Towards dynamic structure of biological complexes at atomic resolution by cryo-EM

Kai Zhang(张凯)

Citation: Chin. Phys. B . 2018, 27(6): 066801. **doi:** 10.1088/1674-1056/27/6/066801

TOPICAL REVIEW — Electron microscopy method for the emergent materials and life sciences

Towards dynamic structure of biological complexes at atomic resolution by cryo-EM

Kai Zhang(张凯)†

The Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

(Received 2 March 2018; revised manuscript received 26 April 2018; published online 30 May 2018)

Cryo-electron microscopy makes use of transmission electron microscopy to image vitrified biological samples and reconstruct their three-dimensional structures from two-dimensional projections via computational approaches. After over 40 years of development, this technique is now reaching its zenith and reforming the research paradigm of modern structural biology. It has been gradually taking over X-ray crystallography as the mainstream method. In this review, we briefly introduce the history of cryo-EM, recent technical development and its potential power to reveal dynamic structures. The technical barriers and possible approaches to tackle the upcoming challenges are discussed.

Keywords: cryo-electron microscopy, protein complexes, three-dimensional reconstruction, dynamic structures, probabilistic conformational spaces

PACS: 68.37.Lp **DOI:** 10.1088/1674-1056/27/6/066801

1. Introduction

We human beings understand the objective world via sensation of external signals. Direct visual perception of the surrounding objects is the most effective way. That is why our eyes are so important for us to enjoy the beauty of nature. However, the human vision is insurmountably limited to a resolution no better than 0.1 mm. For more than tens of thousands of years since the birth of mankind, this had always been the ultimate limit. The invention of light microscope, almost all of a sudden, substantially changed this situation around 400 years ago. Nevertheless, for several centuries the resolution had been limited by the wavelength of light at around 400 nm $\sim$ 700 nm. It was not until the early 1930s that the birth of electron microscope suddenly expanded human vision to an even higher level with unprecedented details as a result of the much smaller wavelength of electrons (e.g. $\sim$ 2 pm at 300 kV).[1]

However, structural studies of biological molecules by electron microscopy (EM) had always been challenging. They are typically very fragile and can be easily damaged by high-energy electron beams.[2] This problem was finally addressed by a powerful technique, the cryo-electron microscopy (cryo-EM).[3,4] In a broader point of view, cryo-EM is a specialized technique of transmission electron microscopy (TEM) focusing on biological molecules, usually protein complexes or nucleic acids.[5] The primary goal is to obtain three-dimensional (3D) density maps from the two-dimensional (2D) projections or diffraction patterns of biomolecules in vitreous ice. If the resolution of the map is good enough, *de novo* modelling then becomes possible, which allows the analysis of biological mechanisms at atomic level. In history, X-ray crystallography had always been the dominant method in structural biology. The low resolution of cryo-EM had been a major bottleneck for its potentially wide application. The situation only began to change in recent years.[3,6]

A typical structural study by cryo-EM contains three major steps:[7] (i) cryogenic sample preparation; (ii) high quality data collection using a TEM; (iii) data processing, reconstruction and structural analysis. All these steps are equally important and crucial for structure determination in practice. Three scientists, Jacques Dubochet, Richard Henderson and Joachim Frank were awarded the Nobel Prize in Chemistry 2017 for their fundamental contributions to the development of cryo-EM technique in these key steps. Every nut and bolt of this technique has been more or less optimized in the past few years.

The modern cryo-EM has been a versatile method that can be used to study biological samples at least in four different ways according to the symmetries of the targets, data collection and processing approaches. These methods mainly include electron crystallography,[8] single particle analysis,[9] helical reconstruction,[10] and tomography[11] in general cases, with possible combination for special application. Electron crystallography requires the 2D crystals of protein complexes.[8] Early high resolution studies were mainly limited to this method as the radiation damage could be compensated by averaging thousands of molecules in the same orientations. However, most protein complexes simply resist crystallization. Single particle cryo-EM[9] has therefore emerged and been the overwhelmingly dominant method as it only requires dispersed molecules in solution. These

†Corresponding author. E-mail: kzhang@mrc-lmb.cam.ac.uk

molecules are usually called 'particles' in cryo-EM and each class of particles must be essentially identical from conventional perspective. Helical reconstruction is used to determine filament-like structures, such as actin filament,[12] microtubule,[13] and other protein fibrils.[14,15] It can benefit from both diffraction patterns and the geometric constraints for orientation determination. In modern data processing, helical reconstruction is usually treated in the same way as normal SPA with additional constraints of the intrinsic symmetry.[16] Electron tomography (ET) is similar to computed tomography (CT) by using X-ray scan. The latter is usually used to get the cross-sectional images of partial or entire human body, while ET does similar things on biological molecules at the nanometer scale. This is the most generalized method that allows structural studies on a much broader range of targets, e.g. the nuclear pore complex,[17,18] the secretion machine,[19] vesicles,[20] mitochondria,[21,22] viruses,[23,24] etc.

After years of development, cryo-EM began to mature along with the recent resolution revolution. The successful application of direct electron detectors (DED)[6,25,26] and improved image processing algorithms[27–29] both played very important roles. Nowadays, cryo-EM has become a multidisciplinary field. Physics, chemistry, mathematics, computing science, engineering, and so on are all running into each other in the same field, with the central goal to determine the biological structures and understand their mechanisms at atomic level. It is not feasible to discuss every little part of the technical details in this review. Therefore, we will try to emphasize how cryo-EM helped scientists understand the biological molecules to higher and higher levels in history, then in a bit more details about the recent technical development; after that, we will also discuss about the real power of cryo-EM for structural studies on macromolecules as dynamic machinery, including personal perspectives.

2. A brief history of cryo-EM

2.1. Early development

The first electron microscope was invented by Max Knoll and Ernst Ruska at the Berlin Technische Hochschule in 1930s.[30] Marton then reported one of the earliest application on biological samples and noticed that the high-energy electron beams destroyed the native structures.[1] For many years, the TEM images had been interpreted in 2D. The birth of 3D electron microscopy was attributed to Aaron Klug and his colleague in 1968.[31] The key mathematical theory that makes the 3D reconstruction possible is the Central Section Theorem, which states that the Fourier transform of a 2D projection is identical to the central slice of the Fourier transform of the 3D object in the same orientation. This implies the possibility to reconstruct a 3D structure from sufficient views of 2D images that can fill up the Fourier coefficients. The goal can be obtained by tilting the stage or using dispersed particles with randomly distributed orientation.

Intrinsic symmetry was a very important factor for 3D reconstruction in early days. Actually, almost all early structural studies of biological samples were based on the 2D crystals or highly symmetric samples, such as filament or icosahedral viruses.[32] These targets were easier at the beginning for several reasons, such as easy orientation determination and reduced radiation damage by averaging identical copies. Before the birth of cryo-EM, the biological samples had been analyzed by staining approaches with heavy atoms.[33–35] Nevertheless, there were clear limitations using such approaches because the structural details were destroyed during staining.

2.2. The infancy of cryo-EM

In 1975, Henderson and Unwin determined the structure of bacteriorhodopsin at 7 Å resolution in its native state as 2D crystals by using glucose to replace the solution.[36] This allowed clear visualization of transmembrane helices and demonstrated the potential power of EM to reveal high resolution structures of biological molecules. But further improvement seemed almost impossible for biological molecules at room temperature at that time. Scientists were therefore groping for new methods that could potentially reduce the radiation damage. Freezing the biological samples using liquid nitrogen looked very promising to preserve high resolution structural information.[37] But there were severe contamination problems using liquid nitrogen and the aqueous solution tended to form ice crystals during freezing.

The real birth of cryo-electron microscopy was attributed to the work by Dubochet and his colleagues in 1980s.[4,38–40] Dubochet developed a most successful method to vitrify biological solution. In his simple but revolutionary method, biological solution is rapidly plunged into liquid ethane or propane and frozen as a thin film on a holey EM grid. The native structure of biological molecules can be successfully preserved in the vitreous ice and the radiation damage is largely reduced. This method has been widely used in modern researches. No doubt, it was the key progress that finally gave the real birth to cryo-EM technique. Fifteen years later after the first publication of bacteriorhodopsin structure,[36] Henderson *et al.* successfully determined the 3.5 Å resolution structure and built the first atomic model of this membrane protein by cryo-EM.[41]

The anticipation of the cryo-EM method for studying general structures was gradually arising in the meantime. Joachim Frank is the pioneer who generalized cryo-EM to study isolated proteins without crystals. Frank's method is exactly the one we are widely using today, the so-called 'single particles analysis (SPA)'. In contrast to Henderson's previous study on bacteriorhodopsin using 2D crystals, Frank aimed for the structure determination using dispersed 'particles' in solution.

In his early career, he found it was possible to align two EM images to high accuracy by cross-correlation.[42] By aligning and averaging the structurally identical particles, signal-to-noise ratio (SNR) could be greatly enhanced which allowed structural analysis in more details.[43,44] In 1981, Frank was able to build up the earliest software package called 'SPIDER'[45] for SPA. These early studies had created the basis and built up the main frameworks of cryo-EM.

2.3. Single particle cryo-EM growing up

In spite of the possibility that EM could be applied to study 3D structures in general case, i.e. the samples in solution, single particle analysis was not a straightforward method at the beginning. Most early high resolution structures were determined using 2D crystals in 1990s.[41,46–48] It was also pointed out that cryo-EM was very likely to become a general method for near-atomic structural studies of non-crystalline proteins[41,49] at that time. However, several limiting factors, such as radiation damage, specimen movement, and charging etc. still remained very challenging before cryo-EM could finally reach its theoretical limit.[49]

In additional to the critical limiting factors that largely affected the image quality,[49,50] computational methods to obtain the initial model and determine the particle orientations used to be very difficult for single particle cryo-EM.

Frank's early study in 1970s demonstrated the possibility to study non-crystalline proteins.[42,44] However, study of isolated particles required more sophisticated computation. The representation of each EM image as a multidimensional vector by van Heel *et al.* created a number of possible methods afterwards, including multivariate statistical analysis (MSA) for classification.[51] At that point, orientation determination in general cases still remained a challenging problem. A key progress was attributed to Harauz and Ottensmeyer, who originated the projection matching algorithm for orientation determination and refinement against an available 3D model.[52] This algorithm has been the most popular method and the basis for most modern programs. However, a reliable initial model is still required to successfully match the calculated projections of the 3D map with the observed projections, i.e., the cryo-EM images. Based on some earlier common line approach on icosahedral reconstruction by Crowther *et al.*,[32] a method for *ab initio* orientation determination on asymmetric objects was developed by van Heel *et al.*, known as the 'angular reconstitution'.[53] At the same time, Radermacher *et al.* developed another well-known method, the 'random conical tilt' (RCT).[54,55]

Once the concept of 3D reconstruction from random projections became clear, many new ideas began to blossom. In the late 1990s, several more programs were developed for general application of single particle cryo-EM, such as IMAGIC,[56] XMIPP,[57] EMAN,[58] FREALIGN,[59] and Bsoft.[60] Many other stand-alone programs were also developed to deal with many technical problems in different parts.

3. Recent technical development

3.1. Direct electron detectors

The most important recent progress in cryo-EM field was probably the successful application of direct electron detectors (DED).[6,26,29,61] Before that, film and charge-coupled device (CCD)[62] were used for electron detection, which usually resulted in sub-nanometer resolution in well-off cases.[14,63,64] Two main factors had hampered further development of these detectors for high resolution structure determination by cryo-EM. One was the low detective quantum efficiency (DQE) and the other was the incapability for recording movie frames. Compared to CCD which has a DEQ of $\sim$ 7–10% at half Nyquist, the film was usually better at high frequency, with a DQE of 30—35% at half Nyquist.[65] This was probably the key reason why most high resolution cryo-EM structures[66,67] were obtained on film before DED, with only a few exceptions on icosahedral viruses.[68–71] The new generation of DED, based on complementary metal–oxide–semiconductor (CMOS) technology, has much better DQE and faster recording speed.[6]

The conceptual innovation of movie recording provided an incisive solution for several critically important limiting factors in cryo-EM, including the beam induced movement.[29,49,50] Due to the fast recording rate, individual electrons can be 'counted' in theory, given the probability of overlapped electrons approaches zero.[61,72] The other great benefit from locating the centers of individual electrons is that physical pixels can be divided into smaller 'logic pixels'. This operation is usually called 'super-resolution' that can further enhance the DQE of modern direct detectors.[61,72]

The year 2013 was marked by a milestone work from Cheng's group, who determined the first cryo-EM structure of the membrane protein TRPV1 at 3.4 Å resolution by using direct detector.[25] This breakthrough inspired many scientists all over the world, not only in cryo-EM field, but also in X-ray crystallography and many other fields. The whole field has kept on expanding since then. Both methodology development and biological applications are exploding now.

3.2. Phase plate

The image contrast is one of the most important limiting factors for structural determination by cryo-EM. It has been well known that in-focus micrographs lack good contrast.[73] A certain amount of defocus is required to enhance the contrast of EM images. Vitrified biological samples in cryo-EM usually require even higher defocuses, typically several microns, to gain sufficient contrast. This in turn makes the CTF determination and correction more challenging for high resolution

information restoration. Longer exposure with higher doses can also enhance the contrast. But it then leads to severe radiation damage and beam-induced movement. Energy filter is another option for higher contrast by getting rid of the inelastic scattering during TEM imaging. However, its potential is limited and does not essentially break out of the technical restrictions from the conventional TEM imaging. Phase plate is an exceptional concept that is possible to achieve tremendous contrast enhancement.

Based on the assumption of weak phase object approximation (WPOA) in traditional TEM imaging, the CTF is a sine-like function in reciprocal space. As a result, the low frequency information is reduced to almost zero, thus greatly bringing down the image contrast. The idea of phase plate is to convert the sine-like function to cosine-like so that low frequency signal can be well preserved. In other words, the aim is to find a proper approach that can be used to impose a 90° phase shift of the electron waves in the back focal plane. Zernike phase plate (ZPP) was firstly designed by Danev and Nagayama[74] to achieve this goal in cryo-EM application. The design could be briefly described as a thin carbon film with a hole in the center. In spite of the enhanced contrast in several good examples,[75,76] the disadvantage clearly emerged in the following years, such as contamination, short lifetime, fringe artefacts, and challenging alignment.

Voltage phase plate (VPP) is a recently-developed new version which consists of a thin ($\sim$ 10 nm) continuous amorphous carbon film constantly heated to $\sim$ 200 °C.[77] The phase shift is typically created 'on the fly' (by accumulation of exposures).[77] This new design successfully overcame several drawbacks that had limited the application of previous ZPP. In the past few years, it has been used to successfully determine several structures of very small proteins to high resolution,[78,79] including the 64 kDa (1 Da = 1 u = 1.66054×10^{-27} kg) myoglobin at 3.2-Å resolution[80] and a class B GPCR–G-protein complex.[81] By using VPP and Cs-corrector as a combination, Fan *et al.* discovered the equivalence between under-focus and over-focus.[82]

The VPP technique is now being widely used for high resolution study on many targets that could have never been imagined to be determined by single particle cryo-EM method. If we regard the application of direct detector in cryo-EM the critical progress for 'resolution revolution', VPP is probably nothing less than the revolution once more.

3.3. Maximum likelihood

In addition to hardware development, substantial progress on data processing and software development also plays equally important roles in cryo-EM. On major challenge in single particle cryo-EM is the orientation. Classical cross-correlation approach, or least square method using normalized images, had several key problems, such as severe overfitting and model-bias. The failure is attributed to the extremely low SNR of cryo-EM images and unreliable references. Maximum likelihood (ML) approach in cryo-EM image analysis was first introduced by Sigworth[83] to improve the accuracy of alignment on noisy images. Instead of assigning a single value for each parameter of individual particles, the basic concept of ML is to assign probabilities for all possible values of the unknown parameters. The objective is to estimate the 'most likely model' that best fits the assumption and observations (i.e., experimental data). The likelihood is usually calculated under the assumption that noise and unknown parameters obey certain distributions, e.g. Gaussian distribution for noise and (X, Y) origins, even distribution for in-plane rotation etc. The target is to maximize the likelihood function, typically in the form of logarithm, which can be achieved in an iterative way, called the 'expectation-maximization' approach. The expectation step assigns the probabilities for undetermined parameters in the presence of a currently available model; the maximization step estimates a new model based on previous expectation of parameter probability distribution.

The ML method was then developed by Scheres *et al.* and extended to tackle a more challenging problem, the conformational heterogeneity.[84] It was originally implemented in the software package XIMPP. Later, a Bayesian approach[27] was implemented in the program RELION[28] for cryo-EM structure determination by optimization of a single function. Because of its user-friendliness and satisfactory results in most cases, this program is now prevailingly popular in the cryo-EM field. One of the limitations in old version of RELION was the cost of computation as a computer cluster was required for general application. The latest version, RELION2,[85] takes the advantage of the Graphics Processing Unit (GPU) to accelerate many of its kernel functions. This significantly relieves the computational burden for many cryo-EM labs. Recently, a much faster software package, cryoSPARC,[86] was released to further accelerate single particle data processing. In most cases, a single GPU workstation is competent to process an entire dataset. The great speed was achieved by taking advantage of GPU and faster algorithms, the stochastic gradient descent (SGD) and branch-and-bound maximum likelihood optimization.

3.4. Movie processing

In addition to better alignment and classification algorithms, the application of DED probably created more opportunities in cryo-EM field, which allows the movie processing and sophisticated weighting schemes. Specimen movement, including both mechanical drift and beam-induce motion, is one of the most important limiting factor for high resolution structural studies.[50] Because of much faster recording speed, movies can now be recorded at a dose below 1 e/Å^2 per frame. Grigorieff's lab first showed that tens of frames from

the same movie stack can be accurately aligned to correct the movement of each individual particle, which dramatically restored high resolution structural information.[29] A follow-up publication demonstrated that the Fourier coefficients of each frame can be precisely weighted according to the accumulated doses and frequencies.[87] This 'dose filtering' approach can be regarded as pre-reconstruction weighting of particle frames. Another post-reconstruction approach is implemented in RELION, called 'particle polishing',[88] in which the weights are typically calculated using the estimated B-factors from each frame reconstruction. Due to the very low electron doses per frame and high noise, motion correction on each particle alone proved to be very challenging and inaccurate. Li *et al.* reported a program called MotionCorr[61] for frame level motion correction, based on least square method. This has been a very useful and successful tool in cryo-EM field, which generate nice estimation of the drift overall and is user-friendly. However, least square fitting usually does not guarantee smooth estimation of the drifts among all frames. Also, local drift correction and dose weighting were not available in previous version of MotionCorr.[61] An improved variant of this program, the MotionCor2,[89] was designed to deal with these problems and released recently. In most practical cases, the results were reported to be better, such as the representative work on TRPV1.[25,90]

3.5. CTF estimation

Apart from the important problem of unknown orientations, the 2D projections of a 3D object are severely modulated by the contrast transfer function (CTF) in TEM.[91,92] In theory, if all the parameters, such as defocus, astigmatism and phase shift could be accurately recorded during imaging, the information could be restored via deconvolution of the CTF. However, this is not the case in cryo-EM for practical reason. Therefore, accurate estimation of the unknown CTF parameters is essential for high resolution reconstruction.

Since the beginning of 2014, we have been developing a GPU-accelerated CTF determination and correction program, called GCTF. The original purpose at the beginning was to overcome the severe orientation preference of dynactin complex[93] by tilting the stage to high degree. Accurate CTF estimation on each particle is the key solution. After a careful investigation, we found that bending of the support layers, beam induced movement, particle position in ice, charging etc. could all lead to the CTF variation. Therefore, we decided to write our own program to do accurate CTF estimation on individual particles. The local CTF refinement strategy with a weighting scheme in both Fourier and real spaces generates accurate local CTF parameters and avoids overfitting of noise. In addition, the very fast speed of GCTF allows the users to perform real-time evaluation of micrographs during the data collection. Many people in the cryo-EM field have benefited from our GCTF program. Since 2014, it has already been widely used via personal communication. Several early applications[93–95] of GCTF were even published before the official publication and release of GCTF.[96]

We will keep developing GCTF and related programs in the future. One plan is to generalize the concept of equiphase averaging (EPA) approach described in early version of GCTF for CTF determination and refinement on phase plate data, tilted micrographs and movie frames etc. Another major plan is to extend GCTF for thick samples in both cryo-EM and cryo-ET.[97] When the sample is considerably thick, we must take the defocus gradient and variation into consideration in order to get higher resolution. In the most simplified model, defocus gradient is considered as a linear function of Z-height. Based on this assumption, one can simply compensate the defocuses in different local region by adding an offset of Z-height. In practice, however, the modulation of structural information of thick biological samples is more complicated. Many other factors are also critically important, such as multiple scattering, uneven accumulation or fluctuation of the charges on samples and on support layers, variable local potentials of the vitrified biological samples (e.g., differences among solution, lipids and proteins etc.), interaction between electron beams etc. There is no simple and magic trick to solve this problem merely by adding an offset of Z-height to the estimated defocus.

What is more important, the projection is no longer a linear integration if the sample is too thick. For very thin samples under the assumption of strict WPOA, one can consider EM images in two different perspectives: i) linear projection first and then CTF modulation; ii) CTF modulation first and then linear projection. They are essentially identical operations because the phase shift of the exit wave is considered as instant effect. The order of the CTF modulation and linear projection can be swapped without any concerns. However, the two operations severely affect each other on a considerably thick 3D object. Back-projection (BP) algorithms or its variants, such as weighted back-projection (WBP), must be based on linear integration. These analytical algorithms are incompetent to decompose the severely CTF-modulated and non-linearly overlapped signals. Whenever a slice at a certain Z-height is CTF-corrected by adding a defocus offset to the averaged defocus estimation, signals from all other slices with different Z-heights are incorrectly deconvoluted by completely wrong CTF information. We proposed to tackle this problem together with algebraic reconstruction techniques (ART) several years ago, particularly the filtered interactive reconstruction technique (FIRT, first developed in 2008).[98] Unlike WBP, the ART algorithm and its variants are perfectly suitable to address such CTF problem on thick 3D objects. We will discuss a bit more about these issues elsewhere in the future.

Of course, there are many other aspects of technical development that contributed to the reformation of modern cryo-EM. It will not be possible to discuss every nut and bolt of the cryo-EM technique in details in this length-limited review. In the following sections, we will try to provide several personal perspectives in cryo-EM.

4. Potential power of cryo-EM

4.1. The many golden opportunities in cryo-EM

It can be expected that methods on all parts of cryo-EM, i.e. sample preparation, data collection and data processing, will be further improved in the near future. More and more types of cryo-EM grids[99–101] will probably be designed to increase the efficiency of sample adsorption, improve the particle distribution and reduce the beam induced movement etc. Equipment for fast sample freezing will be further improved to capture intermediate states for time-resolved structural studies at second or millisecond scale. Microscopes themselves can also be improved in multiple aspects, such as better mechanical stability, more coherent beams and faster imaging speed etc. Direct detectors will be further optimized with built-in counting algorithms and much improved DQE. In spite of the dramatic progress of VPP,[79] we believe there is still ample room for optimization so that an almost perfect 90° phase shift can be imposed to the electron waves without additional background. If this goal were archived, the micrograph contrast of even more challenging targets, such as tomographic tilt series on sub-cellular context, can be further enhanced by many times. This will possibly allow more accurate alignment of individual particles inside the cell with very low electron doses, say 1 e/Å^2–2 e/Å^2.

In addition to the kernels of cryo-EM, the modern technology innovation in all other related fields will also make tremendous differences. Computers will be faster and cheaper to allow daily structure determination on individual workstations or even personal desktops. The GPU revolution will further reduce the computational cost and the latest cards (e.g. NVIDIA Titan V) can easily exceed the performance of an entire computer cluster 10–15 years ago. The classical image processing algorithms can still be improved, while more intriguing opportunities may emerge from other modern concepts like 'deep learning' and 'compressive sensing' etc. Together with these advances, highly automatic data collection, structure determination and analysis are already on the agenda.

What we can imagine is that conventional structure determination by cryo-EM will gradually become a routine, easy and straightforward task. Structures of protein complexes will probably be determined on-the-fly and even in batches. Systematic screening of chemical compounds for drug design will be expectable as well. All these intriguing prospects are arousing more and more excitements in this field. Many exciting, inspiring and unexpected results will be exploding in the near future, for instance, structures determined at super-high resolution, network contacts among a whole bunch of protein complexes, complete structural mechanisms of the entire pathway of a specific system, *in situ* structural studies in sub-cellular or even cellular context, dynamic structure determination on extremely flexible targets, so on and so on.

In classical cryo-EM technique, the reconstruction is always performed in 3D. With the fast development of cryo-EM, now is the time to seriously consider about determining dynamic structures of protein complexes at high resolution. Not only the structures *in vitro* will be suitable for dynamic analysis, but also the *in situ* systems will be deeply explored. That is the real power of cryo-EM. Single particle analysis and electron tomography will probably be both useful for different purposes. In the past few decades, they have been developed separately; in the following years, there will be great opportunities to combine multiple techniques together to answer really fundamental biological questions, probably in the cellular context. We would expect to see more breakthroughs in both the resolution and complexity of the biological targets. The really cool expectation is that we are now ready to visualize the molecules as dynamic machinery in their native states and even in the cell. As there are too many interesting possibilities to be discussed in details, we will have to focus on the dynamic structure analysis in the following parts of this review as a representative point for the future cryo-EM technique.

4.2. Dynamic structure determination by cryo-EM

Before we discuss what can be done to achieve these goals, we need to clarify several concepts so as not to mislead the beginners in this field. First of all, what do we mean by dynamic structure determination after all? It should be clarified that dynamic structure by cryo-EM is not the same concept as 'direct visualization' of the active objects, just like we observe the living cells under light microscopy. Because all molecules in cryo-EM have been frozen into the vitreous ice, they are essentially immobile if the tiny thermal motion at atom level is ignored. When we mention the dynamic structural analysis or 4D reconstruction from the 2D projections by cryo-EM, we mostly mean computational approach to recover and sequence all conformational snapshots of a certain biological molecule or complex. These snapshots approximate a continuous and smooth conformational space given sufficient sampling. How could it be possible? The answer is actually very simple. Just think about a perfect dancer, who is repeating exactly the same dance again and again. You take photos randomly and capture a certain gesture on each acquisition. If you are allowed to take millions or even billions of pictures whenever you want to, it will be very likely to cover the whole choreography. Structural biologists are able to do the same thing on molecules by cryo-EM. On each micrograph, you can take pictures for tens or

hundreds of individual proteins showing different conformations at different orientations. Given sufficient particle numbers, it becomes possible to recover the whole conformational space. With much better and faster equipment, such goal is no doubt expectable in the future.

We should also emphasize that the dynamic structural analysis by cryo-EM is also different concept from molecular dynamics (MD) simulation. The MD simulation is basically theoretical prediction of the behaviors of a molecule in solution, while dynamic structure determination by cryo-EM is completely experimental analysis. We do use computational approach to recover the dynamic structural information in cryo-EM, but there is nothing to be simulated. The computation in cryo-EM is only a tool to process the experimental data. The aim of cryo-EM is to obtain all possible conformations just from experimental data alone. Although it is possible to take advantage of MD simulation in the intermediate state of data analysis (e.g. to predict the preferred motion), no prior knowledge from theoretical prediction should be imposed in the final reconstruction. This is very important to make sure that the dynamic structural information is essentially experimental via computational approaches. We will discuss in more details about the possible approaches for practical implementation in the following section.

5. Approaches for dynamic structural analysis by cryo-EM

5.1. A brief summary of considerable factors

Successful analysis of a highly heterogeneous system to reveal the protein dynamics requires preservation of structural information as much as possible. This is an even more challenging objective than obtaining a static high resolution map for atomic model building by cryo-EM. There are many aspects of cryo-EM technique that can lead to the loss of structural information. These can be briefly classified into three major groups: sample preparation, imaging and computation. The information loss in each major group is also a combined effect of multiple sub-factors. We briefly summarize what factors can be improved or need to be considered.

(I) Sample preparation

This includes purification of the targets, the optimization of solution environment and the cryo-EM grid preparation. They are all very important and can make big differences depending on the specific application. Both compositional and conformational heterogeneity can be introduced during the protein purification step or by post-purification treatment, such as cross-linking, considerable times of concentration or dilution, freezing and thawing (for storage purpose), changes of salt or pH or temperature, detergent or other chemical compounds added to the solution and so on. The sample quality can also be severely affected during the freezing, including the types of grid used, treatment of the grid and conditions for freezing. These factors can significantly change the absorption and orientation distribution of the protein complexes. The air-water interface is probably the most disastrous region for the preservation of biological structures.[2,102] The strong surface tension of the thin solution film introduced during the blotting can lead to severe distortion of protein complexes, sometimes completely destroying the targets. We need functional heterogeneity and conformation distribution to finally reveal the dynamic structures for understanding the mechanisms. But on the other hand, we must try all our best to avoid the artificial or non-physiological heterogeneity. Otherwise, all the rest work on the structural analysis will be merely in vain. Therefore, careful sample preparation is almost the basis for all the following study that guarantees the quality of cryo-EM structures.

In addition to the quality of structural information, factors that can affect the quantification of the conformation distribution of the targets should be taken into serious consideration as well. Almost all the aforementioned factors can make more or less differences. In the future, if we want to disclose the elaborated mechanisms of a certain complex at quantifiable precision, this is very important to make sure the final computational result is a true reflection of the nature itself rather than artificial or biased. So far, there are almost no literatures discussing such quantitative effects during experiments and its effects on the following computational analysis. Further investigation and efforts are needed to address these issues. Obviously, to obtain a dynamic structure with quantifiable precision probably requires much more data than what is needed for a static 3D structure.

In most cases, the biological sample solution can be regarded as an equilibrium system. All conformations are assumed to co-exist in the system and they can be captured in the same grid simply by normal freezing methods using equipment like Vitrobot etc. Computational approach is sufficient to reveal the conformational distribution from such a system given sufficient data. We do not need to resort to time-resolved techniques experimentally. However, if we want to make clear what is happening to the whole system during an irreversible reaction, i.e., the target is not in an equilibrium state, 'time-resolved freezing methods' must be required. This is still a very challenging topic at the current stage since most biochemical reactions are too fast to be controlled for structural studies. The methods like Spotiton[103] and mixing-spraying[104] are representative and inspiring approaches, but still not suitable to capture structures in action instantly. More elaborate methods by taking advantage of special materials or extreme conditions may be worth thinking about in the future.

(II) Cryo-EM imaging

Recent development in cryo-EM has manifested the power of direct electron detectors to record much improved

micrographs. We will soon expect even better products to be launched into market. The DQE curve of future detectors is very likely to be near 1.0 at almost all frequencies except near-Nyquist. Is it possible to record perfect images using such perfect detectors? The answerer is probably very disappointing. It is definitely impossible using current cryo-EM technique. Even a perfect detector would not allow 100% recording of the structural information from biological samples. The key reason is that most information is actually lost before detection rather than during detection, if a modern direct detector is used for data collection. In order to clarify this issue, we define two consecutive processes during the entire TEM imaging, the 'pre-detection' and 'detection'. The pre-detection represents the period of time from the moment electrons are emitted by the FEG to the point they have penetrated the biological samples, ready to be sensed by detectors. The detection refers to the process from the moment electrons hit the surface of the detector to the point they are sensed, recorded and quantized as digital signals. The DQE of direct detectors only connotes how much information we can preserve during the detection process. It has nothing to do with the loss of information during the pre-detection process.

The information loss during pre-detection is attributed to many factors, such as the irreversible physical changes of biological samples during imaging, random thermal motion of molecules, imperfect optical system of microscope, mechanical instability, image formation on the vulnerable biological targets with considerable sizes and all other factors that violate the analysable theories. The beam-induced movement, ionizing radiation, accumulating or fluctuating charges, etc. can all contribute to irreversible physical changes of the samples. Thermal motion of molecules is inevitable even if the cryo-EM grids are preserved under liquid nitrogen temperature. In addition to the overall movement of large patches induced by electron beams, thermal motion of individual water molecules can be severely enhanced by electron beams. Therefore, the 'signals' of water molecules are turned into pure noise because of beam-enhanced thermal motion. This eventually brings more troubles for later motion correction. The thickness of biological samples can be a problem under the assumption of weak-phase object approximation. The defocus gradient, dynamic scattering and effective contrast all become considerably serious problems on large samples (e.g. >100 nm), particularly when the required resolution becomes higher. At a lower voltage (e.g. 200 kV instead of 300 kV), the effect is even stronger. Not only the structural information of the specimen itself is attenuated, but also the electron beams can be affected by the structural changes and charging effect during imaging, thus further reducing the image quality. These factors are nothing related to detectors. We believe there is still vast space to be improved during the pre-detection process.

Once scientists are able to make a clear mathematical description of the physical process during cryo-EM imaging, it might be very likely to recover as much information as possible by well-designed computational approaches afterwards. In principle, it would be possible to recover almost all structural information from the highly noisy, severely modulated and radiation-damaged cryo-EM images. If not, it is mainly because we have not established a perfect physical model with a perfect estimation of all required parameters. That is exactly our future endeavour. We hope in the future the major uncertainty left will be the quantum effect of biological molecules during TEM imaging. But that is probably far beyond the objective we can think about at current stage.

(III) Computation

We will discuss possible computational approaches separately.

5.2. Manifold perspective for dynamic structure analysis by cryo-EM

In order to get continuously moving structures from cryo-EM data, we must make a clear picture on the intrinsic property of functional conformational changes of biological complexes. Two simplified mathematical models can be applied to deal with the heterogeneous datasets. In the classical model, the conformational changes are regarded as movement of multiple rigid bodies. In the other more generalized model, the conformational changes are interpreted as elastic deformation. Both can be useful in different cases. The first model is simpler for programing. The latter model seems a bit more complicated, but represents the generalized case. To be more precise, the latter actually incorporates the multibody model as one special case.

In the case of multiple bodies with discrete conformations or assemblies, conventional classification methods can be very useful. Although there are a variety of methods that already exist to deal with conformational heterogeneity, the concepts are all deeply rooted in 3D analysis. Most of the clustering methods are essentially designed to tackle the discrete conformations. The entire dataset is typically classified into several individual classes and then refinement is simply performed on each dataset in the same way as homogenous dataset. Examples include separation of the 8-fold and 9-fold symmetric chaperonins[63] and dynactin complex with the p150 projection density.[93] The popular program RELION has shown its great power to sort out many discrete conformations and is currently widely used in the cryo-EM field.

Local refinement of the orientations for each individual domain will also be very helpful. The goal can be archived by applying a mask on the interesting region. When the targeted local density is big enough, the calculated local projection and experimental 2D images of the entire molecule can

be accurately aligned to improve the local resolution. However, when the interesting region becomes smaller and smaller, false peaks will emerge because of the extremely noisy image and the non-zero correlation with the projections from other uninteresting regions of the 3D map. This significantly affects the accuracy and correctness of the local orientation. The signal subtraction approach was proposed to partially address this problem.[105,106] The approach is implicitly based on the assumption that the alignment of the major part of a flexible structure is accurate enough in the presence of small local flexibility. The cross-correlation (or any other types of similarity functions) can then be successfully maximized (or minimized) on the local region by avoiding its correlation with the signals of the major region. This approach has been successfully applied in both cases of local flexibility[107,108] and symmetry-mismatch reconstruction.[71,105] In practice, each individual subunit or domains or any other types of user defined local regions could be iteratively improved by subtraction and re-subtraction of different local regions to further improve the local alignment accuracy. Overfitting of local alignment is probably inevitable when an interesting region becomes too small. In such a case, local classification without alignment might generate better results, but still not fully guaranteed.

More realistically, no protein complex can be a perfect multibody system; they are intrinsically elastic and eternally dynamic. For such a reason, signal subtraction approach is not suitable in the general cases of continuous conformational changes as the implicit assumption of multibody does not hold any more. The subtraction approach becomes even challenging when the complexity of the system become huge and completely uncontrollable. The concept of manifold can be borrowed from mathematics to perfectly describe such conformational changes of protein complexes with some modifications. A manifold in mathematics can be regarded as the generalized concept of curves or curved surfaces and locally resembles Euclidean space. The concept of manifold itself is nothing new and can be traced back to the beginning of 19th century. The application of this basic mathematical concept was first introduced to natural science by physicists to study the curved space-time of the universe. The manifold methods in image processing was only proposed less than 20 years ago.[109] In cryo-EM, it only began to draw the attention from structural biologists a few years ago.[110] The concept can now be used to well describe the continuous motion of biological macromolecules.

It should be noted that even if the basic concept is similar, there are always different perspectives to understand the same natural phenomena. The manifold concept that describes protein motion in cryo-EM can be interpreted in several apparently distinct ways. In one point of view, a cryo-EM image with N pixels (typically 10^6–10^9) is mapped to a single point in an N-dimensional Euclidian space. Different points in this high-dimensional space represent different conformations. The continuous motion of a protein can then be regarded as a smooth manifold embedded in this N-dimensional Euclidian space. This perspective has both advantages and disadvantages. The concept is clear, simple and makes sense. Machine learning algorithms for manifold embedding in high-dimensional space is well developed in computer science and there are many available resources to implement it as well. However, one big issue is that some very useful information of cryo-EM images is discarded or reduced when a 2D image is completely discretized as a single point in N-dimensional space. For example, the local continuity and smoothness in the original 2D image are both lost. It therefore becomes very challenging to apply geometric constraints among certain regions of a density map in high-dimensional space. Dimensional reduction approaches have to be exploited to handle these problems. The most important disadvantage is that the manifold embedding approach in high-dimensional space lacks a good sense of 'geometrical intuition' for human beings to understand the natural motion of proteins.

5.3. Probabilistic conformational space

We would rather suggest a different perspective, the probabilistic conformational space (PCS), which is also based on the manifold approach for cryo-EM structural analysis in the case of continuous motion. It is essentially a low-dimensional manifold assigned with a probability function for each point of time, plus some additional structures (to be discussed later). Multiple time axes (say 4–10) are necessary to describe more complicated motions. Typically, the time axes are continuously and smoothly curved. Such perspective does not necessarily require the embedding of manifold into a high-dimensional space, but tries to reveal the implicit properties of the manifold itself, such as local curvature and deformation parameters. We believe the best approach is to directly build a target function between the PCS and the observed cryo-EM images, instead of classification, sub-classification, sub-sub-classification, and so on.

If we regard the dynamic structure determination by cryo-EM as a unified target, then the aim is to assign each observed 2D image with a group of parameters $(x, y, \alpha, \beta, \gamma, t)$ by minimizing the following target function:

$$\min \sum_i^N (P_{\alpha,\beta,\gamma} \cdot D_t \cdot CTF \cdot A_3 \cdot \rho(x,y,z) - O_{tx,ty,\theta} \cdot A_2 \cdot I_i(x,y)).$$

Where ρ is the density map in 3D and I_i is the i-th observed (experimental) 2D image from the entire N particles; A_2 and A_3 are symmetry operators (including local symmetry) in 2D and 3D respectively; $O_{tx,ty,\theta}$ is a translation and rotation operator in 2D; t is a point of time in multiple dimensions, including 1D and isolated points; CTF refers to the contrast transfer function applied on the symmetrized 3D map; D_t is a time-dependent deformation operator with a probability assigned to

each point of time; $P_{\alpha,\beta,\gamma}$ is the projection operator with the Euler angles α,β,γ. Note all the necessary information to reveal the PCS is implicitly implied in ρ and D_t.

This point of view can also be interpreted as a generalized concept of multibody approach in cryo-EM data analysis, in the case of infinitely many rigid bodies, each of which is infinitely small. It is apparently not feasible to divide a map into too many rigid bodies. But the concept of PCS is competent to deal with such problem. All the rigid bodies along the time axes share 100% mutual information no matter how small they are. This extremely useful information can be implicitly estimated and fully exploited in the PCS. The concept of 3D reconstruction is then extended to the question of how to estimate the PCS. The classical concept of linear projections of a 3D density map is also generalized to be a non-linear integration in PCS. We regard this integration as non-linear because the projection paths are curved. The classification question will no longer exist, because discrete conformations are simply regarded as a special probability distribution, the Dirac delta function and its variants.

In practice, the reconstruction of the low-dimensional manifold and estimation of probability distribution might not be accurate due to the very low SNR in cryo-EM datasets. An optimized weighting scheme will be very helpful for better reconstruction. We believe there will be great opportunities for such optimizations based on the PCS concept. The following properties will be very useful if they are taken into consideration and treated properly.

I) Density conservation

This means that the integration of the density map for a specific molecule is constant for any conformation. The assumption might not be useful for the whole molecule as the cryo-EM images are always scaled and normalized to the same level in practical data processing. The real power is that any partial structure holds this property as well, which can be extremely useful to constrain the relative scales of local density maps among all individual regions. In theory, each region can be infinitely small. In practice, because of the difficulty in extracting useful information from very small region in the highly noisy cryo-EM maps, each region may require at least several residues.

II) Homeomorphism and homotopy equivalence

This property is based on the fundamental nature of proteins whose conformational changes are usually smooth and reversible (we do not discuss sub-atom or quantum effect at current stage). For any two conformations of exactly the same protein complex, there exists a continuous mapping that preserves all the topological properties and has an inverse mapping. And for each point in a protein density map, there also exists a smooth path (or curve) from any conformation to another. To be more concrete and specific, a protein can smoothly change its conformation from any state A to any other state B and *vice versa* (not necessary to be exactly the inverse path). All conformations of a protein are topologically identical, recyclable and reproducible in a smooth manner. This mathematical model also implies that once a protein conformation has ever existed in the past, it is likely to appear again after sometime, no matter how long it will take and how small the probability is.

III) Isometry

We define two types of isometric deformation. One is called 'strong isometry' to describe the fact that the lengths of all chemical bonds of proteins are relatively conserved during the dynamic conformation changes. Therefore, the distance between two atoms is a constant by tracing the backbone in any conformation. The chemical bond vibration at quantum level can be completely ignored at the resolution of 2 Å–3 Å by using current cryo-EM method. This kind of isometry can be concretely interpreted as 'invariant sequence distance', which is very different from the variable Euclidian distance between two atoms in 3D dimensional space during a conformational change. However, for a fully folded protein, the Euclidian distance between two atoms in a local region can also be roughly confined, though not as strict as the sequence distance. We call this property the 'weak isometry'.

These aforementioned properties are mathematically very useful for analysing dynamic structures by cryo-EM. They can significantly simplify the computational approaches if properly exploited.

5.4. Tripartite information combination: experiment, computation and theory

We have briefly discussed the factors that may affect the dynamic structure analysis by cryo-EM. It has been clearly demonstrated in both history and recent cryo-EM development that experiment, computation and theory are all critically important in practice. They will all contribute to dynamic structural analysis by cryo-EM. Experimentally, we need decent functional evidences or proofs from biochemistry, molecular biology, cell biology, genetics, and probably physiological analysis as well. For instance, each nucleotide binding state of dynein motor itself is very complicated and flexible;[107] the intermediate conformational changes from one state to another are even more difficult to be captured. It would be more efficient to lock or constrain such flexibility to some extent by using biochemical approaches as initial investigation. It is also very useful to investigate the processive movement by light microscope. Although the resolution is not as good as cryo-EM, light microscope is almost free of radiation damage and can be used to visualize the living cell or tissue at nanometer resolution with a much larger field of view. This useful information is far beyond the reach of cryo-EM at the current stage. With the information being accumulated, it will then gradually become realistic to study the whole functional

cycle. All types of experimental data can be converted to well-designed structural constraints in the following computational analysis. Computational information itself is contributed by two distinct resources, the cryo-EM reconstruction and MD simulation. These two can be well combined to generate better results. Cryo-EM maps can be used to constrain and regulate the MD simulation, while the latter can help cryo-EM predict the most likely motion and deformation involved in the continuous conformational changes. However, none of such constraints can be useful without a solid support from successful theories. Unlike mathematics and physics, the theoretical analysis in biological systems is so far still the weakest point, partially because of their huge complexity. Therefore, when we are being intoxicated by solving all kinds of fancy biological structures, more and more efforts are required to establish reliable and useful theories for daily experimental and computational analysis. A deep analytical understanding of the complicated biological imaging will be very helpful, particularly the extremely large samples. Ultimately, we will have to combine the tripartite information in order to obtain dynamic structures: experiments, computation and theories.

6. Summary and discussion

Cryo-EM is now becoming a thriving and prosperous field. We will soon be able to reveal the 3D structures from almost all kinds of protein complexes everywhere in the world and even from outer space. Scientists have longed to realize such a dream for hundreds of years since the discovery of the basic unit of life, the cell. The cryo-EM technique, after several decades of arduous development, has now brought golden chances to make the dream possible and realistic. Scientists have never given up the strong desire to expand our knowledge on more detailed structures of life. Without the help of light microscope, we human beings could have never been able to take a glimpse of the silhouette of a single cell. Electron microscope turned out as a more powerful tool that allowed us to directly visualize the exquisite organelles of cells at much higher resolution. At the beginning, TEM was only suitable for imaging the stained biological samples at 2D level. Then the birth of cryo-EM created great chances to visualize native structures at 3D level. But the resolution used to be very low compared to other methods such as X-ray crystallography and NMR. The recent application of direct electron detector rapidly changed the rules of this game and reshuffled the whole field of structural biology. Now, we can routinely determine 3D structures at near-atomic resolution which allows *de novo* model building as good as crystallography. And much beyond, we can now easily classify many distinct conformations of protein complexes in solution. We clearly see how cryo-EM is being developed as a progressively powerful tool for structural studies. Needless to say, the development of cryo-EM will keep on exploding and never stop. We are seeing such a golden time when many opportunities for multidisciplinary researches are being created. At this circumstance, the cryo-EM has been entrusted with an important mission to reveal the real mystery of life at atomic details, probably as dynamic machinery in the future. Such a mission is so important that we scientists will probably not rely on any single evidence by single approach. The tripartite information from experiments, computation and theories must all be unified as a whole to answer really fundamental questions. The paradigm of structural biology is doomed to be reformed again and again. We long to visualize the atomic details of biological macromolecules as dynamic machinery inside the cell while they are functioning. This must come true in the future.

References

[1] Marton L 1934 *Nature* **133** 911
[2] Glaeser R M 2016 *Methods Enzymol.* **579** 19
[3] Cheng Y 2015 *Cell* **161** 450
[4] Dubochet J, Adrian M, Chang J J, *et al.* 1988 *Q. Rev. Biophys.* **21** 129
[5] Zhang K, Zhang Y, Hu Z, *et al.* 2010 *Acta Biophys. Sin.* **26** 533
[6] Mcmullan G, Faruqi A R and Henderson R 2016 *Methods Enzymol* **579** 1
[7] Zhang K and Liu Z 2018 *Sci. China-Life Sci.* **61** 368
[8] Fujiyoshi Y 2011 *J. Electron. Microsc.* (Tokyo) **60** Suppl 1 S149
[9] Frank J 2016 *Microscopy* (Oxford) **65** 3
[10] Egelman E H 2015 *Arch. Biochem. Biophys.* **581** 54
[11] Baumeister W, Grimm R and Walz J 1999 *Trends Cell Biol.* **9** 81
[12] Mentes A, Huehn A, Liu X, *et al.* 2018 *Proc. Natl. Acad. Sci. USA* **115** 1292
[13] Zhang R, Alushin G M, Brown A, *et al.* 2015 *Cell* **162** 849
[14] Zhang Y, Wang W, Chen J, *et al.* 2013 *Structure*
[15] Pang X, Fan J, Zhang Y, *et al.* 2014 *Dev. Cell* **31** 73
[16] Egelman E H 2007 *J. Struct. Biol.* **157** 83
[17] Kabachinski G and Schwartz T U 2015 *J. Cell Sci.* **128** 423
[18] Fernandez-Martinez J, Kim S J, Shi Y, *et al.* 2016 *Cell* **167** 1215
[19] Hu B, Lara-Tejero M, Kong Q, *et al.* 2017 *Cell* **168** 1065
[20] Sun S, Zhang K, Xu W, *et al.* 2009 *Prog. Biochem. Biophys.* **36** 729
[21] Nicastro D, Frangakis A S, Typke D, *et al.* 2000 *J. Struct. Biol.* **129** 48
[22] Davies K M, Strauss M, Daum B, *et al.* 2011 *Proc. Natl. Acad. Sci. USA* **108** 14121
[23] Zhao G P, Perilla J R, Yufenyuy E L, *et al.* 2013 *Nature* **497** 643
[24] Wan W, Kolesnikova L, Clarke M, *et al.* 2017 *Nature* **551** 394
[25] Liao M F, Cao E, Julius D, *et al.* 2013 *Nature* **504** 107
[26] Bai X C, Fernandez I S, Mcmullan G, *et al.* 2013 *Elife* **2** E00461
[27] Scheres S H 2012 *J. Mol. Biol.* **415** 406
[28] Scheres S H 2012 *J. Struct. Biol.* **180** 519
[29] Brilot A F, Chen J Z, Cheng A, *et al.* 2012 *J. Struct. Biol.* **177** 630
[30] Ruska E 1987 *Biosci. Rep.* **7** 607
[31] De Rosier D J and Klug A 1968 *Nature* **217** 130
[32] Crowther R A, Amos L A, Finch J T, *et al.* 1970 *Nature* **226** 421
[33] Brenner S and Horne R W 1959 *Biochim. Biophys. Acta* **34** 103
[34] Hall C E, Jakus M A and Schmitt F O 1945 *J. Appl. Phys.* **16** 459
[35] Huxley H E and Zubay G 1961 *J. Biophys. Biochem. Cytol.* **11** 273
[36] Henderson R and Unwin P N 1975 *Nature* **257** 28
[37] Taylor K A and Glaeser R M 1974 *Science* **186** 1036
[38] Dubochet J and Mcdowall A W 1981 *J. Microsc.* **124** 3
[39] Dubochet J, *et al.* 1982 *J. Microsc.* **128** 219
[40] Adrian M, Dubochet J, Lepault J, *et al.* 1984 *Nature* **308** 32
[41] Henderson R, Baldwin J M, Ceska T A, *et al.* 1990 *J. Mol. Biol.* **213** 899
[42] Frank J 1975 *Ultramicroscopy* **1** 159
[43] Frank J, Verschoor A and Boublik M 1981 *Science* **214** 1353
[44] Frank J, Goldfarb W, Eisenberg D, *et al.* 1978 *Ultramicroscopy* **3** 283
[45] Frank J 1981 *Ultramicroscopy* **6** 343
[46] Kuhlbrandt W, Wang D N and Fujiyoshi Y 1994 *Nature* **367** 614
[47] Murata K, Mitsuoka K, Hirai T, *et al.* 2000 *Nature* **407** 599

[48] Wang D N and Kuhlbrandt W 1991 *J. Mol. Biol.* **217** 691
[49] Henderson R 1995 *Q. Rev. Biophys.* **28** 171
[50] Henderson R 1992 *Ultramicroscopy* **46** 1
[51] Van Heel M and Frank J 1981*Ultramicroscopy* **6** 187
[52] Harauz G and Ottensmeyer F P 1983 *Ultramicroscopy* **12** 309
[53] Van Heel M 1987 *Ultramicroscopy* **21** 111
[54] Radermacher M, Wagenknecht T, Verschoor A, *et al.* 1987 *J. Microsc.* **146** 113
[55] Radermacher M, Wagenknecht T, Verschoor A, *et al.* 1986 *J. Microsc.* **141** Rp1
[56] Van Heel M, Harauz G, Orlova E V, *et al.* 1996 *J. Struct. Biol.* **116** 17
[57] Marabini R, Masegosa I M, San Martin M C, *et al.* 1996 *J. Struct. Biol.* **116** 237
[58] Ludtke S J, Baldwin P R and Chiu W 1999 *J. Struct. Biol.* **128** 82
[59] Grigorieff N 1998 *J. Mol. Biol.* **277** 1033
[60] Heymann J B 2001 *J. Struct. Biol.* **133** 156
[61] Li X M, Mooney P, Zheng S, *et al.* 2013 *Nat. Methods* **10** 584
[62] Faruqi A R and Henderson R 2007 *Curr. Opin. Struct. Biol.* **17** 549
[63] Zhang K, Wang L, Liu Y X, *et al.* 2013 *Protein Cell* **4** 432
[64] Huo Y W, Hu Z J, Zhang K, *et al.* 2010 *Structure* **18** 1270
[65] Mcmullan G, Chen S, Henderson R, *et al.* 2009 *Ultramicroscopy* **109** 1126
[66] Zhang X, Jin L, Fang Q, *et al.* 2010 *Cell* **141** 472
[67] Grigorieff N and Harrison S C 2011 *Curr. Opin. Struct. Biol.* **21** 265
[68] Cheng L P, Sun J C, Zhang K, *et al.* 2011 *Proc. Natl. Acad. Sci. USA* **108** 1373
[69] Yang C W, Ji G, Liu H R, *et al.* 2012 *Proc. Natl. Acad. Sci. USA* **109** 6118
[70] Cheng L P, Huang X X, Li X M, *et al.* 2014 *Virology* **450–451** 174
[71] Liu H R and Cheng L P 2015 *Science* **349** 1347
[72] Li X M, Zheng S Q, Egami K, *et al.* 2013 *J. Struct. Biol.* **184** 251
[73] Zernike F 1955 *Science* **121** 345
[74] Danev R and Nagayama K 2001 *Ultramicroscopy* **88** 243
[75] Dai W, Fu C, Raytcheva D, *et al.* 2013 *Nature* **502** 707
[76] Murata K, Liu X, Danev R, *et al.* 2010 *Structure* **18** 903
[77] Danev R, Buijsse B, Khoshouei M, *et al.* 2014 *Proc. Natl. Acad. Sci. USA* **111** 15635
[78] Danev R and Baumeister W 2016 *Elife* **5**
[79] Danev R, Tegunov D and Baumeister W 2017 *Elife* **6**
[80] Khoshouei M, Radjainia M, Baumeister W, *et al.* 2017 *Nat. Commun.* **8** 16099
[81] Liang Y L, Khoshouei M, Radjainia M, *et al.* 2017 *Nature* **546** 118
[82] Fan X, Zhao L Y, Liu C A, *et al.* 2017 *Structure* **25** 1623
[83] Sigworth F J 1998 *J. Struct. Biol.* **122** 328
[84] Scheres S H, Gao H, Valle M, *et al.* 2007 *Nat. Methods* **4** 27
[85] Kimanius D, Forsberg B O, Scheres S H, *et al.* 2016 *Elife* **5**
[86] Punjani A, Rubinstein J L, Fleet D J, *et al.* 2017 *Nat. Methods* **14** 290
[87] Grant T and Grigorieff N 2015 *Elife* **4** E06980
[88] Scheres S H W 2014 *Elife* **3**
[89] Zheng S Q, Palovcak E, Armache J P, *et al.* 2017 *Nat. Methods* **14** 331
[90] Gao Y A, Cao E, Julius D, *et al.* 2016 *Nature* **534** 347
[91] Wade R H 1992 *Ultramicroscopy* **46** 145
[92] Scherzer 1949 *J. Appl. Phys.* **20** 20
[93] Urnavicius L, Zhang K, Diamant A G, *et al.* 2015 *Science* **347** 1441
[94] Zhu L, Wang X X, Ren J S, *et al.* 2015 *Nat. Commun.* **6** 8316
[95] Brown A, Shao S, Murray J, *et al.* 2015 *Nature* **524** 493
[96] Zhang K 2016 *J. Struct. Biol.* **193** 1
[97] Zhang X and Zhou Z H 2011 *J. Struct. Biol.* **175** 253
[98] Chen Y, Zhang Y, Zhang K, *et al.* 2016 *J. Struct. Biol.* **195** 49
[99] Russo C J and Passmore L A 2014 *Science* **346** 1377
[100] Razinkov I, Dandey V, Wei H, *et al.* 2016 *J. Struct. Biol.* **195** 190
[101] Yu G M, Li K P and Jiang W 2016 *Methods* **100** 16
[102] Glaeser R M 2016 *Nat. Methods* **13** 28
[103] Jain T, Sheehan P, Crum J, *et al.* 2012 *J. Struct. Biol.* **179** 68
[104] Chen B, Kaledhonkar S, Sun M, *et al.* 2015 *Structure* **23** 1097
[105] Huiskonen J T, Jaalinoja H T, Briggs J A, *et al.* 2007 *J. Struct. Biol.* **158** 156
[106] Ilca S L, Kotecha A, Sun X, *et al.* 2015 *Nat. Commun.* **6** 8843
[107] Zhang K, Foster H E, Rondelet A, *et al.* 2017 *Cell* **169** 1303
[108] Bai X C, Rajendra E, Yang G, *et al.* 2015 *Elife* **4**
[109] Roweis S T and Saul L K 2000 *Science* **290** 2323
[110] Dashti A, Schwander P, Langlois R, *et al.* 2014 *Proc. Natl. Acad. Sci. USA* **111** 17492