Maximum-likelihood Multi-reference Refinement for Electron Microscopy Images

Sjors H. W. Scheres[1], Mikel Valle[1], Rafael Nuñez[1], Carlos O. S. Sorzano[1,2] Roberto Marabini[3], Gabor T. Herman[4] and Jose-Maria Carazo[1,3]*

[1]*Biocomputing Unit, Centro Nacional de Biotecnología Campus Universidad Autónoma Cantoblanco, 28049, Madrid Spain*

[2]*Escuela Politécnica Superior Universidad San Pablo-CEU Boadilla del Monte, 28668 Madrid, Spain*

[3]*Escuela Politécnica Superior Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

[4]*The Graduate Center, City University of New York, New York, NY 10016, USA*

*Corresponding author

A maximum-likelihood approach to multi-reference image refinement is presented. In contrast to conventional cross-correlation refinement, the new approach includes a formal description of the noise, implying that it is especially suited to cases with low signal-to-noise ratios. Application of this approach to a cryo-electron microscopy dataset revealed two major classes for projections of simian virus 40 large T-antigen in complex with an asymmetric DNA-probe, containing the origin of simian virus 40 replication. Strongly bent projections of dodecamers showed density that may be attributed to the complexed double-stranded DNA, while almost straight projections revealed a twist in the relative orientation of the hexameric subunits. This new level of detail for large T-antigen projections was not detected using conventional techniques. For a negative stain dataset, maximum-likelihood refinement yielded results that were practically identical to those obtained using conventional multi-reference refinement. Results obtained using simulated data suggest that the efficiency of the maximum-likelihood approach may be further enhanced by explicitly incorporating the microscope contrast transfer function in the image formation model.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: maximum-likelihood; multi-reference refinement; single-particles; 2D-alignment; classification

Introduction

Electron microscopy is increasingly being used for structural characterization of large macromolecular complexes. In particular, cryo-electron microscopy, where macromolecules are rapidly frozen in a thin layer of vitreous ice, has allowed structural characterization of large biological assemblies almost in their native state. Due to the requirement of a low electron dose to minimize radiation damage, and the low contrast between biological matter and ice, the recorded images typically suffer from large amounts of noise. Especially in high-resolution studies, where close-to-focus images are recorded, projections with extremely low signal-to-noise ratios (SNRs) are relatively common.

In a pre-processing step, the low signal-to-noise ratios can be improved by two-dimensional (2D) averaging over multiple images. Such averaging is only justified if all images correspond to the same 2D-structure and if their relative rotations and translations are known. To determine the unknown orientations, a range of different 2D-alignment methods is available.[1] However, these methods typically require that the set of images is structurally homogeneous, and it is very common for electron microscopy datasets to contain more than one different 2D-structures. Examples of different 2D-structures that may be present within one dataset are projections of a given molecule from different directions, projections of different conformations of a given molecule, or even projections of different molecules in the specimen preparation.

To separate structurally heterogeneous image sets into homogeneous classes, multiple classification methods are at hand.[2,3] However, since most classification methods require that the images are aligned beforehand, alignment of heterogeneous sets of images represents a chicken-and-egg problem. The use of rotational and translational

invariants is a possible solution to this problem.[4,5] But, especially in cases with high noise levels, the relatively low information content of many of these invariants may impose problems upon classification. An alternative solution is the iterative use of 2D-alignment and classification methods, which has been termed "alignment through classification". Multiple variants of this approach have been reported, employing a range of various classification and alignment techniques.[6–8]

A more straightforward approach that intrinsically combines alignment and classification into a single, iterative process is multi-reference refinement.[9] In multi-reference refinement the complete set of images is compared to a predefined number of reference images, which are assumed to represent the structural diversity among the data. Each experimental image is aligned with respect to all references and is assigned to one of them based on a similarity measure. Averaging over the subsets so obtained serves to calculate new, improved references for the next iteration. This process stops when no more migration of particles between subsets is observed and the averages within each subset do not change any more.

Many different similarity measures for alignment and classification have been proposed,[10–14] and most of them may be used in multi-reference refinement. Probably the most widely used similarity measure for multi-reference refinement is based on a maximum cross-correlation (CC) criterion. A drawback of the employed CC function is that, due to the high noise levels in EM images, it typically suffers from many false maxima and the assignment of orientation and reference may become incorrect.[15] Furthermore, the final partition into different subsets has been reported to suffer from a strong dependency on the initial seeds (i.e. the initial reference images). Especially for particles with low signal-to-noise ratios, initial preferences in the starting images may be amplified, leading to biased results.[16,17]

An alternative approach to CC refinement is based on maximum-likelihood (ML). ML refinement aims at finding the most likely model given the observed data. This likelihood is measured by the probability that the observations would be made, given the current model. This model does not only consist of the structure(s) to be recovered, but also includes a formal error model. It can be shown that, for infinitely large datasets, maximizing the likelihood yields less biased estimates of the model parameters than those provided by alternative methods.[18] In the related field of X-ray crystallography, ML approaches have resulted in significantly improved structure refinement procedures.[19]

A ML approach to electron microscopy image refinement was first introduced by Sigworth for the alignment problem with a single 2D-reference.[20] In this approach, the parameters describing the relative orientations of the particles are treated as hidden variables, which are integrated out in the

likelihood calculation. Instead of taking discrete decisions based on small differences in CC, probability-weighted averages over all possible assignments are calculated. The error model for calculating these probabilities is based on the assumptions of independent and zero-mean Gaussian noise in the experimental projections. Furthermore, probability calculations include the prior probability distribution of the alignment parameters, assuming random rotations and Gaussian-distributed translations, centred at the origin. Using simulated data that complied with these assumptions, ML refinement was observed to yield better models with less bias than conventional CC alignment, especially for images with low SNRs.

Here, we introduce an extension of single-reference ML image refinement to the multi-reference case. Furthermore, we report the first application of ML image refinement to experimental (cryo-) electron microscopy data. The experimental situation differs substantially from the theoretical assumptions in the actual implementation of the ML approach. Aberrations in the optical system of the electron microscope lead to artefacts in the imaging process. This results in a point-spread function (the direct-space equivalent of the contrast transfer function, or CTF) of finite width, which gives rise to correlations between the intensity values of nearby pixels. Consequently, the assumption of independent noise does not hold for experimental data. We discuss results obtained with ML and conventional CC multi-reference refinements of both cryo-electron microscopy and negative stain data. We will focus on the former type that due to its high levels of noise represents the more challenging case, and for which ML with its implicit noise model is expected to be particularly well-suited. In addition, we use results obtained for synthetic data with simulated microscope-induced artefacts to further assess the potential of ML refinement of electron microscopy images.

Mathematical Background

The multi-reference model

Each image in the observed dataset $\mathbf{X}$ is assumed to be a rotated and translated copy of one of K underlying structures, to which independent Gaussian noise is added. The data are modelled as follows:†

$$X_i(\phi_i) = A_{k_i} + R_i \qquad (1)$$

where $X_i \in \mathfrak{R}^M$ is the ith observed image ($i=1,\dots,N$) of M pixels; $A_k \in \mathfrak{R}^M$ is an estimate for the kth ($k= 1,\dots,K$) underlying structure and A_{k_i} indicates

which structure is the correct one for image X_i; $X_i(\phi_i)$ describes that image X_i is mapped onto a transformed coordinate system and ϕ_i defines the transformation (both rotation and translation) that maps the image onto the standard position of A_{k_i}; $R_i \in \mathfrak{R}^M$ are images of independent, zero-mean Gaussian noise with standard deviation σ.

We aim to optimize the likelihood $P(\mathbf{X}|\Theta)$, i.e. the probability of observing dataset $\mathbf{X}$ given a model with parameter set Θ. Maximizing the likelihood is equivalent to maximizing its logarithm $L(\mathbf{X}|\Theta)$. Assuming independence between the individual images, this function can be written as a sum of log-likelihoods for all images X_i. Furthermore, since the assignments of k_i and ϕ_i are mutually exclusive, the probability of observing one image can be expanded as a summation over all assignments. The log-likelihood can thus be written as:

$$L(\mathbf{X}|\Theta) = \sum_{i=1}^{N} \log P(X_i|\Theta)$$

$$= \sum_{i=1}^{N} \log \sum_{k=1}^{K} \int_{\phi} P(X_i|k, \phi, \Theta) P(k, \phi|\Theta) \mathrm{d}\phi \quad (2)$$

where $P(X_i|k,\phi,\Theta)$ is the conditional probability of observing image X_i, given k, ϕ and Θ; and $P(k,\phi|\Theta)$ is the probability density function of k and ϕ.

Given k and ϕ, we model each pixel value of the rotated and translated image $X_i(\phi)$ as the value of the corresponding pixel in A_k, to which zero-mean Gaussian noise with standard deviation σ is added (*cf.* equation (1)). The probability density function for each pixel value can thus be expressed as a Gaussian. Furthermore, we assume independence between the probability density functions of all individual pixels j, so that the conditional probability of observing the entire image given k and ϕ can be written as a multiplication of M Gaussians:

$$P(X_i|k, \phi, \Theta)$$

$$= \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{([A_k]_j - [X_i(\phi)]_j)^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left(-\frac{||A_k - X_i(\phi)||^2}{2\sigma^2}\right) \quad (3)$$

where $[\cdot]_j$ denotes the jth pixel in an image; and $||\cdot||^2$ is used to indicate the squared norm of an image.

We model the probability density of k by the estimated proportions α_k of the kth structure in the data ($\alpha_k \geq 0$ and $\sum_{k=1}^{K} \alpha_k = 1$). Assuming that particle-picking has yielded roughly centred particles with uniformly distributed rotations, we model the probability density of ϕ by a two-dimensional Gaussian distribution centred at the origin. Thereby, the probability density function of k and ϕ is given by:

$$P(k, \phi|\Theta)\mathrm{d}\phi$$

$$= \alpha_k \frac{1}{2\pi\xi^2} \exp\left(-\frac{q_x^2 + q_y^2}{\xi^2}\right) \frac{\mathrm{d}q_\alpha}{2\pi} \mathrm{d}q_x \, \mathrm{d}q_y \quad (4)$$

where $\phi = (q_\alpha, q_x, q_y)$, with q_α being its rotational component, and q_x, respectively q_y its translational components in x and y; ξ is the estimated standard deviation in the assumed Gaussian distribution of residual origin offsets.

Equations (1), (3) and (4) form the underlying model in our approach. Its parameter set Θ thus consists of: images A_k, standard deviation in the noise σ, proportions of underlying structures α_k, and standard deviation in the origin offsets ξ. The optimization task at hand is to find those parameters that maximize the log-likelihood as defined in equation (2).

The EM algorithm

Maximization of the log-likelihood would be trivial if we knew beforehand the values of k_i and ϕ_i for all images. Unfortunately, these are unknown and the data are thus incomplete. instead, we use the expectation-maximization (EM) algorithm to maximize the log-likelihood.[21] In the *E*-step of this algorithm, a lower bound to the likelihood is built, based on the current estimates for the model parameters ($\Theta^{(n)}$, n being the iteration number). The best lower bound is given by:

$$Q(\Theta, \Theta^{(n)}) =$$

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \int_{\phi} P(k, \phi|X_i, \Theta^{(n)}) \log \{P(k, \phi|\Theta)P(X_i|k, \phi, \Theta)\}\mathrm{d}\phi$$

$$(5)$$

where $P(k,\phi,|X_i,\Theta^{(n)})$ is the probability of k and ϕ, given the observed image and the current estimates for the model parameters. Using Bayes' theorem, this probability is calculated by:

$$P(k, \phi|X_i, \Theta^{(n)})$$

$$= \frac{P(X_i|k, \phi, \Theta^{(n)})P(k, \phi|\Theta^{(n)})}{\sum_{k'=1}^{K} \int_{\phi'} P(X_i|k', \phi', \Theta^{(n)})P(k', \phi'|\Theta^{(n)})\mathrm{d}\phi'} \quad (6)$$

In the subsequent *M*-step of the algorithm, we maximize $Q(\Theta, \Theta^{(n)})$ to obtain a parameter set for the next iteration ($\Theta^{(n+1)}$), which increases the likelihood. For this purpose, we set partial derivatives to each of the parameters equal to zero and solve for the corresponding variable. For images A_k we get:

$$A_k^{(n+1)} = \frac{\sum_{i=1}^{N} \int_{\phi} P(k, \phi|X_i, \Theta^{(n)})X_i(\phi)\mathrm{d}\phi}{\sum_{i=1}^{N} \int_{\phi} P(k, \phi|X_i, \Theta^{(n)})\mathrm{d}\phi} \quad (7)$$

Similarly, estimates σ for the standard deviation of the noise, and ξ for the standard deviation in the origin offsets are calculated as:

$$\sigma^{(n+1)} =$$

$$\sqrt{\frac{1}{NM} \sum_{i=1}^{N} \sum_{k=1}^{K} \int_{\phi} P(k,\phi|X_i,\Theta^{(n)})||A_k^{(n)} - X_i(\phi)||^2 \, d\phi}$$

(8)

$$\xi^{(n+1)} = \sqrt{\frac{1}{2N} \sum_{i=1}^{N} \sum_{k=1}^{K} \int_{\phi} P(k,\phi|X_i,\Theta^{(n)})(q_x^2 + q_y^2)d\phi}$$

(9)

For maximization of $Q(\Theta,\Theta^{(n)})$ with respect to α_k, we introduce a Lagrange multiplier to constrain $\sum_{k=1}^{K} \alpha_k = 1$, yielding the following solution:

$$\alpha_k^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} \int_{\phi} P(k,\phi|X_i,\Theta^{(n)}) \, d\phi \qquad (10)$$

The EM algorithm proceeds by using the newly derived parameter set $(\Theta^{(n+1)})$ to calculate the lower bound for the next iteration.

Results

We used various types of structurally heterogeneous data to validate ML multi-reference refinement and tested its relative performance with respect to the conventional CC approach.[9] The order in which we present our results is from the most challenging and biologically relevant to the more controllable data. We first present results obtained for a cryo-electron microscopy dataset, then for a negative stain dataset, and finally for simulated data of simplified phantom structures. Both experimental datasets were recorded recently in the context of our ongoing work on prokaryotic and eukaryotic helicases. The high levels of noise in the cryo-electron microscopy data severely limited the application of conventional alignment and classification tools, whereas existing tools could be applied to classify the negative stain data.

Application to cryo-electron microscopy data

Both the ML and CC[9] multi-reference refinement protocols were applied to a cryo-electron microscopy dataset comprising 7590 projections of simian virus 40 (SV40) large T-antigen (Figure 1(a) and Materials and Methods). These molecules have been previously observed to assemble as hexamers or as dodecamers,[22,23] and for the latter different lengths and curvatures have been reported.[24,25] The specimen under study here contained dodecameric large T-antigen molecules in complex with the origin of SV40 replication. The 107 base pairs long, asymmetric DNA-probe of this complex was designed to extend 23 base pairs from one side of the helicase complex (Figure 1(b)). Using

various conventional 2D-alignment and classification techniques, no DNA-density was observed (results not shown).

To avoid any bias in the starting seeds of the multi-reference refinement, initial estimates A_k were obtained by calculating the average of equally sized, random subsets of the unaligned images. Starting from the same five references, both ML and CC refinement converged to images that may be interpreted as straight and bent projections of dodecamers, as well as projections of hexamers or disordered dodecamers (Figure 1(c)). Refinement using seven references yielded a similar picture (not shown). In this case, both methods yielded two very noisy references with relatively low probability weights or numbers of assigned images, suggesting that five references may be sufficient to reflect the structural diversity in these data.

The first three references from the ML and CC refinements with five references are grossly related and may be interpreted as projections of dodecamers. However, important differences of potential biological relevance were observed between the results of the two methods. The ML-refined images, and in particular references two and three, appear less fuzzy than their CC-refined counterparts. As a consequence, the ML-refined images may be interpreted as a strongly bent projection (reference one) and more-or-less straight projections (references two and three), whereas such clear differences are not appreciated among the CC-refined images. Furthermore, among the ML-refined references two different projections of the hexameric subunits can be distinguished (most clearly in reference two), while the fuzzy appearance of the CC-refined images does not allow such a distinction. Another difference between the results of both methods concerns density that may be attributed to the complexed double-stranded DNA. In particular for reference one from the ML-refinement, relatively weak, elongated density of 10–15 Å width was observed to protrude approximately 40 Å from the lower hexameric subunit. For the CC-refined projections no such density was observed.

For references four and five, the results of both methods are less related. References four and five from the CC refinement may be interpreted as a projection of a dodecamer and a projection of a hexamer or a disordered dodecamer, respectively. For the ML-refinement, reference four may represent a hexamer or a disordered dodecamer, while reference five represents an overall gradient in the (wrapped-around) background density level. Such a structure was not observed among the CC-refined images. Detailed analysis of the experimental data showed that the signal in many images indeed suffers from a gradient in the background density (for example the image on the right-hand side of Figure 1(a)). While these images mainly contribute to this particular reference in the ML refinement, they may "pollute" any reference in the CC refinement, which may explain their relatively fuzzy appearance.
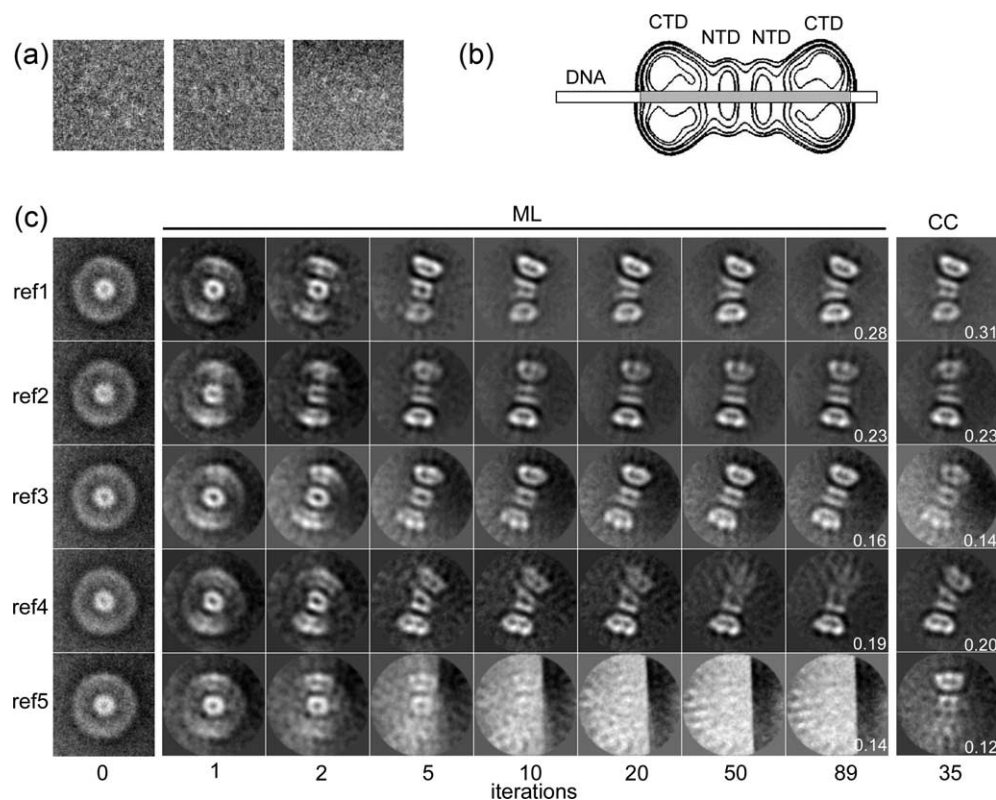
Figure 1. Application to cryo-electron microscopy data. (a) Three cryo-electron microscopy images of large T-antigen molecules in complex with the origin of SV40 replication. (b) Schematic representation of a model for the proposed interaction of large T-antigen dodecamers (isolines) with the asymmetric DNA-probe (rectangle). Large T-antigen assembles in two hexameric rings around the DNA. The C-terminal and N-terminal domains of the hexameric subunits are indicated (CTD and NTD, respectively), and the part of the DNA that is protected from proteolysis upon formation of a complex with large T-antigen is shown in grey. This part also contains the (asymmetric) origin of SV40 replication. The DNA protruding from the left-hand side of the complex is 23 base pairs long. (c) ML and CC five-reference (ref1-5) refinement runs. ML refinement was stopped after 89 iterations; CC refinement after 35 iterations. Shown are the initial seeds of both runs (left column), the ML-refined structures after iterations 1, 2, 5, 10, 20, 50 and 89 (middle block), and the CC-refined structures after iteration 35 (right column). White numbers in the reference images indicate the corresponding values for α_k (equation (10)) in the ML case, and the relative amount of image assigned to the corresponding reference in the CC case.

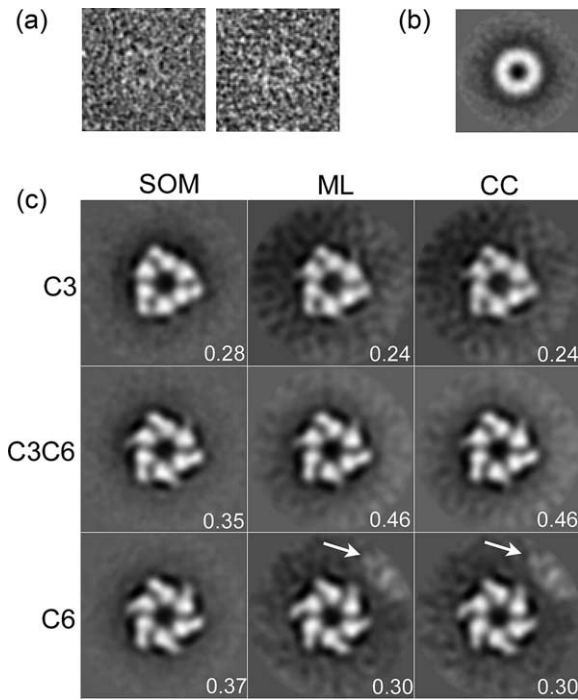## Application to negative stain data

We also applied ML and CC multi-reference refinement to a well-characterized, negative stain electron microscopy dataset. These test data consisted of 2120 top views of *Bacillus subtilis* bacteriophage SPP1 replicative helicase G40P particles (see Figure 2(a) and Materials and Methods). This helicase has been reported to assemble in three quaternary organizations with different rotational symmetries: a C3, a C6 and an intermediate C3C6 form.[26,27] We used self-organizing maps[25] of rotational spectra[28] to analyse these data, resulting in a well-characterized test set comprising only G40P top views, of which 28% corresponded to the C3 state, 35% to the C3C6 and 37% to the C6 state.

Again, bias-free seeds for the multi-reference refinements were obtained by averaging over three equally sized random subsets of the unaligned images (Figure 2(b)). For these data, ML and CC refinement converged to very similar solutions, starting from the same three references (Figure 2(c)).

Even features in the surrounding background density of the different structures as obtained by the two methods showed a strong resemblance (indicated with arrows in Figure 2(c)). From the resulting images, the C3, C6 and C3C6 quaternary states are readily recognizable, although the refined C3 images exhibit a somewhat elevated sixfold component.

Test on simulated data

To further assess the potentials of ML refinement and its relative performance with the CC approach, we tested both methods on synthetic data. A structurally heterogeneous test set was simulated by applying 1000 random rotations and translations to two phantom images of 32×32 pixels (Figure 3(a)). These images were obtained by projection of phantom structures consisting of six and seven spheres, respectively. The applied rotation angles were uniformly distributed and the applied translations were taken from a Gaussian distribution

Figure 2. Application to negative stain data. (a) Two negative stain electron microscopy images of replicative helicase G40P. (b) One example of the three initial reference images. (c) Structures as obtained using classification of rotational spectra with self-organizing maps (SOM), ML or CC multi-reference refinement. These images were aligned with respect to a common reference, and correspond to the three quaternary states of G40P with different rotational symmetries: C3, C3C6 and C6. ML refinement was stopped after 32 iterations; CC after 38. White numbers in the reference images indicate the corresponding values for α_k (equation (10)) in the ML case, and the relative amount of image assigned to the corresponding reference in the CC case. Arrows are used to indicate non-relevant features in the background density, which coincide for the ML and CC-refined C6-structures.

centred at the origin and with a standard deviation of two pixels. Multiple variations of this test set were generated by adding different amounts of independent, Gaussian noise or applying a CTF-simulated filter (Figures 3(b) and 4(b)).

The quality of the refinement solutions was assessed in two ways: (i) inspection of the refined images; and (ii) calculation of a figure-of-merit (FOM). This FOM, based on the differences in rotational symmetry of the underlying phantom images, was calculated as follows:

$$\text{FOM} = \arg\max\left[\frac{C_{\text{ref1}}^6 + C_{\text{ref2}}^7}{2}, \frac{C_{\text{ref1}}^7 + C_{\text{ref2}}^6}{2}\right] \quad (11)$$

where $C_{\text{ref1}}^6 \in [0,1]$ is the sixfold component of the rotational spectrum[28] of reference image number 1, C_{ref2}^7 the sevenfold component of reference number 2, etc.

In a first experiment, we applied ML and CC
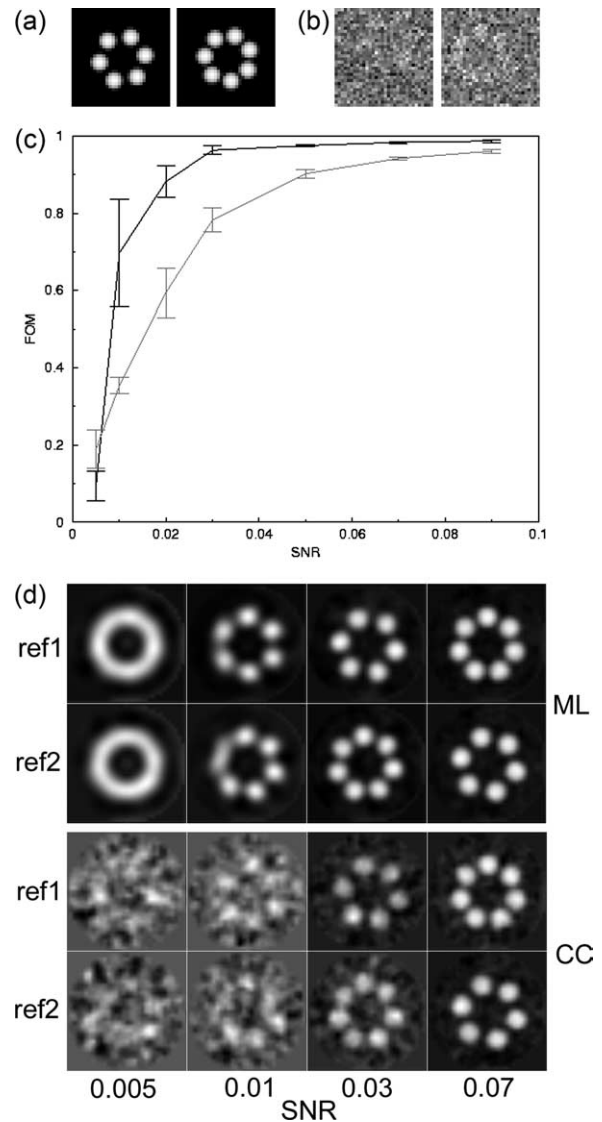
Figure 3. Application to unfiltered, phantom data. (a) The two phantom images that were used to create simulated, structurally heterogeneous datasets. (b) Two examples of rotated and translated copies of the phantom images to which independent Gaussian noise was added, resulting in a SNR of 0.07. SNRs were calculated as the variance of the signal divided by the variance of the added noise. (c) Average and standard deviation values of the FOMs (see main text) as obtained for six independent runs of ML (black) and CC (grey) two-reference refinements, using unfiltered phantom data with increasing SNRs. (d) Gallery of structures as obtained using ML and CC refinement at various SNRs. Shown are both references (1 and 2) for the third best out of six ML refinement runs (i.e. close to the median result) and the same structures as resulting from CC refinement.

refinement to unfiltered data with increasing amounts of independent, zero-mean Gaussian noise (Figure 3). For these data, which are in accordance with the ML noise model, ML refinement yielded significantly better results than CC. Although both methods yielded similar solutions with identical orientations for data with (quadratic)
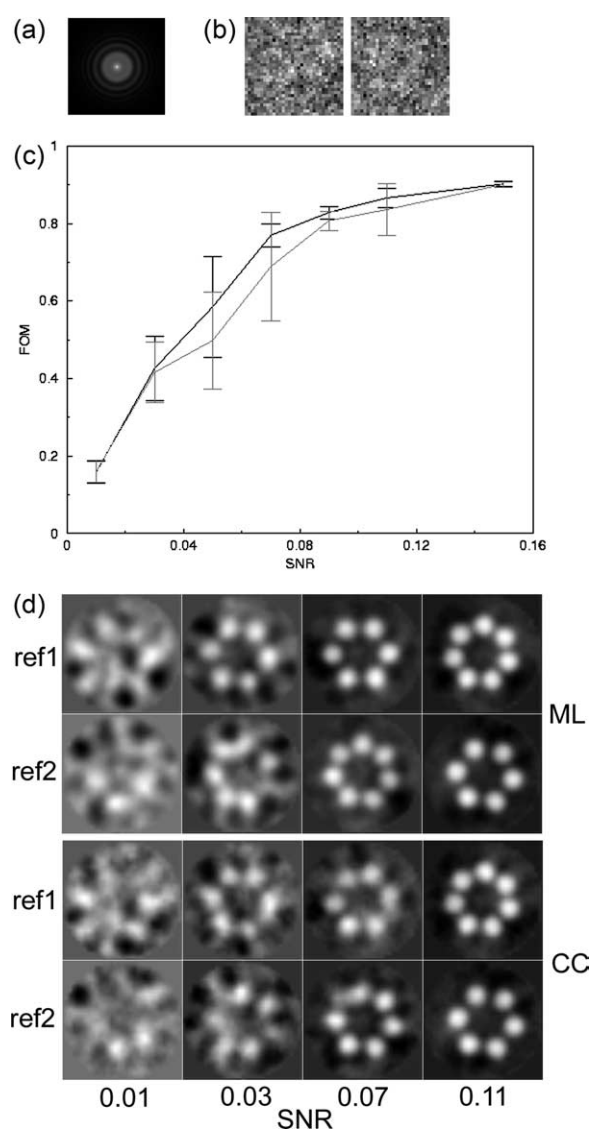
Figure 4. Application to filtered, phantom data. (a) Amplitudes of the CTF-filter that was applied to the phantom data. This filter was estimated for the presented cryo-electron microscopy dataset, using ARMA models and subsequent parametric modelling.[38] (b) Two examples of filtered phantom images with SNRs of 0.07. SNRs were calculated as the variance of the filtered signal divided by the variance of the filtered noise. (c) Average and standard deviation values of the FOMs (see main text) as obtained for six independent runs of ML (black) and CC (grey) two-reference refinements, using filtered phantom data with increasing SNRs. (d) Gallery of structures as obtained using ML and CC refinement with filtered data at various SNRs. Shown are both references (1 and 2) for the third best out of six ML refinement runs (i.e. close to the median result) and the same structures as resulting from CC refinement.

yielded reasonable to good solutions, while CC refinement failed. At even lower SNRs neither method was capable of reaching a satisfactory solution.

In a second experiment, we investigated the relative performance of ML and CC refinement for data where the assumption of independent noise no longer holds. To simulate data with noise characteristics similar to those of experimental data, we estimated the CTF for the above described cryo-electron microscopy dataset on SV40 large T-antigen (Figure 4(a)). The resulting filter was applied to the phantom data at various SNRs (see Figure 4(b)). For these filtered data, both ML and CC refinement required higher SNRs (i.e. >0.08) than for the unfiltered data (Figure 4(c) and (d)). Furthermore, the observed superiority of ML over CC for unfiltered data was reduced, although ML refinement still yielded better results in the SNR-range between 0.03 and 0.09. Again, smoother background densities were observed for the ML-refined images compared to the CC-refined ones, but also this difference was not as large as for the unfiltered data.

Discussion

ML and CC multi-reference refinement

We introduced a maximum-likelihood approach to multi-reference refinement. This approach aims at finding the most likely model that describes the experimental data. This model not only includes the multiple structures to be refined, but also a formal description of the noise and of the distributions of the alignment and classification parameters. The corresponding target function was derived from the single-reference case as presented previously.[20] Whereas, steepest descent optimization was applied in the single-reference case, we optimize our target function by means of the expectation-maximization algorithm,[21] which also addresses the estimation of the distribution of the different models as implied by the experimental data.

We compared the ML approach with conventional CC multi-reference refinement using a maximum shift criterion in the translational search. An alternative way of including prior knowledge about the origin offsets has been proposed for the single-reference case and involves introducing a bias towards the estimated probability density function of the translations.[20] We also tested this alternative approach for the multi-reference refinements presented here (results not shown). For both the phantom and the negative stain data, this yielded practically identical results as obtained using a maximum shift criterion. However, for the cryo-electron microscopy data significant improvement over the maximum shift criterion was obtained, although the results were still inferior to those obtained using ML. This indicates that part of the advantage of the ML approach lies in the way

SNRs of 0.07 and higher, the background of the ML-refined images was cleaner than that of the CC-refined images. The latter was also reflected in a somewhat higher FOM for these ML-refined images. Much larger differences were observed at lower SNRs. In the range of 0.01–0.03, ML still

how prior information about the origin offsets is handled.

Application to experimental data

Application of ML multi-reference refinement to cryo-electron microscopy data revealed biologically relevant differences among large T-antigen projections that were not observed using conventional alignment and classification tools. The two major classes of a ML five-reference refinement showed strongly bent and (almost) straight projections of dodecameric molecules. Relatively large differences were observed between the two hexameric subunits of the straight projection, while the two hexamers of the bent projection appeared very similar. Considering that these images are projections of molecules with hexagonal symmetry, the different appearance of the subunits in the straight projection may be interpreted as a different rotation in the plane of the image. Such a rotational twist of the hexameric subunits along the central axis of the molecule was suggested previously to play a role in the so-called "iris"-mechanism of DNA unwinding.[29] Former 3D-electron microscopy experiments on large T-antigen dodecamers[24] also suggested a rotational mismatch between the hexameric subunits, but never with such clarity as observed in the ML-refined projections. Furthermore, the strongly bent ML-refined projections showed additional density that may correspond to the complexed double-stranded DNA. The appearance of this weak density, protruding from the C-terminal side of one of the hexameric subunits, is in agreement with the expectations based on the asymmetric design of the DNA probe. In addition, it agrees with the model for DNA-interaction as based on previous electron microscopy experiments[30] and the crystal structure of the hexameric C-terminal domain of large T-antigen.[29] CC refinement of the same seeds yielded images that were more fuzzy and more similar to each other in terms of curvature. This suggests that the alignment and separation into different classes was not as successful as in the ML refinement and that the resulting images are thus averages over sub-optimally aligned, different projections.

The higher SNR of the presented negative stain dataset allowed analysis by conventional classification tools. ML refinement of three references yielded similar classes as obtained using self-organizing maps of rotational spectra, validating the novel approach. However, practically identical results were obtained using CC refinement. This indicates that for these data the two approaches may be nearly equivalent, as is expected from the theory for data with low levels of noise. That this may indeed be the case is strengthened by the observation that even for non-relevant background densities similar features were obtained with both methods.

Since the refinement approaches presented are based on expectation maximization protocols, both ML and CC multi-reference refinement will converge towards the nearest local maximum in their target function.[21] Therefore, the starting seeds of the refinement may play an important role in their convergence. In this context, an interesting observation was made when using average images of pre-aligned subsets of the experimental images as initial seeds for the refinement process (results not shown). Although pre-alignment of the images may seem more intuitive than using the averages of unaligned subsets, both ML and CC refinements performed significantly worse when starting from pre-aligned models. In this case, for the cryo and for the negative stain data, the refinement of different references yielded rather similar models that did not reflect the structural heterogeneity in the data. This indicates that pre-alignment may introduce a bias in the initial seeds that cannot be easily removed by either of the two methods. For CC multi-reference refinement similar observations have been reported previously,[16,17] and apparently also ML refinement may suffer from bias in the initial models. However, the results presented here show that better results are obtained when using completely unbiased seeds, which can be obtained by calculating the average image for random subsets of the unaligned images. In this context we also note the work of Brink *et al.*,[31] who avoided bias in the initial seeds of a 3D multi-reference refinement by application of random perturbations to one common volume.

Application to simulated data and future implications

To assess the potentials of ML multi-reference refinement in more detail, we tested the developed methodology also on simulated data with different amounts of noise. For data in agreement with its assumptions of independent, zero-mean Gaussian noise, ML multi-reference refinement is clearly superior to the CC approach. A different picture was observed for data where the assumption of independent noise is no longer met. Application of a CTF-filter, which was estimated for the presented cryo-electron microscopy dataset, yielded phantom data with realistic correlations between the pixel values. In this case, ML still performed somewhat better than CC refinement, although the benefits of ML refinement with filtered data were much smaller than with the unfiltered data. Since the noise in experimental images is typically filtered, further enhancement of the efficiency of ML refinement is expected from the introduction of an appropriate model for correlated noise. We note that in the case of correlated noise the assumption of independency between the probability density functions of the individual pixels no longer holds, and that the multiplication of these probabilities in equation (3) is thus no longer valid. As for the current ML approach, its benefits for experimental data are expected to be largest when the correlations between nearby pixel values are small. This

is in particular the case when close-to-focus images are recorded.

Further use in the field may demonstrate the effectiveness of the approaches presented for different experimental conditions. Therefore, the ML as well as the CC protocols have been implemented in the publicly available image processing package Xmipp†,[32] ensuring easy access for the electron microscopy community. We note that, besides the importance of the initial references as described above, the number of references to be refined is an important variable in these calculations. This number should be sufficiently high to reflect the structural heterogeneity in the data, but not too high, in order to avoid potential refinement of artefacts. In practice, most cases will require various runs with different numbers of references, as were also performed for the presented case on large T-antigen. Other model parameters, describing the noise, the origin offsets and the distribution of different models among the experimental data, are determined during the optimization process. This self-learning character and the intrinsic combination of alignment and classification in a single iterative process make these multi-reference refinement protocols promising candidates for (semi-) automated approaches to the 2D-analysis of structurally heterogeneous data.[3]

Finally, we observe that different views of a 3D-structure may be considered as multiple references in 2D. Therefore, the methodology presented may be applied to generate class averages of different views in an angular reconstitution approach to 3D-reconstruction.[33,34] What is more, by extension of the theory presented here, the benefits of 2D-ML refinement may be carried over to the refinement of (multiple) 3D-structures. Together with an incorporation of correlated noise in the ML formulation, this might greatly enhance the quality of electron microscopy structure analysis, which is a topic for future research.

Materials and Methods

Large T-antigen purification and electron microscopy

Large T-antigen was expressed and purified as described.[35] A dsDNA fragment comprising the origin of replication of SV40 (*Ori*) was produced by PCR amplification from two primers (oligo-EcoRI 5′-GAATT CCCGGGGGATCCGGTCGAC-3′ and oligo-HindIII 5′-AA GCTTTCTCACTACTTCTGGAATAGC-3′), using the pOR1 plasmid as a template.[36] The resulting dsDNA probe was 107 bp long and included the *Ori*, as well as an extension of 23 bp at the AT-rich region side. Double-hexamer complexes, assembled on the *Ori*, were formed as reported previously.[24] After vitrification of the specimen on carbon-coated grids, images were collected in a Tecnai G^2 field emission gun microscope at a magnification of 62,000× and an accelerating voltage of 200 kV. Micrographs were digitized on a Zeiss-Integraph

(Photoscan TD model) scanner with a 2.2 Å step size. Down-sampling by linear interpolation yielded a final pixel size of 4.4 Å. For each micrograph, the CTF was estimated and corrected by flipping the sign of the Fourier components. A total number of 7590 particles was selected and extracted from these micrographs.

G40P purification and electron microscopy

Bacteriophage SPP1 G40P was expressed and purified as described.[37] Glycerol removal was performed by dialysis against a buffer containing 50 mM Hepes (pH 7.0), 50 mM NaCl, 50 mM MgCl$_2$, 1 mM ATP and 1 mM dithiothreitol. Subsequently, a 0.7 μM G40P solution was adsorbed onto carbon-coated holey grids previously glow-discharged, stained with 2% (w/v) uranyl acetate, and visualized in a Jeol 1200 EX-II transmission electron microscope at 60,000× magnification and an accelerating voltage of 80 kV. Micrographs were digitized in a Zeiss-Integraph (Photoscan TD model) scanner at a pixel size of 4.1 Å. A total number of 13,475 particles was selected and extracted from the micrographs. A self-organizing map algorithm (KerDenSom[25]) was applied to analyse the variability among the rotational power spectra[28] of all translationally aligned particles. Discarding all particles without apparent three or 6-fold symmetry, a final population of 2120 top views was selected. This test set comprised all three quaternary architectures as observed previously for G40P[26]: 589 particles with C3, 740 particles with C3C6, and 791 particles with *C*6 symmetry.

Multi-reference refinement schemes

All subsequent image operations were performed using the image processing package Xmipp,[32] in which both ML and CC multi-reference refinement protocols were implemented. ML refinement was carried out by iterative evaluation of equations (7)–(10); CC multi-reference refinement[9] was performed using a complete search of all rotations and all translations up to a maximum allowed shift. Rotations were performed using bilinear interpolation and rotation angles were sampled every five degrees. Fast Fourier transforms were employed to calculate cross-correlations, resulting in periodic boundary conditions for translations, being sampled every pixel. Prior to multi-reference refinement, all images were normalized to average zero and standard deviation one. Given this normalization, initial estimates for the standard deviation of the noise in the ML approach were set to one. Initial values for α_k were set to correspond to uniform distributions of the different models. Initial estimates for the standard deviation in the origin offsets were set to two pixels for the G40P and phantom data, while for the LTA-data a value of ten pixels was used to reflect an increased uncertainty in particle picking due to the high levels of noise. The maximum allowed shifts in the CC refinements were set to 20 pixels for the LTA data, ten pixels for the G40P data, and six pixels for the phantom data. Initial reference images were calculated in a bias-free way by calculating the average of equally-sized, random subsets of the unaligned images. Refinements were stopped when, for all references, the average squared intensity differences resulting in a single iteration dropped below 0.005% of the average squared intensity of the image.

The CC algorithm was implemented striving for maximum resemblance to the ML protocol. This resulted in an implementation that is sub-optimal in terms of computing speed, requiring almost as much CPU as the

ML implementation. Required computing times for ML refinement depend on the quality of the images, and are of order $O(N+K)$ for N experimental images and K references, and of order $O(\sqrt{M} \log \sqrt{M})$ for M pixels. For the 2120 images of 80×80 pixels of the negative stain dataset, ML refinement of three references took approximately 1.5 hours/iteration on a 1 GHz Alpha processor. For the 7590 images of 90×90 pixels of the cryo-electron microscopy dataset, ML refinement of five references took approximately 14 hours/iteration. Part of these calculations was performed in a parallelized fashion. (Parallelization is straightforward, since within a single iteration the probability calculations of different experimental images are independent.)

References

1. Joyeux, L. & Penczek, P. A. (2002). Efficiency of 2D alignment methods. *Ultramicroscopy*, **92**, 33–46.
2. Bonnet, N. (2000). Artificial intelligence and pattern recognition techniques in microscope image processing and analysis. *Advan. Imaging Electron Phys.* **114**, 1–77.
3. Bonnet, N. (2004). Some trends in microscope image processing. *Micron*, **35**, 635–653.
4. Schatz, M. & van Heel, M. (1990). Invariant classification of molecular views in electron micrographs. *Ultramicroscopy*, **32**, 255–264.
5. Schatz, M. & van Heel, M. (1992). Invariant recognition of molecular projections in vitreous ice preparations. *Ultramicroscopy*, **45**, 15–22.
6. Harauz, G., Boekema, E. & van Heel, M. (1988). Statistical image analysis of electron micrographs of ribosomal subunits. *Methods Enzymol.* **164**, 35–49.
7. Marabini, R. & Carazo, J. M. (1994). Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys. J.* **66**, 1804–1814.
8. Heymann, J. B., Conway, J. F. & Steven, A. C. (2004). Molecular dynamics of protein complexes from four-dimensional cryo-electron microscopy. *J. Struct. Biol.* **147**, 291–301.
9. van Heel, M. & Stoffler-Meilicke, M. (1985).

Characteristic views of *E. coli* and *B. stearothermophilus* 30 S ribosomal subunits in the electron microscope. *EMBO J.* **4**, 2389–2395.
10. Unwin, P. N. & Klug, A. (1974). Electron microscopy of the stacked disk aggregate of tobacco mosaic virus protein. I. Three-dimensional image reconstruction. *J. Mol. Biol.* **87**, 641–656.
11. Salzman, D. B. (1990). A method of general moments for orienting 2D projections of unknown 3D objects. *Comput. Vision Graphics Image Process.* **50**, 129–156.
12. Marabini, R. & Carazo, J. M. (1994). Practical issues on invariant image averaging using the Bispectrum. *Signal Process.* **40**, 119–128.
13. Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *J. Mol. Biol.* **277**, 1033–1046.
14. Stewart, A. & Grigorieff, N. (2004). Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*, **102**, 67–84.
15. Sorzano, C. O., Jonic, S., El-Bez, C., Carazo, J. M., De Carlo, S., Thevenaz, P. & Unser, M. (2004). A multi-resolution approach to orientation assignment in 3D electron microscopy of single particles. *J. Struct. Biol.* **146**, 381–392.
16. Boekema, E. J., Berden, J. A. & van Heel, M. G. (1986). Structure of mitochondrial F1-ATPase studied by electron microscopy and image processing. *Biochim. Biophys. Acta*, **851**, 353–360.
17. Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* **12**, 1303–1309.
18. Rice, J. A. (1995). *Mathematical Statistics and Data Analysis*, Duxbury Press, Belmost, CA.
19. Brunger, A. T., Adams, P. D. & Rice, L. M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.* **8**, 606–611.
20. Sigworth, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122**, 328–339.
21. Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum-likelihood from incomplete data *via* the EM algorithm. *J. Roy. Stat. Soc. ser. B.* **39**, 1–38.
22. VanLoock, M. S., Alexandrov, A., Yu, X., Cozzarelli, N. R. & Egelman, E. H. (2002). SV40 large T antigen hexamer structure: domain organization and DNA-induced conformational changes. *Curr. Biol.* **12**, 472–476.
23. Valle, M., Gruss, C., Halmer, L., Carazo, J. M. & Donate, L. E. (2000). Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol. Cell. Biol.* **20**, 34–41.
24. Gomez-Lorenzo, M. G., Valle, M., Frank, J., Gruss, C., Sorzano, C. O., Chen, X. S. *et al.* (2003). Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *EMBO J.* **22**, 6205–6213.
25. Pascual-Montano, A., Donate, L. E., Valle, M., Barcena, M., Pascual-Marqui, R. D. & Carazo, J. M. (2001). A novel neural network technique for analysis and classification of EM single-particle images. *J. Struct. Biol.* **133**, 233–245.
26. Barcena, M., San Martin, C., Weise, F., Ayora, S., Alonso, J. C. & Carazo, J. M. (1998). Polymorphic quaternary organization of the *Bacillus subtilis* bacteriophage SPP1 replicative helicase (G40 P). *J. Mol. Biol.* **283**, 809–819.
27. Pascual, A., Barcena, M., Merelo, J. J. & Carazo, J. M.

(2000). Mapping and fuzzy classification of macromolecular images using self-organizing neural networks. *Ultramicroscopy,* **84**, 85–99.

28. Crowther, R. A. & Amos, L. A. (1971). Harmonic analysis of electron microscope images with rotational symmetry. *J. Mol. Biol.* **60**, 123–126.

29. Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J. A. *et al.* (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature,* **423**, 512–518.

30. Mastrangelo, I. A., Hough, P. V., Wall, J. S., Dodson, M., Dean, F. B. & Hurwitz, J. (1989). ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. *Nature,* **338**, 658–662.

31. Brink, J., Ludtke, S. J., Kong, Y., Wakil, S. J., Ma, J. & Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure (Camb),* **12**, 185–191.

32. Sorzano, C. O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H., Carazo, J. M. & Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204.

33. Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy,* **21**, 111–123.

34. Penczek, P. A., Zhu, J. & Frank, J. (1996). A common-lines based method for determining orientations for $N>3$ particle projections simultaneously. *Ultramicroscopy,* **63**, 205–218.

35. Simanis, V. & Lane, D. P. (1985). An immunoaffinity purification procedure for SV40 large T antigen. *Virology,* **144**, 88–100.

36. DeLucia, A. L., Deb, S., Partin, K. & Tegtmeyer, P. (1986). Functional interactions of the simian virus 40 core origin of replication with flanking regulatory sequences. *J. Virol.* **57**, 138–144.

37. Ayora, S., Weise, F., Mesa, P., Stasiak, A. & Alonso, J. C. (2002). *Bacillus subtilis* bacteriophage SPP1 hexameric DNA helicase, G40P, interacts with forked DNA. *Nucl. Acids Res.* **30**, 2280–2289.

38. Velazquez-Muriel, J. A., Sorzano, C. O., Fernandez, J. J. & Carazo, J. M. (2003). A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment. *Ultramicroscopy,* **96**, 17–35.