

Subtraction of liposome signals in cryo-EM structural determination of protein—liposome complexes

Shouqing Li(李首卿), Ming Li(李明), Yumei Wang(王玉梅), and Xueming Li(李雪明)
Citation: Chin. Phys. B, 2024, 33 (**8**): 088702. DOI: 10.1088/1674-1056/ad4cdb

What follows is a list of articles you may be interested in

New insight into the mechanism of DNA polymerase I revealed by single-molecule FRET studies of Klenow fragment

Rokshana Parvin, Qi Jia(贾棋), Jianbing Ma(马建兵), Chunhua Xu(徐春华), Ying Lu(陆颖), Fangfu Ye(叶方富), and Ming Li(李明)
Chin. Phys. B, 2022, 31 (**8**): 088701. DOI: 10.1088/1674-1056/ac6ed5

The determinant factors for map resolutions obtained using CryoEM single particle imaging method

Yihua Wang(王义华), Daqi Yu(余大启), Qi Ouyang(欧阳颀), Haiguang Liu(刘海广)
Chin. Phys. B, 2018, 27 (**12**): 128702. DOI: 10.1088/1674-1056/27/12/128702

Structural biology revolution led by technical breakthroughs in cryo-electron microscopy

Chang-Cheng Yin(尹长城)
Chin. Phys. B, 2018, 27 (**5**): 058703. DOI: 10.1088/1674-1056/27/5/058703

Bio-macromolecular dynamic structures and functions, illustrated with DNA, antibody, and lipoprotein

Lu Gou(缑璐), Taoli Jin(金桃丽), Shuyu Chen(陈淑玉), Na Li(李娜), Dongxiao Hao(郝东晓), Shengli Zhang(张胜利), Lei Zhang(张磊)
Chin. Phys. B, 2018, 27 (**2**): 028708. DOI: 10.1088/1674-1056/27/2/028708

Comparison of ligand migration and binding in heme proteins of the globin family

Karin Nienhaus, G. Ulrich Nienhaus
Chin. Phys. B, 2015, 24 (**12**): 128705. DOI: 10.1088/1674-1056/24/12/128705

Subtraction of liposome signals in cryo-EM structural determination of protein–liposome complexes

Shouqing Li(李首卿)[1,2], Ming Li(李明)[1,2,†], Yumei Wang(王玉梅)[1,8,‡], and Xueming Li(李雪明)[3,4,5,6,7,§]

[1] *Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *University of Chinese Academy of Sciences, Beijing* 100049, *China*
[3] *Key Laboratory for Protein Sciences of Ministry of Education, School of Life Sciences, Tsinghua University, Beijing* 100084, *China*
[4] *State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, Beijing* 100084, *China*
[5] *Tsinghua–Peking Joint Center for Life Sciences, Beijing* 100084, *China*
[6] *Beijing Frontier Research Center for Biological Structure, Beijing* 100084, *China*
[7] *School of Life Sciences, Tsinghua University, Beijing* 100084, *China*
[8] *Beijing Branch of Songshan Lake Materials Laboratory, Beijing* 100190, *China*

(Received 8 April 2024; revised manuscript received 28 April 2024; accepted manuscript online 17 May 2024)

Reconstituting membrane proteins in liposomes and determining their structure is a common method for determining membrane protein structures using single-particle cryo-electron microscopy (cryo-EM). However, the strong signal of liposomes under cryo-EM imaging conditions often interferes with the structural determination of the embedded membrane proteins. Here, we propose a liposome signal subtraction method based on single-particle two-dimensional (2D) classification average images, aimed at enhancing the reconstruction resolution of membrane proteins. We analyzed the signal distribution characteristics of liposomes and proteins within the 2D classification average images of protein–liposome complexes in the frequency domain. Based on this analysis, we designed a method to subtract the liposome signals from the original particle images. After the subtraction, the accuracy of single-particle three-dimensional (3D) alignment was improved, enhancing the resolution of the final 3D reconstruction. We demonstrated this method using a PIEZO1-proteoliposome dataset by improving the resolution of the PIEZO1 protein.

Keywords: cryo-EM, protein–liposome complexes, liposome signal subtraction, 2D classification averaging

PACS: 87.64.–t, 87.80.–y, 87.64.Ee, 87.85.jf **DOI:** 10.1088/1674-1056/ad4cdb

1. Introduction

Membrane proteins play key roles in many biological processes and are research hotspots of structural biology. Cryo-EM single-particle analysis (SPA) is an essential tool for structural determination of membrane proteins. Due to the hydrophobic nature of the transmembrane regions, the membrane proteins must be solubilized ahead of cryo-EM SPA. Common methods involve encapsulating the hydrophobic regions of the membrane proteins with detergents[1–3] or reconstituting the membrane proteins into liposomes.[4–6] In the latter method, membrane proteins are embedded within the phospholipid bilayer, more closely mimicking their physiological state, thereby facilitating the study of structures sensitive to membrane composition.[7,8] Consequently, this approach has been increasingly adopted in recent years. However, the strong signals exhibited by liposomes under electron microscopy often obscure the structural signals of the embedded proteins, thereby complicating the structural analysis of proteins.

To mitigate the impact of liposome signals on the structural analysis of reconstituted proteins, a common approach is to subtract the membrane structure signals from micrographs before determining the protein structure. For spherical liposomes embedding membrane proteins, the structural characteristics of liposomes are relatively consistent. Wang *et al*. modeled spherical liposomes,[9] allowing for the direct removal of liposome density from the original particle images. Additionally, by utilizing the projection position information of the membrane protein on spherical liposome images, partial spatial orientation information of the membrane protein can be deduced, aiding subsequent orientation estimation.[10] However, ideal physical modeling does not account for the contrast variance of different liposomes in the raw micrographs. To address this issue, Jensen *et al*. performed higher-order singular value decomposition on particle images to construct a subspace for liposomes and then projected each particle image onto this liposome subspace to obtain a residual subspace representing individual liposome differences, thus achieving a 2D statistical model for each liposome.[11] For liposome particles deviating from spherical shapes, in further work, Jensen *et al*. unfolded the membrane structure of each particle image using polar coordinate transformation,[12] which enables the liposome statistical model to do nonlinear shape alignment with

[†]Corresponding author. E-mail: mingli@iphy.ac.cn
[‡]Corresponding author. E-mail: wangym@iphy.ac.cn
[§]Corresponding author. E-mail: lixueming@tsinghua.edu.cn

© 2024 Chinese Physical Society and IOP Publishing Ltd

the membrane. This method allows for the acquisition of structural characteristics of each liposome, reducing the flaws introduced by incorrect subtraction. However, to better learn the contrast and shape characteristics of liposomes, this method requires taking two micrographs for the same particle: one with low defocus to preserve the high-frequency detail information of the membrane protein and another with high defocus to ensure the liposome is visible.

Herein, we propose a liposome signal subtraction method based on 2D classification average images to reduce the interference of liposome signals in the 3D alignment of SPA. This method leverages images obtained from the 2D classification averaging of protein–liposome complexes, enabling the attenuation of liposome signals in individual particles, thereby enhancing the structural resolution of membrane proteins embedded within liposomes. In this paper, using the PIEZO1 proteoliposomes as a case study, we analyzed signal distribution patterns of the 2D classification average images in the frequency domain. Based on this analysis, we identified the main distribution regions of the liposome signals for membrane signal subtraction. Subsequently, by optimizing the scaling factor and the number of classes, the density map of the center cap region and the distal blade region in the PIEZO1 protein were improved.

2. Methods

The 2D classification averaging is a commonly used analysis method in cryo-EM SPA, which averages common structural features in experimental single-particle micrographs to enhance them. In this work, we first analyzed signal distribution characteristics of the 2D classification average images of protein–liposome complexes in the frequency domain. Based on the analysis, we identified the liposome signals and designed a method for subtracting liposome signals in cryo-EM micrographs.

2.1. Liposome signals in 2D classification average images

To analyze liposome signals in the presence of embedded proteins, we utilized a dataset of PIEZO1 proteoliposomes.[8] The PIEZO1 protein is primarily composed of a central cap region and three peripheral propeller blade regions (Fig. S1(a) in supplementary information). Inside the blades, transmembrane helices are aligned in parallel (Fig. S1(b) in supplementary information), while the bottom of the blades features supporting beam structures (Fig. S1(c) in supplementary information). The PIEZO1 protein can be reconstituted into liposomes under two conformations: a flattened state and a curved state. In the flattened state, the cap region is oriented extracellularly (Figs. S2(a) and S2(b) in supplementary information, and the three propeller blades unfold in an almost parallel arrangement, stretching the liposome into a "D" shape (Fig. S2(c) in supplementary information). In the curved state, the cap region is oriented intracellularly (Figs. S2(d) and S2(e) in supplementary information), with the three propeller blades encircling the central cap region, keeping the liposome in an almost spherical shape (Fig. S2(f) in supplementary information).

Under two different conformational conditions, the resolution of liposomes is different from that of PIEZO1 proteins in the 2D classification average images of PIEZO1 proteoliposomes. In the results of the curved state, the transmembrane regions of the PIEZO1 proteins display the highest resolution in the 2D classification average images, revealing clear secondary structures (Fig. 1(a)). The resolution of the cap region is slightly lower than that of the transmembrane region but is higher than that of the liposomes. This indicates that the alignment of the PIEZO1 proteins was not significantly affected by the liposome signals during the 3D alignment. However, in the results of the flattened state, both the PIEZO1 proteins and the liposomes exhibit low resolution, represented by the blurry densities of the transmembrane and cap regions (Fig. 1(b)). This suggests that the presence of liposome signals affected the alignment accuracy of the PIEZO1 proteins. In the work of Yang *et al*., the resolution of the cap and distal blade regions in the flattened conformation was lower than that in the curved conformation,[8] which confirms the analysis.

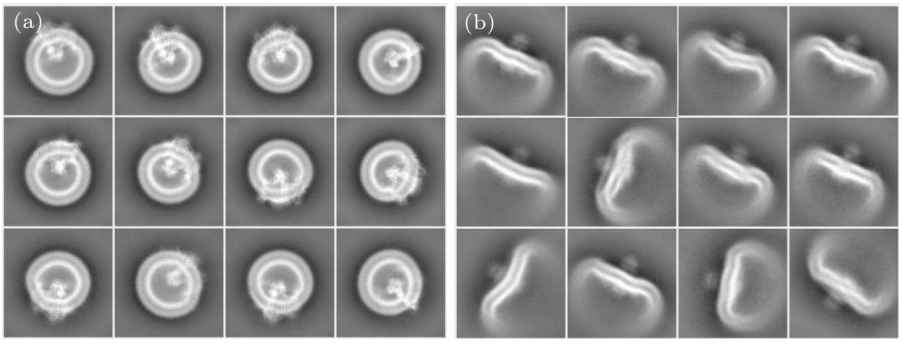


Fig. 1. Schematic representation of 2D classification average images of PIEZO1 proteoliposomes. (a) 2D classification average images of PIEZO1 proteoliposomes in the curved state, (b) 2D classification average images of PIEZO1 proteoliposomes in the flattened state.

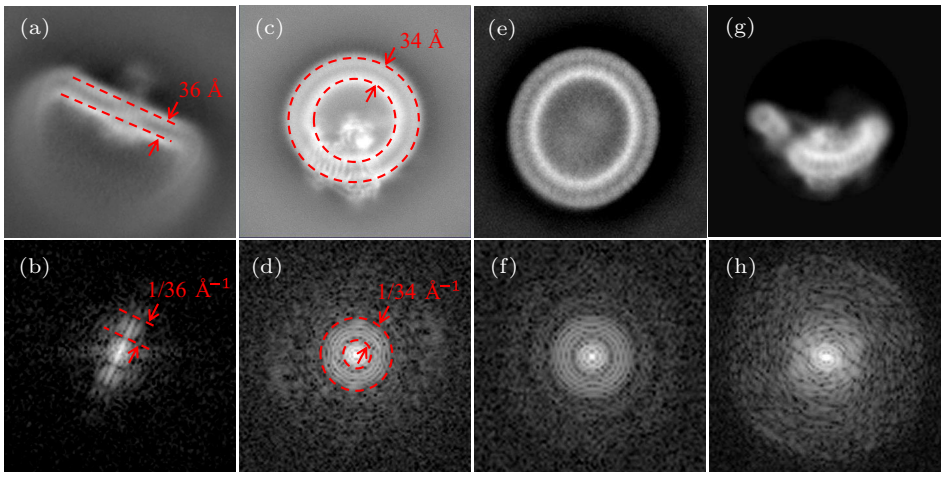


Fig. 2. Single-particle 2D classification average images of different proteins and their corresponding power spectra. (a) 2D classification average image of PIEZO1 proteoliposomes in the flattened state, (b) the low-frequency region of the power spectrum from (a) exhibits signal characteristics of parallel membranes, (c) 2D classification average image of proteoliposomes in the curved state, (d) the low-frequency region of the power spectrum from (c) displays signal characteristics of curved membranes, (e) 2D classification average image of liposomes, (f) the power spectrum from (e) shows concentric ring-like signal distribution characteristics, (g) 2D classification average image of PIEZO1 proteins, (h) The power spectrum from (g) demonstrates dispersed signal characteristics. The red dashed lines in the images denote the spacing of geometric shape features.

In the power spectra of 2D classification average images of PIEZO1 proteoliposomes, characteristic patterns of the membrane structure can be identified. For the flattened conformation, the power spectrum of the 2D classification average image exhibits parallel stripes. These stripes are perpendicular to the membranes in the 2D classification average images, and the spacing between the parallel stripes is approximately the reciprocal of the distance of the phospholipid bilayers (Figs. 2(a) and 2(b)). In the curved conformation, the power spectrum of the 2D classification average image shows periodic ring-like features and the spacing of the ring-like stripes is approximately the reciprocal of the spacing of the phospholipid bilayers (Figs. 2(c) and 2(d)). These numerically consistent relationships indicate that the low-frequency patterns in the power spectrum are characteristic signals introduced by liposomes. In the power spectrum of liposomes without PIEZO1, we also observed the same concentric ring-like signal distribution (Fig. 2(f)), further confirming the signal characteristics of the curved membrane structure in the power spectrum. Unlike the membrane structure, in the power spectrum of the 2D classification average image, the structural signals of the PIEZO1 protein are diffusely distributed in the entire frequency domain (Fig. 2(h)).

We further inspected the distribution characteristics of various structural signals by frequency in the power spectrum. In the power spectra of 2D classification average images of complexes in both conformations, signals from the liposome membranes predominantly occupy the low-frequency region. This is attributed to the phospholipid molecules within the membrane lacking a regular structure, leading to the details of liposomes being averaged out during 2D classification averaging, and only the low-resolution contour features in the low-frequency region are enhanced. On the contrary, the structural signal of the PIEZO1 protein is uniformly distributed across a broader frequency range compared to that of the membrane structure.

2.2. Subtracting liposome signals using 2D classification average images

The characteristics of the power spectrum mentioned above provide an opportunity for subtracting liposome signals from the original particles. The signals to be subtracted are from a region in the Fourier transform of the 2D classification average images, in which the membrane signals are dominant. This subtraction aims to weaken the membrane signals in the original particles, thereby reducing their interference in the alignment during the reconstruction of membrane proteins. Although signal subtraction inevitably removes part of protein signals, our subsequent experiments demonstrate that the attenuation of some low-frequency protein signals does not significantly impact the alignment of the protein. The following sections describe the process of subtraction.

2.2.1. Particle selection for liposome signals subtraction

To improve the accuracy of signal subtraction for membranes, choosing the particles suitable for signal subtraction is always essential. The particle selection is first based on the quality of the 2D classification average images. If the membrane structure features in a 2D classification average image are unclear or chaotic (Fig. S3 in supplementary information), all the corresponding particles in this class should be marked as dirty particles, and will not undergo the subtraction. Furthermore, the selection of particles can be based on some stability indicators during 2D classification averaging. Two indicators are frequently used: the similarity score between a particle and the corresponding 2D classification average image, and the number of times that a particle changes classes dur-

ing the classification process.[13] A higher similarity score and fewer class changes both imply a higher similarity between the particle and its 2D classification average image, making it more suitable for subtraction. The threshold for the similarity scores and the class changes needs to be adjusted based on the effectiveness of liposome signal subtraction. Particles below this threshold are marked as dirty data and are excluded from the subtraction process.

Considering the dirty particles marked after the above selection process still contain membrane protein signals, although we do not subtract the membrane signals from the dirty particles, they still participate in the final 3D reconstruction. The proportion of dirty particles ($\sim 5\%$ in this PIEZO1-proteoliposome dataset) in the overall dataset is not high, therefore the possible alignment errors induced by the liposome signals in the dirty particles don't have an obvious impact on the reconstruction results. In our test, removing dirty particles reduced the final map quality (Fig. S4 in supplementary information). Therefore, we recommend to include the dirty particles.

2.2.2. Intensity normalization of the 2D classification average images

Given that the signals of the original particles and their 2D classification average images may be in different intensity ranges, it is necessary to normalize the signal intensity of the 2D average images to match that of the original particles to successfully subtract the membrane structure signals.[14,15] In the Fourier transform of the 2D classification average images, we keep the phase unchanged but adjust its amplitude by referencing the amplitude of corresponding pixels in the Fourier transform of the original particles. The adjustment sets the amplitude in the low-frequency region, where liposome signals dominate, to the average amplitude of the original particles in the same frequency range as given below:

$$F'_{\mathrm{avg}}(\boldsymbol{u}_{ij}) = \begin{cases} \alpha \cdot \mathrm{CTF} \cdot \dfrac{(1/N)\sum_{|\boldsymbol{u}|\in[|\boldsymbol{u}_{ij}|,|\boldsymbol{u}_{ij}|+P)} |F_{\mathrm{raw}}(\boldsymbol{u})|}{|F_{\mathrm{avg}}(\boldsymbol{u}_{ij})|} \cdot F_{\mathrm{avg}}(\boldsymbol{u}_{ij}), & |\boldsymbol{u}_{ij}| \in [0, u_{\mathrm{l}}), \\ 0, & |\boldsymbol{u}_{ij}| \notin [0, u_{\mathrm{l}}), \end{cases} \tag{1}$$

where F_{raw} and F_{avg} represent the Fourier transforms of the original particles and the 2D classification averaged images, respectively, $\boldsymbol{u}_{ij}$ represents the spatial frequency of (i, j) pixel of the Fourier spectrum, P is the pixel size of the Fourier spectrum and N represents the number of pixels in the frequency range of $[|\boldsymbol{u}_{ij}|, |\boldsymbol{u}_{ij}| + P)$. CTF denotes the contrast transfer function of the original particle. u_{l} denotes the boundary of the main distribution area of the membrane signals, and α represents the scaling factor of the 2D classification average image. Both u_{l} and α are hyperparameters that need to be determined through experimentation. The scaling factor α is used to adjust the weight of the signals from the 2D classification average image during the subtraction process.

2.3. The overall workflow

Taking PIEZO1 proteoliposomes in both flattened and curved states as examples, we demonstrated the specific procedure for subtracting liposome signals from the original particles (see Protocol in supplementary information). Initially, we perform 2D classification averaging on the particles and select particles suitable for liposome signal subtraction based on 2D classification results (Fig. 3(a)). Next, we identify the

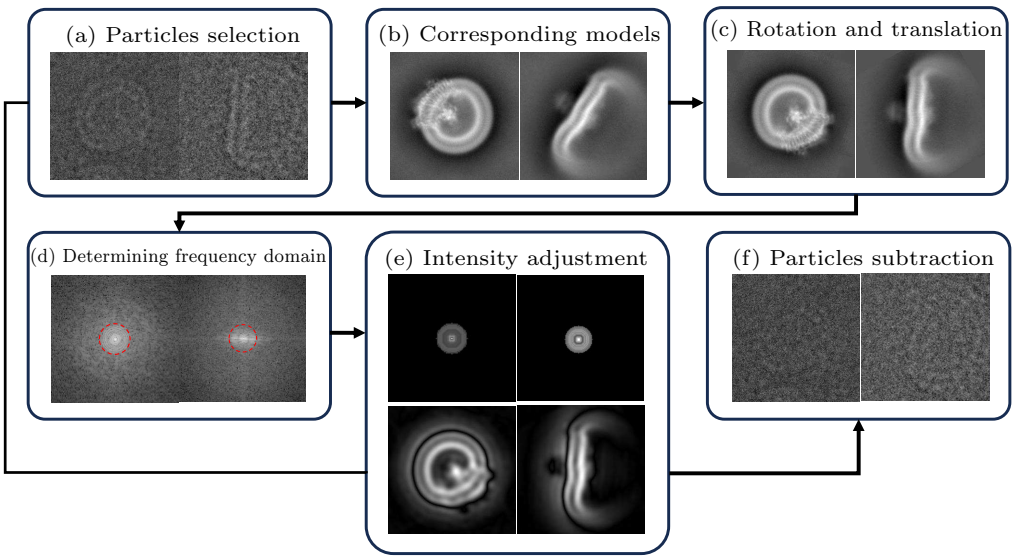


Fig. 3. Schematic diagram illustrating the process of liposome signal subtraction using 2D classification average images. Each box represents a step in the process, and arrows indicate the sequence of operations.

2D classification average image corresponding to each particle (Fig. 3(b)) and align the 2D classification average image to the original particle (Fig. 3(c)). Subsequently, we determine the range of the frequency domain to be used for subtraction (Fig. 3(d)) and normalize the signal intensity in the frequency domain (Fig. 3(e)). Finally, we subtract the processed 2D classification average images from the original particles (Fig. 3(f)). In the present work, all operations for liposome signal subtraction and protein 3D reconstruction were carried out by THUNDER software,[16] and the local resolution of proteins was evaluated by ResMap software.[17]

3. Results

3.1. Determination of the boundary u_l and the scaling factor α

In the step of intensity normalization, two crucial hyperparameters in Eq. (1), the boundary u_l and the scaling factor α can significantly impact the performance of liposome signal subtraction. The setting of the boundary u_l may have two choices: the spatial frequency corresponding to the resolution of the 2D classification, and the spatial frequency at which membrane structure features are observable in the power spectrum. For the dataset of PIEZO1-liposome complexes, these spatial frequencies are $1/10$ Å^{-1} and $1/19$ Å^{-1}, respectively. First, we subtracted signals within $1/10$ Å^{-1} using a scaling factor α of 1.0 (Fig. 4(a)). However, in the final reconstruction of PIEZO1, the density map showed significant discontinuities, and noise densities wrapped around the entire PIEZO1 density (Fig. 4(b)). The densities of the cap and distal blade domains of PIEZO1 were completely fragmented and could not cover the protein structure model (PDB: 7WLU) (Fig. 4(c)). Then, we tried to subtract signals within the frequency domain of $1/19$ Å^{-1} under a scaling factor α of 1.0 (Fig. 4(d)). The 3D reconstruction result of PIEZO1 was significantly improved. The resolution of the transmembrane blade regions was reported at 4.8 Å. The overall density of the protein was reconstructed in high quality, without fragmented densities surrounding it (Fig. 4(e)). The density of the cap region and the distal blade regions well cover the structural model (Fig. 4(f)). Comparing the above two results, we determined that the liposome signals mainly distribute within $1/19$ Å^{-1}. In the frequency range between $1/19$ Å^{-1} and $1/10$ Å^{-1}, the signals may predominantly originate from the membrane protein. Subtracting signals within this area actually removes more membrane protein signals, leading to the fragmentation of the overall density in the reconstruction.

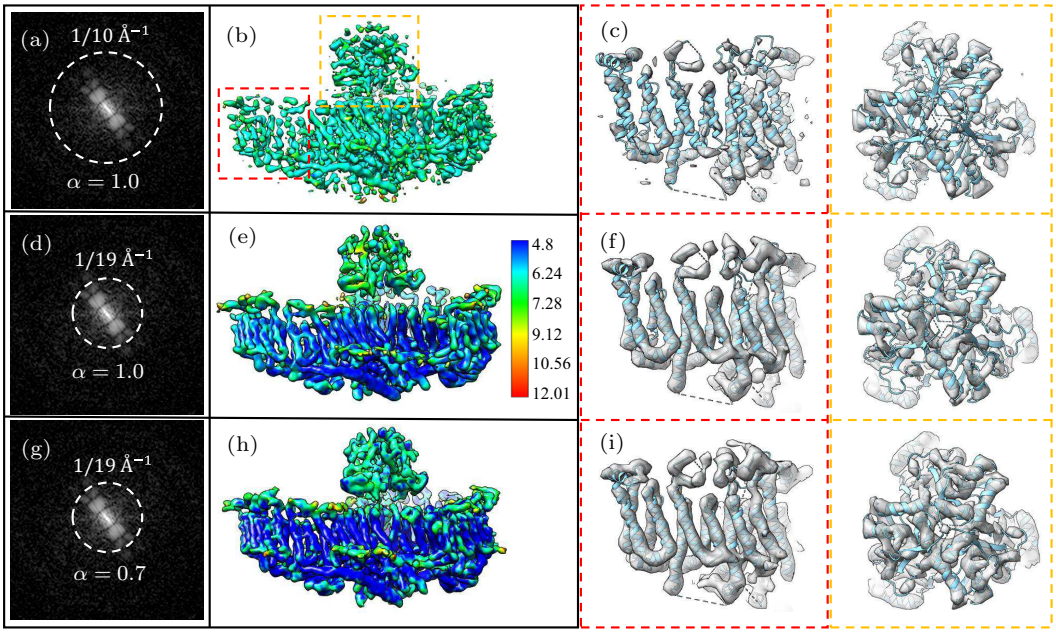


Fig. 4. The impact of selecting the boundary u_l and the scaling factor α on the performance of liposome signal subtraction. (a) The boundary u_l is set to $1/10$ Å^{-1}, and the scaling factor α is set to 1.0. (b) 3D refinement results of particles after lipid signal subtraction, with colors on the 3D model representing local resolution information. (c) Detailed view of the PIEZO1 protein model in panel (b), where the gray area is the reconstruction result, and the blue ribbon represents the 7WLU model from the protein database. The red dashed box shows the distal blade region of the PIEZO1 protein, and the yellow dashed box displays the cap region of the PIEZO1 protein. (d)–(f) The boundary u_l is set to $1/19$ Å^{-1}, and the scaling factor α is set to 1.0. (g)–(i) The boundary u_l is set to $1/19$ Å^{-1}, and the scaling factor α is set to 0.7.

To determine the scaling factor α, we typically try a series of values from 0 to 1.0 at intervals of 0.1 and compare the resolutions of the PIEZO1 reconstructions under different scaling conditions. Based on the boundary spatial frequency of $1/19$ Å^{-1} chosen above, we identified the optimal scaling factor α as 0.7 (Fig. 4(g)). In the final reconstruction, the resolution of the cap domain reached 6.5 Å, and the transmembrane blade domain was also enhanced (Fig. 4(h)); the densi-

ties of the transmembrane helical structures in distal blade domain and the cap domain were also improved (Fig. 4(i)). The purpose of introducing the scaling factor α is to depress the liposome signal to be removed and preserve more membrane protein signals during the subtraction. Selecting an appropriate scaling factor α plays a crucial role in optimizing liposome signal subtraction.

3.2. The number of classes for 2D classification averaging

The number of classes for 2D classification averaging can significantly impact the performance of liposome signal subtraction. First, the number of classes affects the consistency of particles allocated to the same class. A smaller number of classes means more particles in each class, maybe leading to poor consistency of particles within a class. Second, fewer particles in a class means less signal in the class average (corresponding to a large number of classes), which reduces the accuracy of the obtained membrane signals in the 2D classification average images. Therefore, a different setting of the number of classes has an essential influence on subtraction. As liposomes and membrane proteins coexist in the 2D classification average images, the process of subtracting the 2D classification average image from the original particle inevitably affects the membrane protein signals. Thus, the subtraction of liposome signals and the preservation of membrane protein signals are a pair of antagonistic factors. An appropriate number of classes for 2D classification averaging should be found to balance these two aspects.

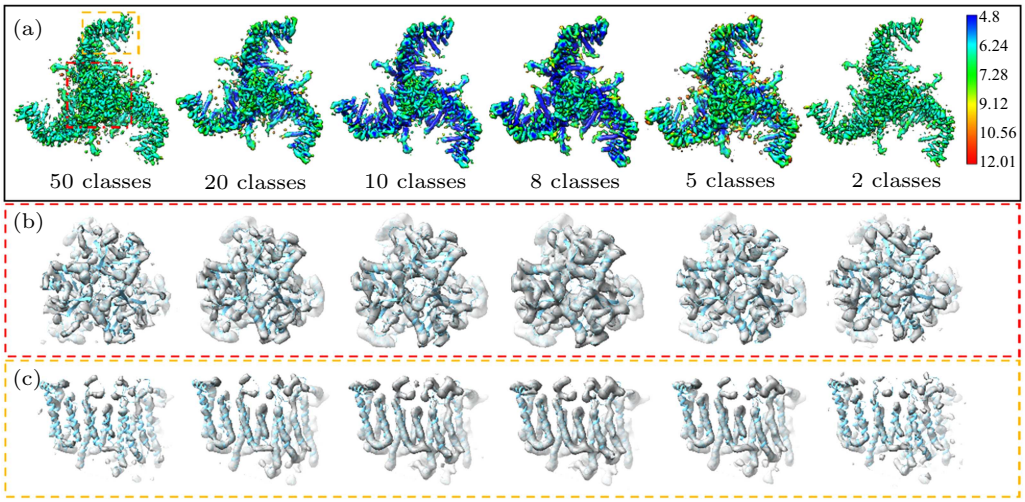


Fig. 5. Comparison of liposome signal subtraction effects under different numbers of classes. (a) Refinement results of particles after liposome signal subtraction under conditions of varying number of classes. From left to right, the class numbers of 2D classification averaging are 50, 20, 10, 8, 5, and 2, respectively, with colors on the 3D model indicating local resolution information. (b)–(c) Comparisons of reconstruction details under different numbers of classes. The gray area represents the 3D reconstruction results, and the blue ribbon is the 7WLU model from the protein database. (b) Comparison of reconstruction details in the cap region of the PIEZO1 protein, corresponding to the red dashed box in panel (a). (c) Comparison of reconstruction details in the distal blade region of the PIEZO1 protein, corresponding to the yellow dashed box in panel (a).

To compare the impact of the number of classes in 2D classification averaging on liposome signal subtraction, we initially classified 1.2×10^5 particles of PIEZO1 proteoliposomes under six different settings of the number of classes, 50, 20, 10, 8, 5, and 2. With the number of classes of 8, we achieved the best resolution of 4.8 Å in the transmembrane blade region and the cap region was at about 6 Å (Fig. 5(a)). The densities of the cap and distal blade regions in the PIEZO1 protein can evenly cover the protein structure model (Figs. 5(b) and 5(c)). However, as the number of classes decreased, the density of PIEZO1 protein became discontinuous, many fragmented densities appeared around the PIEZO1, and the resolution of the transmembrane region also decreased (Fig. 5(a)). The densities of the cap and distal blade regions became fragmented, leaving the protein structure model exposed (Figs. 5(b) and 5(c)). Similarly, as the number of classes increased, the overall resolution of PIEZO1 protein also decreased, and the surroundings became noisy (Fig. 5(a)). Meanwhile, the densities of the cap and distal blade regions could not cover the protein structure model (Figs. 5(b) and 5(c)).

Until the number of classes is set to a saturation value, indicated by the appearance of empty classes, increasing the number of classes leads to finer 2D classification averaging (Fig. S5 in supplementary information). The averaged images of liposomes closely resemble the shape characteristics of the liposomes in individual particles, allowing for the perfect removal of liposome signals. However, the PIEZO1 in the 2D classification average image also becomes clearer, which means that more PIEZO1 signals might be subtracted together with the liposome signals during the subtraction. As the number of classes decreases, each class contains more PIEZO1 proteoliposomes with varied shapes, resulting in more blurred liposomes and PIEZO1 (Fig. S5 in supplementary information). At this point, although the subtraction operation can leave the signals of PIEZO1 largely intact, the removal of liposome signals introduces more imperfections, which will con-

tinue to interfere with the alignment of membrane proteins. A common selection criterion for an appropriate number of classes is that the liposome part of each 2D classification average image has clear structural features, but the membrane protein part remains relatively blurred.

3.3. Reconstruction after liposome signals subtraction

For the structural determination of protein–liposome complexes, a widely used method was iteratively carrying out 3D classification. However, the 3D classification is often interfered with by liposome signals. We compared the result of the 3D classification and the liposome signal subtraction method using the flattened conformation of the PIEZO1–liposome complex.

Initially, we performed a global refinement using 1.2×10^5 particles in the dataset, followed by a local refinement with a mask of the PIEZO1 protein. Ultimately, the reconstruction of PIEZO1 reached an overall resolution of approximately 8.4 Å, in which the local resolution of the transmembrane blade region was reported nearing 6.2 Å, and the cap region at 9.1 Å (Fig. 6(a)). During the alignment, the membrane region dominated the alignment, indicated by the better reconstruction at the liposome and the blade region (Fig. 6(a)) than that at the cap region of PIEZO1 protein (Fig. 6(b)).

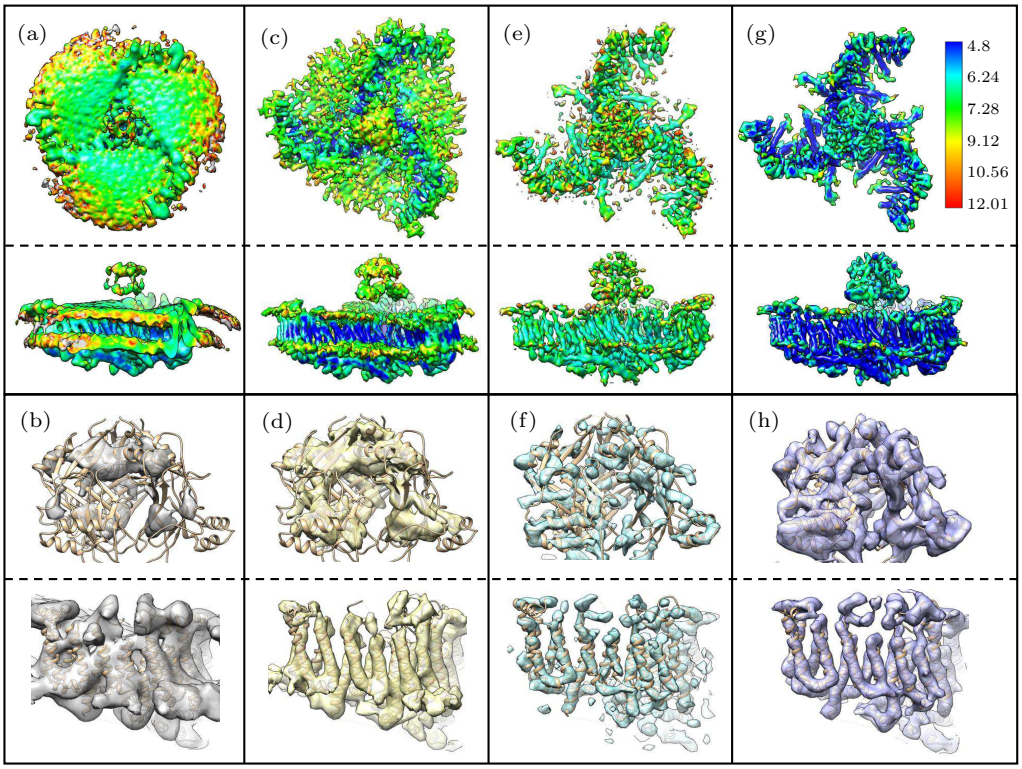


Fig. 6. Comparison of PIEZO1 reconstruction using 3D classification averaging and liposome signal subtraction methods. (a) Density map of the PIEZO1 protein reconstructed using 1.2×10^5 particles. Above the dashed line is the top view, and below the dashed line is the side view, with colors on the 3D model indicating local resolution information. (b) Comparison of the 3D model in panel (a) with the protein structure model (PDB: 7WLU). Above the dashed line is the cap region of the PIEZO1 protein, and below the dashed line is the distal blade region of the PIEZO1 protein. (c)–(d) The reconstruction using 3.5×10^4 particles selected through 3D classification averaging. (e)–(f) The reconstruction using 3.5×10^4 particles after liposome signal subtraction. (g)–(h) The reconstruction using 1.2×10^5 particles after liposome signal subtraction.

Subsequently, we iteratively performed 3D classification with only local alignment on the 1.2×10^5 particles. 3.5×10^4 particles were then selected from the class with the highest resolution, and subjected to further local refinement with a shaped mask of PIEZO1. Finally, the overall resolution of the PIEZO1 protein was reported at 6.8 Å, with the local resolution of the transmembrane blade domain at 5.2 Å and the cap domain at 7.8 Å (Fig. 6(c)). The improvement was observed in the blade region. On the contrary, the density of the liposome became weaker than that of the blade regions (Fig. 6(c)). The density of the cap region was improved relative to the previous reconstruction, but was still not connected to the other region of the PIEZO1, and the resolution was still much lower than that of the blade region (Fig. 6(d)). 3D classification focusing on a local region could partially suppress the impact of liposome signals, but the liposome signals remained unchanged and relatively strong. Consequently, the alignment was still focused on the transmembrane region.

To reduce the influence of membranes, we carried out membrane subtraction on the 3.5×10^4 particles selected in the above process and performed the local refinement. Although the overall resolution of the reconstruction map was still around 7.1 Å, and the local resolution of the transmembrane blade region decreased to 6.3 Å, the resolution of the

cap region was enhanced to 7.5 Å (Fig. 6(e)). Additionally, the density of the liposome in the reconstruction map virtually disappeared (Fig. 6(e)), indicating the effective removal of liposome signals. Although the reported resolution of the cap region remained unchanged, its density map was improved. However, the resolution of the blade region overall declined, and many fragmented densities appeared around the PIEZO1 protein (Fig. 6(f)). A possible reason for this might be that the alignment was no longer focused on the transmembrane region, and the removal of liposome signals may also lead to the partial loss of the protein signal.

Finally, we expanded the number of particles, performing membrane subtraction on all of the initial 1.2×10^5 particles, and performed local refinement based on the orientation parameters obtained in the above processing. The overall resolution of the PIEZO1 protein was improved to 5.8 Å, with the local resolution of the transmembrane blade region reaching 4.8 Å, and the cap region at 6 Å (Fig. 6(g)). The quality of the density map in the cap region was significantly improved, clearly displaying the structural features of the main chain and some side chains. At the same time, the resolution of the distal blade region was enhanced, and the continuity of the density was improved (Fig. 6(h)). Although the operation of subtracting the 2D classification average images from the original particles altered the signals of the membrane proteins, the above results indicate that the increase in data volume can compensate for this deficiency, thereby allowing us to obtain better 3D reconstruction results. In summary, the liposome signal subtraction method proposed in this paper improved the overall resolution of the PIEZO1 protein by approximately 3 Å, with an improvement in the density of both the blade and cap domains from absent to present.

4. Discussion

In the alignment of cryo-EM SPA, the similarity comparison is the core step accomplished through iterative comparison of particle images and model projections. For membrane proteins reconstituted in liposomes, the membrane of liposome is involved in the original particle image but excluded in the model. Furthermore, the intrinsic flexibility and high contrast under cryo-EM imaging obscure the alignment by reducing the similarity between particle images and model projections as well as dominating the alignment, resulting in a low resolution of the reconstruction.

This work proposed a method that subtracts liposome signals using 2D classification average images to reduce the interference of liposome signals, thereby diminishing their impact on the alignment of membrane protein structure signals. Based on the analysis using a PIEZO1-proteoliposome dataset, the liposome signals in the low-frequency region of the 2D classification average images are stronger than those of the PIEZO1 protein. Reducing the liposome signals can improve the accuracy of the alignment. Although subtracting the 2D classification average images from the original particle images also removes part of the target protein signals in the low-frequency region, the influence on the reconstructed map was subtle in our test and could be highly compensated by involving a large number of particles. To balance the influence of the subtraction and remaining protein signals, we introduced the scaling factor as a hyperparameter so that the user can optimize the subtraction. Meanwhile, errors are inevitably introduced into the modified particle images during the subtraction operation. To minimize the influence, the refinement should be limited to local orientation optimization, and the global alignment must be done ahead of the subtraction.

The present method performed well in improving the resolution of the flattened PIEZO1-proteoliposome, but there are still spaces to improve the method. First, the operation of liposome signal subtraction is strongly depending on several hyperparameters, including the boundary u_1 and the scaling factor α and the number of classes in the 2D classification averaging. These hyperparameters are not universal for different datasets, and the determination of them is a procedure of trial and error. Second, the improvement of the method requires that the resolution of the liposomes in a 2D classification average image is better than that of the proteins embedded in them. Otherwise, too much protein signal might be removed during the subtraction. Therefore, this method is suitable for the proteins embedded in the liposomes with some flexibility.

The 2D classification averaging is a routine step in SPA. The method of liposome signal subtraction eliminates the need for additional prior information or experimental approaches to specifically remove signals of certain structures and is effectual in improving the resolution of membrane proteins reconstituted in liposomes.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 32241023 and 92254306) and the Fund from the Tsinghua–Peking Joint Center for Life Sciences, and Beijing Frontier Research Center for Biological Structure. We acknowledge the Tsinghua University Branch of China National Center for Protein Sciences Beijing for providing facility support in computing and cryo-EM instruments. We also thank Institute of Physics, Chinese Academy of Science & Beijing Branch of Songshan Lake Materials Laboratory for our cryo-EM work.

References

[1] Garavito R M and Ferguson-Miller S 2001 *J. Biological Chem.* **276** 32403
[2] Seddon A M, Curnow P and Booth P J 2004 *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666** 105

[3] Henry G D and Sykes B D 1994 *Meth. Enzymol.* **239** 515
[4] Tilley S J, Orlova E V, Gilbert R J C, Andrew P W and Saibil H R 2005 *Cell* **121** 247
[5] Wang L and Sigworth F J 2009 *Nature* **461** 292
[6] Lin Y C, Guo Y R, Miyagi A, Levring J, MacKinnon R and Scheuring S 2019 *Nature* **573** 230
[7] Yao X, Fan X and Yan N 2020 *Proc. Natl. Acad. Sci. USA* **117** 18497
[8] Yang X, Lin C, Chen X, Li S, Li X and Xiao B 2022 *Nature* **604** 377
[9] Wang L, Bose P S and Sigworth F J 2006 *Proc. Natl. Acad. Sci. USA* **103** 18528
[10] Wang L and Sigworth F J 2010 *Meth. Enzymol.* **481** 147
[11] Jensen K H, Sigworth F J and Brandt S S 2016 *IEEE Trans. Image Process.* **25** 540
[12] Jensen K H, Brandt S S, Shigematsu H and Sigworth F J 2016 *J. Structural Biology* **194** 49
[13] Gong X, Qian H, Zhou X, Wu J, Wan T, Cao P, Huang W, Zhao X, Wang X, Wang P, Shi Y, Gao G F, Zhou Q and Yan N 2016 *Cell* **165** 1467
[14] Joseph A P, Lagerstedt I, Jakobi A, Burnley T, Patwardhan A, Topf M and Winn M 2020 *J. Chem. Inf. Model.* **60** 2552
[15] Fernández-Giménez E, Martínez M, Sánchez-García R, Marabini R, Ramírez-Aportela E, Conesa P, Carazo J M and Sorzano C O 2021 *Journal of Structural Biology* **213** 107780
[16] Hu M, Yu H, Gu K, Wang Z, Ruan H, Wang K, Ren S, Li B, Gan L, Xu S, Yang G, Shen Y and Li X 2018 *Nat. Methods* **15** 1083
[17] Kucukelbir A, Sigworth F J and Tagare H D 2014 *Nat. Methods* **11** 63