

Introduction

Macromolecular crystallization in the structural genomics era

Progress in molecular biology and its applications to human health have, for the past quarter-century, been crucially dependent on atomic-resolution structural knowledge of proteins, nucleic acids, and other macromolecular complexes and assemblies. Beyond the transforming impact that three-dimensional structure has had on fundamental research in biochemistry and biology, macromolecular structure has proven to be of formidable value in biotechnology as well. Here it has provided the essential knowledge required to apply the strategy of structure-based drug design to the creation and discovery of novel drugs and pharmaceutical products. It serves as the basis for powerful approaches now being utilized in small, emerging biotechnology companies as well as major pharmaceutical enterprises to identify lead compounds to treat a host of human ailments, veterinary problems, and crop diseases in agriculture.

Of equal importance to biotechnology, but which also requires knowledge of three-dimensional structure, is the genetic engineering of proteins. Although recombinant DNA techniques provide the essential synthetic role that permits modification of the proteins, structure determination supplies the analytical function. It serves as a structural guide for intelligent and purposeful changes, in place of random and chance amino acid substitutions. Direct visualization of the structural alterations that are introduced by mutation offers new directions for chemical and physical enhancements.

Redundancies in structural elements and motifs emerging from the expanding collection of structurally known proteins suggest that the number of architectures and substructures that naturally occur are not only finite, but even manageable. Ultimately, all macromolecular structures may be classified and cataloged according to polypeptide folds. Once all, or most, of the folds which are utilized by nature are known, then this will provide predictive insight, based primarily on amino acid sequence, of the structures and functions of unknown proteins. The sequences of many proteins from a wide range of organisms are now being defined on a broad front through massive sequencing efforts, such as the human genome project, carried out in both the public and the private sectors. Extension of these genome projects to the three-dimensional structural level

of the proteome is the next logical step. Indeed, this effort, under the broad rubric of structural genomics, is now well under way.

At present, and in the foreseeable future, the only technique that can yield atomic-resolution structural images of biological macromolecules is X-ray diffraction analysis as applied to single crystals. While other methods may produce important structural and dynamic data, for the purposes described above, only X-ray crystallography is adequate. As its name suggests, application of X-ray crystallography is absolutely dependent on crystals of the macromolecule, and not simply crystals, but crystals of sufficient size and quality to permit collection of precise diffraction data. The quality of the final structural image is directly determined by the perfection, size, and physical properties of the crystalline specimen. Hence the crystal becomes the keystone element of the entire process and the ultimate determinant of its success.

When crystallizing proteins for X-ray diffraction analysis, one is usually dealing with pure often exceptionally pure macromolecules, and the objective is to grow only a few large, perfect crystals. It is important to emphasize that while the number of crystals needed may be few, often the amount of protein available may be severely limited. This in turn places grave constraints on the approaches and strategies that may be used to obtain those crystals. While new methodologies such as synchrotron radiation and cryocrystallography have driven the necessary size of specimen crystals consistently downward, they have in no way alleviated the need for single crystals of high perfection.

In assembling the articles found in this volume of the *Journal of Structural Biology*, an attempt has been made to call forward those principles, strategies, and technologies that currently shape our thinking regarding the challenges posed by macromolecular crystallization. This salient issue, and the associated problem of producing suitable protein samples for crystallization, are now at the very center of the structural genomics enterprise. Overcoming the obstacles posed by these problems promises to contribute more, in all likelihood, than advances in any other aspect of the structural genomics process. It is the fervent hope of the editor and the authors of this compendium that these papers will

offer the novel ideas, strategies, and engineering applications that may point the ways to assured success.

The papers in this volume may be classified according to four broad categories, though clearly, the multidisciplinary nature of the crystallization process invites much overlap and integration. There are four papers that deal with the physical theory, physical–chemical phenomena, assumptions, and available data pertinent to understanding in a quantitative manner the crystallization process as it applies to proteins, nucleic acids, and their complexes. These are the papers by Chernov, Garcia-Ruiz, Mühlig et al., and McPherson et al., though important contributions to this area are found as well in the paper by Wilson.

There are four eminently practical papers by Bergfors dealing with crystal seeding, Hanson et al. with cryocrystallography, Wilson with light scattering, and Dale et al., who address the question of genetically engineering proteins to better crystallize. There are three papers that might be cataloged as dealing with troublesome or emerging areas of macromolecular crystallization. One of these is by Caffrey, who focuses on membrane protein crystallization, principally from lipid matrices, and another is a comprehensive and insightful review of the unique problems presented by nucleic acid crystal growth from Golden and Kundrot. A final paper in this group, by van der Woerd et al., examines the possibilities of carrying out nanoliter-scale crystallization experiments using microfluidics.

Appropriate to the dominant theme of macromolecular crystallization in the structural genomics era, there are eight papers from academic laboratories, structural genomics centers, and private companies focused on the problems attending what has now come to be known as “high-throughput crystal growth.” Two of these, by

Goulding and Perry, and by Loll focus particularly on issues of protein expression, the former on conventional proteins in *Escherichia coli* and the latter on the efficient expression and purification of membrane proteins. A third paper, by Rupp, examines the design of crystallization trials and matrices and statistical approaches to identifying successful conditions.

Four papers in the area of high-throughput crystallization come from structural genomics centers and companies founded on its premise. These are by Hosfield et al., Luft et al., DeLucas et al., and Hui and Edwards. Their papers describe comprehensive and detailed strategies, both conceptual and engineering, for rapidly and efficiently producing proteins and crystallizing them. They explore the applications of robotics, microfluidics, nanoliter-crystallization trials, data acquisition and mining, and the design of trial matrices. Finally in this group is a novel, high-throughput approach based on crystallization in gels by Ng et al.

The distinctions between the material and ideas found in these four categories are ill defined, if they exist at all, but are simply intended to point the reader in a fruitful starting direction. There is a unity in the fundamental ideas and assumptions of all of these papers, but there is a broad diversity in objective, focus, and emphasis, and a wide variation in the application of the principles to real problems.

Alexander McPherson
Department of Molecular Biology and Biochemistry
University of California at Irvine
560 Steinhaus Hall
Irvine, CA 92697-3900, USA
E-mail address: amcphers@uci.edu