Chapter 1
Macromolecular Crystallographic Computing

Kostas Bethanis
Agricultural University of Athens, Greece

Petros Giastas
Hellenic Pasteur Institute, Greece

Trias Thireou
Agricultural University of Athens, Greece

Vassilis Atlamazoglou
Agricultural University of Athens, Greece

ABSTRACT

Structural genomics or structural proteomics can be defined as the quest to obtain the three-dimensional structures of all proteins. Single-crystal X-ray crystallography provides the most direct, accurate and in most of the cases the only way of forming images of macromolecules. Using crystallography, three-dimensional images have been made of thousands of macromolecules, especially proteins and nucleic acids. These give detailed information about their activity, their mechanism for recognizing and binding substrates and effectors, and the conformational changes which they may undergo. This chapter presents the basic crystallographic procedure steps and a thorough survey of the computational software used most frequently by protein X-ray crystallographers. The determination of the structure of 2[4Fe-4S] ferredoxin from Escherichia coli. is examined as a case study of implementation of these steps and programs. Finally, some of the perspectives of the field of computational X-ray crystallography are noted showing the future developments in the ceaseless evolution of new methods and proliferation of new programs.

INTRODUCTION

Macromolecules are the principal non-aqueous components of living cells. Among the macromolecules (proteins, nucleic acids, and carbohydrates), proteins are the largest group. Enzymes are the most diverse class of proteins because nearly every chemical reaction in a cell requires a specific enzyme. To understand cellular processes, knowledge of

DOI: 10.4018/978-1-60566-768-3.ch001

Copyright © 2010, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

the three-dimensional structure of enzymes and other macromolecules is vital. Two techniques are widely used for the structural determination of macromolecules at atomic resolution: X-ray diffraction of crystals and nuclear magnetic resonance (NMR). While NMR does not require crystals and provides more detailed information on the dynamics of the model in question, it can be used only for biopolymers with a molecular weight less than 20,000. X-ray crystallography can be applied to compounds with molecular weight up to at least 10^6. For many proteins, the difference is decisive in favour of X-ray diffraction (Drenth J., 1994). The pioneering work by Perutz and Kendrew on the structure of hemoglobin and myoglobin in the 1950's led to a slow but steady increase in the number of proteins whose structure was determining using X-ray diffraction. The introduction of sophisticated computer hardware and software dramatically reduced the time required to determine a structure while increasing the accuracy of the results. In recent years, recombinant DNA technology has further stimulated interest in protein structure determination. A protein that was difficult to isolate in sufficient quantities from its natural source can often be produced in arbitrarily large amounts using expression of its cloned gene in a microorganism. Also, a protein modified by site-directed mutagenesis of its gene can be created for scientific investigation and industrial application. Here, X-ray diffraction plays a crucial role in guiding the molecular biologist to the best amino acid positions for modification. Moreover, it is often important to learn what effect a change in a protein's sequence will have on its three-dimensional structure. Chemical and pharmaceutical companies have become very active in the field of protein structure determination because of their interest in protein and drug design.

As of January 2008, the Protein Data Bank (PDB) (Berman H. M., *et al.*, 2000), the world's largest repository of macromolecular models obtained from experimental data (called *experimental* models), contains more than 40,000 protein and nucleic-acid models determined by X-ray crystallography (Berman H. M., 2008). However, it should be noted that because many proteins appear in multiple forms -for example, wild types and mutants, or solo and also as part of protein-ligand or multiprotein complexes- the number of unique proteins represented in the PDB is only a fraction of the total number of models. In addition, the PDB holds roughly 4500 models, mostly proteins of fewer than 200 residues, that have been solved by NMR spectroscopy, which provides a model of molecule in solution, rather than in crystalline state. Finally, there are *theoretical* models, either built by analogy with the structure of known proteins having similar sequence, or based on simulations of protein folding (see chapter 'Describing methodology and applications of an *in silico* protein engineering approach'). Theoretical models are also available from databases other than PDB. All methods of obtaining models have their strengths and weaknesses, and they coexist happily as complementary methods (Rhodes G., 2006).

SINGLE CRYSTAL X-RAY CRYSTALLOGRAPHY: THE BASIC STEPS

Crystallography provides the most direct way of forming three-dimensional images of molecules. The most common experimental means of obtaining a detailed model of a large molecule, allowing the resolution of individual atoms, is to interpret the diffraction of X-rays from many identical molecules in an ordered array like a crystal. This method is called *single-crystal X-ray crystallography*. The three-dimensional structures of macromolecules, especially proteins and nucleic acids, give detailed information about their activity, their mechanism for recognizing and binding substrates and effectors, and the conformational changes which they may undergo. They show graphically the evolutionary relationships be-

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/macromolecular-crystallographic-computing/39601

Related Content

Statistical Analysis for Radiologists' Interpretations Variability in Mammograms
Ahmad Taher Azar (2012). *International Journal of Systems Biology and Biomedical Technologies (pp. 28-46).*
www.irma-international.org/article/statistical-analysis-radiologists-interpretations-variability/75152

An Optimized Semi-Supervised Learning Approach for High Dimensional Datasets
Nesma Settouti, Mostafa El Habib Daho, Mohammed El Amine Becharand Mohammed Amine Chikh (2018). *Applying Big Data Analytics in Bioinformatics and Medicine (pp. 294-321).*
www.irma-international.org/chapter/an-optimized-semi-supervised-learning-approach-for-high-dimensional-datasets/182952

Text Mining for Biomedicine
Sophia Ananiadou (2009). *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration (pp. 1-9).*
www.irma-international.org/chapter/text-mining-biomedicine/23052

Clinical Data Linkages in Spinal Cord Injuries (SCI) in Australia: What Are the Concerns?
Jane Moon, Mary P. Galeaand Megan Bohensky (2015). *Big Data Analytics in Bioinformatics and Healthcare (pp. 392-405).*
www.irma-international.org/chapter/clinical-data-linkages-in-spinal-cord-injuries-sci-in-australia/121468

Sentiment Based Information Diffusion in Online Social Networks
Mohammad Ahsan, Madhu Kumari, Tajinder Singhand Triveni Lal Pal (2018). *International Journal of Knowledge Discovery in Bioinformatics (pp. 60-74).*
www.irma-international.org/article/sentiment-based-information-diffusion-in-online-social-networks/202364