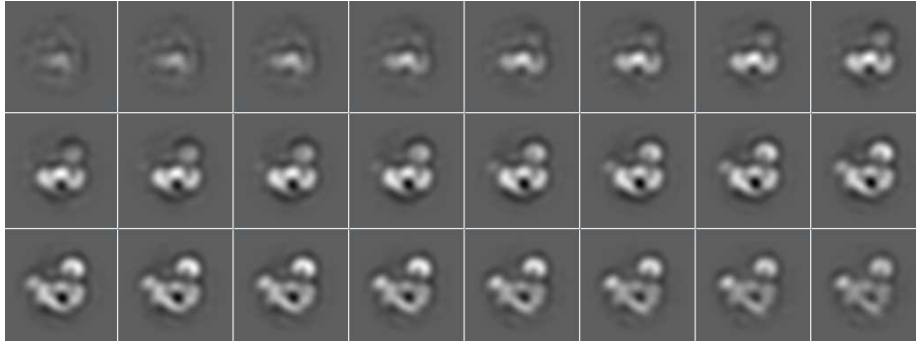# 6   Supplementary figures

Figure 1: Control map: Central slices of the control map of C3b computed from 32,595 images and a 3D classification procedure (see main text for details).
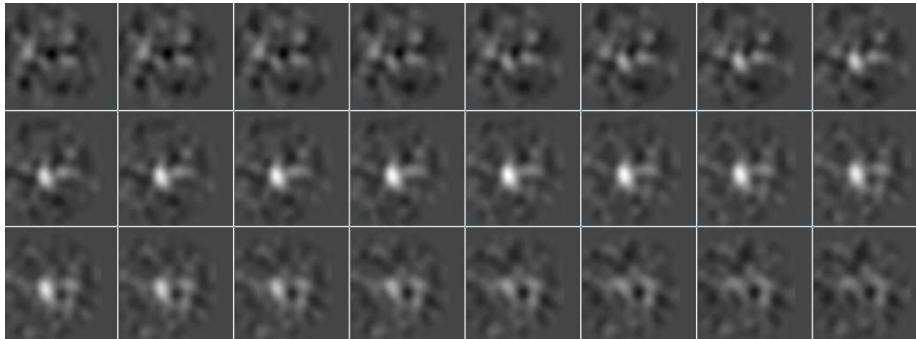

Figure 2: Standard RCT workflow for small-sized 2D classes (20 image pairs): Central slices of the structure of C3b using standard RCT, as implemented in Spider. Compare them to Suppl. Fig. 1.
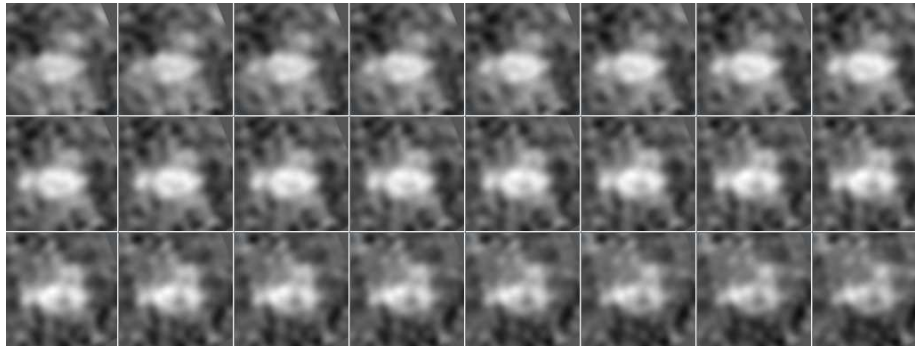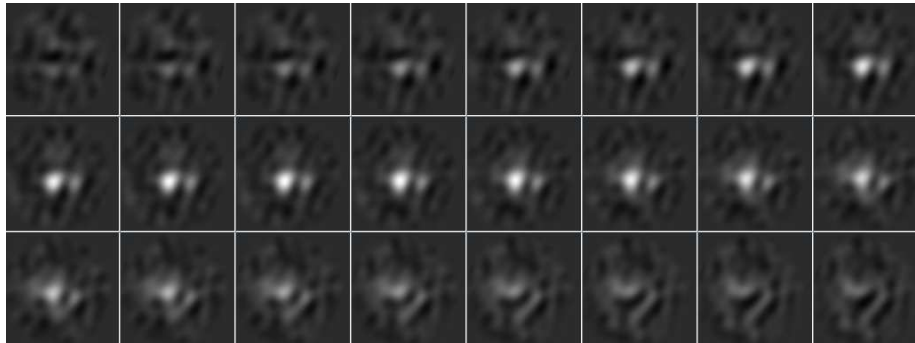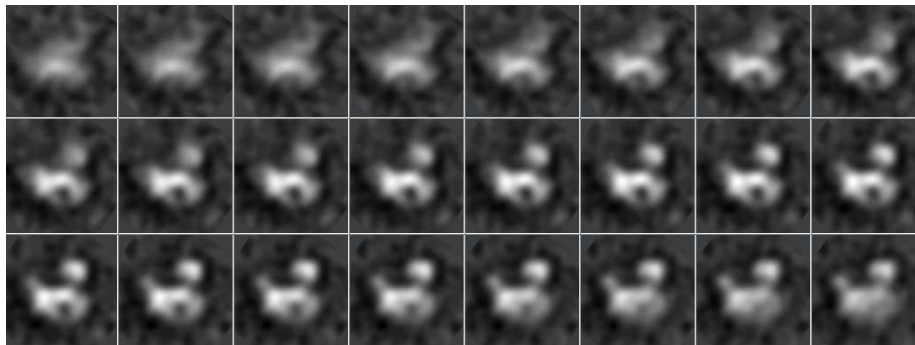
Figure 3: Modified RCT workflow for small-sized 2D classes (13 image pairs): Central slices of the structure of C3b using modified RCT, as implemented in Xmipp. Compare them to Suppl. Fig. 1.

Figure 4: Standard RCT workflow for mid-sized 2D classes (79 image pairs): Central slices of the structure of C3b using standard RCT, as implemented in Spider. Compare them to Suppl. Fig. 1.

Figure 5: Modified RCT workflow for mid-sized 2D classes (61 image pairs): Central slices of the structure of C3b using modified RCT, as implemented in Xmipp. Compare them to Suppl. Fig. 1.
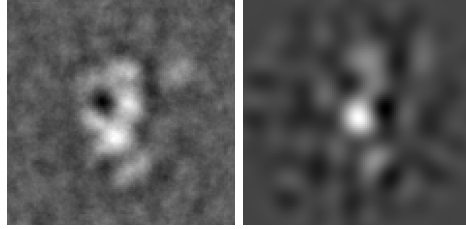
Figure 6: Left: Class average of 20 untilted projections calculated with Spider. Right: Central slice of the corresponding RCT reconstruction calculated with standard RCT, as implemented in Spider.
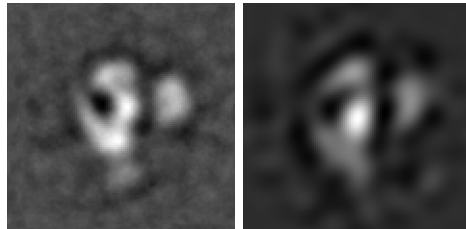

Figure 7: Left: Class average of 79 untilted projections calculated with Spider. Right: Central slice of the corresponding RCT reconstruction calculated with standard RCT, as implemented in Spider.
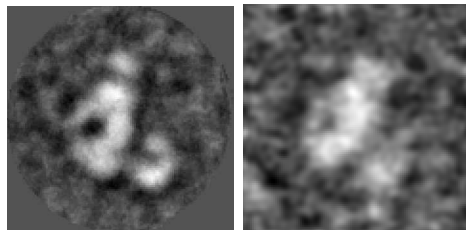

Figure 8: Left: Class average of 13 untilted projections calculated with Xmipp. Right: Central slice of the corresponding RCT reconstruction calculated with modified RCT, as implemented in Xmipp.
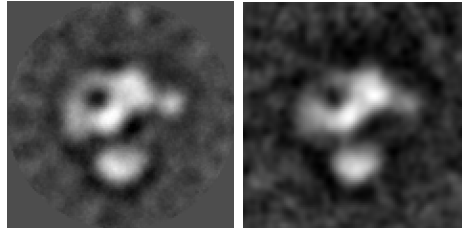
Figure 9: Left: Class average of 61 untilted projections calculated with Xmipp. Right: Central slice of the corresponding RCT reconstruction calculated with modified RCT, as implemented in Xmipp.

Bibliography

References

[Guckenberger(1982)] Guckenberger, R., 1982. Determination of a common origin in the micrographs of titl series in three-dimensional electron microscopy. Ultramicroscopy 9, 167–174.

[Hauer et al.(2013)] Hauer, F., Gerle, C., Kirves, J.-M., Stark, H., Feb 2013. Automated correlation of single particle tilt pairs for random conical tilt and orthogonal tilt reconstructions. J Struct Biol 181 (2), 149–154.
URL http://dx.doi.org/10.1016/j.jsb.2012.10.014

[Hegerl et al.(1991)] Hegerl, R., Pfeifer, G., Pühler, G., Dahlmann, B., Baumeister, W., May 1991. The three-dimensional structure of proteasomes from thermoplasma acidophilum as determined by electron microscopy using random conical tilting. FEBS Lett 283 (1), 117–121.

[Radermacher(1988)] Radermacher, M., 1988. Three-Dimensional reconstruction of single particles from random and nonrandom tilt series. J. Electron Microscopy Technique 9, 359–394.

[Scheres et al.(2005)] 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. Molecular Biology 348, 139–149.

[Shatsky et al.(2014)] Shatsky, M., Arbelaez, P., Han, B. G., Typke, D., Brenner, S. E., Malik, J., Glaeser, R. M., Jul 2014. Automated particle correspondence and accurate tilt-axis detection in tilted-image pairs. J. Structural Biology 187 (1), 66–75.
URL http://dx.doi.org/10.1016/j.jsb.2014.03.017

[Sorzano et al.(2010)] 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. Structural Biology 171, 197–206.

[Sorzano et al.(2004)] Sorzano, C. O. S., Marabini, R., , Velázquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H. W., Carazo, J. M., Pascual-Montano, A., 2004. XMIPP: A new generation of an open-source image processing package for electron microscopy. J. Structural Biology 148, 194–204.

[Voss et al.(2009)] Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S., Carragher, B., May 2009. Dog picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. J Struct Biol 166 (2), 205–213.

[Zampighi et al.(2004)] Zampighi, L. M., Kavanau, C. L., Zampighi, G. A., 2004. The kohonen self-organizing map: a tool for the clustering and alignment of single particles imaged using random conical tilt. J. Structural Biology 146, 368–380.

[Zheng et al.(2007)] Zheng, S. Q., Kollman, J. M., Braunfeld, M. B., Sedat, J. W., Agard, D. A., Jan 2007. Automated acquisition of electron microscopic random conical tilt sets. J. Structural Biology 157 (1), 148–155.