Journal of Structural Biology 181 (2013) 149–154

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

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

Automated correlation of single particle tilt pairs for Random Conical Tilt and Orthogonal Tilt Reconstructions

Florian Hauer [a], Christoph Gerle [b], Jan-Martin Kirves [a], Holger Stark [a,*]

[a] Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[b] Career Path Promotion Unit for Young Life Scientists, Kyoto University, Bldg. E, Yoshida Konoe Cho, Sakkyo-Ku, 606-8501 Kyoto, Japan

ARTICLE INFO

Article history:
Received 21 August 2012
Received in revised form 19 October 2012
Accepted 24 October 2012
Available online 7 November 2012

Keywords:
TEM
Single-particle cryo-electron microscopy
Random Conical Tilt
Orthogonal Tilt
Automated particle selection

ABSTRACT

One of the major methodological challenges in single particle electron microscopy is obtaining initial reconstructions which represent the structural heterogeneity of the dataset. Random Conical Tilt and Orthogonal Tilt Reconstruction techniques in combination with 3D alignment and classification can be used to obtain initial low-resolution reconstructions which represent the full range of structural heterogeneity of the dataset. In order to achieve statistical significance, however, a large number of 3D reconstructions, and, in turn, a large number of tilted image pairs are required. The extraction of single particle tilted image pairs from micrographs can be tedious and time-consuming, as it requires intensive user input even for semi-automated approaches. To overcome the bottleneck of manual selection of a large number of tilt pairs, we developed an algorithm for the correlation of single particle images from tilted image pairs in a fully automated and user-independent manner. The algorithm reliably correlates correct pairs even from noisy micrographs. We further demonstrate the applicability of the algorithm by using it to obtain initial references both from negative stain and unstained cryo datasets.

1. Introduction

Recent consensus in the field of Single-Particle Electron Microscopy highlights the necessity to consider the structural heterogeneity of the examined biomacromolecules in the analysis. Structural heterogeneity, which can be constituted by conformational rearrangements or compositional impurities, impairs high resolution Cryo-EM reconstructions. Additionally, structural heterogeneity is in most cases innately linked to the function of a biomacromolecular complex. Approaches which are based on alignment and classification of Random Conical Tilt (RCT) (Radermacher et al., 1987) or Orthogonal Tilt Reconstruction (OTR) (Leschziner and Nogales, 2006) 3Ds have been recently proposed. With these approaches, it is possible to create a set of initial models for cryo-EM which are bias-free and represent the structural heterogeneity of the dataset (Sander et al., 2010). In comparison to techniques which use subsamples of the dataset (Penczek et al., 2006), these approaches can be used for all datasets, even if there is no previous structural information available and the structural heterogeneity comprises large parts of the complex. A routine application of these approaches, however, is still impractical because of the necessity of a large number of initial RCT or OTR 3D reconstructions for which, in turn, requires a large number of single particle images which are recorded at two different angles (*tilt pairs*). Semi-automated approaches for single particle tilt pair extraction exist (Sorzano et al., 2004; Voss et al., 2009), but require detailed user input for every micrograph pair, where an initial set of seed pairs has to be manually selected, or where the geometric relation between zero-tilt and tilted specimen micrograph has to be determined. Due to the low SNR, so far, the selection of tilt pairs recorded under cryo conditions generally has to be done manually. Being a strenuous and time-consuming process, the interactive selection of tilt pairs severely limits the routine applicability of RCT/OTR based approaches (Sander et al., 2010; Scheres et al., 2009). To overcome this bottleneck, we developed a software, termed '*MaverickTilt*', that determines tilt pairs from independent particle coordinates from images which have been recorded during RCT/OTR analysis. The algorithm is fully automated and does not require user input, and is highly robust against correlation errors and noise. Our algorithm delivers reliable results for RCT datasets recorded both under negative stain and unstained cryo conditions.

2. Material and methods

In RCT or OTR, the same specimen area is recorded twice, with the specimen being tilted in the second image with respect to the

Abbreviations: TEM, transmission electron microscopy; RCT, Random Conical Tilt reconstruction; OTR, Orthogonal Tilt reconstruction; SNR, signal-to-noise ratio; HBC, homogeneous barycentric coordinates.

* Corresponding author. Fax: +49 551 201 1197.
E-mail address: hstark1@gwdg.de (H. Stark).

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2012.10.014

first around an experimentally defined angle (usually ranging between 45° and 70°). Typically, the transformation of a point X_{ut} with the coordinates x_{ut} and y_{ut} from the original tilt state of the specimen to the point X_t with the coordinates x_t and y_t can be described by an affine transformation. The underlying affine transformation combines rotations around the three axes of the three-dimensional coordinate system around the angles θ_x, θ_y and θ_z with a translation t and can thus be denoted as

$$\begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y 0 & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ut} \\ y_{ut} \\ 1 \end{bmatrix} \quad (1)$$

In (1), the three-dimensional affine transformation of a two-dimensional plane is described. Since the electron microscopy grid has a certain topology, the two-dimensional case described in (1) does not give an accurate description of the practical problem. Yet, the affine transformation of coordinates of particles on an EM grid can be approximated by (1) in the case of points which are close to each other since here, the topology of the EM grid occurs only in a very small dimension, thus, the problem can be approximated as being two-dimensional.

Any algorithm seeking to correlate particle positions on an untilted versus a tilted micrograph faces a set of challenges. First, due to instability of the cryo-EM specimen holders and the underlying tilt geometry, the illuminated areas in the untilted versus the tilted specimen are not identical; instead, they are overlapping only in certain areas which are unknown. Second, automated particle detection algorithms (for example Chen and Grigorieff, 2007; Ludtke et al., 1999; Woolford et al., 2007) will potentially select different particles on untilted and tilted specimen areas. These differences in results of automated particle selection are even greater when specimens were recorded under cryo conditions. Furthermore, the centers of the selected particles in both micrographs suffer on a jitter effect caused by the particle picking software as well as due to the differential ice thickness and grid topology. Moreover, the axes of tilt which determine the underlying affine transformation in (1) are not known. For a fully automated determination of correlating particles in the untilted versus tilted micrograph, it thus becomes necessary to compare all points in both untilted and tilted micrograph as potential tilt pairs. In order to achieve comparability of particle coordinates, every particle center coordinate (referred to hereafter as *point*) is regarded within its environment of closely surrounding points. Since a set of close particle center coordinates on an EM micrograph can be approximated to be in a plane, within a set of four close points, every point P can be represented by

$$P = \alpha P_1 + \beta P_2 + \gamma P_3 \quad (2)$$

where

$$\alpha + \beta + \gamma = 1 \quad (3)$$

Here, α, β and γ are called homogenous barycentric coordinates, and P_1, P_2, and P_3 denote particle center coordinates close to and in the same plane as P. Homogeneous barycentric coordinates (HBCs) are affine invariant, meaning that they are conserved under any combination of affine transformations. Thus, HBCs can be used for an affine invariant representation of particle coordinates on an EM grid that are close to each other.

2.1. Identification of an initial corresponding set of tilt pairs

For the identification of an initial set of corresponding tilt pairs, every particle center coordinate in both untilted and tilted micrograph is represented by the HBCs which are calculated by a combination of surrounding particles. As a standard, *MaverickTilt* calculates the HBCs from all possible combinations (disregarding sequence) of three out of the seven closest particle coordinates and the respective particle coordinates. Thus, the algorithm always compares sets of four points, P and surrounding points P_1, P_2 and P_3 For each possible set of combinations, HBCs α, β and γ are calculated according to (2). Subsequently, α, β and γ and the corresponding points P_1, P_2 and P_3 are reordered by sorting according to numerical order of the HBCs (smallest to largest). Depending on the number of particle center coordinates used as input, the number of combinations can be very high, so the transposed coordinate sets are prescreened in order to reduce the complexity for further processing. From each set of HBCs from untilted particle center coordinates, the barycentric error $error_{bary}$ with respect to each set of HBCs from all tilted particle center coordinates is calculated:

$$error_{bary} = (\alpha_{ut} - \alpha_t)^2 + (\beta_{ut} - \beta_t)^2 + (\gamma_{ut} - \gamma_t)^2 \quad (4)$$

where α_{ut}, β_{ut} and γ_{ut} denote the HBCs of the untilted particle coordinate and α_t, β_t and γ_t denote HBCs of the tilted particle coordinates, respectively. For each untilted coordinate set P_{ut} consisting of P, P_1, P_2 and P_3 according to (2) only up to twenty coordinate sets P_t of tilted particles with the lowest $error_{bary}$ are stored for further processing. From this coordinate set combination, the underlying affine transformation is estimated by the simplified form

$$P_t = AP_{ut} + t \quad (5)$$

The real-number parameters A and t from (5) are determined by a least-square fitting algorithm and applied to each P_{ut} and up to 70 surrounding coordinates to construct the coordinate set U which is then compared to all coordinates in the tilted micrograph V. The coordinate sets U,V are compared by the *enhanced Hausdorff distance* H_{EH} (Gope and Kehtarnavaz, 2007) which is calculated according to

$$H_{EH}(U,V) = \max(h_{EH}(U,V), h_{EH}(V,U)) \quad (6)$$

where

$$h_{EH}(U,V) = \frac{1}{m-\psi}\sum_{u\in U} d(u.V) \quad (7)$$

and

$$d(u,V) = \min_{v\in V}||u - v|| \quad (8)$$

In (7), ψ is equivalent to the number of cases in which more than one point in U has the smallest Euclidian distance to a distinct point in V according to (8), and m is the number of points in U. By calculating the enhanced Hausdorff distance between the two affine invariant point sets, the correspondence information (which point in the untilted point set is most likely to correspond to another point in the tilted point set) can be obtained. The corresponding set of four points P_{ut} and P_t for which (5) resulted in transposing parameters for which the smallest enhanced Hausdorff distance was calculated is accepted as the initial set of tilt pairs.

2.2. Iterative identification of tilt pairs

After an initial corresponding set of four tilt pairs is identified, these tilt pairs are used to iteratively extend the number of identified tilt pairs over the whole dataset. In the first step of each iteration, the particle coordinates of any particle to which no tilt mate is known and which are closest to any particle coordinate of a par-

ticle to which a tilt mate is known are calculated. From the closest four known tilt pairs, the coordinates are used to estimate the underlying parameters of the affine transformation according to (5). These parameters are then used to predict the position of the particle in the tilted micrograph. Additionally, the closest three coordinates of any known tilt pairs in the untilted and tilted image are used to calculate the HBCs of the same particle coordinates in the untilted image. The HBCs are then used to calculate the position of the particle in the tilted micrograph from the coordinates of the tilt pairs of the closest three coordinates in the untilted micrograph. From the two independently predicted coordinates, the closest actual coordinates in the tilted micrograph are calculated, and if both prediction methods predict the same point, the error of prediction is calculated, otherwise, the prediction is discarded. If the distance between the closest coordinate to the predicted coordinate is minimally twice as small as the distance to the second closest coordinate and if the barycentric error of the newly predicted tilt mates according to (4) is smaller than 0.1, the new match is accepted. Here, the cutoff of 0.1 was chosen and has worked well in a number of positive tests. This process is iterated throughout the whole dataset until no new tilt pairs can be identified.

The algorithm was implemented in the '*MaverickTilt* software' using Python 2.6 and NumPy (Oliphant, 2007).

2.3. Tests on simulated data

To assess the accuracy and effectiveness of the software as well as to estimate the susceptibility of the approach to noise, the software was tested against simulated data. For the simulation of RCT/OTR particle center coordinate data, a defined number of coordinates was randomly distributed on a two-dimensional grid of defined maximal dimensions in order to represent particle distributions on the untilted micrograph. To account for the tilting of the specimen in the electron microscope, an affine transformation as described in (1) was applied to the coordinates representing the untilted micrograph. In practice, automated or semi-automated selection of particle center coordinates from micrographs will lead to errors in selecting the accurate particle center. Additionally, three-dimensional grid topology will lead to a deviation of the particle center coordinates from a simulated two-dimensional plane transform. This combined effect will be referred to as 'de-centering noise'. For the simulation of de-centering noise, all particles in the dataset were shifted in random *x*- and *y*-directions by a predefined maximal de-centering noise value. When automatically selecting particles from the same specimen area in the untilted and the tilted micrograph, there will always be a fraction of selected particles in one micrograph which is not present on the other micrograph. This is due to the fact that the automated particle selection software potentially selects a different set of particles and contaminants (i.e. ice, ethane, etc.) on each of the two corresponding micrographs. Additionally, the illuminated areas are partially non-overlapping because of the tilt geometry and shifts of the sample stage in the electron microscope during tilting. This contributing factor is referred to as 'coordinate noise'. Coordinate noise was simulated by adding non-correlating random coordinates to each untilted and tilted coordinate dataset.

2.4. Tests on experimental data

To test the applicability of the software on experimental data, tilt pairs of negatively stained V-type ATPase of *Thermus thermophilus* and unstained *cryo* tilt pairs of the *Escherichia coli* 70S ribosome were recorded. V-type ATPase was purified as previously described (Toei et al., 2007) and submitted to the GraFix protocol (Kastner et al., 2008). GraFix-treated particles were stained with 2% uranyl acetate and low dose images were recorded on a 4 k × 4 k CCD camera in a Philips CM200 FEG using an acceleration voltage of 160 kV at 121,000-fold magnification. Micrographs were recorded at an electron dose of 10–15 $e^-/Å^2$, the specimen holder was tilted around −45° and the second image was taken at an electron dose of 10–15 $e^-/Å^2$. From micrographs of both zero-tilt and tilted specimens, particles were selected using the SIGNATURE software (Chen and Grigorieff, 2007). Particle coordinates obtained from SIGNATURE were correlated using the *MaverickTilt* Software and around 42000 Tilt pairs were obtained from 27 micrographs. Following standard procedures (Sander et al., 2010), 650 RCT 3Ds were reconstructed from the dataset, using around 70 particles per reconstruction. Iterative 3D alignment and weighted averaging was done subsequently (Sander et al., 2010). After 3D alignment, 3D MSA and classification was applied to calculate 3D averages representing structural heterogeneity in the dataset. Unstained cryo grids of ribosomes were recorded on an 4 k × 4 k CCD camera in a FEI Titan Krios using an acceleration voltage of 300 kV at 75,000-fold magnification. The specimen holder was tilted around −45°, and images were recorded at an electron dose of 10–20 $e^-/Å^2$. Subsequently, the holder was reset to 0° and the previously recorded specimen areas were recorded at the same electron dose. From both zero-tilt and tilted micrographs, particles were selected using the boxer software from EMAN 1.9 (Ludtke et al., 1999). Using the *MaverickTilt* software, 20,915 Tilt pairs were obtained from 100 micrographs, from which 350 RCT 3D volumes were reconstructed as previously described (Sander et al., 2010), again using around 70 particles per reconstruction. 3D alignment, weighted averaging and classification was done like for the V-type ATPase (Sander et al., 2010). For image processing, the IMAGIG-5 software package was used (van Heel et al., 1996).

3. Results and discussion

3.1. Simulated data

For tests with simulated data, 500 random coordinates were plotted within a range of 0–5,000 pixels. To simulate tilting of the specimen, the coordinates were rotated for 45° around the *y*-axis and random shift of maximally 500 pixels in *x*- and *y*-direction was applied. When testing the '*MaverickTilt*' software on simulated data, the effects of de-centering noise and coordinate noise were first tested separately (Fig. 1A and B). De-centering noise was added as a random *x*- and *y*-shift of maximally ±60 pixels in increments of 5 pixels (Fig. 1B). For each parameter set, 10 independent datasets were analyzed following the methodology and terminology proposed earlier (Langlois and Frank, 2011). Here, we focus on three parameters: precision of the correlations which is defined as the number of true positives divided by the number of predicted positives (Langlois and Frank, 2011), the recall of the correlations which defines the fraction of true positives over all positives present in the dataset (Langlois and Frank, 2011) and the selection rate which is defined as the fraction of all correlations (true and false positives) over the number of elements present in the dataset. The result of the independent analysis of de-centering noise and coordinate noise is shown in Fig. 1A and B. For coordinate noise, the algorithm is relatively robust, with a precision above 0.86 even when 100% noise coordinates are added, i.e. 50% of the dataset being entirely uncorrelated. In this case, only around every tenth particle is assigned incorrectly. De-centering noise is introduced mostly by incorrect determination of the particle centers during automated particle selection. Here, the precision drops below 0.95 rapidly after shifts of up to 20 pixels, which would usually correspond to the radius of the molecules in micrographs. Notably, in both cases, recall and selection rates drop even more

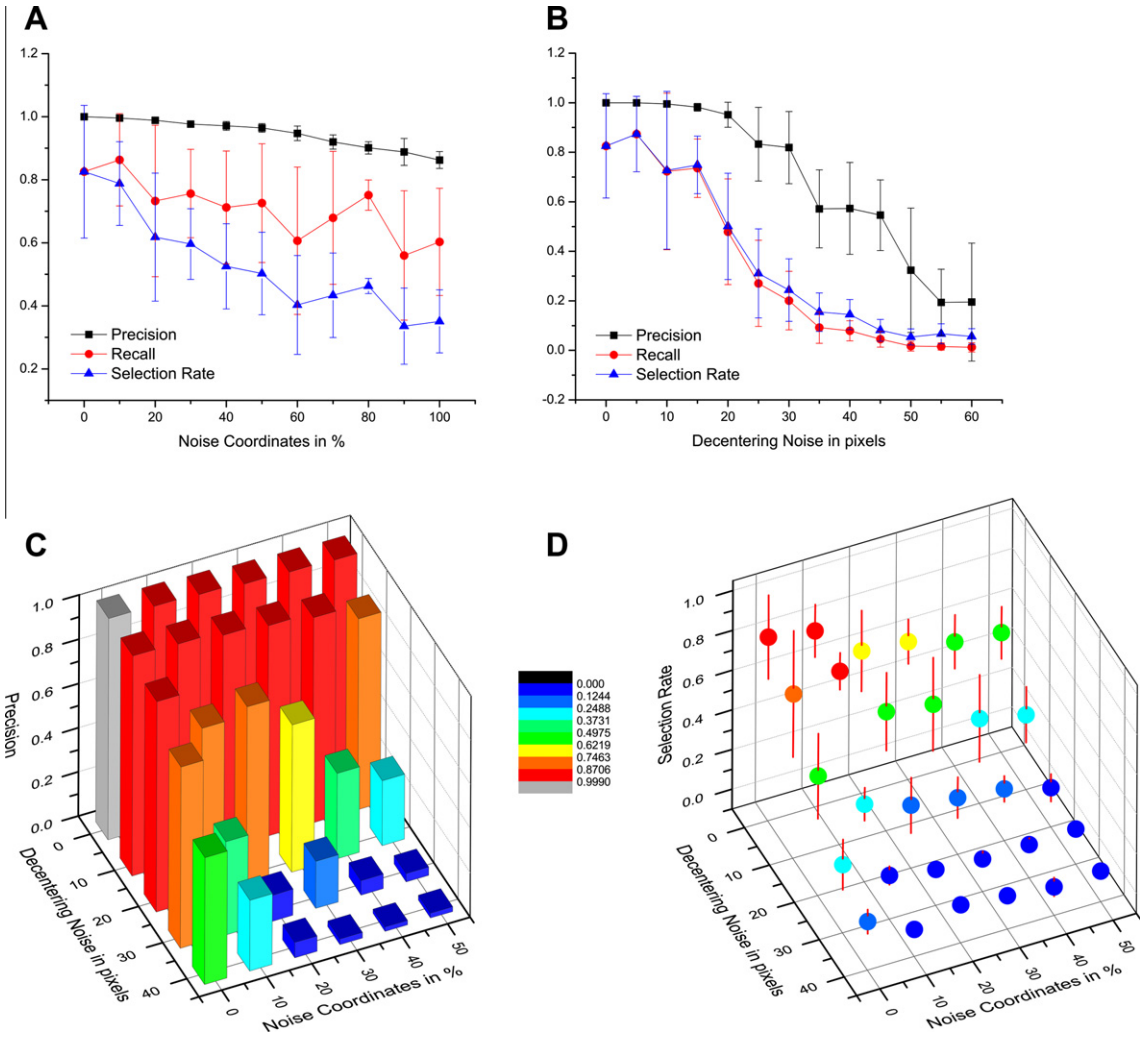

Fig. 1. Tests on simulated data. (A) Effects of uncorrelated noise coordinates on the performance of the *MaverickTilt* software. (B) Effects of decentering noise introduced by incorrect centering of the particle picking software. Precision and recall drop below tolerable limits starting from a centering error of around 20 pixels. (C) Combined effects of noise coordinates and decentering noise on the precision of the *MaverickTilt* software. (D) Selection Rate of the *MaverickTilt* software under the combined influence of noise coordinate and decentering noise. Error bars are indicated by vertical lines through data points. The selection rate invariantly drops below 0.37 when precision drops below 0.87, thus allowing identification of false positive correlations.

rapidly than precision. To assess the combined effects of de-centering noise and noise coordinates, the above mentioned parameters were combined to a maximum 50% noise coordinates and 40 pixels maximum shift in *x*- and *y*-direction (see Fig. 1C). For each parameter set, five independent datasets were analyzed. When plotting the selection rate of the *MaverickTilt* software (Fig. 1D), it can be seen that for every parameter set for which the percentage of correctly correlated particles is below 90%, the selection rate drops below 40%. Furthermore, the standard deviation of the percentage of retrieved coordinates (vertical bars in Fig. 1D) becomes negligible if the percentage of correctly correlated coordinates is lower than 15%. Thus, a low number of retrieved coordinates serves as a reliable measure for incorrect correlation of a dataset.

3.2. Experimental data

The Maverick Tilt software was tested on a dataset of negatively stained V-type ATPase from *Thermus thermophilus* and a dataset from an unstained *cryo*-preparation of the *E. coli* 70S ribosome. For both datasets, automated procedures for particle selections were used (Chen and Grigorieff, 2007; Ludtke et al., 1999). Taken together, the user input for the selection of tilt pairs required less than 30 min in both cases. In order to validate the correctness of the tilt pair prediction, around 70 particles were used for the reconstruction of a single RCT 3D volume. With this relatively high number of input projections, the SNR in the single reconstructions is high enough to allow for a visual inspection of 3Ds at this early stage (Fig. 2C and E). Additionally, using fewer 3Ds for 3D alignment and classification leads to a speed-up in the alignment process without having noticeable impact on the final consensus model (Sander et al., 2010). An overview of the results of the analyses of both datasets is given in Fig. 2. For the V-type ATPase, there were a total of 45 micrograph pairs with 10,000–12,000 square pixels, containing 1000–3000 particle coordinates per micrograph. As a result, coordinates from 42,000 particles were retrieved, most of which were separated from other particles instead of being located in local aggregates (see Fig. 2A). 650 RCT 3Ds could be reconstructed from the retrieved tilt pairs, which clearly showed features of V-Type ATPases (Fig. 2C). In these reconstructions, the difference between cytosolic V_1 and membrane-bound V_o domains can be clearly discerned. When calculating the consensus structure after 40 rounds of 3D alignment (Sander et al., 2010), the cytosolic V_1 and membrane-bound V_o domains are still discernible (Fig. 2F). The central and peripheral stalk connections of the molecule are

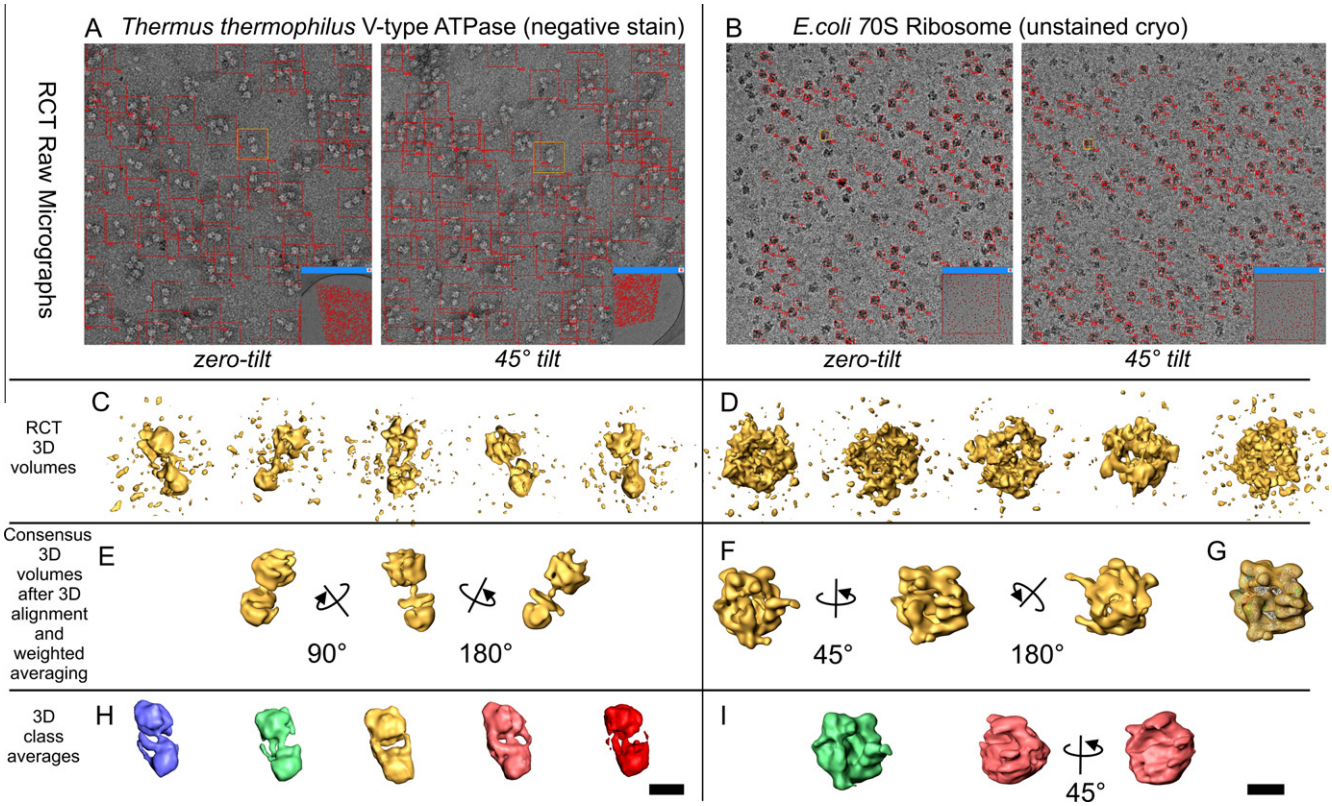

Fig. 2. Application examples of automated tilt pair selection. Scale bar: 10 nm. (A) Example raw micrographs of negatively stained V-type ATPase from *Thermus thermophilus*. A total of ≈42,000 tilt pairs were used for analysis. C. Example RCT 3D volumes of the V-type ATPase calculated from 65 tilt pairs each on average. (F) Consensus structure of the V-type ATPase after 3D alignment of all RCT volumes. The cytosolic V_1 domain and the membrane-bound V_o domain can be clearly distinguished. (I) 3D class averages of aligned RCT 3D volumes of the V-type ATPase. (B) Example raw micrographs of *E. coli* 70S ribosomes under unstained cryo conditions. A total of 20,915 tilt pairs were used for analysis. (D) Example 3D RCT volumes of 70S ribosomes, each calculated from 65 tilt pairs on average. (G) Weighted average 3D structure of all RCT reconstructions of the *E. coli* 70S ribosome. Structural features are highlighted (L1: L1 ribosomal stalk, L7/L12: L7/L12 ribosomal stalk, CP: central protuberance). (H) *E. coli* 70S crystal structure (PDB IDs 3R8N and 3R8S) fitted into the weighted average obtained from RCT 3Ds. (J) 3D class averages of the aligned RCT reconstructions, showing 70S ribosomes (green) and 50S ribosomes (red).

absent, indicating structural heterogeneity of these domains. These features, however, reappear in different conformations when applying 3D MSA and classification to all aligned 3D volumes (Fig. 2I). Additionally, a differential positioning of the upper cytosolic V_1 domain versus the lower membrane domain V_o can be observed. For the *E. coli* 70S ribosome, 20,915 tilt pairs were correlated from 100 micrograph pairs with 4096 square pixels, containing 200–500 particle coordinates per micrograph. From the correlated tilt pairs, a total of 350 RCT reconstructions were calculated. Example RCT volumes are shown in Fig. 2E. The consensus structure after 3D alignment is shown in Fig. 2G. Here, the structure of the *E. coli* 70S ribosome is reconstructed with variable domains like the L1 and L7/L12 stalk and the toe being present, reproducing the reconstruction presented earlier by Sander et al. (2010). The crystal structure of the *E. coli* 70S can be docked into the reconstructed density in order to validate the accuracy of the reconstruction (Fig. 2H, PDB ID 3R8N and 3R8S). Example 3D class averages for the aligned RCT 3D reconstructions show both 70S and 50S ribosomes (Fig. 2J), demonstrating the successful disentanglement of compositional heterogeneity of the dataset. Without discussing the structural findings of these reconstructions in further detail, we demonstrate the applicability of the *MaverickTilt* software on both negatively stained and unstained cryo RCT datasets.

3.3. Comparison with manual annotation of tilt pairs

To demonstrate the usability of the *MaverickTilt* software, we made a comparison of a manually annotated tilt pair data set and a tilt pair data set correlated automatedally. Tilt pairs of an unstained cryo dataset were manually selected and centered. For the automated correlation, particles were selected from micrographs using EMAN's boxer software (Ludtke et al., 1999), and subsequently correlated with *MaverickTilt*. Results were compared by calculating the Euclidian distance between coordinates acquired with both methods, and particles were evaluated to be identical if the distance of their center coordinates were less than half the particle's radius. Over 20 micrographs, the *MaverickTilt* algorithm found on average 80% of the total number of tilt pairs manually identified, of which 96% were identical to the manually selected tilt pairs on average. Two percent of the compared coordinates did not have the same tilt mates, an error which might have arisen either from incorrect manual indexing or the automated particle selection of the center of two aggregated particles as a single particle. In any case, an error of 2% in the tilt pair correlation will not be noticeable in the RCT/OTR analysis of large datasets and may be within the standard error of correlation during manual assignment of tilt pairs.

The applicability of the *MaverickTilt* software as well as the amount of user input required depends largely on the previous step of semi-automated or automated particle selection. We have demonstrated the use of two different software packages here. There are, however, many more software packages available. *MaverickTilt* delivers usable results even with an average de-centering noise of 20 pixels (see Fig. 1). For most automated or semi-automated particle selection setups with particle diameters of up to 200 Å and pixel sizes around 5 Å/pixel, this would mean that

the determined particle coordinate is one particle radius away from the actual particle center. Thus, the *MaverickTilt* software handled even suboptimal results from previous automated or semi-automated particle selection routines. Setting up the *MaverickTilt* software takes less than a minute once the coordinates are available, so the amount of user input required solely depends on the input and interaction required by the previous particle picking step.

The computational effort of automated tilt pair correlation increases exponentially with the number of input coordinates. Thus, the automated correlation of very large micrographs (>2000 particles per zero-tilt/tilted micrograph) can become time-consuming (up to 3 days for the correlation of two micrographs with 5000 particle coordinates each) even compared with manual correlation, where the time expense grows linearly with the number of particles. The required user input, however, remains the same, and parallelization can help to overcome the bottleneck of an extended calculation time. In its current implementation, the *MaverickTilt* software makes use of multiprocessor environments. Although further parallelization is possible, it was not necessary for the datasets tested.

Since the algorithm will always come up with a minimal set of 4 particle pairs even in case of uncorrelated input, the automated identification of false positives is crucial for the applicability of the algorithm. When recording a RCT or OTR dataset, typically, a larger set of zero-tilt and tilted micrographs are recorded. For the recording of the entire dataset, usually, the tilt angle of the specimen holder as well as the image rotation angle of the microscope remains the same. Once these requirements are met, they can be exploited for an additional detection and correction of falsely correlated tilt pairs. As a first step of the angular refinement of correlated tilt pairs, the transformation between the zero-tilt particle coordinates and the coordinates from the tilted micrographs are determined following the methodology proposed by Voss et al. (2009). In this approach, the specimen tilt angle and the untilted and tilted image rotation angles are determined via a least squares refinement of the actual and predicted particle coordinates (Voss et al., 2009). From all correlations of all micrographs, a set of weighted consensus angles is calculated as a weighted average of all angles, where each singular angle is weighted by the calculated RMSD and the number of particles which were correlated for the respective micrograph pair. Correlation sets from micrographs where either tilt angle is more than one standard deviation away from the set of weighted consensus angles is deleted, allowing for the automated elimination of false positives.

4. Conclusions

We present an approach to correlate coordinates from zero-tilt and tilted micrographs, where coordinates from both micrographs can be determined independently and then be correlated without any further user input. This approach overcomes the bottleneck of time-consuming interactive correlation of tilt pairs for RCT and OTR. Together with the automated protocols for the acquisition of tilt pairs (Yoshioka et al., 2007) and particle selection (Chen and Grigorieff, 2007; Ludtke et al., 1999; Woolford et al., 2007), the *MaverickTilt* software enables the routine application of RCT or OTR analysis to large single-particle EM datasets. The *MaverickTilt* software reliably automates the initial process of tilt pair correlation for RCT-based approaches which allow the determination of bias-free initial references to represent the structural heterogeneity in the dataset (Sander et al., 2010; Scheres et al., 2009). With the removal of this initial burden and bottleneck, RCT-based approaches become applicable for an extended variety of macromolecular complexes, resulting in more reliable initial models which allow the *ab initio* analysis of structural heterogeneity of a single-particle EM dataset.

The *MaverickTilt* software has been implemented in Python 2.6 and requires NumPy, allowing a cross-platform usage and integration into existing frameworks. In its current implementation, the software handles coordinates from EMAN 1.9 (Ludtke et al., 1999), IMAGIC-5 (van Heel et al., 1996) and SPIDER (Frank et al., 1996). The software has a self-explanatory command-line interface in which the input files and input format of the coordinate files have to be specified. The software allows for optional scaling and axis swapping of the output files. The software is available from F.H. and H.S. upon request.

Acknowledgments

70S Ribosomes of *E. coli* were kindly provided by Andrey Konevega and Marina Rodnina. This work was carried out partly in the group of Yoshinori Fujiyoshi at the University of Kyoto, Japan. F.H. was supported by a postdoctoral short term fellowship from the Japanese Society for the Promotion of Sciences (JSPS).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2012.10.014.

References

Chen, J.Z., Grigorieff, N., 2007. SIGNATURE: a single-particle selection system for molecular electron microscopy. J. Struct. Biol. 157, 168–173.
Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., et al., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190–199.
Gope, C., Kehtarnavaz, N., 2007. Affine invariant comparison of point-sets using convex hulls and Hausdorff distances. Pattern Recogn. 40, 309–320.
Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., et al., 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods 5, 53–55.
Langlois, R., Frank, J., 2011. A clarification of the terms used in comparing semi-automated particle selection algorithms in Cryo-EM. J. Struct. Biol. 175, 348–352.
Leschziner, A.E., Nogales, E., 2006. The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. J. Struct. Biol. 153, 284–299.
Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97.
Oliphant, T.E., 2007. Python for scientific computing. Comput. Sci. Eng. 9, 10–20.
Penczek, P.A., Yang, C., Frank, J., Spahn, C.M., 2006. Estimation of variance in single-particle reconstruction using the bootstrap technique. J. Struct. Biol. 154, 168–183.
Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1987. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. J. Microsc. 146, 113–136.
Sander, B., Golas, M.M., Luhrmann, R., Stark, H., 2010. An approach for de novo structure determination of dynamic molecular assemblies by electron cryomicroscopy. Structure 18, 667–676.
Scheres, S.H., Melero, R., Valle, M., Carazo, J.M., 2009. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. Structure 17, 1563–1572.
Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., et al., 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. 148, 194–204.
Toei, M., Gerle, C., Nakano, M., Tani, K., Gyobu, N., et al., 2007. Dodecamer rotor ring defines H+/ATP ratio for ATP synthesis of prokaryotic V-ATPase from *Thermus thermophilus*. Proc. Natl. Acad. Sci. USA 104, 20256–20261.
van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., 1996. A new generation of the IMAGIC image processing system. J. Struct. Biol. 116, 17–24.
Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., Carragher, B., 2009. DoG picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 166, 205–213.
Woolford, D., Ericksson, G., Rothnagel, R., Muller, D., Landsberg, 2007. SwarmPS: rapid, semi-automated single particle selection software. J. Struct. Biol. 157, 174–188.
Yoshioka, C., Pulokas, J., Fellmann, D., Potter, C.S., Milligan, R.A., et al., 2007. Automation of random conical tilt and orthogonal tilt data collection using feature-based correlation. J. Struct. Biol. 159, 335–346.