

Published in final edited form as:

*J Struct Biol.* 2014 March ; 185(3): 295–302. doi:10.1016/j.jsb.2014.01.004.

## Automatic cryo-EM particle selection for membrane proteins in spherical liposomes

Yunhui Liu<sup>a,b</sup> and Fred J. Sigworth<sup>a,\*</sup>

<sup>a</sup>Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT 06520, USA

<sup>b</sup>School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China

### Abstract

Random spherically constrained (RSC) single particle reconstruction is a method to obtain structures of membrane proteins embedded in lipid vesicles (liposomes). As in all single-particle cryo-EM methods, structure determination is greatly aided by reliable detection of protein “particles” in micrographs. After fitting and subtraction of the membrane density from a micrograph, normalized cross-correlation (NCC) and estimates of the particle signal amplitude are used to detect particles, using as references the projections of a 3D model. At each pixel position, the NCC is computed with only those references that are allowed by the geometric constraint of the particle’s embedding in the spherical vesicle membrane. We describe an efficient algorithm for computing this position-dependent correlation, and demonstrate its application to selection of membrane-protein particles, GluA2 glutamate receptors, which present very different views from different projection directions.

### Keywords

cross-correlation; template matching; RSC reconstruction; glutamate receptor; particle picker

## 1. Introduction

Random Spherically Constrained (RSC) single particle reconstruction (Jiang et al., 2001; Wang and Sigworth, 2010) is a useful approach for studying the structures of membrane proteins in membrane environments. The proteins are reconstituted into spherical lipid vesicles (liposomes), which in turn are frozen in vitreous ice and imaged by electron cryomicroscopy (cryo-EM). As in all single-particle methods, the automatic “picking”, that is selection, of protein particles in micrographs is an important first step in three-dimensional structure determination. The embedding in spherical lipid vesicles presents both disadvantages and advantages. Vesicles are large (~30 nm diameter) but ideally have only one or two protein particles embedded each one, yielding a low overall density of particles in a micrograph. This low density has made the problem of particle selection particularly important, as in our experience a typical 4k × 4k pixel micrograph contains only about 50

© 2014 Elsevier Inc. All rights reserved.

\*Corresponding author: Fred.Sigworth@yale.edu (Fred J. Sigworth).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

particles. This means that many micrographs must be taken and processed to obtain large datasets. Further, the low density of particles means that the relative density of “non-particles” such as frost balls or other contaminants is high. A further disadvantage is that, to avoid distortion of vesicles, the ice layer must be relatively thick, larger than the diameter of the vesicles; thus the high contrast that is observed with very thin ice layers in cryo-EM specimens cannot be exploited.

On the other hand, an advantage arises from the fact that no particles exist in regions between vesicles, and at a given position relative to the center of a vesicle an embedded particle presents a restricted set of views. We sought to design a particle-picking algorithm that efficiently exploits these geometric constraints.

A popular approach to automatic particle picking involves the search of a micrograph using one or more templates (reference images) using the local correlation function (Roseman, 2003; Zhu et al., 2004; Voss et al., 2009) sometimes combined with other criteria for improved detection (Chen and Grigorieff, 2007; Langlois et al., 2011). The local correlation function is also known in the image-processing literature as the normalized cross-correlation (NCC). The program described here computes the NCC, and another selection criterion derived from the NCC, using reference images that are projections of an initial 3D model for the particles to be discovered. Due to the geometric constraint of embedding a protein particle in a spherical vesicle, two of the three Euler angles of orientation of a particle can be determined, within a fourfold ambiguity, from the position of the particle’s projection in the micrograph. Here we adapt an eigenimage-expansion strategy for multiple-template searching (Sigworth, 2004) to exploit these constraints.

As a test case we consider the GluA2 glutamate receptor, a 400 kDa membrane protein. We simulate images using its known X-ray crystal structure (Sobolevsky et al., 2009), and search for particles in cryo-EM images of GluA2 reconstituted into lipid vesicles.

The following notation is used in this paper. An upper-case variable denotes a matrix or an image; a bold-face variable denotes a vector, while lower-case variable denotes a scalar. The squared norm or “power” of an image  $X$  having  $n_p$  pixels with values  $x_i$  is defined as

$$|X|^2 = \sum_{i=1}^{n_p} x_i^2,$$

and  $C \circ X$  denotes operator  $C$  operating on  $X$ .

## 2. Methods

### 2.1. Spherically-constrained projection geometry

We denote the position of a particle on a spherical vesicle by angles  $\alpha$ ,  $\beta$ ,  $\gamma$  (Fig. 1) which, unlike the standard definition of Euler angles for electron microscopy (Heymann et al., 2005) are defined as three successive rotations of the *particle* rather than rotations of the coordinate system. The projection direction is the  $-\mathbf{z}$  axis and the projection plane is the  $(\mathbf{x}, \mathbf{y})$  plane. The definition of the angles is different from the one we used previously (Jiang et al., 2001) and is designed such that a projection image of a particle with  $\alpha = 0$  has the membrane-protein particle oriented along the  $\mathbf{y}$ -axis, as is customary in the depiction of membrane-protein structures. The two angles  $\alpha$  and  $\beta$  can be determined from the geometrical relationship between the particle and the vesicle. Suppose in the projection image the distance from the vesicle center to the particle’s center of mass is  $r$  and the

particle's center of mass is at location  $(x, y)$ ; then the two angles can be described by  $x = r\sin\alpha$ ,  $y = r\cos\alpha$ ,  $r = (r_{\text{ves}} \pm \delta)\sin\beta$  where  $r_{\text{ves}}$  is the radius of the three-dimensional vesicle (the distance from the vesicle center to the middle of the lipid bilayer membrane) and  $\delta$  is the radial distance between the particle's center of mass and the middle of the bilayer. The two possible signs for  $\delta$  denote two incorporation directions of the membrane-protein particle, inside-out or right-side out.

A particle may also be embedded in the upper or lower hemisphere of the vesicle. In Fig. 1 we could say that the particle is embedded right-side out and in the upper hemisphere. Rotations about the membrane normal, through the angle  $\gamma$ , represent the final degree of freedom of the particle. Fig. 2 illustrates projections of the GluA2 glutamate receptor, which has C2 symmetry.

## 2.2. Vesicle membrane subtraction

The RSC method requires that the vesicle-membrane density be subtracted. This is necessary since the vesicles do not all have the same radius, precluding the reconstruction of the entire vesicle with the protein particle. Further, the high density of the vesicle "edge" falls on different parts of the particle image, depending on the particle's  $\beta$  angle, so it is essential that the density be removed. The processing of a micrograph starts with the determination of the center and radius of each of the perhaps 100 vesicles it contains. With a micrograph downsampled to about 1 nm/pixel, the initial vesicle-finding is done using a simple NCC search with a set of model images of vesicles, typically having radii of 12 to 25 nm with a step size of 1 nm. The step size is chosen to be small compared to the thickness of the vesicle membrane, which is about 5 nm. The position and amplitude of the peaks of the NCC yield an approximate least-squares fit of the vesicle model, which is then refined by a nonlinear least-squares fit of position and radius, one vesicle at a time, to a scalable model (Wang et al., 2006). The end results are the center and radius of each vesicle, determined with  $\sim 0.1$  nm precision, and an estimate of the contrast in the image of a vesicle. Fig. 2 shows the simulated image of a vesicle, along with images of protein particles as would be obtained after membrane subtraction. It is the membrane-subtracted micrograph which is searched for protein particles.

## 2.3. Reduced reference set and matched filter

For simplicity let us consider the problem of identifying a single membrane-protein particle in a micrograph that contains one vesicle, after the vesicle-membrane density has been subtracted from the image. This image  $X$  can be modeled from the scaled, noiseless particle image  $S_j$ :

$$X(\mathbf{x}) = aS_j(\mathbf{x} - \mathbf{x}_0) + N \quad (1)$$

where  $a$  is the image amplitude,  $\mathbf{x}_0$  is the centered position, and  $N$  is independent, zero-mean Gaussian noise.

The underlying particle image  $S_j$ , is obtained as:

$$S_j = W \circ C \circ P_j \circ M \quad (2)$$

where  $M$  is the 3D model density,  $P_j$  is operator for projection along the  $j^{\text{th}}$  projection direction,  $C$  is the contrast transfer function (CTF) and  $W$  is a pre-whitening filter (Huang and Penczek, 2004; Sigworth, 2004). The application of a pre-whitening filter to an image makes its noise approximately white and therefore better corresponds to the model (1). The use of a pre-whitening filter is not common in cryo-EM image processing, but the common

practice of high-pass filtering of cryo-EM images has a similar effect, dampening the excess low-frequency noise that is found in images.

We assume that a set of these underlying images,  $S_k$ ;  $k = 1..n$  constitute the set of references that we will use for searching for a particle. To find a particle we first calculate, at pixel location  $\mathbf{x}$ , the cross-correlation of the image  $X$  with a reference  $S_k$

$$R_k(\mathbf{x}) = \sum_{\mathbf{t}} S_k(\mathbf{t}) \cdot X(\mathbf{t} + \mathbf{x}) \quad (3)$$

where the sum is taken over all pixel locations  $\mathbf{t}$  in the reference image. We will simplify this calculation by making an eigenimage expansion of the references, as in (Sigworth, 2004). Each of the  $S_k$  is approximated by a linear combination of  $m$  orthonormal eigenimages  $U_i$  obtained by singular value decomposition,

$$\tilde{S}_k = \sum_{i=1}^m U_i b_{ik} \quad (4)$$

where  $b_{ik}$  are linear combination coefficients.

Fig. 3A illustrates the first 20 eigenimages computed from a set of almost 4000 references. Successive rows of Fig. 3B compare a few of the approximated images  $\tilde{S}_k$  with the true references  $S_k$  when the expansion uses  $m = 20$  terms. As shown in Eq. (15) in (Sigworth, 2004), the squared peak-to-noise ratio of the output of a correlation detector is proportional to  $\sum_{i=1}^m b_{ik}^2$  which rapidly approaches its maximum value as  $m$  increases. Fig. 3C shows that the choice  $m = 20$  yields at least 80% of the maximum value for each of the references; we take this value to be sufficient.

We can calculate the cross-correlations of the image  $X$  with eigenimages  $U_i$ :

$$Q_i(\mathbf{x}) = \sum_{\mathbf{t}} U_i(\mathbf{t}) X(\mathbf{t} + \mathbf{x}) \quad (5)$$

and with these, for any single pixel position  $\mathbf{x}_0$ , the cross-correlation of image  $X$  with a reference  $S_k$  can be approximated by

$$\tilde{R}_k(\mathbf{x}_0) = \sum_{i=1}^m Q_i(\mathbf{x}_0) b_{ik} \quad (6)$$

where the  $b_{ik}$  are the same coefficients as in Eq. (4).

In our particle-picking algorithm we consider each vesicle separately, and search for particles that may be embedded in that vesicle. The geometric constraint of a particle embedded in a spherical vesicle means that, at a given position in the micrograph, we do not need to search with all possible particle orientations, say 4000 of them; instead, we need consider only a small subset, say 32 orientations. The challenge is to efficiently compute the cross-correlation with the appropriate small subsets of orientations.

To impose the geometric constraint of particle embedding, for each pixel position  $\mathbf{x}_0$  in the neighborhood of the vesicle we choose references  $S_k$  where

$$k \in \kappa_l(\mathbf{x}_0) \quad (7)$$

that is, for this pixel position  $\mathbf{x}_0$ , there is a set of allowable indices  $\kappa$  which is indexed by  $l$ . The indices  $\kappa_l(\mathbf{x}_0)$  refer to references with  $\alpha$  and  $\beta$  appropriate to that position in the image, but different  $l$  values identifying different particle orientations: right-side out or inside-out, upper or lower hemisphere, and various  $\gamma$  values. If there are  $n_\gamma$  different  $\gamma$  values to be tested, then there are  $n_l = 4n_\gamma$  references that are appropriate to a given position in the projection plane. Fig. 4A is an example of the position-dependent lookup table for  $l = 1$  that gives the indices  $\kappa_1(\mathbf{x}_0)$  in the example used here.

The set of position-dependent coefficients can be written as a matrix at each pixel position having the elements

$$B_{il}(\mathbf{x}_0) = b_{i\kappa_l(\mathbf{x}_0)} \quad (8)$$

and using this notation the cross-correlation  $\tilde{R}_l^c(\mathbf{x}_0) \equiv \tilde{R}_{\kappa_l(\mathbf{x}_0)}(\mathbf{x}_0)$ , obtained from Eq. (6) but constrained to the relevant references at each pixel position, is given by

$$\tilde{R}_l^c(\mathbf{x}_0) = \sum_{i=1}^m Q_i(\mathbf{x}_0) B_{il}(\mathbf{x}_0). \quad (9)$$

Thus a vector containing the  $n_l$  values of the cross-correlation with the relevant references, at a given pixel position  $\mathbf{x}_0$ , can be evaluated by the product of a vector  $Q(\mathbf{x}_0)$  and a matrix  $B(\mathbf{x}_0)$ . This is the way in which the cross-correlation can be computed against references that vary with pixel position, and is the key for the efficient search for particles under the geometric constraints.

#### 2.4. Normalized cross-correlation

Returning to Eq. (4) for approximating the  $k^{\text{th}}$  reference image, the orthonormality of the eigenimages  $U_i$  implies that the power in a reference image is given by

$$\tilde{\sigma}_k^2 = |\tilde{S}_k|^2 = \sum_i b_{ik}^2 \quad (10)$$

or, in view of the geometric constraint, we write it as  $\tilde{\sigma}_l^2(\mathbf{x}_0)$ .

The normalized cross-correlation (NCC) of an image with the references can then be computed as

$$\tilde{N}_l(\mathbf{x}_0) = \frac{\tilde{R}_l^c(\mathbf{x}_0)}{\tilde{\sigma}_l(\mathbf{x}_0) \sigma_X(\mathbf{x}_0)} \quad (11)$$

where  $\sigma_X$  is the standard deviation of the image  $X$  in the vicinity of  $\mathbf{x}_0$ . At each position, we find the best index  $l = \lambda(\mathbf{x}_0)$  by maximizing the NCC:

$$\lambda(\mathbf{x}_0) = \arg \max_l \tilde{N}_l(\mathbf{x}_0). \quad (12)$$

Thus  $\lambda(\mathbf{x}_0)$  gives the index of the reference that best matches the image at the position  $\mathbf{x}_0$ . Its determination does not depend on the particular value of  $\sigma_X(\mathbf{x}_0)$ .

## 2.5. Signal amplitude as a selection criterion

Many particle-selection algorithms use the values of the NCC to distinguish particles from non-particles in a micrograph. A problem with the NCC is that the local standard deviation  $\sigma_X(\mathbf{x}_0)$  of the image must be estimated, but the magnitude of  $\sigma_X$  depends on the degree of filtering of the image. Thus determining  $\sigma_X$  requires an arbitrary choice of filtering. We have chosen instead to estimate the value of the scaling amplitude  $a$  in the image model (Eq. (1)) to identify valid particles.

An absolute calibration of image contrast is available from the vesicle images. Because the composition of a lipid bilayer membrane is known, a quantitative model of its three-dimensional inner potential  $V(\mathbf{r})$  can be computed (Wang et al., 2006), and the weak-phase approximation allows the contrast in the image of a vesicle to be computed as

$$S_{\text{ves}} = a_{\text{ves}} C \circ P_z \circ V(\mathbf{r}) \quad (13)$$

where  $P_z$  is the projection operator,  $C$  is the contrast-transfer function, and the amplitude factor  $a_{\text{ves}}$  describes the ratio of image contrast to integrated inner potential. In ideal weak-phase imaging  $a_{\text{ves}} = 2\sigma_e$  where the interaction parameter  $\sigma_e$  is the electron phase shift due to the integral of the inner potential along the electron path; for 200 keV electrons  $\sigma_e = 7.3$  mrad/volt-nm (Kirkland, 1998). Fitted values of  $a_{\text{ves}}$  are typically near the theoretical values, but vary by 20–30% over a micrograph.

On the basis of the image contrast  $a_{\text{ves}}$  and a model for a membrane-protein particle, we can predict the value of  $a$  for the particle. Specifically, if the 3D model density  $M$  in Eq. (2) is expressed in units of the inner potential in volts, then  $a$  should be equal to  $a_{\text{ves}}$ . Meanwhile, from the micrograph an estimate of  $a$  for a putative particle is obtained as

$$\tilde{a}(\mathbf{x}_0) = \frac{\sum_{\mathbf{t}} \tilde{S}_{\kappa_\lambda}(\mathbf{t}) X(\mathbf{t} + \mathbf{x}_0)}{\sum_{\mathbf{t}} \tilde{S}_{\kappa_\lambda}^2(\mathbf{t})} = \frac{\tilde{R}_\lambda^c(\mathbf{x}_0)}{\tilde{\sigma}_\lambda^2(\mathbf{x}_0)}. \quad (14)$$

Because each vesicle image provides a contrast reference, we can search for  $\tilde{a}(\mathbf{x}_0)$  in the proper range in the vicinity of each vesicle and use its value as a criterion in distinguishing particles from non-particles. The overall scheme for particle picking in the vicinity of one vesicle is diagrammed in Fig. 5.

## 2.6. Processing of an entire micrograph

The preceding sections have described the search for a particle embedded in a single vesicle. The processing of an entire micrograph starts with the determination of the center, radius and contrast of each of perhaps 100 vesicles. The fitted models of all vesicles of interest are then subtracted from the micrograph, as demonstrated in Fig. 6A–C.

A pre-whitening filter is then constructed for the subtracted micrograph. This is done by computing the average power spectrum for the lowest-variance regions of the micrograph, and the filter transfer function is computed to render the spectrum flat, as demonstrated in Fig. 6D and E. Meanwhile a large set of projections of the 3D search model is computed, and each projection is filtered by the combination of the pre-whitening filter and the CTF. By singular-value decomposition (SVD) the eigenimages are computed (Fig. 5B) along with

the coefficients  $b_{ik}$  as in Eq. (4). For a typical micrograph containing 100 vesicles the application of the CTF and calculation of the SVD takes only a few seconds, less computer time than the 100 cross-correlation searches. Therefore it is not expensive to accommodate images with differing CTFs by simply repeating these computation steps for each image.

At this point the vicinity of each vesicle position is searched in turn. The first step in the search is to resample the reference lookup table (Fig. 4A) according to the radius of the vesicle being searched. Then the procedure of Fig. 5C is carried out to compute the best reference assignment  $\lambda(\mathbf{x}_0)$  and amplitude  $\tilde{a}(\mathbf{x}_0)$  at each pixel position according to Eq. (12) and Eq. (14). All of these computations are performed in the batch preprocessing program `rsPickingPreprocessor` that takes about 0.1 s to search the vicinity of each vesicle. The actual detection of particles is subsequently carried out from stored values of  $\tilde{a}(\mathbf{x}_0)$  in the interactive program `SimpleRSPicker` where the user has control over the allowed range of  $\tilde{a}$  values. These and the other programs in our RSC image-processing pipeline are available as open source at [github.com/fsigworth](https://github.com/fsigworth).

## 2.7. Rejection of non-particles

The most common source of false-positive “particles” in our micrographs are small, approximately spherical ice balls. These are not prominent in conventional defocus-contrast images but appear as high-contrast objects in images taken with a phase plate, which preserves low-spatial-frequency information. We work with micrographs in which a low-defocus image is merged with one or more subsequent images taken with a high defocus up to about  $16\mu\text{m}$ . With such merged images we find that it is practical to further boost low-frequency information to flatten the CTF to frequencies below  $0.01\text{ nm}^{-1}$ , yielding an image very similar to a phase-plate image. Fig. 6D shows the transfer function of such a low-frequency boost filter, and the resulting contrast-transfer function. Fig. 6F shows the result of operation with this filter on a vesicle-subtracted micrograph. The ice balls appear as very dense objects that are readily detected. Using this image a user-set threshold identifies regions to exclude from the search for particles.

## 3. Results

### 3.1. Experimental and simulated data

As a test of the algorithm, we consider images of the tetrameric GluA2 glutamate receptor reconstituted into lipid vesicles containing the lipids POPC, DOPS and cholesterol in the ratio 8:1:1. Vesicles were adsorbed to a thin ( $\sim 5\text{nm}$ ) carbon film overlaying a holey-carbon film and plunge-frozen. Micrographs were taken on a Tecnai F20 microscope and recorded with a Gatan Ultrascan 4000 CCD camera at an effective pixel size of  $1.7\text{\AA}$ . The images shown here were merged from three exposures at defocus values of approximately 2, 5 and  $17\mu\text{m}$ . In the X-ray structure of this receptor (Sobolevsky et al., 2009) a competitive inhibitor is bound, and we find that the cryo-EM structure of the reconstituted, inhibitor-bound receptor is very similar and shows the same C2 symmetry.

To construct simulated images and reference images we first computed a 3D map  $M$  of the inner potential of the solvated glutamate receptor as described (Shang and Sigworth, 2012). The solvation layer was computed for the entire protein complex, including the transmembrane region. Within the protein boundary in the transmembrane region the density of a slab of the membrane model was additionally subtracted to model the effect of the vesicle membrane subtraction. To create simulated images from the projections of this map we added both shot noise and “specimen noise”, a noise component that is filtered by the CTF (Zeng et al., 2007; Baxter et al., 2009) and which in our case came mainly from the thin carbon support film. The contrast-transfer function that was employed in calculating

both the image and the specimen noise included as a factor the detective quantum efficiency of the CCD camera.

### 3.2. Particle picking

Fig. 7 shows results from the automatic particle picker. The top panels show the results from a simulated image under conditions where the false discovery rate was below 2% and the recall rate was greater than 98% in a set of 1000 simulated particles; see (Langlois and Frank, 2011) for the definitions of these metrics. The histogram in Fig. 7A<sub>5</sub> shows that the distribution of  $\tilde{a}$  for the true particles (black bars) forms a roughly normal distribution, while the frequency of finding spurious particles (white bars) increases steeply as the threshold value of  $\tilde{a}$  is reduced. At this noise level the population of true particles is almost completely separated from the false-positives. The separation increases further when the noise is reduced (Fig. 7A<sub>7</sub>).

The processing of an actual micrograph is demonstrated in Fig. 7B. All but one of the detected particles (boxes) visually matches the reference. The one questionable particle (more likely a small ice ball) that was detected is marked with an orange box. Although there is no “ground truth” in this case, the histogram of  $\tilde{a}$  values (Fig. 7B<sub>4</sub>) appears to show a separation of true and false particles like that seen in the simulation.

## 4. Discussion

We describe here a geometry-aware particle picker for lipid-vesicle-embedded membrane proteins. At each pixel position the particle-selection algorithm tests for the existence of a particle whose center of mass lies on that position. In this process only those reference images (projections of a 3D particle model) are considered that are appropriate to that pixel location in view of the constraint that the particle is oriented normal to the membrane surface. The expansion of each reference image in a set of basis images makes for the very efficient calculation (Eq. (9)) of the cross-correlation while using position-dependent sets of references.

The GluA2 particles, which are considered here as a test case, are particularly challenging for template-based particle detection because their extended shape presents very different 2D projections depending on the viewing direction. Indeed, the discrimination of true from false particles is much worse if a single average reference projection is used (compare Fig. 7A<sub>5</sub> and Fig. 7A<sub>7</sub>). This arises because one term is insufficient, while ~20 terms in the eigenimage expansion capture most of the power in the thousands of possible references (Fig. 3C). Using this expansion, and tested with simulated images and actual micrographs, the algorithm provides reliable detection of these 400 kDa membrane-protein particles in images taken with a conventional CCD camera.

## Acknowledgments

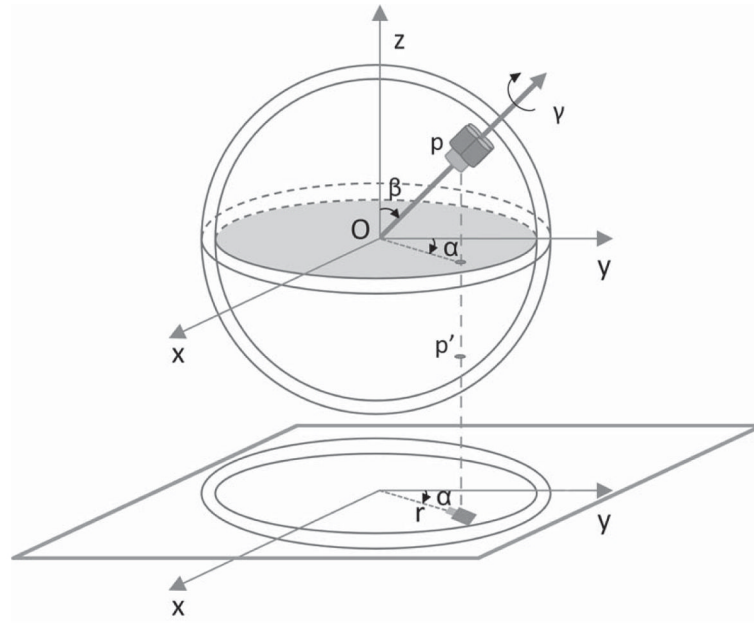
We thank K. Duerr and E. Gouaux (Vollum Institute, Portland, OR) for the purified GluA2 protein, Y-Y. Yan and Y. Yang for reconstituting it into lipid vesicles, and H. Shigematsu for cryo-EM imaging. We are grateful to H. Tagare and A. Barthel for helpful discussions. This work was supported by NIH grant R01NS21501 to F.S.

## References

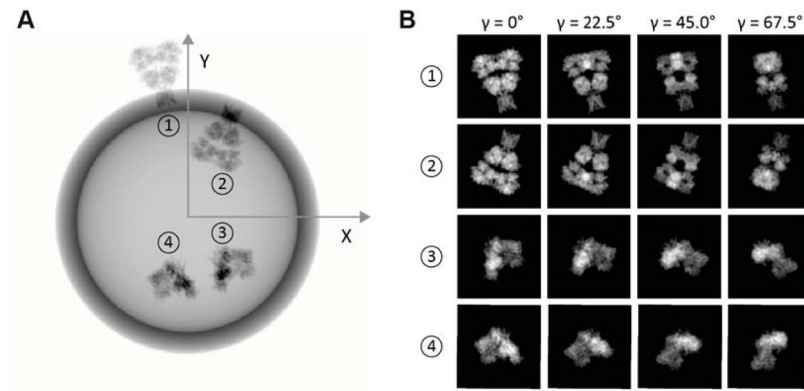
Jiang QX, Chester DW, Sigworth FJ. Spherical Reconstruction: A Method for Structure Determination of Membrane Proteins from Cryo-EM Images. *J Struct Biol.* 2001; 133(2–3):119–131.10.1006/jsbi.2001.4376 [PubMed: 11472084]



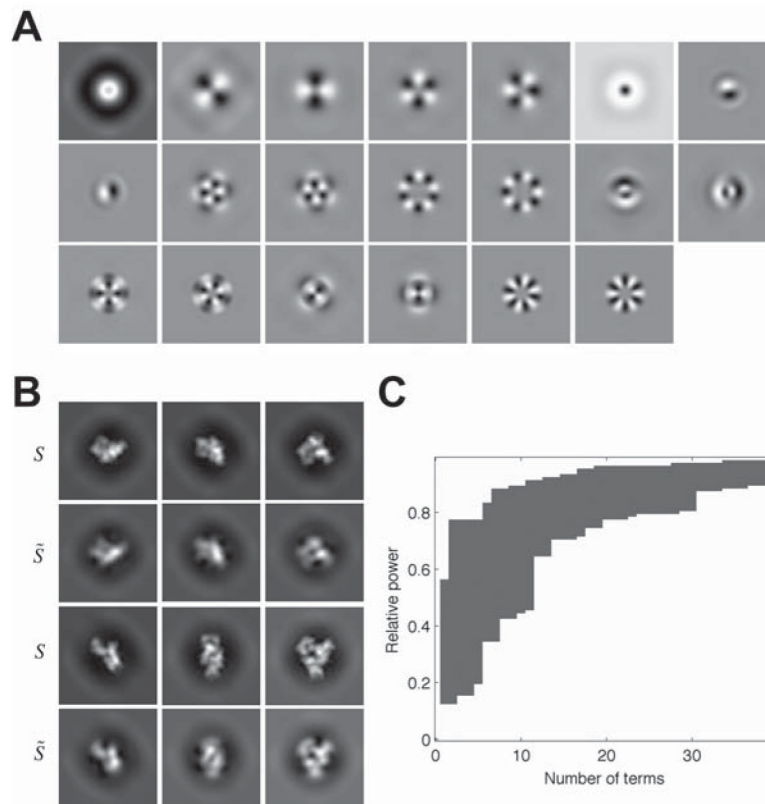
- Wang L, Sigworth FJ. Liposomes on a streptavidin crystal: a system to study membrane proteins by cryo-EM. *Methods Enzymol.* 2010; 481:147–164. [10.1016/S0076-6879\(10\)81007-9](https://doi.org/10.1016/S0076-6879(10)81007-9) [PubMed: 20887857]
- Roseman AM. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy.* 2003; 94:225–236. [http://dx.doi.org/10.1016/S0304-3991\(02\)00333-9](http://dx.doi.org/10.1016/S0304-3991(02)00333-9). [PubMed: 12524193]
- Zhu Y, Carragher B, Glaeser RM, Fellmann D, Bajaj C, Bern M, Mouche F, de Haas F, Hall RJ, Kriegman DJ, Ludtke SJ, Mallick SP, Penczek PA, Roseman AM, Sigworth FJ, Volkman N, Potter CS. Automatic particle selection: results of a comparative study. *J Struct Biol.* 2004; 145:3–14. <http://dx.doi.org/10.1016/j.jsb.2003.09.033>. [PubMed: 15065668]
- Voss N, Yoshioka C, Rademacher M, Potter C, Carragher B. DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol.* 2009; 166(2):205–213. <http://dx.doi.org/10.1016/j.jsb.2009.01.004>. [PubMed: 19374019]
- Chen JZ, Grigorieff N. SIGNATURE: A single-particle selection system for molecular electron microscopy. *J Struct Biol.* 2007; 157(1):168–173. <http://dx.doi.org/10.1016/j.jsb.2006.06.001>. [PubMed: 16870473]
- Langlois R, Pallesen J, Frank J. Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. *J Struct Biol.* 2011; 175:353–361. <http://dx.doi.org/10.1016/j.jsb.2011.06.004>. [PubMed: 21708269]
- Sigworth FJ. Classical detection theory and the cryo-EM particle selection problem. *J Struct Biol.* 2004; 145(1–2):111–122. [10.1016/j.jsb.2003.10.025](https://doi.org/10.1016/j.jsb.2003.10.025) [PubMed: 15065679]
- Sobolevsky AI, Rosconi MP, Gouaux E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature.* 2009; 462(7274):745–756. [10.1038/nature08624](https://doi.org/10.1038/nature08624) [PubMed: 19946266]
- Heymann JB, Chagoyen M, Belnap DM. Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J Struct Biol.* 2005; 151(2):196–207. [10.1016/j.jsb.2005.06.001](https://doi.org/10.1016/j.jsb.2005.06.001) [PubMed: 16043364]
- Wang L, Bose PS, Sigworth FJ. Using cryo-EM to measure the dipole potential of a lipid membrane. *Proc Natl Acad Sci U S A.* 2006; 103(49):18528–18533. [10.1073/pnas.0608714103](https://doi.org/10.1073/pnas.0608714103) [PubMed: 17116859]
- Huang Z, Penczek PA. Application of template matching technique to particle detection in electron micrographs. *J Struct Biol.* 2004; 145:29–40. <http://dx.doi.org/10.1016/j.jsb.2003.11.004>. [PubMed: 15065671]
- Kirkland, EJ. *Advanced Computing in Electron Microscopy*. 1. Plenum Press; New York: 1998.
- Shang Z, Sigworth FJ. Hydration-layer models for cryo-EM image simulation. *J Struct Biol.* 2012; 180(1):10–16. <http://dx.doi.org/10.1016/j.jsb.2012.04.021>. [PubMed: 22609687]
- Zeng X, Stahlberg H, Grigorieff N. A maximum likelihood approach to two-dimensional crystals. *J Struct Biol.* 2007; 160(3):362–374. <http://dx.doi.org/10.1016/j.jsb.2007.09.013>. [PubMed: 17964808]
- Baxter WT, Grassucci RA, Gao H, Frank J. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *J Struct Biol.* 2009; 166(2):126–132. <http://dx.doi.org/10.1016/j.jsb.2009.02.012>. [PubMed: 19269332]
- Langlois R, Frank J. A clarification of the terms used in comparing semi-automated particle selection algorithms in Cryo-EM. *J Struct Biol.* 2011; 175:348–352. <http://dx.doi.org/10.1016/j.jsb.2011.03.009>. [PubMed: 21420497]



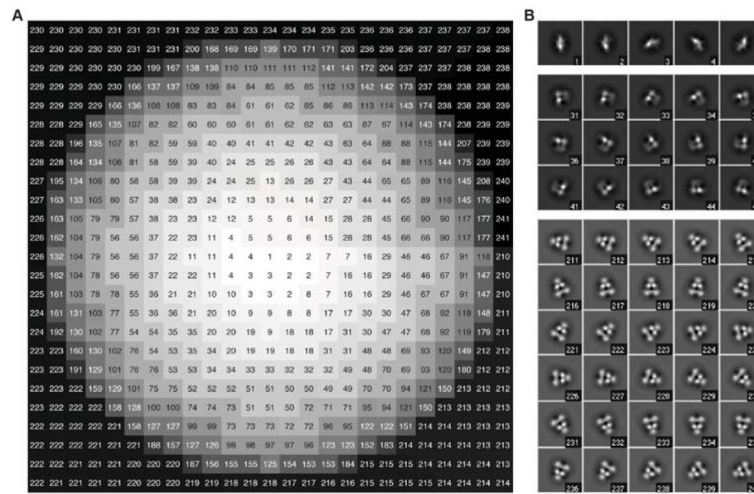
**Fig. 1.** Spherical-constraint projection. The electron beam travels in the  $-z$  direction. The position of the particle's projection image constrains the Euler angles  $\alpha$  and  $\beta$ , while  $p$  and  $p'$  denote two different embedding positions in the vesicle which correspond to the same center of mass of the projection image. The standard Euler angles  $\varphi$ ,  $\theta$  and  $\psi$  defined by (Heymann et al., 2005) are related to these according to  $(\varphi, \theta, \psi) = (\gamma - 90^\circ, \beta, \alpha + 90^\circ)$ .



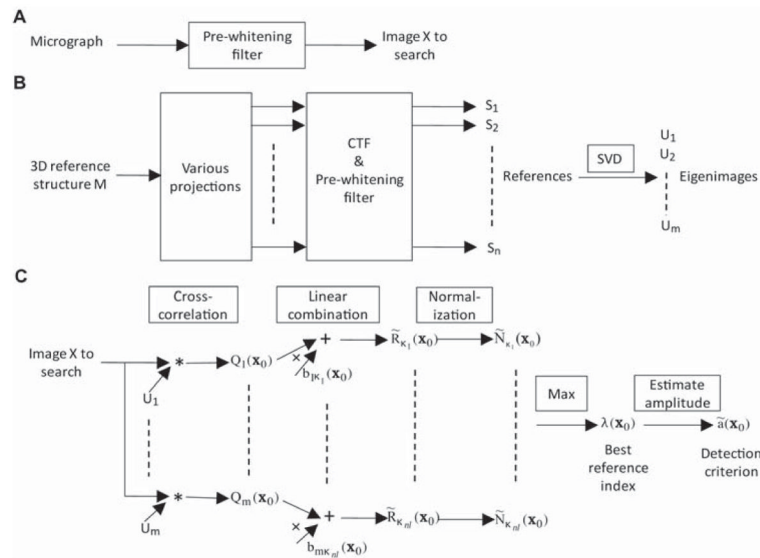
**Fig. 2.** Simulated projections of a membrane protein particle with no CTF applied. Projections were computed from the GluA2 crystal structure, PDB ID: 3KG2. (A) particle at position 1:  $\alpha = -11^\circ$ ,  $\beta = 86^\circ$ , embedded right-side out on the upper hemisphere; position 2:  $\alpha = 23^\circ$ ,  $\beta = 86^\circ$ , inside-out on the upper hemisphere; position 3:  $\alpha = 128^\circ$ ,  $\beta = 24^\circ$ , right-side out on the upper hemisphere; position 4:  $\alpha = -166^\circ$ ,  $\beta = 156^\circ$ , right-side out on lower hemisphere. The radius of the vesicle is 80 pixels at the scale of 0.3 nm/pixel. (B) projections of particles at 1, 2, 3 and 4 with the same values of  $\alpha$  and  $\beta$  but with various  $\gamma$  values, here shown with reversed contrast. The density of the transmembrane region of the particles is shown to be reduced, as is the case after subtraction of membrane density.



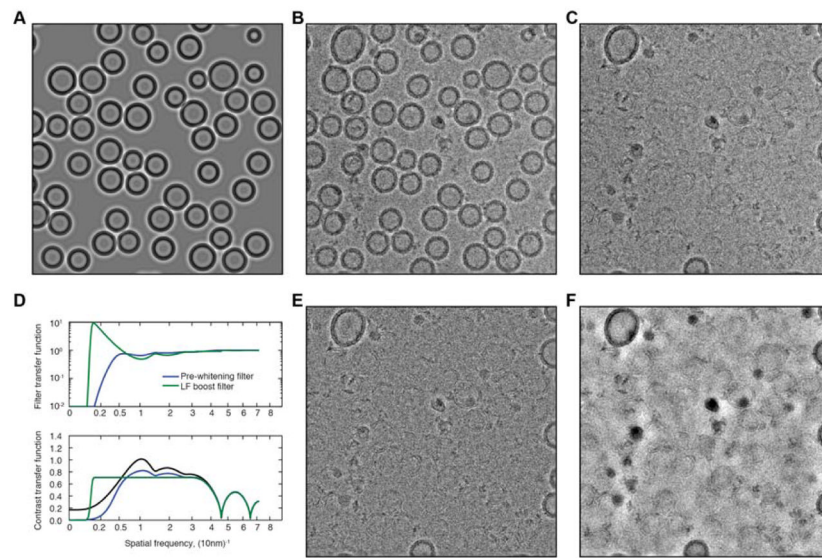
**Fig. 3.** (A) The first 20 eigenimages computed from the projections in the GluA2 reference set, after application of the CTF and filtering appropriate to the micrograph shown in Figs. 6 and 7. (B) Shown are six reference images  $S$  (the first and the third rows) and the corresponding reconstructions  $\hat{S}$  using  $m = 20$  eigenimages (the second and the fourth rows). (C) The range of relative power  $|\hat{S}_k|^2/|S_k|^2$  for all 3856 references enumerated by  $k$ , plotted as a function of the number of terms  $m$  in the eigenimage expansion. Choosing  $m = 20$  gives a relative power of 80% or more for all of the references. The set of references was obtained by approximately uniformly-spaced projections of the 3D model with angular steps of  $12^\circ$ ,  $15^\circ$  and  $22.5^\circ$  for  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively.



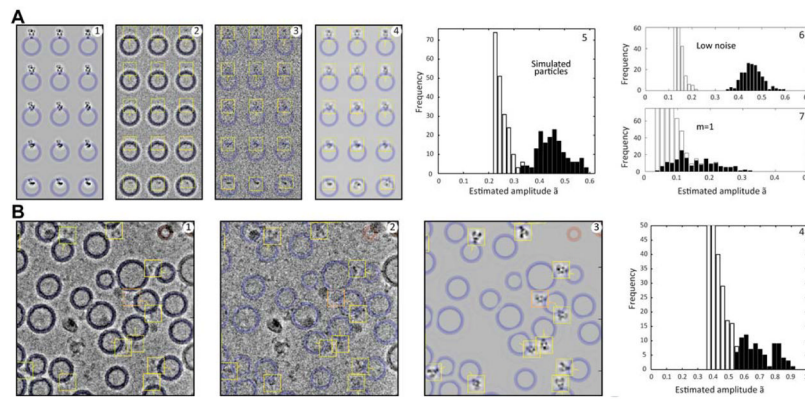
**Fig. 4.** (A) An example of the position-dependent lookup table for  $\kappa_{\gamma}(\mathbf{x}_0)$  for a vesicle of radius 11 pixels having a particle whose center of mass lies 1 pixel outside the membrane center. Values are given for the case  $l = 1$ , which corresponds to the particle being right-side out, on the upper hemisphere, and oriented with  $\gamma = 0$ . Each number is the index of the appropriate reference corresponding to that pixel. (B) A few of the references are illustrated. The reference for a “bottom view” of the particle has index 1, while side views of particles have indices of 210 to 241 as the angle  $\alpha$  takes on 32 values in a 360 degree range.

**Fig. 5.**

Particle picking process (A) The micrograph is processed by a pre-whitening filter, which is chosen to yield a constant power spectrum of noise in the image. If  $P(f)$  is the circularly-averaged power spectrum of particle-free regions of the micrograph, then the filter's transfer function is chosen to be proportional to  $P^{-1/2}$ . (B) Preprocessing of references. From the 3D reference structure a set of 2D projections (~4000 in all) is computed, and each is filtered by the CTF and pre-whitening filter to get the reference set  $S$ . Singular value decomposition is then applied to get the eigenimage set  $U$  (~20 images). (C) The detection process is diagrammed for the neighborhood of one vesicle. Cross-correlations  $Q_i$  are calculated between the filtered image and the set of eigenimages by use of FFTs. The subsequent steps are illustrated for a single pixel position  $\mathbf{x}_0$ . Letting  $k_l(\mathbf{x}_0)$  index the references relevant to the position  $\mathbf{x}_0$  on the image of the vesicle, we compute the estimated cross correlation  $R_{k_l}$  with the appropriate references. Maximizing the normalized cross-correlation yields best index  $l = \lambda(\mathbf{x}_0)$  at each position. The estimated amplitude, computed according to Eq. (14) at this position, is used as a detection criterion.



**Fig. 6.** Processing of a micrograph. (A) Modeled images of vesicles, obtained by fitting to a micrograph. (B) A portion of the micrograph obtained by merging three exposures, each with a dose of  $10 e^-/\text{\AA}^2$  and with defocus values of 1.9, 5.1 and  $17.0 \mu\text{m}$ . (C) Residual after subtraction of the modeled vesicles from the micrograph. (D) Transfer functions of the circularly-symmetric pre-whitening filter and low-frequency boost filter (top) and overall contrast-transfer functions after application of the filters (bottom). The original CTF (black curve) has three low-frequency peaks due to the contributions from the three defocus levels; information from the second and third exposures was used only up to their first CTF zero. (E) The residual image after application of the pre-whitening filter. (F) The residual image after application of the low-frequency boost filter instead. Shown in each case is a  $2\text{k} \times 2\text{k}$  pixel region of a  $4\text{k} \times 4\text{k}$  micrograph. The width of the region corresponds to  $350\text{nm}$  at the specimen. Roughly the upper right quadrant of these images is shown with detected particles in Fig. 7B.



**Fig. 7.**

Automatic particle detection from simulated and experimental images. (A<sub>1</sub>) Simulated merged images of GluA2 particles with the same CTF and pre-whitening filter as employed with the experimental micrograph in Fig. 6. (A<sub>2</sub>) The image after adding white Gaussian noise. (A<sub>3</sub>) The residual after subtracting fitted vesicles. The blue rings indicate the location of the subtracted vesicle density. (A<sub>4</sub>) The best-matching references  $S_k$ . A total of 3856 references were calculated from projections of the X-ray structure with an angular sampling of about  $15^\circ$ ;  $n_l = 32$  references (8 values of the angle  $\gamma$ ) were considered at each pixel position, and  $m = 20$  terms were used in the image expansions. Yellow boxes indicated the detected particles. (A<sub>5</sub>) A histogram of the estimated amplitude values  $\tilde{a}$  for true particles (black bars) and false positives (white bars). The white bars extend only to the value of 0.21 which was the lowest threshold that was tested; a threshold of 0.35 for  $\tilde{a}$  would result in nearly perfect discrimination between true and false particles and is the threshold used in particle detection in panels A<sub>1</sub>–A<sub>4</sub>. (A<sub>6</sub>) The result of automatic particle detection when the noise standard deviation is reduced to half; the false and true particle peaks are widely separated. (A<sub>7</sub>) A histogram showing the poor detection results when a single reference is used, that is when the number of terms  $m$  is set to 1. (B<sub>1</sub>) A portion of the micrograph shown in Fig. 6 with automatically-detected particles marked. Yellow boxes indicate correctly-identified particles, and an orange box indicates a false particle that was detected, probably an ice ball. (B<sub>2</sub>) is the corresponding residual image with vesicle membranes subtracted. (B<sub>3</sub>) The best-matching references are shown. (B<sub>4</sub>) The histogram of signal amplitudes in the entire micrograph, with black bars indicating the 66 visually-identified true particles. A threshold of 0.53 for  $\tilde{a}$  results in apparently ideal particle discrimination. The references shown in Fig. 4B are members of the set of references used for particle selection in both parts A and B of this figure.