# Image processing and computing in structural biology

Jiang, L.

## Citation

Jiang, L. (2009, November 12). *Image processing and computing in structural biology*. Retrieved from https://hdl.handle.net/1887/14335

# Chapter 1 Introduction

## 1.1 Structural biology, cryo-EM and image processing

Structural biology is a branch of life science which focuses on the structures of biological macromolecules, investigating what the structure looks like and how alterations in the structure affect the biological functions.

This subject is of great interest to biologists because macromolecules carry out most of the cellular functions, which exclusively depend on their specific three-dimensional (3D) structure. This 3D structure (or tertiary structure) of molecules depends on their basic sequence (or primary structure). However, the 3D structure cannot be calculated directly from the sequence. In order to understand the complicated biological processes at the cellular level, it is therefore essential to determine the 3D structure of molecules.

The research of structural biology is intimately relevant to human health. A healthy body requires the coordinated action of billions of indispensable proteins. Each protein has a unique molecular shape that exactly fits its particular function. Determining the 3D structures of key proteins and viruses at the atomic level is an important and often vital strategic step to find the reasons behind many human diseases. This step can help us clarifying the role of the shape of proteins and their complexes (including viruses) in health and disease. Structure determination of viruses is thus a persistent hot topic of research. Figure 1 shows an example of the structure of cytoplasmic polyhedrosis virus (CPV).

Biomolecules, even the so called macromolecules, normally have a tiny size measured in tens of nanometers or less. Such molecules are too small to see with the light microscope. The techniques that can reach atomic resolution mainly include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron cryo-microscopy (cryo-EM). X-ray crystallography has been able to tackle large complexes, but is limited to complexes that can form crystals and NMR is only suitable for smaller macromolecules and complexes. This leaves a large number of challenging structures that cannot be resolved using the X-ray and NMR techniques. Especially for large complexes that resist crystallogenesis, electron cryo-microscopy (cryo-EM, or cryo-electron microscopy) is a viable alternative. This technique is a combination of transmission electron microscopy (TEM) and cryo-equipment.

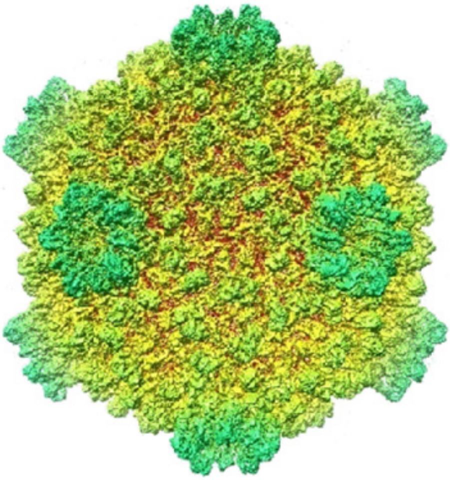

Figure 1. *Structure of cytoplasmic polyhedrosis virus (CPV) with a resolution of 3.88Å, obtained by using cryo-EM single particle reconstruction (EMDataBank id: EMD-1508, Yu et al., 2008), the highest resolution achieved so far by using cryo-EM. CPV belongs to the virus family of Reoviridae. Reoviridae can affect the gastrointestinal system (e.g. Rotavirus) and respiratory tract. Reovirus infects humans often; it is easy to find Reovirus in clinical specimens.*

TEM is suitable for looking into the molecules in atomic detail. The cryo-EM technique provides a way to observe the real “native state” structure as it exists in solution by freezing the samples extremely fast in a layer of vitreous ice. Freezing reduces electron damage, allowing a higher dose of electron exposure to gain better signal-to-noise ratio (SNR) images. Cryo-EM is thus the obvious choice to study large biomolecular complexes. The enormous potential of cryo-EM in biological structure determination has already been realized since the early 1990’s (for a review, see R. Henderson, 2004).

Transmission electron microscopy has two modes available: image mode and diffraction mode (Figure 2). The 3D structure of a molecule cannot be obtained directly by TEM, but must be reconstructed using computational methods. In structural biology, two new methods using TEM are still developing: three-dimensional cryo-electron microscopy (3DEM, also known as single particle reconstruction) and electron diffraction (or electron crystallography). 3DEM uses the image mode of TEM, and electron crystallography uses the diffraction mode.

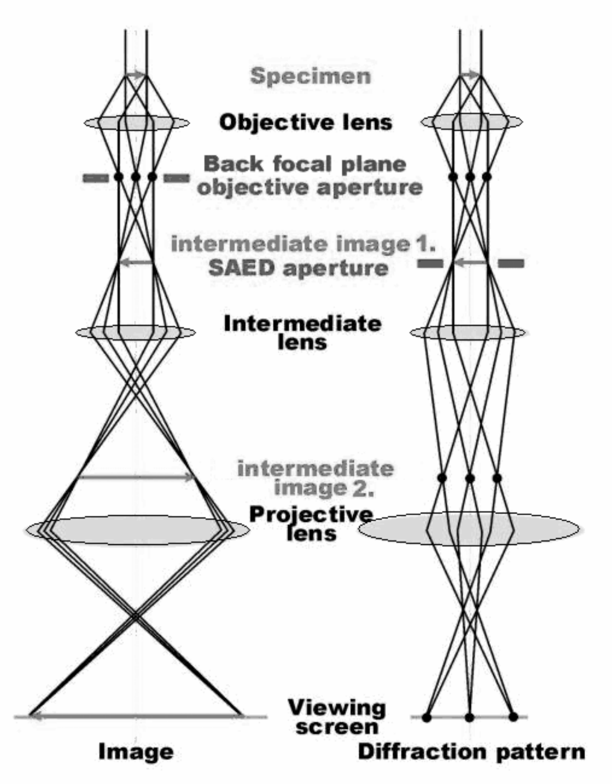

Figure 2. *Image and diffraction modes of transmission electron microscopy (Williams & Carter, 1996). SAED: Selected Area Electron Diffraction. The objective lens forms a diffraction pattern in the back focal plane and generates an image in the image plane (intermediate image 1.). Diffraction pattern and image are both present in TEM. The intermediate lens decides which of them appears in the plane of the second intermediate image (intermediate image 2.) and is projected on the viewing screen. It is easy to switch between image and diffraction modes by adjusting the intermediate lens.*

3DEM requires the reconstruction of a macromolecular 3D model from large amount of noisy 2D projection images (e.g. Figure 3A) of a specimen. Electron crystallography is a method to gain and analyze diffraction patterns (images in Fourier space, e.g. Figure 3B) of crystals (1D, 2D or 3D crystals) for the reconstruction of 3D structure in Fourier space, similar as the technique used in X-ray crystallography. To see the true 3D structure underlying the recorded data, sophisticated image processing and computing are indispensable for either method.

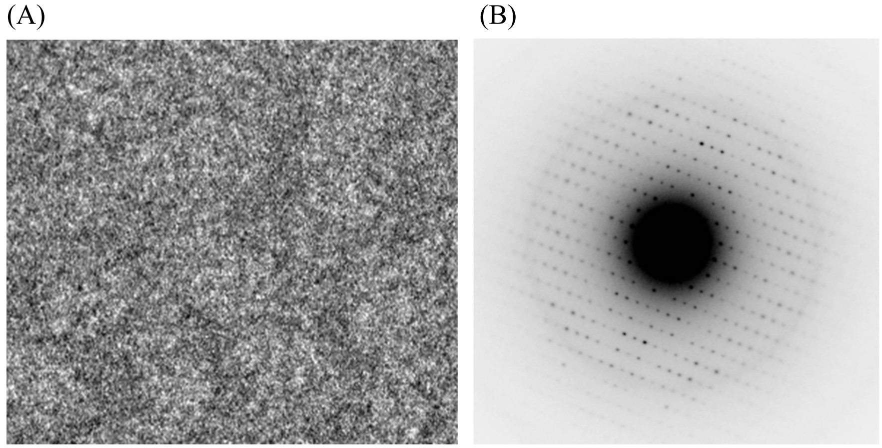

Figure 3. *Examples of a micrograph of single particles (A) and of a electron diffraction pattern of frozen nano-crystal of a protein (lysozyme) (B).*

Image processing and computing methods are essential for solving structures of macromolecules. Both X-ray crystallography and the NMR require the power of computing. Images from electron microscopy (Figure 3) also need image processing to reconstruct the 3D structure. For EM images, computing methods for 3DEM single particle reconstruction is still developing rapidly. They utilize many computer image processing techniques.

In image mode, TEM is affected by the instrumental aberration problem and the image is distorted by the contrast transfer function (CTF). Aberration correction of the CTF is one of the major tasks of image processing in 3DEM. In diffraction mode, diffraction patterns from electron microscopy actually represent a Fourier lattice. Analysing this data also needs complicated procedures of image processing and computing.

Electron crystallography of 3D crystals is not new in inorganic chemistry and material science, but it is a new in biochemistry. There is no existing way to obtain a 3D structure from the diffraction images of 3D protein crystals (though there are a few successful cases with 2D and 1D crystals). But in theory, a set of random diffraction images from one species of 3D protein crystals may include sufficient information to reconstruct their atomic structure. No matter which methods are to be used, complicated image processing procedures and time consuming computing are indispensable to calculate a 3D structure from the EM micrographs.

This thesis mainly focuses on the image processing techniques of 3DEM and electron crystallography, and solves biological problems based on the 3D structures I determined.

## 1.2 Nano-techniques in structural biology, X-ray, NMR, electron diffraction and 3DEM

X-ray diffraction, NMR spectroscopy, and 2D/1D electron crystallography involve measurements of vast numbers of identical molecules at the same time. Most of the solved atomic structures use micro-crystals and X-ray diffraction. In a crystal, all molecules are in the same conformation and binding state. Their uniform orientation and ordered arrangement enable the X-ray diffraction.

The wavelength of the electron beam generated in a TEM is much shorter than that of the radiation which is usually used in X-ray crystallography. E.g. for 300KeV TEM, the wavelength is ~0.019Å; for 200KeV, ~0.025Å. X-rays used for atomic structure determination have wavelengths between 2 Å and 0.5 Å. Theoretically, electrons diffraction therefore has a higher resolution limit than X-ray diffraction.

But electron diffraction suffers from the dynamic diffraction problem, caused by the strong interactions between electrons and the matter. Only single layer crystals (2D crystal) or helical arrays (1D crystals) have been investigated successfully with electron diffraction. In this thesis, nano-crystals (3D protein crystals with nano-scale size) and the new technique of precession of the electron beam were used to reduce dynamic scatter to acquire the electron diffraction patterns (e.g. Figure 3B).

In the current practice in electron diffraction a single nano-crystal must be selected in image TEM mode and then the microscope must be switched to diffraction mode. This is not possible in X-ray diffraction, limiting this technique to the study of the micro-crystals or powders of nanocrystals. Another apparent advantage of using nano-crystals is: it is much easier to grow nano-crystals than to obtain micro-crystals with micrometer-scale size (Georgieva et al., 2007).

In high-resolution 3D EM single particle reconstruction, crystals are not necessary. Particles embedded in vitreous ice can have random orientations and arrangements. Very small amounts of sample are required for a 3D reconstruction, compared to the amount required for growing a crystal. Besides, in 3DEM, structural homogeneity or integrity is more important than purity, as opposed to X-ray crystallography and NMR, in which sample purity is essential (Zhou, 2008).

Generally speaking, both different experimental and computational methods have their advantages and disadvantages:

Advantages of X-ray crystallography method:

- Well-established techniques and software
- Highest atomic resolution structure achieved

Disadvantages of X-ray crystallography method:

- Difficult to grow crystals
- Single conformation or binding state, as a result of the crystal constraints
- Difficult to solve in presence of disorder

Advantages of 3DEM:

- No need to crystallize
- No phase problem

- Small amount of materials needed
- Easy for large molecules, up to 2000 Å
- All “native” functional states in solution can be captured in principle

Disadvantages of 3DEM:

- large computational cost
- limited resolution, highest resolution thus far ~4 Å (Yu et al., 2008)
- less developed for different conformational states

Advantages of electron diffraction:

- Can handle nano-size crystals
- Growing nano-crystals is relative easier
- Small amount of materials needed
- Strong diffraction with matter at an atomic resolution
- Share lots of common knowledge with well-developed X-ray diffraction techniques

Disadvantages of electron diffraction:

- Dynamic scattering
- Electron beam damage
- Manual data acquisition is less automated

Other technologies such as powder diffraction and tomography are also relevant for structure determination, but only have limited applications due to the low resolution that can be achieved.

## 1.3 Basics of 3DEM single particle reconstruction

3DEM single particle reconstruction is the reconstruction of a macromolecular 3D structure from a set of cryo-EM projection images. In a micrograph (e.g. Figure 3A), the molecules exist in the form of single isolated particles, randomly distributing in a layer of vitreous ice. Thousands to hundreds of thousands of noisy images of individual molecules are needed to calculate the 3D structure.

Biomolecules are highly susceptible to radiation damage when exposed to the electron beam. In order to decrease the damage, images are obtained with a low dose of exposure and by using electron cryo-microscopy. Nevertheless, the technique results in extremely noisy images. Averaging method is needed to calculate a high signal-to-noise ratio (SNR) structure from these noisy images. All the molecules must have the same inner conformation to within the resolution limit of the reconstruction, otherwise the averaging is meaningless.

To start a single particle reconstruction, all we need is cryo-EM micrographs of randomly distributed particles and reconstruction software (e.g. IMAGIC, SPIDER, EMAN). A typical reconstruction process shows as Figure 4.

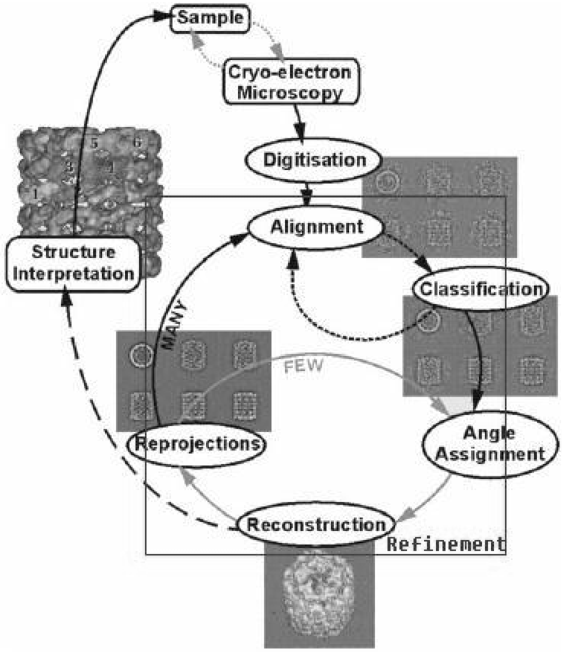

Figure 4. *The diagram of single particle reconstruction process of IMAGIC (van Heel et al., 2000)*

Generally speaking, image processing of 3DEM includes several steps:
(1). Single particle selection

Normally, only about 500 particles can be selected from a single EM micrograph, but a typical 3DEM reconstruction needs more than tens of thousand of particles. The micrographs are very noisy images due to the low dose exposure of cryo-EM. It is too difficult for a person to select large amount of particles required for 3DEM manually.

Automated or semi-automated software was created in need to accelerate this task. The software, Cyclops, designed in our group (Plaisier et al., 2007) includes an automated function to select single particles. Different methods are available in the software to locate the potential particles, such as the methods of local average, local variance and cross-correlation. In this thesis, I describe my contribution to this program in chapter 2.

(2). Filtering & centering

An optional pre-processing step is to filter the particle images with a low pass filter, erasing the high frequency noise (as well as a little information detail).

Mislocated intensities caused by the contrast transfer function (CTF) of the electron microscope have to be phase corrected at this step. That is so called CTF phase correction. Further amplitude correction will be needed in a later step for full CTF correction. In chapter 3 of this my thesis I discuss a novel approach to these corrections.

Particles are centered in several alignment cycles, in which the cross correlation between each individual image and the overall average image (of a given data set) is calculated.

(3). Classifying & averaging

A classification step is required to assign particles to different classes, in which the projections are assumed to be taken from the same view/angle. One of the classification methods is Multivariate Statistical Analysis (MSA) (van Heel et al., 2000). In this method, Principal Components Analysis (PCA) is applied to solve the problem caused by high noise in the images, after a time consuming procedure named multi-reference alignment (or reference supervised alignment). Another classification method is reference supervised classification if coarse starting model is available. A particle image is compared with all the reference images and is then assigned to the class corresponding to the most similar reference image.

Subsequently, an average image is calculated for each class to get high signal-to-noise ratio image. CTF amplitude correction is normally performed in this stage.

(4). 3D reconstruction

According to the common-line projection theorem, two different 2D projections of the same 3D object must have a 1D line projection in common. Relative Euler angles can be assigned for each average image in an angular reconstruction.

Once the Euler angles are assigned, average images are back projected to get a 3D model. This procedure is normally done in Fourier space, because every projection image is a section of 3D model in Fourier transform Back projection can be conveniently implemented by inserting the image in Fourier space and then transferring back to get the real space model.

(5). Refinement

The reconstructed 3D model resulting from the first iteration usually has a sub-optimal resolution. An iterative refinement aiming at higher resolution is then necessary. The rough model is re-projected in many directions, providing a set of reference images. Chapter 2 of my thesis describes an optimal sampling of rotational space to generate a minimal set of reference images with a maximal covering of potential orientations. The set of reference images is used in the subsequent iterative alignment and classification steps.

Refinement is the most time consuming step in 3DEM. For instance, on Pentium 4/1.6G/Linux PC,

1500 particles need ~5 hours per iteration; 2500 particles need ~11 hours per iteration. So how about 100,000 particles? And more particles if an even higher resolution is required? It may need days, weeks, or even longer. So, most state-of-art reconstructions are carried out on a parallel computing facility such as a supercomputer or a computer cluster.

Although there is a reasonably wide choice in software for 3D reconstruction (such as EMAN, SPIDER, IMAGIC, etc.), the method of 3DEM is still developing rapidly, since the cryo-EM technique started booming in the most recent 10 years. The main difficulties of this method are: low resolution, high noise, time consuming calculations and semi-automated software, still leave enough space for improvement.

## 1.4 Basics of electron diffraction and structural reconstruction

When the electron beam in a TEM passes through a thin (e.g. <100 nm) crystalline layer, the electrons scatter and interfere with each other and (if the microscope is set to the proper mode)a diffraction pattern can be observed on a fluorescent screen or be recorded on film, image plate (e.g. Figure 3B) or a CCD camera.

The constructive interference of the electrons observed as spots in the diffraction pattern can be expressed by the Bragg's law (Bragg, 1913):

$n\lambda=2d\cdot\sin\theta$,

Here, n is a given integer. λ is the wavelength of electrons. d is the spacing between the planes in the atomic lattice. θ is the angle between the incident beam and the scattering planes. Figure 5 explain both the constructive and destructive interferences.

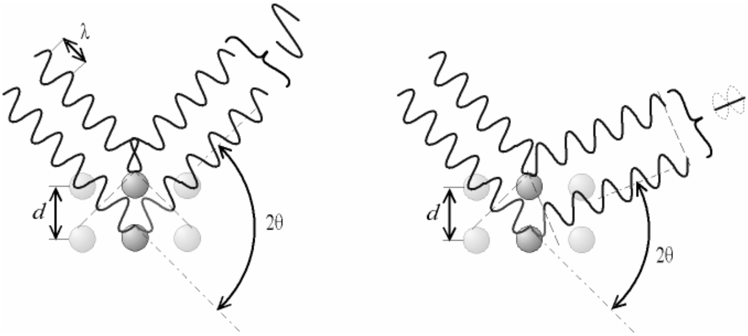

Figure 5. *According to the 2θ deviation, the phase shift causes constructive (left figure) or destructive (right figure) interferences. The interference is constructive when the phase shift is a multiple of 2π. (From Wikipedia)*

Theoretically, diffraction patterns are Fourier transformations of their projection images on the Ewald sphere. If the phases of the diffraction patterns from a crystal are known, these patterns are mathematically equivalent to the projection images, hence they can be used to reconstruct the atomic structure.

Electron diffraction is widely used in material science for analyzing the structure of metals and alloys. In structural biology, the application is still limited to, for instance, structure analysis of 2D and 1D crystals. Up to now, there is no existing way to obtain a 3D structure from the diffraction images of a 3D protein crystal. The difficulties mainly lie in:

(i) The mathematical equivalence between (phased) electron diffraction patterns and their corresponding projection structures are compromised by the multiple scattering of electrons (dynamic diffraction). Even when the thickness of the sample is less than 100 nm, dynamic diffraction still affects the data.

(ii) Protein crystals are susceptible to radiation damage caused by the electron beam. Some researchers are trying to solve the structure of 3D nano-crystals by using tilt series, which is similar to the technique of tomography in diffraction mode as is prevalent in X-ray crystallography. Unfortunately, this is not (yet) suitable for the beam-sensitive protein crystals with current electron detection methods.

(iii) Electron diffraction in TEM still needs lots of manual intervention. For example, locating the crystals in image mode and tilting the sample manually are time-consuming operations. Compared to highly automated X-ray diffraction experiments, electron diffraction is still extremely tedious.

In the research described in chapter 5, nano-crystals and the new technique of precession of the electron beam were used to reduce the dynamic diffraction problem. Clear electron diffraction patterns could be acquired for structure determination. To solve the atomic structure from the electron diffraction patterns of protein nano-crystals, following steps are required:

(i) Background removal and spot location

Firstly, center the diffraction images and remove the strong background caused by the undiffracted electron beam. A Patterson map can be used for centering. If a beam stop exists, its shadow should be taken into account. Then one needs to locate diffraction spots, extract their coordinates and calculate the intensities of the spots in the pattern.

(ii) Unit cell determination

Finding the unit cell parameters from randomly oriented diffraction patterns is essential for structure determination. Existing algorithms from X-ray crystallography and tilt series are not usable, as only single shots of crystals can be recorded, hence a new algorithm had to be created to deal with the multiple patterns with unknown orientation from multiple crystals.

(iii) Indexing

The randomly distributed orientation angles need to be determined, using the found unit cell in step two. The reflections of every electron diffraction image are thus indexed.

(iv) Intensity integration and subsequent steps in structure determination and refinement

When the indices and their corresponding locations on the diffraction pattern are known, methods from X-ray crystallography can be used to reconstruct the 3D spot lattices in reciprocal space. Phase recovery and iterative refinement are essential for determining the atomic structure.

## 1.5 Outline of this thesis

Chapter 2 to chapter 4 focus on single particle analysis, which includes both the methods employed in the single particle reconstruction and the practical 3DEM reconstruction of the macromolecular model of a 50S ribosomal complex. In **chapter 2**, new modules in cryo-EM, automated carbon masking and quaternion based rotation space sampling, are presented. The new modules were implemented and tested in Cyclops software. In **chapter 3**, a novel approximation method of CTF amplitude correction for 3D single particle reconstruction is described. This new method yields higher resolution models compared with to traditional CTF correction methods and shows better convergence in practice. **Chapter 4**, reports 3DEM reconstructions (with a highest resolution of 10Å) of macromolecular ribosomal complexes of stalled 50S ribosomal particles. They sow how Hsp15 rescues heat-shocked, prematurely dissociated 50S ribosomal particles. This 3DEM reconstruction project (the first project in my Ph.D research period) required reconstructing multiple asymmetric macromolecules. Until now, it is still very challenging work to determine EM models of asymmetric complexes at such resolutions.

In chapter 5 and 6, I describe progress in analysing the random electron diffraction images of 3D protein crystals. In **chapter 5**, the second main topic of my Ph D research, discusses a brand new approach to structure determination compared to the traditional X-ray and NMR technologies. A new algorithm to determine unit cells from a set of randomly oriented diffraction patterns is presented here. Unit cell determination is the first step to solve a structure in crystallography. **Chapter 6** describes the implementation of these algorithms and includes a user manual of the EDiff software, which is used for searching unit cell parameters and indexing well-oriented patterns.

Finally, **chapter 7** gives a summary and concludes with future perspectives of my research.

## References

Bragg, W.L. (1913). "The Diffraction of Short Electromagnetic Waves by a Crystal", *Proceedings of the Cambridge Philosophical Society*, 17, 43–57.

Georgieva, D.G., Kuil, M.E., Oosterkamp, T.H., Zandbergen, H.W., Abrahams, J.P. (2007). Heterogeneous crystallization of protein nano-crystals. *Acta Crystallogr. D* 63, 564-570.

Henderson R. (2004). Realizing the potential of electron cryo-microcopy. *Q. Rev. Biophys*. **37**, 3-13.

Plaisier J.R., Jiang L., Abrahams J.P., (2007). Cyclops: New modular software suite for cryo-EM. *J. Struct. Biol.* 157, 19-27.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., Patwardhan, A., (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* 33, 307-69.

Williams, D.B., Carter, C.B. (1996). Transmission electron microscopy: a textbook for materials science. New York: Plenum Press. ISBN 030645324X.

Yu, X., Jin, L., Zhou, Z.H., (2008). 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *NATURE* 453, 415-419.

Zhou, Z.H., (2008). Towards atomic resolution structural determination by single particle cryo-electron microscopy. *Curr. Opin. Struc. Biol.* 18, 218-228