Archives of Biochemistry and Biophysics 581 (2015) 39–48

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Review

Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data

J.M. Carazo [a,*], C.O.S. Sorzano [a,c], J. Otón [a], R. Marabini [b], J. Vargas [a]

[a] Biocomputing Unit, National Center for Biotechnology (CSIC), c/Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain
[b] Escuela Politécnica Superior, Universidad Autónoma de Madrid, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain
[c] Bioengineering Lab., Universidad CEU San Pablo, Campus Urb. Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, Spain

ARTICLE INFO

Article history:
Received 29 January 2015
and in revised form 11 May 2015
Available online 22 May 2015

Keywords:
Cryo electron microscopy
Image processing

ABSTRACT

The Transmission Electron Microscope provides two-dimensional (2D) images of the specimens under study. However, the architecture of these specimens is defined in a three-dimensional (3D) coordinate space, in volumetric terms, making the direct microscope output somehow "short" in terms of dimensionality. This situation has prompted the development of methods to quantitatively estimate 3D volumes from sets of 2D images, which are usually referred to as "three-dimensional reconstruction methods". These 3D reconstruction methods build on four considerations: (1) The relationship between the 2D images and the 3D volume must be of a particularly simple type, (2) many 2D images are needed to gain 3D volumetric information, (3) the 2D images and the 3D volume have to be in the same coordinate reference frame and (4), in practical terms, the reconstructed 3D volume will only be an approximation to the original 3D volume which gave raise to the 2D projections. In this work we will adopt a quite general view, trying to address a large community of interested readers, although some sections will be particularly devoted to the 3D analysis of isolated macromolecular complexes in the application area normally referred to as Single Particle Analysis (SPA).

Introduction

Our field of work is the experimental resolution of the three-dimensional structure of macromolecular complexes using the Transmission Electron Microscope (TEM)[1] under cryogenic condition, an area also known as cryo EM. Within this broad topic, we will focus on three-dimensional reconstruction techniques, which is one of the basic steps in the structural resolution process. Note that cryo EM is experiencing a profound "revolution" nowadays thanks to several key technological and methodological advancements, such as the advent of Direct Electron Detectors and new image processing methods. We refer to other contributions in this Special Issue to properly review the state of the art in this field, so that in the following we focus on the crucial step of how to obtain three-dimensional quantitative information from TEM images.

The search towards always richer information is intrinsic to the human being. Indeed, there are many situations in which a certain type of information is experimentally measured, but our real interest goes beyond these measurements and it pertains to another property "related" to them. In other words, we measure "something", but we are interested in "something else". In a very broad sense, these cases are usually referred to as "inverse problems", which can be expressed in a more formal way as

$$g = Hf \tag{1}$$

with g being our measurements, f being our desired property, and H describing the physical process that links our measurement with the desired property. Since we want to obtain f from g, we have to invert, or "reverse", H leading to

$$f = H^{-1}g \tag{2}$$

and thus the name of "inverse problems".

Quite naturally, our ability to obtain f from g will greatly depend on the inversion properties of H. In the case of Transmission Electron Microscopy (TEM), g refers to sets of 2D images collected at the microscope, f to the 3D structure of our specimen, and H conveys the detailed information on how the electron microscope interacts with the specimen under study, producing concrete sets of 2D images. Once H is known, we have to find the conditions under which H^{-1} can indeed be realized, first from a somehow

* Corresponding author. Fax: +34 913720112.

[1] *Abbreviations used:* TEM, Transmission Electron Microscope; SPA, Single Particle Analysis; FT, Fourier Transform; FFT, Fast Fourier Transform; ML, maximum likelihood; CTF, Contrast Transfer Function.

http://dx.doi.org/10.1016/j.abb.2015.05.003
0003-9861/

abstract mathematical perspective, and then as practically implemented in a computer.

The procedure described above is very general, and it applies not only to electron microscopy but to most areas of biomedical imaging. However, the work with isolated macromolecular complexes, normally referred to as Single Particle Analysis (SPA), introduces some crucial differences with respect to other imaging modalities. Indeed, in a typical biomedical imaging application in a clinical context, we have a well-defined and unique f, the patient, from whom a number of images (X-ray radiographies) are going to be collected in order to calculate a 3D map. However, a macromolecular complex is a very dynamic entity, so the probability is large to have in our sample under investigation not only one f, but a whole sets of different f's, corresponding to different conformational states, giving rise to a mixed population of g's. Clearly, the formulation above has to be extended to take into account this situation. Furthermore, a large number of applications in EM are characterized by uncertainties about the way images have been collected, besides always been affected by heavy noise.

This paper is organized in the following manner. In Section '2D images and 3D volumes: Basic relationships' we will review the basics of the way electrons interact with the specimen in the microscope, producing a 2D image. In practical terms, we will be dealing with the characterization of H. We will also address some of the most common strategies to collect sets of images. Section 'From 2D images to 3D volumes: Reconstruction methods' will then concentrate on ways to invert H, and these will be the different reconstruction methods. At this stage we will present the way the 3D reconstruction process is performed in practice, introducing the notion of a "3D reconstruction workflow", particularized to SPA; this topic will be covered in Section 'A typical 3D reconstruction workflow in SPA'. Quite naturally, any reconstruction process starts with a detailed characterization of the initial experimental images, which will then be addressed in Section 'Ch aracterizing the initial experimental images'. However, we have already indicated that the simple mathematical framework of $g = Hf$ has to be extended to accommodate for the conformational flexibility of macromolecular complexes, besides a large number of experimental uncertainties and noise. This topic will be covered in Section 'From 2D images to 3D volumes: A posteriori projection assignment and classification'. Further elaborating on extensions of the basic reconstruction framework, we will briefly discuss the case of more elaborated H's, typical of certain demanding applications; this will be covered in Section 'On more complicated relationships: When simplification breaks'. Finally, we will present a general discussion in Section 'Discussion and conclusion'.

2D images and 3D volumes: Basic relationships

In this Section we will address three main questions: (1) Which is the relationship between the 3D volume of the specimen under investigation and its associated 2D images?, (2), Is one image enough to obtain a 3D reconstruction, and if this is not the case, how many are needed? and, (3), In practical terms, how 2D images are collected?

A Transmission Electron Microscope works by using highly accelerated electrons as "light source", and focusing these electrons onto an image thanks to electromagnetic lenses. Typical accelerating voltages are in the order of 200 kVolts, producing electrons with associated wavelengths of about 2.5 pm. It is quite clear that, as an instrument, the imaging limitations of the electron microscope are not due to the (very small) wavelength being used (much less than one thousandth of an Å), but to imperfections of the electromagnetic lenses (naturally, the specimen itself may introduce additional limitations, such as those related to dose sensitivity, the material surrounding the sample of interest, or beam induced movement). Electrons interact with the biological specimen under study as negatively charged particles, providing experimental information on the three-dimensional Coulomb (i.e., electrostatic) potential of the specimen. Considering that the typical atomic composition of macromolecular complexes is formed by elements of relatively low atomic number (like carbon, oxygen or hydrogen), and that the specimens themselves are small (a ribosome is in the order of 250Å, as an example), it is normally considered that the interaction between the accelerated electrons and the biological matter is very weak. So weak, in fact, that only some of the electrons going through the specimen interact with some of its atoms, and that the result of this interaction is "only" a change of the associated phase of the electron (they are not absorbed or, in general, loose energy). Under these simplified conditions, it is possible to model electron microscopy images as if the whole three-dimensional structure of the specimen would be "condensed" into an image perpendicular to the electron direction; in other words, as if the whole Coulomb potential would be "summed" (integrated) along the direction of the electron beam into each point of the resulting image. We refer to images formed in this "condensed" manner as "projection images" of the specimen under study (the reader is referred to Hawkes [8], Hawkes and Kasper [9], Frank [7] for further details). We can express the concepts presented above in a simple mathematical way as [2]:

$$\text{EM\_Image} = \text{Projection}(\text{biological specimen}) \tag{3}$$

where "Projection" is an operation performing a summation (line integral) along the electron beam trajectory.

$$g = \textit{line\_integral}(f) \tag{4}$$

Once understood how images are formed, we may start thinking about how the three-dimensional process can take place. Indeed, the field of 3D reconstruction from 2D images may be regarded, at first glimpse, as somehow "magic", and it is not at all obvious that a whole "dimension" can be gained from lower dimensionality data by some mathematical procedure. The question is so fascinating that back in 1917, with no concrete experimental application in mind whatsoever, the Austrian mathematician Johann Radon derived a way to perform this process under a certain set of conditions (a translation in English of this fundamental work can be found in Radon [19]). The first and most critical one was that the lower dimensionality data had to be obtained as line integrals over the higher dimensionality space. Translated into a 2D/3D case, it required that the 2D images had to be projections of the 3D volume, which is exactly the relationship that exists (within approximations) between transmission electron microscopy images and the 3D biological specimen under investigation, as we have presented in previous paragraphs. Radon inversion formula certainly established the feasibility of performing the 3D reconstruction process, but the actual answer was not very practical, since it required an infinite number of noiseless projection images to perform the inversion.

A simple way to have a very practical understanding of the relationship between 2D projection images and its associated 3D volume is to formulate the case in Fourier space. We refer to Fourier space as the range of a very well-known operation known as the

[2] Formally, the projection equation can be written as

$$g(\mathbf{s}) = \text{Proj}\{f(\mathbf{r})\}(\mathbf{s}) = \int_{-\infty}^{\infty} f(H^T\mathbf{s} + z\mathbf{e}_3)dz$$

where $H^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{s} \in \mathbb{R}^2$ is a 2D coordinate in the image, $\mathbf{r} \in \mathbb{R}^3$ is a coordinate in the 3D volume, and $\mathbf{e}_3 = (0,0,1)^T$ is the z-axis.

"Fourier transform", such as when applied to either the 3D Coulomb potential of a biological specimen or to any of its projection image in 2D, it creates a new 3D volume or a new 2D image in the so-called "Fourier space" (on the contrary, we will refer to the space where the original Coulomb potential and its 2D projection exists as "real space"). The Fourier transform is a fundamental tool in the analysis of many processes, and we will be making a limited use of it in the next paragraphs and sections. In a mathematical way, the 3D Fourier Transform (abbreviated FT) of a volume v, noted as V, is obtained from v as

$$V(\mathbf{R}) = FT\{v(\mathbf{r})\} = \int_{\mathbb{R}^3} v(\mathbf{r}) e^{-\langle \mathbf{R},\mathbf{r}\rangle} d\mathbf{r} \tag{5}$$

where $\mathbf{r} \in \mathbb{R}^3$ is the 3D spatial coordinate in real space, and $\mathbf{R} \in \mathbb{R}^3$ is the 3D spatial frequency. In much the same way, the Fourier Transform of a 2D image i, noted as I, is obtained as

$$I(\mathbf{S}) = FT\{i(\mathbf{s})\} = \int_{\mathbb{R}^2} i(\mathbf{s}) e^{-\langle \mathbf{S},\mathbf{s}\rangle} d\mathbf{s} \tag{6}$$

where $\mathbf{s} \in \mathbb{R}^2$ is the 2D spatial coordinate in real space, and $\mathbf{S} \in \mathbb{R}^2$ is the 2D spatial frequency.

For the particular case in which the 2D images are projections of a 3D volume, it is possible to prove (reviewed in Kak and Slaney [12], Sorzano et al. [25]) a fundamental relationship that exists between V and I (that is, the Fourier Transform of v and i), known as the "central slice theorem". Indeed, while a projection image i has condensed information about the whole volume v, I (the Fourier Transform of i) has information of only a slice of V (the Fourier Transform of v). Even more, the direction in which I slices V is the same projection direction that generated i from v, as shown in Fig. 1.

It is now simple to have an intuitive answer to our previous question of "Is one image enough to obtain a 3D reconstruction, and if this is not the case, how many are needed?". Indeed, the case of having only one projection image is the one shown in Fig. 1 in Fourier space, where we only have one plane, one "slice", going through V. Clearly, with just one plane we cannot have a good understanding of the volume V, and many more slices in multiple directions are needed to properly estimate it. How many projections images are needed is a more difficult question to answer in a quantitative manner.

Let consider the simple case in which the specimen is rotated inside the microscope column around a fixed axis while images are being acquired (Fig. 2); it is assumed that images are taken at a constant angular increment of θ. In principle, θ could be as small as desired but, in practical terms, decreasing θ implies collecting more images, and hence to increase the total radiation dose received by the specimen. Clearly, a compromise must be found between θ and acceptable dose. Still, intuitively, decreasing θ makes more precise our coverage of V, so the just referred limit in θ must translate into a loss of precision in our estimate of V. A general formula to calculate the number of images required to perform a 3D reconstruction to a certain resolution R (by resolution we refer to the capacity to see fine details in a volume) was provided by Crowther et al. [4] as:

$$R = D\theta \tag{7}$$

where D is the diameter of the reconstructed specimen, and θ is the above described constant angular increment.

Let us now explore the question of how many images are needed from a digital computing point of view, where all magnitudes are finite and discrete (we have pixels or voxels, not continuous densities), we will particularize a number of equations to this situation, starting with Eqs. (1) and (2) (reviewed in Carazo [3]). Indeed, in the expression $g = Hf$, when g is a set of M images of dimension $N \times N$ pixels, f is a volume of dimension N^3 voxels and H is a linear operator, H transforms into a matrix of dimension $N^2M \times N^3$. If we further follow the convention of presenting pixels and voxels in lexicographic order (that is, piling the first line of the image on top of the second line, and so on, followed by one image after the other), the expression $g = Hf$ becomes

$$\begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{N^2M} \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N^3} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N^3} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N^2M,1} & h_{N^2M,2} & \cdots & h_{N^2M,N^3} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N^3} \end{pmatrix} \tag{8}$$

In other words, we have a linear system of N^3 unknowns for which we need at least N^3 equations (assuming all equations are independent, which is not really the case, although we will treat them that way as a simplification). Considering that the total number of equations is given by the size of each image $(N \times N)$ multiplied by the number of images M, we need at the very least N images, so that $N \times N \times N = N \times N \times M$. However, this simple calculation assumes that there is no noise in the images, which is not at all the case. Noise translates into making the equations inconsistent, so that,

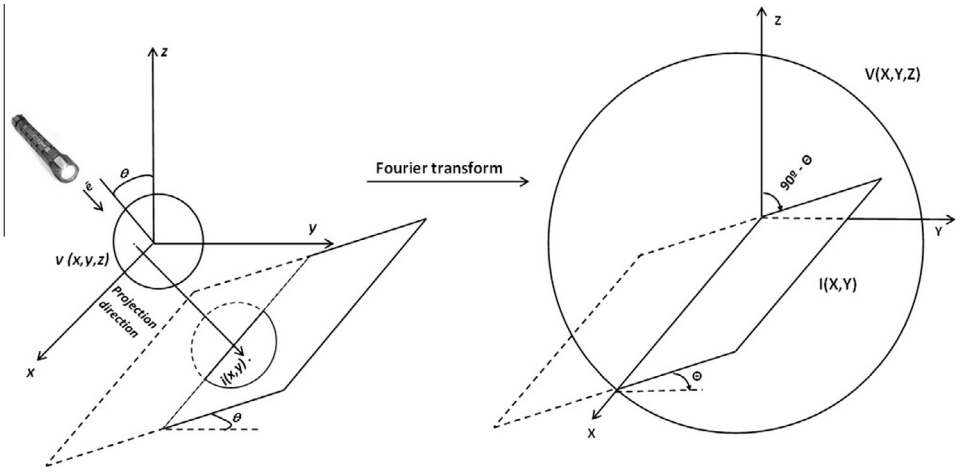

Fig. 1. Relationship between a 2D projection image (i) and its corresponding 3D volume (v). On the left hand side we present the geometry in real space, while the same geometry in Fourier space is shown on the right hand side (note that the Fourier Transform V of a finite volume v would extend over all Fourier space, although, for simplicity, we have represented it as a very large sphere V). In general, notation in Fourier is in capital letters, while it is in small letters in italic for real space magnitudes.

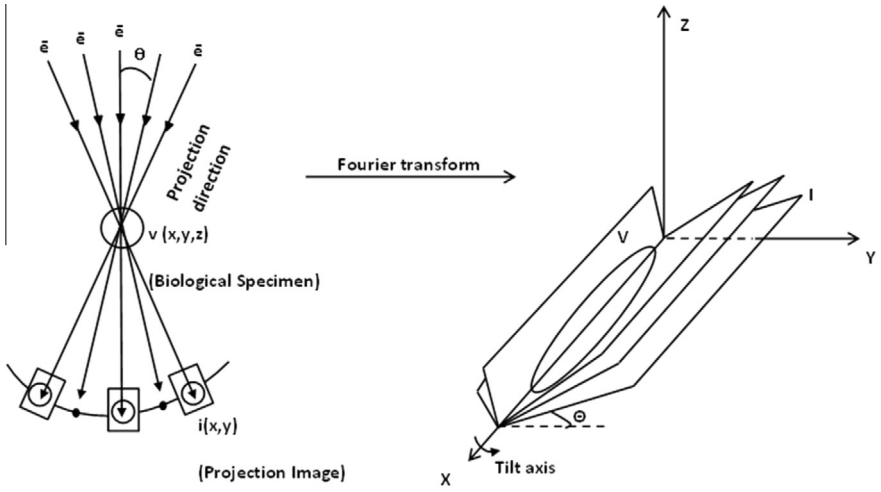

Fig. 2. Single-axis tilt image collection geometry, both in real space – left hand side– and Fourier space – right hand side.

for instance, 2 + 2 will not be equal to 4, but to some random number close to 4. Intuitively, if the equations are inconsistent, we will need many more images than $N \times N \times N$. But, how many more? Remember that the reconstruction process is an inverse problem, and that in order to calculate f we must invert, in some way, H. This was, in fact, Eq. 2, which stated that $f = H^{-1}g$. In this formulation it is clear that, in a general setting, we need as many images as required to properly invert H which, quite naturally, lead us to focus on the nature of H, that is intimately linked to the third question we wanted to address in this section, (3), In practical terms, how 2D images are normally collected? In the following we are going to focus on two common strategies to experimentally collect sets of images and their impact on H.

The first data collection scheme we will present is known as single axis tilt, and it is the most common one in Tomography, while the second one will not involve any tilting at all, and it is mostly used in high resolution studies of macromolecular complexes.

The single axis tilt geometry has already been introduced (Fig. 2), and it is used not only in TEM, but in most clinical applications of tomographic techniques as well. Images are collected by tilting the specimen at known angles around a fixed axis. The size of each of the images is generally large, of several thousand pixels in the x and y dimensions, while the volume to be reconstructed has usually the same dimensions than the images along x and y, but a fraction of them along the z axis. Let consider the case of images of 4000 pixels side, and a volume of 4000 pixels in x and y, but only 1000 pixels along the z directions. The system of equations $g = Hf$ will then be of dimensions $4000 \cdot 4000 \cdot M \times 1$ for g (M being the number of images), $4000 \cdot 4000 \cdot 1000 \times 1$ for f, and $(4000 \cdot 4000 \cdot 1000 \times 4000 \cdot 4000 \cdot M)$ for H. Following the rationale presented above, we would require at least M = 1000 images; however, dose considerations on the specimen force us to decrease the number of images to around 100. Furthermore, unfortunately these 100 images cannot be obtained at equidistant angles, but there is a limit to the maximum tilt angle achievable in the TEM, so that tilting by more than 60 or 70 degrees is impossible (both because of the mechanics of the goniometer and because the images would no longer be projection images). Therefore, the system of equations will be badly undetermined, which will reduce the precision (the resolution) of our reconstructions, at the same time that produces instabilities, requiring some form of smoothing on the volume.

Let now consider the no tilting geometry mentioned before (Fig. 3), specially designed to study purified macromolecular complexes, an area of study normally referred to as Single Particle Analysis (SPA). The key concept is that the sample is formed by, essentially, multiple copies of the same complex (in blue in Fig. 3), trapped in a solid matrix of amorphous ice at random orientations. If we now focus on the subimages containing just one complex, each of them can be regarded as a projection image of the specimen under study at some random projection direction. Note that there are multiple differences with the previously presented geometry. Now we refer as the specimen to the molecular complex under study, occupying only a small fraction of the EM field of view, while before the specimen did not have any restriction and

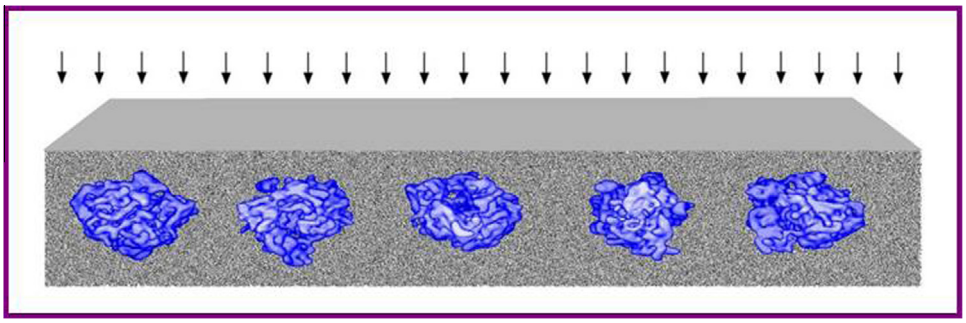

Fig. 3. Conceptual representation of a biological sample of a certain macromolecular complex (in blue) under observation at the TEM. Specimens are suspended on a thin layer of amorphous ice in arbitrary orientations. Electrons are coming from above (arrows). Note that each specimen would generate a projection image from a different projection direction.

occupied the whole large image. Putting some numbers to this new collection geometry, the vector g will be formed by M images of a size around 250×250 pixels, while the volume will be typically of the same dimensions than the images, of dimension around $250 \times 250 \times 250$ pixels. As before, we would expect the minimum number of images to be 250. However, there are two fundamental differences with the previous case: (1) The projection direction for each complex is not known *a priori*, introducing 5 more degrees of freedom per image in the system, three rotations and two translations and, (b), We aim at high resolution, which forces us to explicitly consider the very noisy nature of our images without introducing smoothing operations into the calculation of the volume. In practical terms, this situation increases the number of required images by orders of magnitude, to the tens of thousands, if not more.

From 2D images to 3D volumes: Reconstruction methods

In this section we will build over the knowledge of H, the projection operator, aiming at "inverting" its effect. That is, to obtain f, the reconstructed 3D volume, from a set of images g obtained following a certain geometry coded in H. In other words, we want to study practical ways to invert H, so as to obtain f by $f = H^{-1}g$ (as a general reference, the reader is referred to the book by Herman [10]).

Series expansion methods

Starting from the algebraic way we have used to represent the EM process in previous sections, probably the most obvious manner to address the reconstruction process is by considering methods to solve systems of linear equations. Remember that we need to estimate as many unknowns as voxels in the volume, from as many equations as pixels are in the set of images. Note, also, that images are noisy (typically, they have much more noise than signal, up to an order of magnitude more in power), so that most equations will be inconsistent. In other words, in some Eqs. 2 + 2 will be equal to 3 and in others to 5, depending on the noise. But, which type of solutions may be adequate for an inconsistent system of equation?. This question was posed about half a century ago by the Hungarian mathematician Cornelius Lanczos [14], who suggested that an acceptable solution should be as much in agreement with all the equations as possible, in some well-defined way, proposing an iterative method to find one such a solution.

Building on Lanczos works, a family of reconstruction methods has evolved under the general name of Series Expansion Methods.

These methods are particularly well suited to deal with arbitrary geometries of data collection as well as to explicitly incorporate *a priori* information about the volume to be reconstructed or about the image formation process. On the other hand, they tend to be slower, in terms of computer time, than other reconstruction approaches.

Backprojection

Leaving the discrete formulation used in previous sections, and returning to the original formulation of the reconstruction problem by Radon, another approach has aimed at solving Radon inverse equation in an approximate manner under a certain set of conditions, leading to the filtered backprojection algorithm.

Backprojection is indeed a very popular reconstruction method in electron microscopy as well as in most clinical biomedical applications, probably due to its simplicity and speed. However, the approach suffers specially when there are gaps in the projection geometry, as it is, for instance, the case of Transmission Electron Tomography. Still, the general performance of backprojection in a large number of experimental applications is remarkable, making it a very successful method.

Fourier-based methods

The Fourier Transform (FT) introduced in Section '2D images and 3D volumes: Basic relationships' can also be used in the reconstruction process, so that V (the Fourier Transform of the specimen Coulomb potential v) can be estimated from the "slices" in Fourier space coming from the Fourier Transform of the experimental images. Indeed, Fig. 2 shows how the different slices in Fourier space tend to fill up volume V. Still, there is a very important issue with the use of the Fourier Transform in reconstruction, in that it is a rather slow operation unless a certain computational method is used, known as the Fast Fourier Transform (FFT). However, the FFT requires the data samples coming from the different slices to be equally spaced, preferably in a Cartesian lattice. Therefore, some form of interpolation is required to go from the set of experimental values to a regular grid, which must be precise and fast.

A very positive value of Fourier-based reconstruction methods is that they tend to be fast and, especially, their computational cost is almost the same irrespective of the number of experimental images. This is so because although the interpolation stage will vary depending on the number of images, all further operations will take exactly the same time. Naturally, the handling of data in Fourier space may be sometimes less intuitive than in real space, including the appearance of some artifacts. Still, the use of these methods is increasing in all fields, and a good comparison of several methods can be found in Abrishami et al. [1].

A typical 3D reconstruction workflow in SPA

Although there are many possible workflows leading to the 3D reconstruction of a biological specimen, in the following we will focus on a particular one, presenting in this way the basic processing steps. The workflow in a typical single particle reconstruction starts by recording micrographs in an electron microscope (see Fig. 4). After being digitized, a screening is performed to select the best initial images (this step will be further addressed in next Section). Micrographs can also be downsampled to improve the signal to noise ratio and accelerate subsequent calculations. Then, particles are selected (or picked) from micrographs, either manually or automatically, extracted and saved for further use. Some preprocessing may be applied while extracting, such as: filters, contrast inversion and others. Particles are usually sorted according to a quality factor to identify possible outliers, like wrongly picked images.

At this point, the gallery of particles may be used as input for 2D classification algorithms, so as to detect possible heterogeneities due to sample contamination, different conformations or different specimens, followed by the calculation of a low resolution initial map. Some images may be discarded and not considered in subsequent steps.

There are several approaches to produce an initial low resolution 3D map. The programs used in this step usually produce a collection of possible 3D maps that are visually inspected. After one of the initial volumes is selected, an iterative refining algorithm will carefully assign projection directions to each of the input images. Some of these refining algorithms can also be used for dealing with heterogeneity, by comparing several initial 3D references with the gallery of particles.

Identifying and analyzing 3D heterogeneity is, however, still technically challenging. Regarding the classification of images into homogeneous data sets – that is, grouping together images

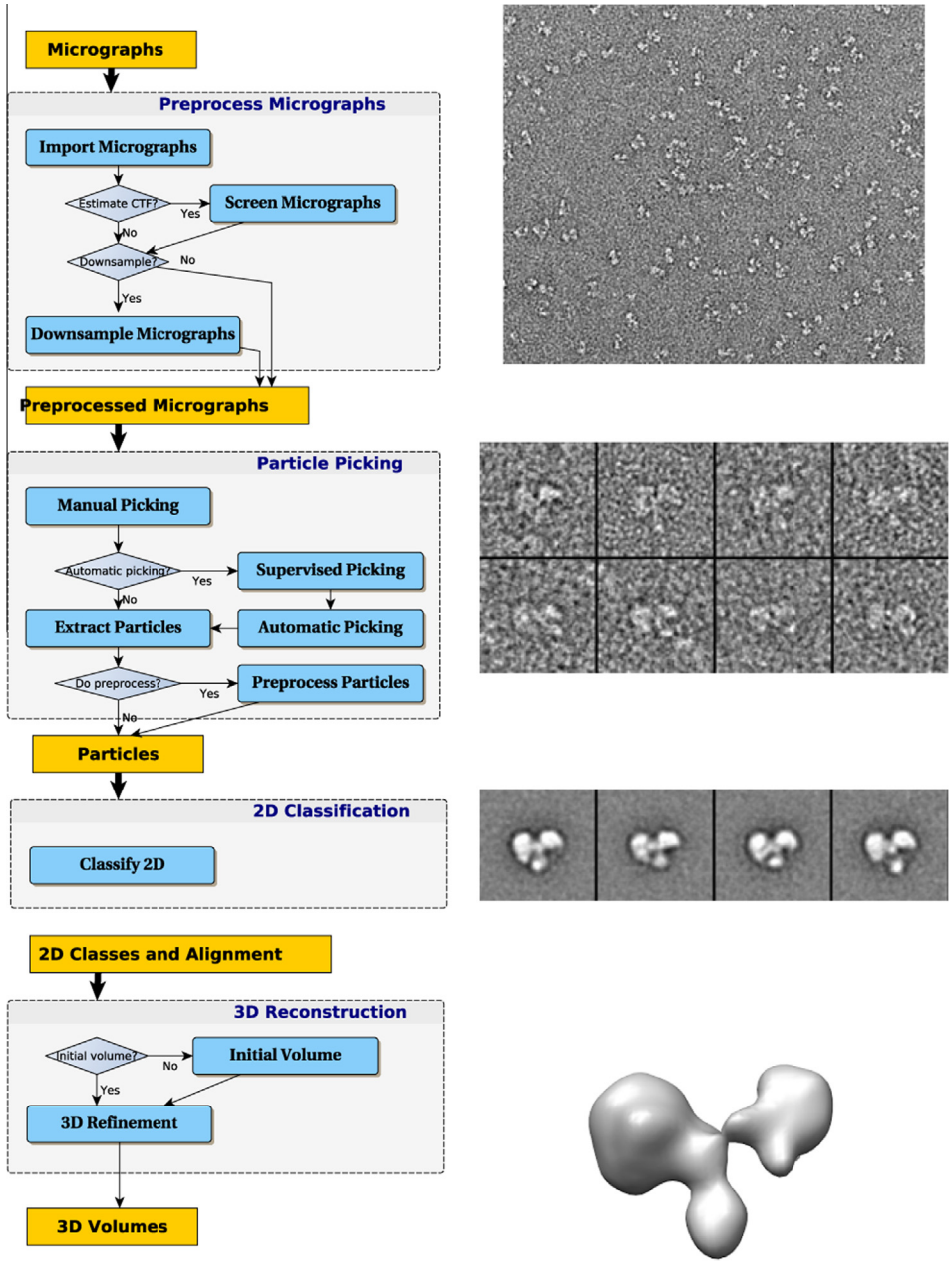

Fig. 4. Left panel:Typical EM workflow. Right panel, from top to bottom: micrograph, gallery of particles, class averages and 3D map. Images courtesy of Nunez-Ramirez et al. [16].

produced by projecting a specific conformation of the specimen under study – many alternatives have been proposed. One of the most popular ones is based on maximum likelihood (ML) [22,21]. ML methods modify the 3D references so that the probability that a given 3D reference would produce a given experimental data set is maximized. When more than one 3D reference is used, each experimental image has a given probability of being produced by each of the available references in each of the different projection directions. This probability depends on (1) the pixel-by-pixel similarity and (2) the parameters needed to transform the experimental image into the reference projection (extent of the shifts and noise statistics). This probability is converted into weights that control the contribution that a particular image will make to each of the new 3D references in each of the projection directions.

This basic workflow has been recently expanded after the development of radiation-hardened CMOS-detectors that can directly detect electrons. These new Direct Electron Detectors – also known as DEDs – can be used to record cryo-EM data with very high signal to noise ratio. Additionally, DED's have also shown how during the 1 to 2 s exposure normally required to acquire a micrograph, biological specimens suffer small movements that blur high-resolution features, these are the so called Beam Induced Movements [2]. Indeed, DEDs record images at a rate of many frames per second, effectively producing a "movie", so that the different movie frames can be compared and aligned in order to detect movements on the Åscale. Once the movements of the particles are detected, they may be reversed in the computer producing much sharper, motion-corrected particles.

Characterizing the initial experimental images

Previous sections have presented the field of 3D reconstruction from EM images in a somehow simplified manner, modelling TEM images as simple projection images of the specimen under study.

Indeed, a further step towards realism is to consider that the projection images are modified, or more precisely, modulated by a certain function that takes into account the effect of the electron microscope on the ideal projection images. The electron microscope, as any experimental image formation system, modifies the ideal projections. These modifications are usually referred to as aberrations. We can easily understand these aberrations, or modifications of the ideal projection images, in a more familiar image formation system, the eye. You may think of the eye of a person affected by myopia, for example. In this case, the image formation system (the eye) affects or modifies the ideal images by a defocus aberration that produces a blurring in the resultant experimental images. Note that myopia can be easily corrected by glasses, but aberrations introduced by an electron microscope are more difficult to correct.

The modifications introduced by an Electron Microscope to the ideal projections are characterized by the Contrast Transfer Function (CTF), which is directly related to the most important aberrations of the microscope, and is defined in Fourier space. The Fourier Transform I of an ideal image i is then related with the Fourier Transform of the experimental image $I\prime$ through the CTF by

$$I'(\mathbf{R}) = I(\mathbf{R})CTF(\mathbf{R}) \tag{9}$$

where $\mathbf{R}$ is the location in Fourier space, also referred as 2D spatial frequency. The CTF depends on the 2D spatial frequency, which means that the action of the electron microscope depends on the size of the object details. The first process after obtaining the experimental micrographs consists in the obtention of the CTF of each of them. The screening of the CTFs permits to reject bad quality micrographs. A reason to discard micrographs may be the presence of strongly asymmetric rings (astigmatism) or rings that fade in a particular direction (drift), see Fig. 5.

From Expression 9, we can observe that in order to obtain I, which is free from the electron microscope aberrations, and perform a proper reconstruction process, the CTF has first to be estimated, and then somehow inverted.

From 2D images to 3D volumes: A posteriori projection assignment and classification

Up to this moment we are presenting a "traditional" reconstruction process, as it is usually addressed in other biomedical fields. However, there are important differences between, for instance, the reconstruction of some brain features of a patient on a fixed and well-known medical scanner, and the structural study of a macromolecular complex on an electron microscope. We will concentrate on two of these differences, the first one is related to our knowledge of the projection geometry, and the second one refers to macromolecular structural flexibility and/or presence of different molecular species.

Common to all reconstruction methods presented in previous sections is the need to know the projection geometry, that is, knowing the directions from which the projection images were obtained, so that H is fully characterized. Indeed, this is not usually an issue on a clinical setting, since the scanner operates acquiring images following a predetermined pattern, normally fixed by the manufacturer. However, in the "no tilt" data collection geometry for the study of macromolecular complexes introduced previously, the projection directions are not known *a priori*, raising the need to find them *a posteriori*. This latter requirement is very specific to TEM applications, and its fulfillment represents one of the most complex processes in the course of a 3D reconstruction. In principle, it is possible to separate the step of estimating the projection directions from the step of estimating the volume [27]. This is done by exploiting an interesting property of projections in Fourier space: any two projections share a line in Fourier space along which Fourier coefficients are the same (note that this line is different for every pair of projections). However, this procedure may be error-prone due to the difficulty of finding the common lines in the presence of noise (recall that the signal to noise ratio is smaller than 0.1). Even if raw projections are gathered into similar groups (2D classes), the amount of local minima in this search of common lines is so large that the probability of finding a relatively correct structure is rather low in practice. A more practical approach iteratively alternates between the two steps, leading to a truly intertwined process: we need the geometry to estimate the volume, and we need a volume to estimate the geometry. In turn, this situation leads to a key issue, as it is the choice of the initial volume to be used in the iterative process, the so called "Initial Model Problem". There are quite a number of methods to estimate an initial model ([27,18,17,20,23,24,6,28], and others), and even there is a Web Service to help in its determination http://scipion.cnb.csic.es/myfirstmap, but it still represents a critical step in many structural studies.

Let us now focus on issues related to the precise definition of the specimen under study. Clearly, in a clinical setting there is a unique specimen under study. Indeed, we aim at, for instance, studying the particular brain of the patient under investigation, which has a unique 3D distribution of densities. On the contrary, in the case of studying macromolecular complexes, we image thousands of allegedly identical particles. However, in practice, they may be intrinsically flexible, alternate between different conformational states, have different binding states, or, simply, not be as pure as initially thought. In all these cases the reconstruction problem has to be extended, so that to model the process of obtaining the projection geometry *a posteriori* as well as the situation in which the experimental set of projection images were coming from n different volumes and not just from one.

Note that the complexity introduced by previous considerations is rather large, since not only the projection direction for each image has to be estimated *a posteriori*, but also the actual volume

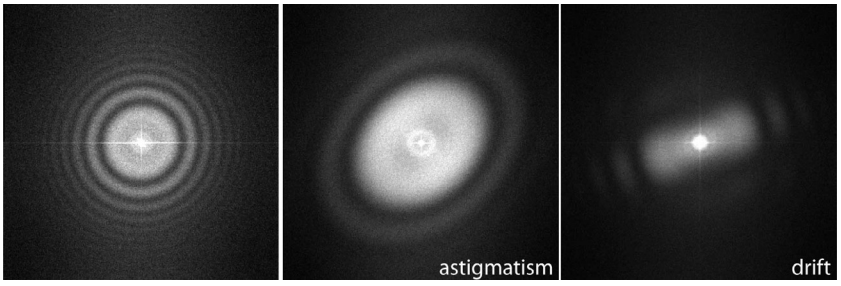

Fig. 5. CTF examples of a good quality micrograph (left), a bad quality micrograph affected by astigmatism (middle), and a bad quality micrograph affected by drift (right) [16].

giving rise to that projection has to be selected from n possible candidates [15,22,26,21]. Furthermore, normally n is not known *a priori*, requiring an additional exploratory work, which may be further complicated by the fact that the macromolecule may not be presenting n distinct conformations, but actually fluctuate on a rather continuous conformational range [11].

The main way to address macromolecular flexibility has been to extend current methods to estimate the projection direction of an image given a certain initial model, to the case in which several models are provided. The image is assigned to the volume that maximizes a certain similarity measure. However, the space in which we have to optimize this similarity measure has many dimensions and, on top of it, all input measurements are extremely noisy, which results in a complex space of solutions characterized by having many local minima. Let us make a simple calculation on how big this solution space is. Let consider that we are reconstructing a volume of size $250 \times 250 \times 250$ pixels from 25,000 images of size 250×250 pixels. The solution space has a dimension that coincides with the number of voxels of the volume and the orientation parameters (5 per image), conforming a space of $250 \cdot 250 \cdot 250 + 25,000 \cdot 6$ variables (3 Euler angles + 2 in-plane shifts + 1 model selection) = 15,625 M + 150 k = 15,775 millions of variables. In practice, it is unfeasible to globally search a so large space, so that we are forced to use methods that are known to stop at local minima. It is now very clear why the Initial Model Problem has been highlighted before, because usually our search will stop at some local optima close to our initial model. Of course, given these considerations, our quest will be towards developing methods to reduce the existence of local optima, moving from a situation like the one in Fig. 6 (top) to the one in Fig. 6 (bottom). Indeed, if we successfully reduce the number of local optima, it will now be much more likely that starting from a number of different initial models, we will still reach the same result, which will be the global optimum, or close to it, even if our searches would still be local. The key question then, is how to accomplish this simplification of the landscape of the objective function.

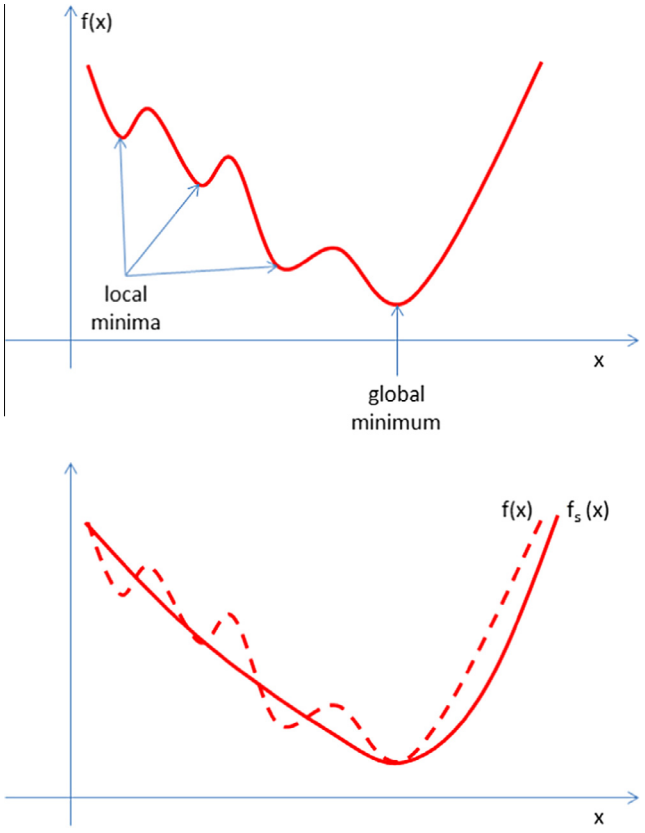

Fig. 6. Top: Difference between local minima and global minimum of an objective function: the global minimum is the minimum value of a given function, while local minima are the smallest values of the function only in a certain neighbourhood. Bottom: The objective function $f(x)$ has been smoothed by a surrogate function $f_s(x)$ whose global minimum is close to the global minimum of $f(x)$.

Let consider a very simple case: we impose the volume to be smooth, which implies that we strongly reduce the number of degrees of freedom since each voxel cannot be independent from its surroundings. We may reduce the number of degrees of freedom to as few as 0.5 or 1 % of the number of voxels and still reconstruct a macromolecule to a resolution of about 15–20 Å. Naturally, many local optima will then disappear, since the dimension of the search has been strongly reduced. This smoothing effect can be obtained by reducing the resolution (the details) of the initial models, which is certainly the most common method used in the field, although other basis, like wavelets or hyperspherical harmonics, may be more suitable to promote a reduced representation of the macromolecules. This process can be further elaborated, incorporating other desired properties of the volume and formulating the problem on a statistical framework. This latter approach is the one followed by the family of Maximum Likelihood methods [22,21], which have effectively opened a new dimension in the study of macromolecular flexibility.

Still, many complexes may not have distinct conformational states, but explore a continuous of conformations, a case that cannot be easily fitted under the considerations presented before. How can they be analyzed? The situation of continuous flexibility is characterized by the macromolecule following a certain conformational trajectory, with the possibility of occupying at a given time any point along this trajectory. This problem has been addressed more recently than the case of discrete states [11,5], being a very active line of research.

A common issue associated to all optimization schema so far, is their high computational cost. Indeed, the task of projection direction and 3D structural class assignment is the one demanding more computational resources nowadays, which can be measured by years of CPU and weeks of clock-wall time for a typical experimental case. Ways to find alternative views, formulating the problem in substantially faster manners, are in real need, being a clear subject of research.

On more complicated relationships: When simplification breaks

In this section we will briefly elaborate on the case in which some of the basic simplifications we have done along this work do no longer hold, as it happens when aiming at very high resolution or working with very thick objects.

In Section '2D images and 3D volumes: Basic relationships' we described that only a small fraction of the electrons that arrive to the specimen interacts with it, being the consequence a variation in the associated phase of the electrons. Moreover, these interacting electrons are scattered following a diverging path from the unscattered electrons straight path. Depending on the scattering pattern characteristics of the specimen, the size of the lens and its distance to the sample, not all the scattered electrons are collected by the lens to be focused onto the camera sensor (see Fig. 7). This loss of electrons produces a contribution in the projection images known as amplitude contrast, which is independent of the phase contrast contribution. This amplitude contrast contribution has its own CTF function, which complements the phase contribution CTF in such a way that at the frequencies where the contributions of the phase CTF is maximal, the amplitude contribution is zero, and viceversa (reviewed in Kirkland [13]). In addition,

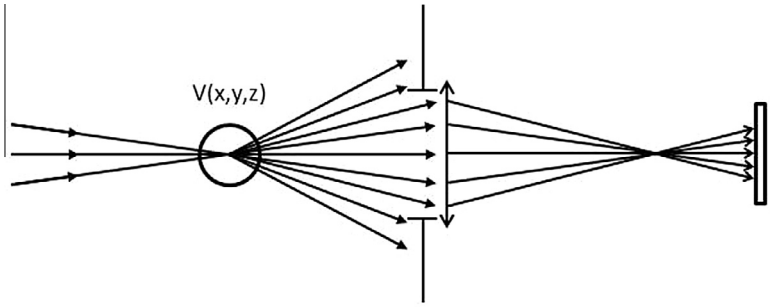

Fig. 7. Simplified model of a Transmission Electron Microscope, where some of the scattered electrons are blocked by the optics, leading to an amplitude contrast regime.

previously we have considered that the interaction between electrons and matter is so weak that the resulting images could be approximated by the integration of the Coulomb potential along the projection axis, leading to projection images. However, this sentence is essentially true only under the assumption of the specimen being very thin. In the case of thick specimens, the contribution to the projection images is no longer in the direct way described in Eq. 3, but in an expression closely related to the well-known Beer–Lambert law:

$$\mathrm{EM\_Image}_{\mathrm{Thick}} = I_0 e^{\mathrm{Projection(biological\ specimen)}}, \quad (10)$$

where I_0 is the projection image measured in the absence of specimen.

Furthermore, the assumption that the specimen is thin enough also makes it to be homogeneous in focus, that is, all its features are affected by the same CTF. Obviously, if this were not to be the case, then a different CTF should be considered for each part of the specimen, and the general framework presented so far would simply not be applicable. But, how small is "small" in previous sentences? In which cases is this complicated situation of real biological importance? Let us answer these questions referring to a recent work on Adenovirus Zhou et al. [29], where it was considered that for this large virus, with a capsid of approximately 1300Å, these effects already had a noticeable effect at about 8Åresolution.

Discussion and conclusion

The capability to obtain quantitative three-dimensional information on molecular complexes and cellular components from sets of transmission electron microscopy images has a deep impact in structural and cell biology, specially when coupled with very powerful sample preservation procedures, such as cryogenic conditions. Indeed, nowadays to reach quasi atomic information from samples containing purified macromolecular complexes directly imaged at the microscope is starting to be more and more common, making this technology very interesting for biochemists and cell biologists at large. However, the mathematical and algorithmic concepts empowering this approach are not that common and certainly not that obvious, but they have to be mastered by anybody wishing to use this approach and extract new biological information. It has been with this idea in mind how we have approached the writing of this work, keeping the equations not only to the bare minimum but to the bare concepts, at the same time that highlighting the core ideas that make possible the whole approach, including discussing about those experimental conditions for which some of the basic mathematical assumptions may start to break.

Acknowledgements

Authors would like to acknowledge economical support from: Comunidad de Madrid through grant CAM (S2010/BMD-2305) and the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638 and BIO2013-44647-R. C.O.S. Sorzano is recipient of a Ramón y Cajal fellowship.

References

[1] V. Abrishami, J.R. Bilbao-Castro, J. Vargas, R. Marabini, J.M. Carazo, C.O.S. Sorzano, Ultramicroscopy (2015) (in press).
[2] A.F. Brilot, J.Z. Chen, A. Cheng, J. Pan, S.C. Harrison, C.S. Potter, B. Carragher, R. Henderson, N. Grigorieff, J. Struct. Biol. 177 (3) (Mar 2012) 630–637.
[3] J.M. Carazo, The fidelity of 3D reconstructions from incomplete data and the use of restoration methods, in: Electron Tomography, Plenum Press, 1992, pp. 117–166.
[4] R.A. Crowther, L.A. Amos, J.T. Finch, D.J. De Rosier, A. Klug, Nature 226 (1970) 421–425.
[5] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinizadeh, H.Y. Liao, J. Pallesen, G. Sharma, V.A. Stupina, A.E. Simon, J.D. Dinman, J. Frank, A. Ourmazd, Proc. Natl. Acad. Sci. U.S.A. 111 (49) (2014) 17492–17497.
[6] H. Elmlund, D. Elmlund, S. Bengio, Structure 21 (8) (2013) 1299–1306, http://dx.doi.org/10.1016/j.str.2013.07.002.
[7] J. Frank, Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State: Visualization of Biological Molecules in Their Native State, Oxford University Press, USA, 2006. URL: <http://books.google.es/books?id=o5EUmAEACAAJ>.
[8] P. Hawkes, Electron tomography, Ch. The Electron Microscope as a Structure Projector, Plenum Press, 1992, pp. 17–38.
[9] P. Hawkes, E. Kasper, Principles of Electron Optics: Wave Optics, vol. 3, Academic Press, 1996.
[10] G.T. Herman, Fundamentals of Computerized Tomography: Image Reconstruction from Projections, Springer, 2009.
[11] Q. Jin, C.O.S. Sorzano, J.M. de la Rosa-Trevín, J.R. Bilbao-Castro, R. Núñez Ramírez, O. Llorca, F. Tama, S. Jonić, Structure 22 (3) (2014) 496–506, http://dx.doi.org/10.1016/j.str.2014.01.004.
[12] A. Kak, M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, 1988.
[13] E. Kirkland, Advanced Computing in Electron Microscopy, vol. 40, 2010.
[14] C. Lanczos, J. Res. Natl. Bur. Stand 49 (1952) 33–53.
[15] A.E. Leschziner, E. Nogales, Ann. Rev. Biophys. Biomol. Struct. 36 (2007) 43–62.
[16] R. Nunez-Ramirez, S. Klinge, L. Sauguet, R. Melero, M.A. Recuero-Checa, M. Kilkenny, R.L. Perera, B. Garcia-Alvarez, R.J. Hall, E. Nogales, L. Pellegrini, O. Llorca, Nucl. Acids Res. 39 (18) (Oct 2011) 8187–8199.
[17] T. Ogura, C. Sato, Posterior Euler angle assignment using simulated annealing 156 (2006) 371–386.
[18] P.A. Penczek, J. Zhu, J. Frank, Ultramicroscopy 63 (1996) 205–218.
[19] J. Radon, P.C. Parks (translator), On the determination of functions from their integral values along certain manifolds, IEEE Trans. Med. Imaging 5 (4) (1986) 170–176.
[20] E. Sanz-García, A.B. Stewart, D.M. Belnap, J. Struct. Biol. 171 (2) (2010) 216–222, http://dx.doi.org/10.1016/j.jsb.2010.03.017.
[21] S.H.W. Scheres, A Bayesian view on cryo-EM structure determination, J. Mol. Biol. 415 (2) (2012) 406–418.
[22] S.H.W. Scheres, H. Gao, M. Valle, G.T. Herman, P.P.B. Eggermont, J. Frank, J.M. Carazo, Nat. Methods 4 (1) (2007) 27–29.
[23] A. Singer, R.R. Coifman, F.J. Sigworth, D.W. Chester, Y. Shkolnisky, J. Struct. Biol. 169 (3) (2010) 312–322.
[24] A. Singer, Y. Shkolnisky, SIAM J. Imaging Sci. 4 (2) (2011) 543–572, http://dx.doi.org/10.1137/090767777.
[25] C.O.S. Sorzano, R. Marabini, J. Vargas, J. Otón, J. Cuenca-Alba, A. Quintana, J.M. de la Rosa-Trevn, J.M. Carazo, Computational methods for three-dimensional

microscopy reconstruction, in: Ch. Interchanging Geometry Information in Electron Microscopy Single Particle Analysis: Mathematical Context for the Development of a Standard, Springer, 2014, pp. 7–42.
[26] C.M.T. Spahn, P.A. Penczek, Curr. Opin. Struct. Biol. 19 (5) (2009) 623–631, http://dx.doi.org/10.1016/j.sbi.2009.08.001.
[27] M. van Heel, Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction 21 (1987) 111–124
[28] J. Vargas, A.-L. Álvarez Cabrera, R. Marabini, J.M. Carazo, C.O.S. Sorzano, Bioinformatics 30 (20) (2014) 2891–2898, http://dx.doi.org/10.1093/bioinformatics/btu404.
[29] Z.H. Zhou, W.H. Hui, S. Shah, J. Jih, C.M. O'Connor, M.B. Sherman, D.H. Kedes, S. Schein, Four levels of hierarchical organization, including noncovalent chainmail, brace the mature tumor herpesvirus capsid against pressurization, Structure 22 (10) (2014) 1385–1398.