



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Structural Biology 142 (2003) 154–161

Journal of
Structural
Biology

www.elsevier.com/locate/yjsbi

High-throughput protein crystallization

Raymond Hui^a and Aled Edwards^{a,b,*}

^a *Affinium Pharmaceuticals, 100 University Avenue, Toronto, Ont., Canada*

^b *Banting and Best Department of Medical Research, University of Toronto, and Ontario Center for Structural Proteomics, University Health Network, Toronto, Ont., Canada*

Received 5 February 2003

Abstract

High-throughput structural biology is a focus of a number of academic and pharmaceutical laboratories around the world. The use of X-ray crystallography in these efforts is critically dependent on high-throughput protein crystallization. The application of current protocols yields crystal leads for approximately 30% of the input proteins and well-diffracting crystals for a smaller fraction. Increasing the success rate will require a multidisciplinary approach that must invoke techniques from molecular biology, protein biochemistry, biophysics, artificial intelligence, and automation.

© 2003 Published by Elsevier Science (USA).

1. Introduction

In the past few years, advances in genomics have spawned large-scale structural biology efforts (structural proteomics) in both academic and pharmaceutical laboratories. There are publicly funded research efforts in structural proteomics in Europe, Japan, and North America. In the United States, the projects aim to provide greater insight into the mechanisms of protein folding (Terwillinger, 2000). In Canada, the projects are designed to use structural biology to derive protein function. In Japan, there are two major projects in large-scale structural biology: the RIKEN Structural Genomics Initiative and the Human Membrane Protein Project at the Biological Information Research Center (Yokoyama et al., 2000). In Europe, there are a number of major initiatives, including the Protein Structure Factory in Germany and the newly formed European initiative, SPINE (Heinemann, 2002).

As was the case for genomics, government-funded projects in structural proteomics along with academic research groups have created spin-off biotechnology companies, including Affinium Pharmaceuticals in

Canada; Structural Genomix, Syrrx, and Plexxikon in the United States; and Astex Technology in the United Kingdom (Dry et al., 2000). Most of the private efforts were launched with overexuberant goals. Recently, the industrial players have adopted a more pragmatic business model that is focused on structure-based drug discovery. In the drug discovery setting, these established high-throughput structural biology engines are being harnessed to create numerous constructs to tackle a small number of intractable proteins. The ability to rapidly create and screen up to hundreds of constructs per target increases significantly the probability of success. With the need for potentially hundreds of constructs per protein as well as around the same number of protein–ligand complexes, the demand for throughput in the private sector is probably equivalent to that required for proteome-based projects.

Whether the objective is to solve the structures of a large number of distinct proteins or to solve structures of variations of a small number of disease targets, X-ray crystallography is the workhorse of structural biology. Indeed, most academic and industrial laboratories rely exclusively on crystallography. The reliance on X-ray crystallography underscores the central importance of protein crystallization, which not only is among the most complex and least understood problems in structural biology but also is the step that has the greatest rate of attrition.

* Corresponding author. Fax: +416-598-1262.

E-mail addresses: aedwards@afnm.com, aled.edwards@utoronto.ca (A. Edwards).

The public and industry-based large-scale structural biological projects which have been running for 2 years or longer are expected to contribute greatly to the protein crystallization literature. Although the literature from these projects is currently restricted to methods for protein crystallization (“how to”), we can soon expect to have information that describes the fruits of these efforts (“how well did the methods/techniques work on large scale”).

In this article, we review the methods and tools used in high-throughput protein crystallization, as well as the initial results from analyzing crystallization data.

2. Protein productionthe pipeline feeding crystallization

Although approaches vary, the steps leading to protein crystallization—cloning, expression, and purification—are common to all large-scale structural biology projects; the complexity and the rates of attrition increase along the pipeline. The DNA-based stages of the protein production pipeline have been readily automated, in part because of the relative simplicity and stability of DNA. Cloning, which consists primarily of liquid handling, has been automated to the point at which many labs have the capacity to mass produce expression constructs 96 at a time, many times per week (Arrowsmith et al., 2003). Testing the levels of recombinant protein expression, while more labor-intensive, has also been adequately mastered such that it can be conducted at high throughput; see for example Lesley (2001).

A completely automated system that can accomplish high-throughput, large-scale purification of proteins for structural biology is not yet commercially available. Systems that yield partially purified proteins do so on a microtiter-plate scale (Albala et al., 2000; Braun et al., 2002) and typically yield less than 1 mg of protein. Companies such as Qiagen and Beckman now offer standard kits for purification on this scale.

Purifying proteins to homogeneity on a larger scale is more challenging to automate. Proteins differ dramatically in their expression, solubility, and purification properties, and this fact has led most scientists to adopt a combination of manual approaches and off-the-shelf liquid chromatography systems, rather than to design completely automated systems. However, there is room for automated approaches in protein purification. There are subsets of proteins (largely those that are produced to very high levels) that can be purified to near homogeneity using generic approaches, such as affinity chromatography.

Various biotechnology companies have developed proprietary, more integrated solutions for larger scale purification. Syrrx of San Diego, California, USA, employs a system that purifies sufficient protein for crystallization and was developed at the Genomics Institute

at the Novartis Research Foundation. It combines robotic centrifugation and sonication with a parallel column chromatography system capable of purifying 96–192 proteins per day (Lesley, 2001). In this system, other steps, including desalting and concentration, are carried out offline before the samples are ready for crystallization. Affinium Pharmaceuticals of Toronto, Ontario, Canada, which champions a “flexible automation” approach, developed the ProteoMax system to process cell extracts to purified, concentrated protein samples suitable for protein analysis. This system clarifies the cell lysate, performs the column chromatography, desalts, and concentrates; in optimal cases, the purified protein is ready for structural studies. By obviating centrifugation and integrating all steps into an automated system, the ProteoMax system can perform the high-throughput purification problem in its totality.

3. Automation of protein crystallization

By enabling high-throughput cloning, expression, and purification, the number of pure protein samples entering crystallization can be staggering, with a number of laboratories now having the theoretical capacity to exceed 1000 different proteins per month. To keep up with this demand, it is critical to employ automated crystallization, crystal visualization, and crystal optimization techniques.

Protein crystallization is currently a process defined by trial, error, anecdotes, philosophy, and best guesses. In large part, this results because of nonstandardized approaches. Different laboratories adopt significantly different approaches to the problem, largely because proteins themselves are by design nonuniform. It is therefore possible that one of greatest contributions of structural proteomics will be to provide crystallization data derived from thousands of different proteins which have been purified and crystallized using standard protocols. These data, when analyzed for trends and patterns, will enable more quantitative assessments of crystallization parameters. In addition, the availability of thousands of purified proteins will provide experimenters with a battery of reagents with which to compare directly the merits of different experimental methodologies with statistically significant numbers (such as sitting vs hanging drop, nanocrystallization vs conventional crystallization). Crystallization scientists can look forward to evidence-based assessment of the various parts of the crystallization process.

Since protein crystallization is accomplished primarily by liquid handling, automation is not a major challenge, particularly when one considers vapor diffusion and microbatch setups. In fact, there are already a number of companies selling dedicated protein crystallization robots, some of which are listed in Table 1.

Table 1
List of suppliers of protein crystallization robots

Company	Product	Description
Decode Biostructures ^a	ROBOHTC	One robot for preparing crystallization conditions (Matrix Maker) and another one for setting drops
Douglas Instruments	Oryx 6	Sitting-drop vapor diffusion trials; microbatch trials; 240 wells/h
Gilson	Cyberlab 240PC	96-well hanging drop setup

^a Formerly Emerald Biostructures, which was acquired first by Medichem and then by Decode.

A number of suppliers provide plates for crystallization. It is noteworthy that most plates have been designed for high-throughput drop setting but not to facilitate image capture. For example, sitting-drop plates, due to the extra layers of plastic, are particularly “unfriendly” to imaging systems. This issue is an important consideration when one employs automatic imaging for protein crystallization (see below).

Many laboratories employ custom crystallization robots derived from customized liquid handlers to entirely proprietary designs, with a few listed in Table 2. In the case of Syrrx, the Agincourt system was originally developed as a joint effort between Lawrence Berkeley National Lab and the Genomics Institute of Novartis Research Foundation to set-nanoliter drops (20–100 nl). By resorting to submicroliter volume, microcrystals can form more rapidly and with reduced use of protein in each trial. In addition, it is also possible to screen more conditions without demanding more protein. The practical limitation of this approach is the resulting crystal size-up to 50 μm .

Microbatch crystallization under oil is also being practiced at high throughput. The Hauptmann–Woodward group (Luft et al., 1999) has developed a system based on 1536-well plates and a customized 384-channel Hydra liquid-handling robot from Robbins Scientific. The robot sets the following experiment in each well: 200 nl of protein and 200 nl of screening solution independently dispensed into 5 μl of paraffin oil. In order to dispense consistently sized drops, this group has developed the technique of controlling the robot to dispense after immersing the needles in oil and then withdrawing the needles out of oil to separate droplets from the needles. Low-speed centrifugation ensures mixing of the protein and buffers and settlement of the mixed drops onto the bottoms of the wells.

At Affinium, in keeping with the flexible automation approach, the Gem crystallization system was developed using an industrial Cartesian robot from Adept Technology (<http://www.adept.com>) customized for versatile protein crystallization. Gem is equipped with 4 degrees of freedom (x , y , z , and yaw), a 750 \times 1000-cm deck, 14 independent pipetting channels, chilled stations, an automated plate sealer, and an integrated database (Crystal Management System or CMS). The Gem robot is quicker and more versatile than both the commercially available protein crystallization robots and most liquid handlers. Gem, which was designed to have a large platform that can hold many different stock solutions, automatically prepares the crystallization screens from the stock solutions, as well as carrying out preprogrammed crystal optimization screens.

4. The inspection problem

For a given protein, it may be necessary to set over 1000 drops before either a diffractable crystal is obtained or the experiment is abandoned. With sufficient investment in robotics such as systems mentioned above, preparing crystal trials is not a problem; however, every crystallization experiment has to be inspected to score the drops, to determine the optimization strategy, and to capture the time at which a crystal appears optimal. Most labs rely on human inspection, but there have been advances in imaging that promise to alleviate the challenge and tedium of inspecting thousands of protein drops every week.

Automated image capture offers various advantages over human inspection, namely ergonomics and throughput. With an image capture system, most pro-

Table 2
Examples of custom automated protein crystallization solutions

Group	Description of custom automated protein crystallization system
Syrrx	Agincourt system with custom-designed robot for setting up submicroliter drops Maximum crystal size 50 μm
Affinium Pharmaceuticals	Gem system—Customized AdeptCartesian robot with x , y , z , and yaw axes Robot integrated with database for semiautomated refinement
Hauptman–Woodward Medical Research Institute	Microbatch-under-oil setup in 1536-well plates Customized Hydra liquid handler (needle washing required)

tein drops can be examined on a computer with the microscope used only for a limited number of cases. Using automated methods, the drops can be inspected more frequently and each observation time-stamped precisely. Moreover, digitized images of protein drops can be analyzed using a myriad of artificial intelligence techniques which, though not yet competitive with human expertise, can be trained to recognize patterns not obvious to the human eye. Just as important is the fact that although all image processing systems have built-in bias, the biases are consistent from system to system. On the other hand, even trained crystallization scientists do not score experiments identically.

It is important to recognize that the data storage requirements of an imaging system are rather daunting for a high-throughput structural biology lab. Assuming 2 MB per image and 10 000 drops which require inspection at least once a day, 20 GB of data will be produced daily.

There are by now a number of off-the-shelf image capture solutions customized for protein crystallization. In addition, a small number of companies offer the crystal detection capability. In Table 3, some representative products are listed.

In addition to off-the-shelf products, structure-based drug discovery companies are also developing proprietary systems with the ability to automatically detect and score crystals.

Notwithstanding these systems and others offering automatic crystal detection and scoring, there is no system in existence that can completely obviate human inspection. Crystals are sufficiently difficult to grow, and hence precious, that until the development of a system guaranteeing 0% false negatives, the human eye will remain the most trusted crystal image system.

The increased throughput capacity of an imaging system over human inspection should be put toward capturing more data not just for scoring individual experiments, but also for contributing to understanding of protein crystallization overall. To this end, an ideal system should have the following features:

- Identification and elimination of clear drops with 0% failure (i.e., no risk whatsoever of losing a crystal);
- Capture of kinetic data—growth of crystal and/or precipitate as a function of time;
- Ability to distinguish crystals from precipitate and find crystals among precipitate in a high percentage of cases;
- Determination of crystal size and morphology;
- Ability to improve the two features above over time based on “experience.”

Finally, an ideal system should store images as persistent data rather than just the processed outcome. This will lend the data to future, ever-improving image processing algorithms and provide the opportunity to increase our understanding of crystallization at a higher rate.

5. Crystallization techniques and success rates

In order to consolidate the experiences of crystallographers, there have been efforts to archive the results of protein crystallization experiments. Gary Gilliland and colleagues have created the Biological Macromolecule Crystallization Database (BMCD), which comprises over 3500 crystals of over 2500 macromolecules, including mostly proteins and a small percentage of DNA (supported by the US National Institute of Standards and Technology—<http://www.bmcd.nist.gov:8080/bmcd/bmcd.html>). Included in this database, which has been compiled from literature sources and submissions from scientists (Gilliland et al., 1994), are crystallization techniques, conditions (pH, protein concentration, and temperature), and screening additives used for every successful crystallization experiment.

The BMCD archives the results from successful crystallization experiments and often ones that have not been repeated independently. These datasets are probably not as useful as those that report both successes and failures. Nevertheless, it is possible to analyze the current databases for trends in the experimental techniques.

Table 3
Examples of off-the-shelf protein crystal imaging solutions

Company	Product	Description
Decode BioStructures	Crystal Monitor workstation	Stereomicroscope Digital camera Voice control
Tritek	CrystalPro	Interface to database and ROBOHTC Microscope with motorized stage
Diversified Scientific	CrystalScore	Automatic image capture and storage Microscope with motorized stage
RoboDesign	RoboMicroscope II	Crystal counting and sizing Automatic drop locating
	CPXO	Automatic focusing Colored image capture and storage Classifies drops as clear, precipitate, crystal, or other

Table 4
Crystallization plates

Company	Type of plate
Decode BioStructures	24-well/96-drop Combiclover sitting-drop plates
Greiner	24-well hanging-drop plates 96-well/288-drop sitting-drop plates, square or round wells for protein drops 60-, 72-, 96-, 384-, and 1536-well microbatch plates
Corning	96-well sitting drop-plates
Hampton Research	48-well hanging-drop plates, greased and ungreased 24-well sitting-drop plates 24-well Q plates for hanging and sitting drops

The following data were extracted from the BMCD: 2454 of the 3704 crystals in the BMCD records were obtained by vapor diffusion (including sitting and hanging drops), 454 by batch or microbatch under oil, 174 by dialysis, 192 by seeding, and the rest by other methods. Even given that the BMCD does not represent a controlled dataset and also includes a small percentage of nonprotein data, these numbers support the common notion that most protein crystals have been grown by vapor diffusion or under oil.

Although the microbatch method has not been compared in large-scale (i.e., statistically significant scale) against the hanging-drop vapor diffusion method, Baldock et al. (1996) did compare the two methods by screening six well-behaved proteins against 48 solution conditions. In all, 84 crystals formed, with 15 deriving from vapor diffusion alone, 17 from microbatch alone, and 26 from both techniques. This study demonstrated that the methods are not entirely identical, but are equally effective. In a separate study, Rupp (2000) reported no significant difference in success rate found in conducting 96 crystal trials, each on three proteins using four different vapor diffusion setups: sitting drop in a plastic Cryschem plate, sitting drop on a siliconized glass coverslip shard, sitting drop on a siliconized glass pedestal, and hanging drop off a siliconized glass coverslip. Although neither experiment included a significant number of samples, and this is a significant deficiency, the results suggest that hanging drop, sitting drop, and microbatch methods will probably produce similar numbers of crystals but may not produce crystals for the same proteins under the same conditions (see Table 4).

6. Increasing the success rate of protein crystallization

6.1. Experimental approaches

Automated approaches will increase the throughput of protein crystal trials. However, it is widely believed that the success of crystal trials is largely dependent on various—as yet unidentified—properties of the protein. For example, there is a positive correlation between its degree of monodispersity in solution and its ability to

crystallize. There is also thought to be a negative correlation between the degree of disorder in the protein and its ability to crystallize (see Koth and Edwards, 2003). The effects on protein crystallization of these two parameters have been described recently. Here, we will review some recent studies that focus on protein-based strategies to increase crystallization success rates.

6.2. The use of orthologues

With the advent of genome sequencing, it is possible to identify close sequence and functional homologues (an orthologue) of a given protein target. One of the genome-based approaches to derive a crystal structure for one of the family of orthologues is to clone, express, purify, and attempt to crystallize a large number of the orthologous proteins. Recently, Savchenko and colleagues (Savchenko et al., 2003) attempted to purify and crystallize 68 different proteins from *Escherichia coli* (*EC*) and their orthologues from *Thermotoga maritima* (*TM*). Fifty-three (*EC*) and 50 (*TM*) proteins could be purified and concentrated to >2 mg/ml under standardized growth and purification procedures. Together, these proteins represented 62 of the 68 pairs. Of the 53 *EC* proteins, 14 (26%) formed crystals. For *TM* proteins, 16/50, or 32%, crystallized. Only 3 proteins crystallized from both sources. Clearly, the use of orthologues will increase the probability of purifying a representative protein and growing a protein crystal.

6.3. Refolding

A large fraction of recombinant proteins are found in the insoluble of bacterial lysates. One of the approaches to obtaining material from these proteins is to purify the protein under denaturing conditions and to renature the protein. The effectiveness of this procedure was addressed by Maxwell and colleagues (submitted for publication). Seventy-five insoluble recombinant proteins (<20 kDa) were purified in denatured form from *E. coli* cell lysates. Of these 75, 25 had previously been characterized after having been purified from the soluble fraction. A standard denaturation/renaturation protocol generated soluble, folded protein (as judged by NMR)

for 22/25 of the proteins known to be soluble when expressed in bacteria and 24/75 of the proteins that could not be produced in the soluble fraction of a bacterial lysate. A number of the previously insoluble proteins crystallized and led to 3D structures. The conclusion is that standard denaturation/renaturation protocols can increase the number of proteins that enter crystal trials and would be expected to produce an increased number of protein crystals.

6.4. Effect of affinity tags

Kimber and colleagues (2003) compared the success rates of protein crystallization trials for given proteins with and without the hexahistidine affinity tag. In all, 49 proteins were put into crystallization trial in both tagged and untagged forms: 7 yielded a crystal only with a tag, 9 without a tag, and 11 in both forms. This suggests that recombinant tags represent another variable in protein crystallization worthy of testing.

7. Knowledge-based approaches

7.1. Screening conditions

While crystallization techniques and automation are important starting considerations, the purpose of protein crystallization trials is to efficiently find useful lead conditions from which crystal size and morphology can be optimized to a diffractable form. The combination of conditions that are typically tested includes protein concentration, pH, and temperature, as well as precipitant type and concentration. The precipitants are usually inorganic or organic salts, small organic molecules such as MPD, alcohols, or long-chain polymers such as polyethylene glycol (PEG).

An analysis of the BMCD shows almost 400 precipitant types used to yield over 3500 crystals. Among these, the most popular constituents include sodium (used for 1082 crystals) and ammonium salts (575 crystals), PEG (611), and various forms of Tris (383).

Table 5 lists a number of commercially available screens, which have been developed based on the two main strategies—wide coverage of known crystallization

space vs sparse matrix experiments deliberately biased with previous experience. The second approach demands less protein and lower management overhead (drop inspection, data archival and analysis, etc.). Furthermore, it can increase in efficiency and effectiveness over time if the screens are continually improved based on crystallization results.

7.2. Data mining of structural proteomics datasets

Data mining approaches can be used to develop more efficient crystallization methods. Crystallographers will be able to use the information from structural proteomics data to initiate iterative cycles of adding further conditions: test and replace the poorest conditions while not interfering unduly with routine crystal production. This is in fact the underlying premise of the sparse matrix strategy (Jankarik and Kim, 1991).

One obvious strategy is to select further candidate conditions close to conditions that have previously proved productive, but in regions of parameter space that are undersampled. In this way, one can experiment with a wide variety of conditions, without increasing the amount of work necessary to unmanageable proportions and without sacrificing the proven productivity of a core set of conditions. Importantly, this strategy should encourage the exploration of conditions that might otherwise not be thought sufficiently “safe” to include in a fixed, general screen.

In an effort to explore the success rates of 48 commonly used conditions (Hampton Crystal Screen 1) in a controlled manner, Kimber et al. (2003) compiled a dataset of 755 different proteins from six bacterial organisms, all purified and crystallized under uniform conditions. Subsequent analysis revealed that of the 338 (45%) the test proteins that yielded crystals, 60% could be crystallized under the same 6 conditions, and 94% could be crystallized under only 24 conditions. In other words, many of the crystallization conditions are noticeably less effective than others.

There are large differences in the number of proteins crystallized per condition. Three hundred thirty-eight proteins were crystallized, 76 of these crystallized under condition 9 (30% PEG 4000 Na citrate, pH 5.6, 0.2 M NH₄ acetate). Although salt is the precipitant in only

Table 5
Commercial crystallization screens

Company	Screen
Decode BioStructures	Wizard I and II screens based on random sparse matrix design
Jena Bioscience	JB Screen 1–10 based on best known successful crystallization ingredients, including different forms of PEG, PPD, alcohol, ammonium sulfate, and other salts
	JB Screen Mixed—240 conditions based on JB Screen 1–10
Hampton Research	Crystal Screens I and II—based on sparse matrix design by Jankarik and Kim (1991)

30% of the cases, the use of salt enables the crystallization of 59.2% of the proteins. The different salts vary in their effectiveness; sodium citrate is represented by a single condition in the screen but is the eighth most productive condition. Citrate has strong metal-chelating abilities; perhaps this feature explains its excellent crystallization properties (McPherson, 1999). All four conditions that use ammonium sulfate as the major precipitant yield substantial numbers of crystals. Interestingly, citrate, sulfate, and formate salts form a cluster that affects protein solubility in similar fashions; they tend to crystallize the same subset of proteins. Phosphate, acetate, and tartrate are much less successful precipitants.

High-molecular-weight PEGs are widely considered the most successful protein crystallization agents (Gilliland et al., 1994; McPherson, 1999). Two-thirds (67.7%) of the proteins crystallized under the 30% of the conditions that consist of PEG 4000 or 8000 at concentrations greater than 18%. The six most productive conditions overall are all in this group. However, there is, in general, a great deal of redundancy among the PEG conditions (Kimber et al., 2003). This region of parameter space appears to be heavily oversampled, a consequence of strong bias toward previously successful experiments built into the original design of the screen.

Interestingly, PEG 400 appears to affect the solubility of proteins in a manner that resembles more the action of a salt than that of a high-molecular-weight PEG. This may reflect the fact that smaller PEGs probably precipitate proteins more by a solvent competition effect than a volume exclusion effect. Also, zinc in conjunction with PEG yielded solubility behavior quite different from that of other salt/PEG combinations, likely reflecting the ability of transition metals to effect strong interactions between proteins by virtue of their ligand chemistry.

7.3. *The concept of minimal screens*

The various conditions of the Hampton crystal screen imply that it is, in its present implementation, less than ideal. Several conditions produce very few crystals, while other groups of conditions, such as those containing high-MW PEGs, are highly correlated. Kimber et al. used this information to identify the set of conditions that would maximize crystals but minimize the number of conditions used. Kimber identified six conditions that crystallized 60% of the proteins successfully crystallized in the full screen and noted that the dispersion of these conditions-high-molecular-weight PEG at acidic, neutral, and basic pH; low-molecular-weight PEG; and two different salts—samples the major crystallization parameters. This analysis differs from those that simply quantify the productivity of each condition. Had the selection of a minimal set of conditions been

done by productivity alone, the “minimal screen” would have been less productive.

8. Conclusions

Process and automation advances in the past few years have resulted in very high throughput structural biology engines throughout the world. Many labs are now capable of cloning, expressing, purifying, and putting into crystal proteins at a previously unforeseen rate. While the steps feeding crystallization have increased in throughput as well as output, crystallization itself has yet to become a predictable process. This will gradually take place with accumulation and analysis of data. The few reports on data analysis of crystallization have yielded a better understanding of screening solutions, techniques, and other conditions, but no correlation of all parameters (for example, will certain screening conditions work more effectively in microbatch vs vapor diffusion?). To build on the early results, more analysis of larger controlled datasets is needed. However, the current databases may not be up to the task. Ultimately, the most useful and revealing databases will archive more than crystallization conditions and discrete experimental outcomes (crystal vs precipitate vs clear drop); they will include protein construct, production methods, and conditions as well as crystallization images.

References

- Albala, J.S., Franke, K., McConnell, I.R., Pak, K.L., Folta, P.A., Rubinfeld, B., Davies, A.H., Lennon, G.G., Clark, R., 2000. From genes to proteins: high-throughput expression and purification of the human proteome. *J. Cell Biochem.* 80, 187–191.
- Arrowsmith, C., Edwards, A., Hui, R., Marino, F., Savchenko, A., Yamazaki, K., Yee, A., 2003. Producing proteins (in press).
- Baldock, P., Mills, V., Stewart, P.S., 1996. A comparison of microbatch and vapor diffusion for initial screening of crystallization conditions. *J. Cryst. Growth* 168, 170–174.
- Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J., 2002. Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci. USA* 99, 2654–2659.
- Dry, S., McCarthy, S., Harris, T., 2000. Structural genomics in the biotechnology sector. *Nat. Struct. Biol.* 7, 946–949.
- Gilliland, G., Tung, M., Blakeslee, D., Ladner, J., 1994. Biological macromolecule crystallization database, version 3.0: new features, data and the NASA archive for protein crystal growth data. *Acta Crystallogr. D* 50, 408–413.
- Heinemann, U., 2002. Establishing a structural genomics platform: the Berlin-based protein structure factory. *Gene Funct. Dis.* 3, 25–32.
- Jankarik, J., Kim, S.H., 1991. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* 24, 409–411.
- Koth, C.M., Edwards, A.M., 2003. *Methods Enzymol.* (in press).
- Kimber, M.S., Vallee, F., Houston, S., Necakov, S., Vedadi, M., Skarina, T., Evdokimova, E., Beasley, S., Christendat, D., Savchenko, A., Arrowsmith, C.H., Gerstein, M., Edwards, A.M.,

2003. Data mining crystallization databases: Knowledge-based approaches to optimize protein crystal screens. *Proteins*, in press.
- Lesley, S.A., 2001. High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expr. Purif.* 22, 159–164.
- Luft, J.R., Rak, D.M., DeTitta, G.T., 1999. Microbatch macromolecular crystallization on a thermal gradient. *J. Cryst. Growth* 196, 447–449.
- McPherson, A., 1999. *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Rupp, B., 2000. High throughput protein crystallization. EMBL practical course on protein expression, purification and crystallization. Available from <http://www-structure.llnl.gov/Xray/tutorial/High_Throughput_EMBL_full.htm#_Toc490635209>.
- Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N., Das, R., Gerstein, M., Arrowsmith, C.H., Edwards, A.M., 2003. Strategies for structural proteomics of prokaryotes: quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins* 50, 392–399.
- Terwillinger, T., 2000. Structural genomics in North America. *Nat. Struct. Biol.* 7, 935–939.
- Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R., Kuramitsu, S., 2000. Structural genomics projects in Japan. *Nat. Struct. Biol.* 7, 943–945.
- Carter, C.W.J., Carter, C.W., 1979. Protein crystallisation using incomplete factorial experiments. *J. Biol. Chem.* 254, 12219–12223.
- Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J., Booth, V., Mackereth, C., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K., Wu, N., McIntosh, L., Gehring, K., Kennedy, M., Davidson, A., Pai, E., Gerstein, M., Edwards, A., Arrowsmith, C., 2000. Structural proteomics of an archaeon. *Nat. Struct. Biol.* 7, 903–909.
- Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y., McPherson, A., 1994. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr. D* 50, 414–423.
- Edwards, A., Arrowsmith, C., Christendat, D., Dharamsi, A., Friesen, J., Greenblatt, J., Vedadi, M., 2000. Protein production: feeding the crystallographers and NMR spectroscopists. *Nat. Struct. Biol.* 7, 970–972.
- Haynie, D.T., 2001. *Biological Thermodynamics*. Cambridge University Press, Cambridge.
- Juriscica, I., Rogers, P., Glasgow, J., Fortier, S., Luft, J., Wolfley, J., Bianca, M., Weeks, D., DeTitta, G., 2001. Intelligent decision support for protein crystal growth. *IBM Syst. J.* 40, 394–409.
- Kingston, R., Baker, H., Baker, E., 1994. Search designs for protein crystallization based on orthogonal arrays. *Acta Crystallogr. D* 50, 429–440.
- Norvell, J., Machalek, A., 2000. Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol.* 7, 931.
- Saridakis, E., Chayen, N., 2000. Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. *Protein Sci.* 9, 755–757.
- Stevens, R.C., 2000. High throughput protein crystallization. *Curr. Opin. Struct. Biol.* 10, 558–563.
- Vitkup, D., Melamud, E., Moulton, J., Sander, C., 2001. Completeness in structural genomics. *Nat. Struct. Biol.* 8, 559–566.

Further reading

- Adams, J.A., Jewell, D., Jorgensen, K., Mickley, M., Newman, J.M., 2002. Image capture and analysis for macromolecular crystal growth using robomicroscope II. *J. Assoc. Lab. Automat.* 7, 36–40.