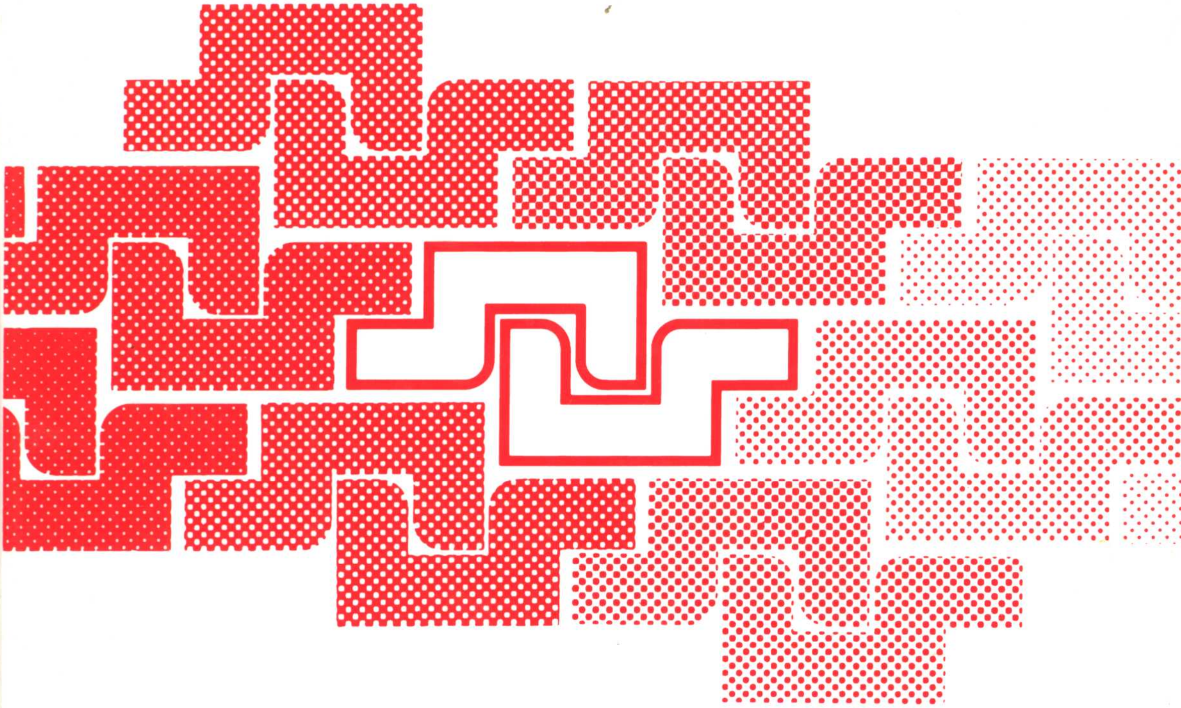

Methods in Macromolecular Crystallography

Edited by
Dušan Turk
Louise Johnson

IOS
Press
Ohmsha

NATO Science Series

Series I: Life and Behavioural Sciences - Vol. 325

Methods in Macromolecular Crystallography

Edited by

Dušan Turk

Department of Biochemistry and Molecular Biology,
Jožef Stefan Institute, Ljubljana, Slovenia

and

Louise Johnson

Laboratory of Molecular Biophysics, University of Oxford, Oxford,
United Kingdom

IOS
Press
Ohmsha

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

Published in cooperation with NATO Scientific Affairs Division

Proceedings of the NATO Advanced Study Institutes on
Methods in Macromolecular Crystallography
and
Chemical Prospectives in Crystallography of Molecular Biology
25 May–4 June 2000
Erice, Italy

ISBN 1 58603 080 9 (IOS Press)
ISBN 4 274 90429 6 C3045 (Ohmsha)
Library of Congress Catalog Card Number: 2001086580

Publisher
IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 620 3419
e-mail: order@iospress.nl

Distributor in the UK and Ireland
IOS Press/Lavis Marketing
73 Lime Walk
Headington
Oxford OX3 7AD
England
fax: +44 1865 75 0079

Distributor in the USA and Canada
IOS Press, Inc.
5795-G Burke Centre Parkway
Burke, VA 22015
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

Distributor in Germany, Austria and Switzerland
IOS Press/LSL.de
Gerichtsweg 28
D-04103 Leipzig
Germany
fax: +49 341 995 4255

Distributor in Japan
Ohmsha, Ltd.
3-1 Kanda Nishiki-cho
Chiyoda-ku, Tokyo 101
Japan
fax: +81 3 3233 2426

LEGAL NOTICE
The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

PREFACE

This volume contains most of the papers presented at an International School of Crystallography within the course "Methods for Macromolecular Crystallography" and some from the course "Chemical Prospectives in Crystallography of Molecular Biology" which were held at the Ettore Majorana Centre for Scientific Culture in late spring 2000 in Erice, Italy. The courses were financed by NATO as Advanced Study Institutes (ASI) and by the European Commission as a EuroSummerSchool.

The courses shared audience and speakers and brought together scientists from all over the world to present, discuss and learn about the fundamentals and the current state of the art of a diffraction based macromolecular structure determination. The unique and secluded environment of Erice helped to create an atmosphere in which formally announced and spontaneously organized discussions as well as computer demos and tutorials were extended far beyond the formal course schedule. The general outline of this volume is similar to the schedule of the methods course: from crystals (McPherson, DeTitta, Garman) throughout synchrotrons (Thompson), detector development (Ealick), data processing (Leslie, Otwinowski), ab initio phasing at high (Sheldrick) and low (Wilson) resolution including non-crystallographic electron density averaging of highly symmetric particles (Rossmann), molecular replacement (Navaza), experimental phase measurements (Weckert), density modification methods (Read) and map improvements (Glykos), interpretation of electron density maps (Jones), towards automatic structure determination (Turk) and dissection of an ultra high resolution structure (Jaskolski). A review of structural biology by the means of electron microscopy (Henderson and Baker) is followed by the structure of hepatitis B virus core shells determined by electron cryomicroscopy (Crowther and Bötther). A possible extension of X-ray diffraction methodology to imaging of non-crystalline specimens is pointed out by Sayre.

Thanks for the success of both courses (according to the analysis of the questionnaire, they made it to the top of the list of the International School of Crystallography meetings) goes to everyone: lecturers, participants from all over the world and the local crew. John Irwin did, with a help from Orazio Mistretta, an excellent job organizing the computer hardware. As all other International Schools of Crystallography also this one was only made possible by the excellent leadership, experience, wisdom and support of Paola Spadon and Lodovico Riva di Sanseverino. Special thanks goto Gregor Gunčar, who adjusted the submitted manuscripts based on Microsoft Word to the publisher standards.

The Course benefited from important financial support from several organizations (International Union of Crystallography; International Union of Biochemistry and Molecular Biology; EMBO Heidelberg, Banco di Sicilia, Sede di Trapani; National Science Foundation, Washington) and companies (AstraZeneca, Molndal; Bristol-Myers Squibb Pharmaceutical; Bruker axs, Karlsruhe; Compaq, Computer Corporation, USA; Cryosystems, Oxford; Hampton Research, Laguna Niguel, CA; Janssen Pharmaceutica N.V., Beerse; MAR Research, Norderstedt; MSC, The Woodlands, TX; Merck & Co., Inc., USA; Novartis, Basel; Novo Nordisk, Bagsvaerd; Pfizer, Kent; Pharmacia-Upjohn Research, Nerviano, Milan; Res. Inst., Princeton, NJ; Roche Discovery, Welwyn Garden City; Schering-Plough Res. Inst., Kenilworth, NJ).

Dušan Turk and Louise Johnson

Contents

Methods in Macromolecular Crystallography
D. Turk and L. Johnson (Eds.)
IOS Press, 2001

Macromolecular Crystal Structure and Properties as Revealed by Atomic Force Microscopy

Alexander McPHERSON
University of California, Irvine
Department of Molecular Biology and Biochemistry
Irvine, CA 92697-3900, USA

Abstract. Atomic force microscopy (AFM) has been used to study protein, nucleic acid, and virus crystals in situ, in their mother liquors, as they grow. From sequential AFM images taken at brief intervals over many hours, or even days, the mechanisms and kinetics of the growth process can be defined. The appearance of both two and three-dimensional nuclei on crystal surfaces have been visualized, defect structures were clearly evident and defect densities of crystals were also determined. The incorporation of a wide range of impurities ranging in size from molecules to micron or larger microcrystals, and even foreign particles were visually recorded. From these observations and measurements a more complete understanding of the detailed character of macromolecular crystals is emerging, one that reveals levels of complexity previously unsuspected. The features of these crystals apparent in AFM images undoubtedly influence the diffraction properties of the crystals and the quality of the images obtained by X-ray crystallography.

AFM can yield images of extraordinary clarity of complex surfaces and objects. It is applicable to fields ranging in size from less than 20 nm up to about 150 μm, and with a spatial resolution on biological, soft materials of about 1 nm, with a height resolution as great as 0.1 nm. Thus it provides precise visual detail over a size range that is beyond most other techniques. Its application extends over dimensions lying between individual macromolecules, which are accessible by X-ray crystallography, macromolecular assemblies amenable to electron microscopy, and includes living cells which can just be seen using light microscopy. Because visualization is carried out in a fluid environment, specimens suffer no dehydration as is generally necessary with electron microscopy, they require no freezing, fixing or staining, indeed, living specimens can be observed over long periods so long as they stay relatively well put. Specimens seem in most cases oblivious even to the presence of the probe tip poking about their surfaces.

The great power of AFM, however, lies not just in its imaging capability, but in the non-perturbing nature of the probe interaction with the surface under study. Because the specimen is unaware of the probe, natural processes, such as growth, continue uninhibited. This allows the investigator to record not simply a single image, but a series that may extend over hours or even days. This is ideal for the study of the growth of macromolecular crystals, which develop over such periods of time. Imaging frequency depends on the scan rate of the probe, and images may be gathered rapidly, within a few seconds, or over several minutes. For macromolecular crystal growth, a relatively slow process, events on the surface impose no requirement for high scan speed. Thus a long series of high quality images are generally accessible to the investigator.

Another property of AFM carried out in fluid cells, is that the media or mother liquor can be changed during the course of experiments without appreciably disturbing the specimen. This is of considerable value in the study of macromolecular crystals because it is often desirous to study the growth process under different conditions of supersaturation. Growth steps are usually visible on the surfaces of crystals, and because advancement is relatively slow, their progression can be readily recorded in a temporal sequence of images. When rates are measured as a function of temperature, salt concentration, supersaturation, or some other influence, then growth step velocities can be used to deduce thermodynamic and kinetic parameters such as the step free energy, and the kinetic constant [1, 2, 3]. In the best of cases, individual virus particles, and even single protein molecules can be observed as they are recruited into advancing step edges.

Macromolecular crystals grow by a variety of mechanisms, some familiar to conventional crystal growth [4, 5, 3], but also by another mechanism, which may be unique. Although kinetic parameters are strikingly different, the underlying physics of the growth processes and the thermodynamic principles are the same as for other crystals [6, 7, 8, 9]. The major differences between macromolecule and conventional crystal growth arise almost entirely from the large sizes and weak interactions of the macromolecules, the liquid environment and the consequent role of water, and the generally higher level of impurities that characterize macromolecular preparations.

What first strikes the investigator using AFM is the complexity, diversity, and variability of macromolecular crystal surfaces. These arise from the different mechanisms and growth processes, and their combinations, the spectrum of defects and dislocations, the roughness of the surfaces produced by impurities and multiple conformers, and the asymmetries and shapes arising from the bonding energies of molecules in different directions in the lattice. AFM has been used to analyze not only the growth but also the dissolution of protein, nucleic acid, and virus crystals and this further serves to delineate the causes that produce and influence this diversity. AFM has been used to provide quantitative descriptions of the crystallization process and to visualize the types, density, and distribution of defects and dislocations throughout crystals.

The purpose of AFM studies is primarily to advance our understanding of the fundamental physics and chemistry underlying the crystallization process. It's second objective is improving the crystallization process in support of macromolecular X-ray crystallography. It is fair to argue that increased understanding of the process may be translated into more effective and efficient crystallization approaches and methods, and these, ultimately, into larger crystals of more proteins, as well as crystals that diffract to higher resolution, have reduced mosaic spread, and withstand the rigors of cryocrystallography [10] and data collection in a more robust fashion.

There is at least some likelihood that AFM may be able to contribute in other ways to X-ray diffraction analyses. Because height information is preserved, for example, the handedness of molecular arrangements arising from screw axes can be deduced. Thus AFM may provide a means of discriminating enantiomorphs, as was done for crystals of fungal lipase [11], something difficult early in a structure analysis. Packing arrangements of molecules within unit cell are sometimes discernable, which may assist in molecular replacement structure solutions. In the case of virus crystallography, the value may be even greater. With AFM, the orientations of individual virus particles, and the capsomere structure of their surfaces may be visible [12]. These may then be used for constructing initial models at low resolution for phase extension. It is important to bear in mind that images are obtained under fully hydrated and unperturbed conditions, thus the images represent the molecules or virus particles as they actually exit in the crystal used for data collection.

A principle that dominates all aspects of crystal growth, macromolecular and otherwise, is the degree of supersaturation of the mother liquor. Virtually all kinetic and thermodynamic parameters vary with supersaturation. This includes the probability of forming critical nuclei, that is, the birth of a new crystal, initiation of new layers on an existing surface, the velocity of step movement on the surface, the incorporation of impurities [13, 14, 3, 1], and a host of lesser properties. Even the particular mechanism employed for growth of a crystal surface is dependent on supersaturation. Supersaturation in turn may, of course, be a function of an array of experimental variables such as salt concentration, macromolecule concentration, temperature, or other physical and chemical factors. It is also dependent on the underlying physical and chemical properties of the macromolecules and the manner by which they interact with one another.

There are four principal mechanisms that have been described from AFM for the development of faces of macromolecular crystals [15]. It should be noted, however, that different faces of a single crystal, being non-identical, might simultaneously employ different mechanisms for development. Even a single face may use more than one mechanism at the same time, and the type of mechanism may change as some experimental variable, such as temperature is altered. Thus, when only one, or a few observations of growth mechanism is available for a particular crystal, this by no means implies that other mechanisms are not operative at other times or under other conditions. Most crystals, it seems, utilize all mechanisms at one time or another, though some one mechanism may be strongly favored.

There are two dominant mechanisms in protein, nucleic acid, and virus crystal growth that serve to generate growth step edges and thereby lead to layer by layer addition of molecules. These are mechanisms that are also important in conventional crystal development, and they are growth by screw dislocation, and growth by the spontaneous appearance on active surfaces of two-dimensional nuclei. Many examples have now been recorded that capture both of these processes. A third mechanism, known as normal growth does not lead to layer by layer addition, but relies on intense random nucleation on active crystals where the surface free energy is unusually low. Though more rarely observed, this mechanism has been recorded for several macromolecules including an intact antibody and ferritin.

An additional mechanism that may be unique to macromolecular crystals, and which has not been described for conventional crystal growth, arises as a consequence of the unique properties of concentrated macromolecular solutions. For virtually all of the protein, nucleic acid, and virus crystals investigated by AFM, the sudden appearance of prominent, multilayer stacks of growth layers has been observed. Often these hillocks, whose characteristic shapes frequently reflect the gross morphology of the entire crystal, are ten to a hundred or more layers in height. Each layer of the stack provides step edges and, therefore, sources for tangential growth and the formation of new layers. Growth by this mechanism, which has been termed growth by three-dimensional nucleation, can in some cases be the dominant growth mechanism [16].

An intriguing question is the origin of these multilayer stacks. One explanation, for which there is now substantial evidence, suggests that they arise from liquid protein droplets that exist in concentrated macromolecular solutions [17, 18, 19], particularly crystallization mother liquors. These liquid protein droplets are composed of hundreds to thousands of molecules, exhibiting short – range order mediated principally by non specific hydrophobic interactions, and random arrangements of hydrogen bonds. Because of the extraordinary concentration of molecules in the droplets, they are locally hyper-saturated.

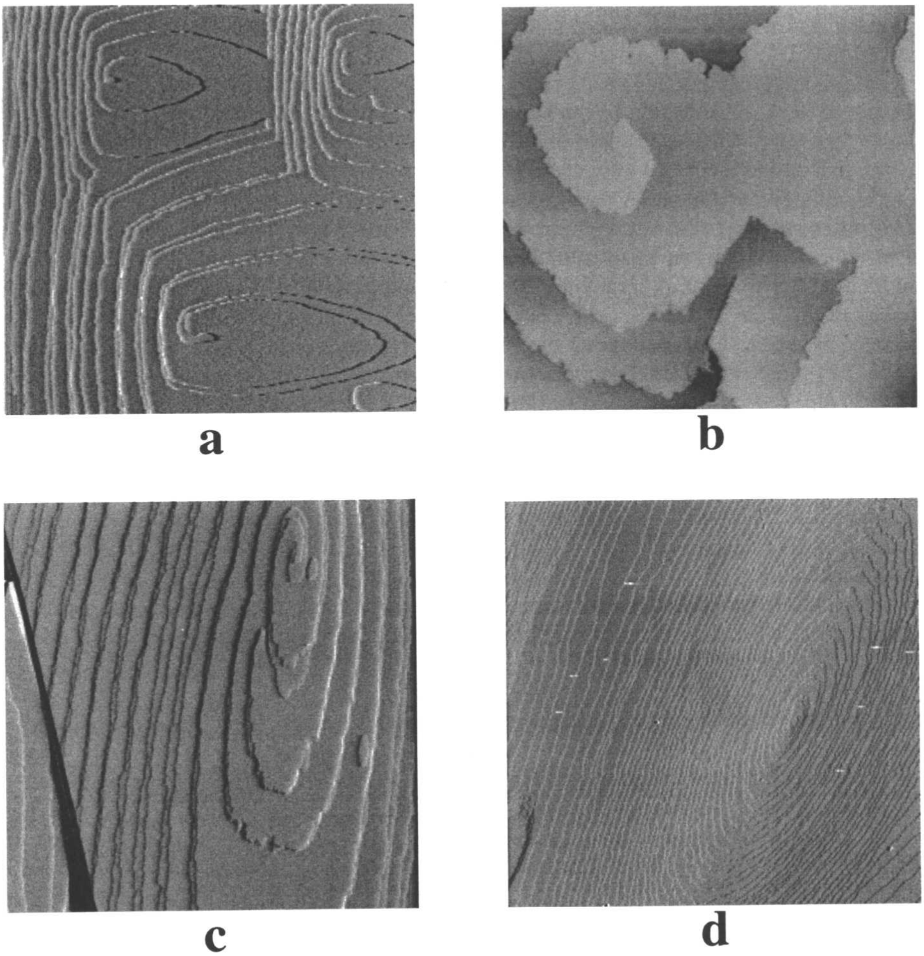

Figure 1. Screw dislocations of different sorts on the surfaces of protein crystals. In (a) two single and one double screw dislocations on a thaumatin crystal, (b) a single right-handed spiral on the surface of a canavalin crystal, (c) a left-handed double screw dislocation on a lysozyme crystal, and in (d) a steep vicinal hillock formed by a screw dislocation source on a lysozyme crystal. Scan areas are (a) 15 x 15 μm^2, (b) 15 x 15 μm^2, (c) 12 x 12 μm^2, (d) 2 x 2 μm^2.

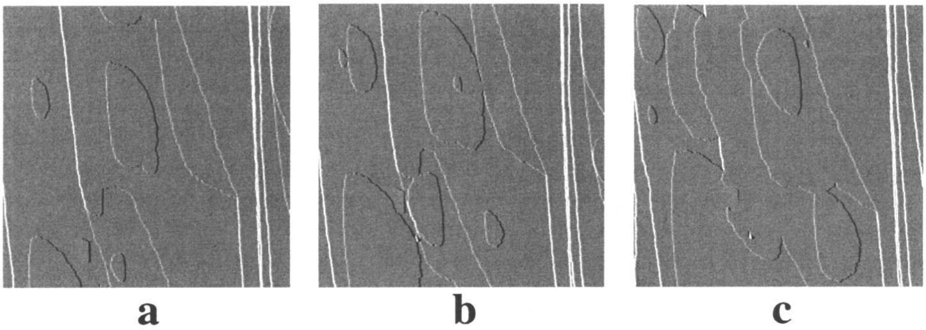

Figure 2. A second sequence of AFM images a different area on the <001> face of a beef liver catalase crystal. The area is 32.5 x 32.5 μm^2, and the interval between images is 12 min. Again, note the sequence of right- and left-handed islands that alternately apears on successive growth layers of the crystal.

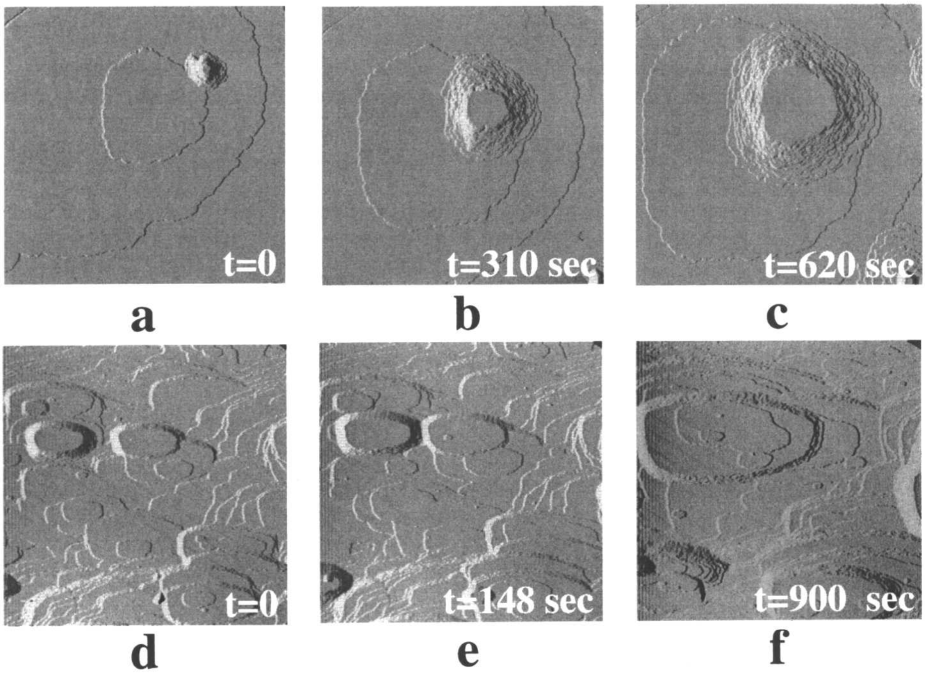

Figure 3. The sudden appearance of a prominent multilayered stack of several tens of growth layers on the surface of a crystal of STMV is seen to develop tangentially in (b) and (c). A large number of three-dimensional, multilayered stacks of growth steps, here in the range of several up to a dozen or more, are seen to appear and develop on the surface of a thaumatin crystal. Scan areas are (a) - (c) 15 x 15 μm^2, (d) - (f) 11.4 x 11.4 μm^2.

When the droplets sediment upon existing crystal surfaces, the lattice serves as an epitaxial substrate to guide and promote crystallization in the molecules above. These form a crystal layer, inspire crystallinity in the molecules above them, and so forth, propagating a continuous series of growth layers, a multilayer stack.

The existence of a liquid protein phase in concentrated protein solutions may have consequences for the physical chemistry and structure of concentrated macromolecular solutions, such as occur inside living cells, far outside the area of crystallization. It may also provide an explanation, or a pathway, not only for the mechanism of crystal growth through three-dimensional nucleation, but also for the spontaneous formation in solution of crystal nuclei having critical size [20].

Levels of impurities and contaminants in macromolecular solutions, despite the greatest care, vastly exceed those in conventional crystal growth solutions [21]. This is unavoidable and unlikely ever to change. Intuitively we might suspect that the kinds of impurities most detrimental to macromolecular crystal growth are those of large size, in the range of nutrient molecules or larger. These, if incorporated into a developing lattice would be most likely produce dislocations, and the kinds of defects that we can clearly see using AFM. Probably the most damaging impurities to the crystal are misoriented, improperly folded, or molecules having alternative conformations including clusters or aggregates of the nutrient molecules, foreign particles such as dust, microcrystals, and other contaminating macromolecules. AFM shows that all of these types of impurities can become incorporated into crystal lattices.

Individual defects, and the overall defect structures [22] present in macromolecular crystals are particularly amenable to visualization by AFM [23]. These show considerable variety, but taken as an ensemble of faults, they suggest the basis of the effect known to X-ray crystallographers as mosaicity. They also suggest why some crystals may appear far more ordered, and diffract to higher resolution that do others. One important finding from AFM studies, where one can simply count defects and dislocations directly, is that macromolecular crystals contain two to four orders of magnitude more faults than do most conventional crystals [21, 23].

Finally, some modest techniques and tools developed for *in situ* AFM of macromolecular crystals seem, by extension, to be useful in X-ray crystallography as well. An example is the seeding approach we use to promote limited crystal growth on surfaces in AFM fluid cells. This relies on inducing nucleation in a 1 μl droplet by vapor diffusion with subsequent flooding by much larger volumes of tempered mother liquors [11]. A second procedure was inspired by *in situ* AFM, and necessitated by the need to carry out sequential AFM and X-ray topography on many of the same crystals in a sample. This approach we refer to as "*in situ* X-ray crystallography" and utilizes crystals grown in droplets by vapor diffusion on small wands that subsequently may be partially dried, treated, or frozen, and thus provide a two dimensional array of crystals displayed on a small surface for X-ray analysis [24].

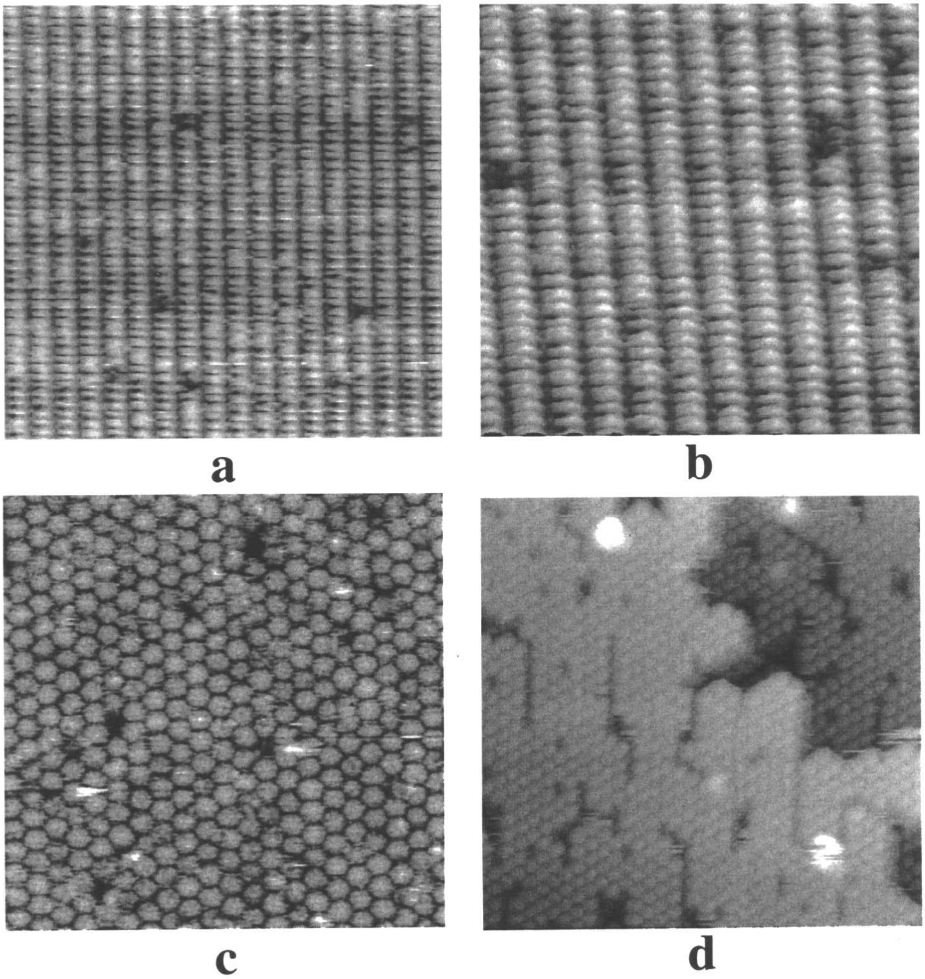

Figure 4. Many unit cells seen on the surfaces of growing macromolecular crystals are partially unfilled or entirely vacant, sometimes over several or even many consecutive cells. These remain unfilled as new layers form over them as revealed by etching experiments. In (a) and (b) are surfaces of thaumatin crystals, in (c) a crystal of BMV, and in (d) an orthorhombic crystal of STMV. Scan areas are (a) 280 x 280 nm^2, (b) 225 x 225 nm^2, (c) 540 x 540 nm^2, (d) 600 x 600 nm^2.

References

[1] Chernov, A.A., Rashkovich, L.N., Smolískii, I.L., Kuznetsov, Yu.G., Mkrtchyan, A.A. and Malkin, A.I. (1988) Growth of Crystals, Vol. 15, 43-91, edited by E.I. Givargizov and S.A. Grinberg, Consultant Bureau, New York.

[2] Chernov, A.A. (1993) Roughening and melting of crystalline surfaces. Prog. Cryst. Growth 26, 195-218.

[3] Chernov, A.A. (1984) Book. Modern crystallography III, Crystal growth. Springer-Verlag, Berlin.

[4] Buckley, H. E. (1951) Book. Crystal Growth, John Wiley and Sons, London.

[5] Burton, W.K., Cabrera, N. and Frank, F.C. (1951) The growth of crystals and the equilibrium structure of their surfaces. Philos. Trans. R. Soc. Land. A243, 299.

[6] Boistelle, R. and Astier, J.P. (1988) Crystallization mechanisms in solution. J. Cryst. Growth 90, 14-30.

[7] Feigelson, R.S. (1988) The relevance of small molecule crystal growth theories and techniques to the growth of biological macromolecules J. Cryst. Growth 90, 1-13.

[8] Feher, G. (1986) Mechanisms of nucleation and growth of protein crystals. J. Cryst. Growth 76, 545-546.

[9] Durbin, S.D. and Feher, G. (1996) Protein Crystallization. Annu. Rev. Phys. Chem. 47, 171-204.

[10] Garman, E.F. and Schneider, T.R. (1997) Macromolecular Cryocrystallography. J. Appl. Cryst. 30, 211-237.

[11] Kuznetsov, Yu.G., Malkin, A.J., Land, T.A., DeYoreo J.J., Barba de la Rosa, A.P., Konnert, J., and McPherson, A. (1997) Molecular resolution imaging of macromolecular crystals by atomic force microscopy. Biophys. J. 72, 2357.

[12] Malkin, A.J., Kuznetsov, Yu. G., Lucas, R.W. and McPherson, A. (1999) Surface processes in the crystallization of turnip yellow mosaic virus visualized by Atomic Force Microscopy. Journal of Structural Biology 127, 35-43.

[13] Rosenberger, F., Vekilov, P.G. Muschol, M. and Thomas, B.R. (1996) Nucleation and Crystallization of Globular Proteins-What We Know and What is Missing. J. Cryst. Growth. 168, 1-27.

[14] Schlichtkrull, J. (1957) Growth rates of protein crystals. Acta Chem. Seand. 11, 439-452.

[15] Malkin, A. J., Kuznetsov, Yu. G., Land, T. A., DeYoreo, J. J. and McPherson, A. (1995a) Mechanisms of growth for protein and virus crystals. Nat. Struct. Biol. 2, No. 11, 956-959.

[16] Malkin, A. J., Land, T. A., Kuznetsov, Yu. G., McPherson, A. and DeYoreo, J. J. (1995b) Investigation of virus crystal growth by in situ atomic force microscopy. Phys. Rev. Letters 75, No. 13, pgs. 2778-2781.

[17] Asherie, N., Lomakin, A. and Benedek, G. B. (1996) Phase diagram of colloidal solutions. Phys. Rev. Lett. 77, 4832-4835.

[18] Lui, C., Lomakin, A. Thurston, G.M., Hayden, D., Pande, A., Pande, J., Ogun, O., Asherie, N. And Benedek, G.B. (1995) Phase separation in multicomponent aqueous-protein solutions. J. Phys. Chem. 99, 454-461.

[19] Kuznetsov, Yu. G., Konnert, J. Malkin, A. J. and McPherson, A. (1999) The advancement and structure of growth steps on thaumatin crystals visualized by atomic force microscopy at molecular resolution. Surface Science 440, 69-80.

[20] Ten Wolde, P.R. and Frenkel, D. (1997) Enhancement of protein crystal nucleation by critical density fluctuations. Science 277, 1975.

[21] McPherson, A, Malkin, A., Kuznetsov, Yu.G., Koszelak, S. (1996) Incorporation of Impurities into Macromolecular Crystals. J. Cryst. Growth. 168, 74-92.

[22] Tiller, W. A. (1991) The science of crystallization: macroscopic phenomena and defect generation. Cambridge University Press, Melbourne Sydney.

[23] Malkin, A. J., Kuznetsov, Y. G. and McPherson, A. (1996) Defect Structure of Macromolecular Crystals. J. Struc. Biol. 117, 124-137.

[24] McPherson, A. (2000) In situ x-ray crystallography. J. of Applied Crystallography, in press.

Methods in Macromolecular Crystallography
D. Turk and L. Johnson (Eds.)
IOS Press, 2001

High Throughput Macromolecular Crystallization:
An Application of Case-Based Reasoning and Data Mining

Igor JURISICA, Patrick ROGERS
Ontario Cancer Institute/Princess Margaret Hospital
610 University Avenue, Toronto, Ontario M5G 2M9, Canada

Janice GLASGOW, Suzanne FORTIER
Queen's University, Departments of Chemistry and Computing and Information Science
Kingston, Ontario K7L 3N6, Canada

Robert COLLINS, Jennifer WOLFLEY, Joseph LUFT and George DETITTA
Hauptman-Woodward Medical Research Institute
73 High Street Buffalo, NY 14203-1196, U.S.A.

1. Introduction

Crystallization continues to be an important bottleneck for structural genomics efforts. Laboratories are facing the challenge of having to crystallize hundreds, possibly thousands, of proteins yearly. Our approach to the challenge is a blend of high throughput wet lab work and computer analysis via case-based reasoning and knowledge-discovery techniques. The thrust of this article is to describe the latter but a brief description of the former will help set the stage for the sophisticated computational efforts.

2. The HTP Search Laboratory

We have outfitted a high throughput crystallization laboratory for the handling of hundreds of proteins a year. The lab includes two pipetting robots ideally suited for microbatch-under-oil [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] crystallization setups and a digital photography robot that can capture the outcomes of the setups. The pipetting robots are modified Robbins Scientific Hydras, one with 96 pipettes, the other with 384 pipettes. Crystallization setups are done in high density microarray plates from Greiner. Each plate contains 1536 wells. Each well contains paraffin oil (4.5 μL), a chemically distinct crystallization "cocktail" (0.2 μL), and protein in a minimal buffer (0.2 μL). Cocktails, i.e., a compound, or compounds, that are known to reduce the solubility of proteins in water, are from five basic groups: inorganic salts at various concentrations; PEGs of various molecular weights in various buffers (pH range 4.5 – 11); PEGs of various molecular weights at various concentrations in combination with inorganic salts at fixed (200 mM) concentrations; very fine concentration screens of three salts (lithium chloride, potassium thiocyanate, and ammonium sulfate); and the commercially available

Hampton Crystal Screens I and II. Crystal Screen I is the commercial formulation of the original Jancarik and Kim (1991) "sparse matrix" screen.

Plates are set up by delivering oil to all the wells using the Hydra 384. This requires four pipetting operations. In a typical run, thirty to fifty plates are prepared with oil. Following the oil delivery the pipetting robot is thoroughly cleaned and then used to dispense cocktails from previously prepared 384 well plates, each having a chemically distinct cocktail. Again, four pipetting operations are required to place cocktails in each of the wells. Once the plate contains oil and cocktail protein stock is added using the Hydra 96. It takes sixteen pipetting operations, or about 7 minutes, to dispense protein to all the wells. Protein is never dispensed with the Hydra 384. This is because each pipette has a dead volume of about 1 μL. While it is acceptable to lose 96 μL of protein stock which, at a typical concentration of 10 mg/mL, represents 1 mg of lost protein, it is unacceptably wasteful to lose nearly 4 mg were the Hydra 384 to be used. Using the Hydra 96 we are able to set up the 1536 wells of a plate with approximately 600 μL of protein stock. Once the protein solution is delivered the syringes can be purged to recover approximately 200 μL of protein stock. At 10 mg/mL, our protein consumption is about 4 mg per plate.

Once the experiments are set up the plates are transferred to a photography robot. This robot can record digitally captured photographs for each well of a plate in about 20 minutes. The robot can accommodate 28 crystallization plates (43,008 crystallization experiments) at a time, and the total time to record all 43,008 outcomes is about 9 hours. Each photograph is saved as a tiff image (320 x 320 pixels in RBG). A working image, which is a minimally compressed jpeg file, is created from the tiff image. Thumbnail sized working images can be viewed on the computer screen 96 at a time. Should a particular outcome be suggestive, the thumb-nail can be clicked to show a larger version of the working image, along with all the information available about the cocktail. Photographs are taken on a regular basis: immediately following setup, one day later, two days later, one week later and two weeks later. It is possible to click a checkbox on the thumb-nail, marking it for historical analysis. In a historical analysis all of the images for the well in question are pulled up on the screen in the order in which they were recorded. While crude, this kind of analysis can suggest if a crystal appeared quickly after setup or not.

Viewed in the conventional way, the robotics lab is a fast, efficient implementation of a screening strategy. We, however, view the approach as a searching strategy. The distinction is not merely one of semantics. In a screening strategy the only useful outcome is a clearly "crystalline" outcome; i.e., large, well-developed crystals; small crystal; microcrystals; or microcrystalline precipitate. In a searching strategy all outcomes are useful. Their utility will become more apparent in the discussion of work ongoing in the computer labs.

3. MAX - An Intelligent Decision-Support System for Crystallization Experiment Design

The primary hypothesis of our research is that past experience can lead us to the identification of initial conditions favorable to crystallization. Faced with the problem of crystallizing a new protein we suggest that successful recipes developed for "similar" proteins provide an optimal starting point for the lab work. Thus, there are two main issues that need consideration: (1) creating a representative repository of past experiences, and (2) finding an effective way to measure "similarity" among proteins. The repository is being constructed using the HTP setup described earlier. A solution to our second challenge is based on a hypothesis that the results of initial solubility experiments can provide a quantitative measure of similarity between proteins with respect to crystallization.

MAX incorporates multiple algorithms (case-based and rule-based reasoning, image processing and knowledge discovery), multiple databases, and a knowledge base. Knowledge in MAX has two forms – experiential (cases) and general principles (rules). The case base stores cases, which are individual experiments with diverse crystallization outcomes (e.g., clear drop, undifferentiated precipitate, amorphous precipitate, crystalline precipitate, microcrystals, phase separation). The general principles could include useful rules acquired from crystallographers, or principles derived using knowledge-mining tools.

After creating a seed crystallization experience repository, we use MAX in two ways: (1) to suggest a crystallization plan for a new protein using case-based reasoning paradigm and (2) to analyze the case base to find underlying principles of crystal growth by using knowledge discovery techniques. The following sections describe these approaches in more detail.

3.1. Case-Based Reasoning

A standard technique for human problem solving is to recall past experiences that are in some way similar to the current situation. These "cases" are then adapted and used to construct a solution for a given problem. Case-based reasoning (CBR) systems are computer programs that incorporate such past experiences as a guide to problem solving. MAX's reasoning engine builds on an efficient and effective CBR system called $\mathcal{TA}$**3** [12, 13, 14].

Cases in MAX capture the problem-solving process of a crystal growth experiment. They contain all of the relevant information about a particular experiment, including input parameters, results of the initial precipitation experiments (including images, extracted image features and outcome classification), and the final results. A case history of a known protein in the case base comprises three components: (1) precipitation reaction index, (2) intrinsic properties of the protein, and (3) the collection of strategies that were employed to crystallize the protein. Although at least one of those strategies had to have succeeded for the protein to be included in the information repository, we also record all the unsuccessful strategies.

The process of using MAX as a decision-support tool is as follows: Run an HTP screening to obtain a precipitation reaction index for a given protein. Use CBR to find the most similar past experiments and adapt their plans to fit the specifics for the current macromolecule. Record any modifications made to the original plan during the crystallization process. These modification can later be used to improve the adaptation process.

The similarity of two proteins is measured as a distance between their precipitation indices. To improve both precision and recall of the CBR retrieval system, we use a two-step process: (1) a binary precipitation index is used to compute the Hamming distance to find a neighborhood of similar proteins; (2) the retrieved subset of the case base is further filtered by differentiating precipitates.

Once relevant cases have been retrieved, the next step in CBR is adaptation, i.e., modifying previous solutions to address the new problem. MAX constructs a solution for the current crystallization problem by using appropriate descriptors from relevant