

Three-Dimensional Electron Microscopy of Macromolecular Assemblies

Joachim Frank, Department of Biochemistry and Molecular Biophysics, and Department of Biological Sciences, Columbia University, 650 W. 168th Street, BB2-221, New York, NY 10032.

In this book, the mathematical principles and working methods of single-particle reconstruction are described – a method designed to retrieve three-dimensional structural information from electron micrographs showing thousands of “copies” of biological molecules trapped in a thin layer of ice. This technique is uniquely suited to obtain three-dimensional images of molecular machines in different functional states, as it dispenses with the need for crystals. The book starts with an introduction of image formation in the electron microscope, which includes the definition of the contrast transfer function. Next averaging techniques and tools for image alignment, multivariate data analysis, and classification are described. An introduction into the mathematical principles underlying reconstruction of an object from its projections is followed by detailed accounts on how projection angles are determined, and how reconstruction is done in practice. The book concludes with a chapter on interpretation of density maps reconstructed, including methods for segmentation as well as fitting and docking of atomic coordinates.

Cryo-electron microscopy, image processing, macromolecular assemblies, molecular machines, single-particle reconstruction, three-dimensional reconstruction

Chapter 1: Introduction

This introductory chapter begins with an appreciation of the unique position of electron microscopy in biological research as it bridges a wide gap between X-ray crystallography and light microscopy. The scope of this book is defined to cover three-dimensional imaging of molecular assemblies that exist in an *in vitro* sample in large numbers with identical or near-identical structure. Only with such samples it is possible to collect large numbers of projection images suitable for averaging and three-dimensional reconstruction. The fact that molecules can be imaged as single, isolated particles embedded in ice (“crystallography without crystals”) makes the techniques described in this book uniquely suited to image molecular machines in their various processing states. In sharp contrast, electron tomography, not covered in this book, is concerned with the three-dimensional imaging of “unique” objects that may be an organelle or slice of a cell. The vision of a unified structural analysis of macromolecules is articulated, which would lead to an integration of results from cryo-EM, X-ray crystallography and NMR, and a cross-fertilization among these disciplines. The chapter concludes by making the point that the development of single-particle reconstruction would not have been possible without the vast increase in computer power seen in the past decades.

Keywords: *Cryo-electron microscopy, Crystallography without crystals, Electron tomography, Light microscopy, Molecular machines, Single-particle reconstruction, X-ray crystallography, Three-dimensional reconstruction.*

Chapter 2: Electron Microscopy of Macromolecular Assemblies

Following an outline of the transmission electron microscope (TEM) and its working principles, the chapter starts with a description of specimen preparation methods for EM imaging, including negative staining, glucose embedment and ice embedment. The principle of image formation in the TEM is described, as it pertains to biological weak phase objects, and in the process the contrast transfer function

(CTF) is introduced. EM images, we learn, are effectively projections of the Coulomb potential distribution of the biological object, convoluted with that function. The chapter closes by describing methods for the determination and computational correction of the CTF.

Keywords: *Contrast transfer function, Coulomb potential, Ice embedment, Glucose embedment, Negative staining, Phase object, Transmission electron microscope*

Chapter 3: Two-Dimensional Averaging Techniques

This chapter starts by listing the common sources of noise in the EM and how they can be addressed by averaging techniques. Conditions for digital sampling, or for the representation of an effectively continuous image by an array of discrete density measurements, are discussed. The concept of image alignment is defined, and alignment is introduced as a precondition for averaging as well as for making any meaningful comparison of experimental images. The cross-correlation function is then introduced as one of the most important tools to achieve alignment. Averages of aligned images are characterized by statistical measures such as variance and signal-to-noise ratio. Measures of resolution are introduced based on a comparison, in Fourier space, of two independent averages from halfsets of the data. Among these are the differential phase residual and the Fourier ring correlation. The chapter ends with a discussion of the resolution-limiting factors and with an outline of rank sum analysis, a method of statistical validation.

Keywords: *Alignment, Cross-correlation function, Differential phase residual, Digital sampling, Fourier ring correlation, Noise, Rank sum analysis, Signal-to-noise ratio, Variance*

Chapter 4: Multivariate Data Analysis and Classification of Images

This chapter is concerned with the need for classification of images to deal with either the co-occurrence of difference view angles or conformational/compositional heterogeneity of the sample, or both. As a start, techniques of multivariate data analysis such as principal component analysis and correspondence analysis are introduced, which sharply reduce the dimensionality of the problem. In the resulting factorial representations of images, variational trends and the presence of clustering can be readily recognized. Application of automated classification is possible in this new space: the chapter describes three techniques of unsupervised classification: K-means, hierarchical ascendant classification, and self-organized maps. The chapter closes by introducing supervised classification, which utilizes similarity to two or more template or reference images to determine class membership.

Keywords: *Correspondence analysis, Heterogeneity, Hierarchical ascendant classification, K-means, Principal component analysis, Self-organized maps, Supervised classification*

Chapter 5: Three-Dimensional Reconstruction

This chapter introduces mathematical principles underlying the relationship between a three-dimensional object and its projections, the projection theorem and Radon's theorem. The shape transform is introduced as an important concept explaining that recovery of the object from a finite set of projections is feasible when the object is of finite extent. It also explains the basis for Crowther's formula, which relates the number of projections to the object's size and the resolution to which it is to be recovered. The next topic is the determination of orientations, either *ab initio*, or by reference to an existing 3D

template (three-dimensional projection alignment). *Ab initio* methods widely used are introduced: random-conical data collection and common lines (or angular reconstitution). Several important reconstruction techniques are covered: weighted back-projection, Fourier interpolation, and iterative methods such as the algebraic reconstruction technique, or ART. Subsequent topics of the chapter include resolution assessment, contrast transfer correction, and reconstructions from heterogeneous datasets by supervised classification.

Keywords: *Algebraic reconstruction technique (ART), Angular reconstitution, Common lines, Crowther's formula, Fourier interpolation, Projection theorem, Radon's theorem, Random-conical reconstruction, Shape transform, Supervised classification.*

Chapter 6: Interpretation of Three-Dimensional Images of Macromolecules

This chapter covers all aspects of interpretation, starting with validation and consistency checks. Methods for visualization of the reconstruction by surface rendering are discussed, along with criteria for the definition of molecular boundaries. Segmentation is described both based on intrinsic properties of the reconstructed density distribution and on the use of tags such as gold labels or antibodies in the experiment. The chapter goes on to describe how a quasi-atomic model may be built by fitting and docking of atomic coordinates into the 3D density map.

Keywords: *consistency checks, docking, fitting, molecular boundaries, quasi-atomic model, segmentation, surface rendering, visualization*

Appendices:

Appendix 1 Some important definitions and theorems: Fourier transform, low- and high-pass filtration, correlation functions, convolution theorem

Appendix 2 Low-pass filters: Gaussian, Fermi, Butterworth

Appendix 3 Bibliography of methods

Appendix 4 Bibliography of structures

Appendix 5 Special journal issues on image processing techniques