Crystallography in the post-genomic era

REVIEW

Isbel Vivar, Nelia López, ✉ Alexis Musacchio

Department of Bioinformatics, Division of Physical Chemistry
Center for Genetic Engineering and Biotechnology, CIGB
Ave. 31 e/ 190 y 158, AP 6162, CP 10 600, Playa, Ciudad de La Habana, Cuba
E-mail: alexis.musacchio@cigb.edu.cu

ABSTRACT

X-ray crystallography is the most powerful technique to determine the three-dimensional structure of proteins. Several steps are required to solve the structure of proteins including cloning and expression of the gene, protein purification, crystallization, X-rays diffraction, data collection and structure determination. The vast amount of data generated by genomic studies has accelerated the development of new methodologies to increase the number of three-dimensional structures solved. This review focuses on recent advances in the field of protein crystallography, highlighting the high-throughput crystallization technologies.

Keywords: Crystallography, crystal screen, X-ray

Biotecnología Aplicada 2007;24:209-214

RESUMEN

La cristalografía en la era post-genómica. La cristalografía de rayos X es la técnica más poderosa para la determinación de la estructura tridimensional de proteínas. Se requieren varios pasos para resolver la estructura de las proteínas, incluyendo el clonaje y la expresión génica, purificación de la proteína, cristalización, difracción de rayos X, colección de los datos y determinación de la estructura. La gran mayoría de los datos generados por estudios genómicos han acelerado el desarrollo de nuevas metodologías para incrementar el número de estructuras tridimensionales resueltas. En esta revisión se abordan los avances recientes en el campo de la cristalografía de proteínas, resaltando las tecnologías de cristalización con alta capacidad de procesamiento.

Palabras clave: Cristalografía, pesquisa de cristales, rayos X

*I*ntroduction

The post-genomic era offers the potential to identify genetic disorders in a host and to design therapies to treat them. Genes encode proteins that can be targets for drugs. The function of these proteins is determined by their three-dimensional (3D) structure. Protein structural information can aid in the development of novel drugs, vaccines and diagnostics. Nowadays, the most powerful technique to determine protein structure is X-ray crystallography, which is totally dependent on the need for highly ordered crystals. Obtaining such crystals is the rate-limiting step to structure determination and the current pressure to produce crystals is greater than ever.

There are five broad areas in the field of macromolecular crystallization where significant advances have occurred: (1) Physical studies and characterization of the crystallization process; (2) Development of new practical approaches and procedures; (3) The use of genetic engineering to improve protein purification and crystallization; (4) The creation of new screening conditions based on information and databases emerging from structural genomics; and (5) Development and implementation of automation, robotics, and mass screening of crystallization conditions using very small amounts of protein. This review focuses on recent advances in high throughput protein crystallization technologies.

*T*he state of the art in protein crystallization

Obtaining atomic resolution 3D structures from protein crystals involves subsequent steps including cloning, protein expression, protein purification, quality assessment, crystallization, X-ray diffraction, data collection and structure determination.

The major obstacle to elucidate the 3D structure of a protein is its crystallization and can be divided into two steps: coarse screening to identify initial crystallization conditions, followed by optimization of these conditions to produce single diffraction-quality crystals. Nowadays, there are no systematic methods to ensure that ordered 3D crystals will be obtained [1, 2]. Table 1 shows that 24.8% of the cloned proteins have been purified, 9.1% crystallized and the crystal structure was solved for only 3.6% of them (http://targetdb.pdb.org/statistics/TargetStatistics.html). These data highlight that even when proteins can be cloned, expressed, solubilized, purified and crystallization trials do yield some crystals, this does not guarantee that the crystals will be good enough for

1. Manjasetty BA, Turnbull AP, Panjikar S, Bussow K and Chance MR. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. Proteomics (2008);8: 612-25.

2. Chayen NE. Optimization techniques for automation and high throughput. Methods Mol Biol (2007);363:175-90.

Table 1. Total number of proteins deposited by SG Centers in TargetDB

Status	Total number of proteins	(%) Relative to "cloned" proteins
Cloned	111306	100.0
Expressed	74670	67.1
Soluble	30922	27.8
Purified	27597	24.8
Crystallized	10143	9.1
Diffraction-quality Crystals	4780	4.6
Diffraction	5091	4.0
Crystal Structure	3954	3.6
In PDB	5413	4.9

✉ Corresponding author

structure determination [3]. A summary of the number of protein structures solved over the years is shown in figure 1 (http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=molType-protein&seqid=100). A big explosion achieved from the nineties can be observed.

Cloning, expression and purification of proteins for macromolecular crystallography

The application of several genetic approaches in protein purification processes have influenced in the growth of the obtained macromolecular crystals. A survey of crystallographic studies (excluding membrane proteins) indicates that over 90% of crystal structures are based on recombinant material [4, 5]. Bacteria, particularly *E. coli*, are preferred as protein producers for biophysical studies [6]. The use of recombinant proteins implies several advantages compared to purification from natural sources, such as: 1) the possibility to change the amino acid sequence; 2) the addition of tags or the fusion to other proteins for affinity purification; 3) the synthesis of isolated domains; 4) the easiness of labeling with Selenemethionine for phasing purposes and 5) the availability of higher amount of proteins.

Approximately 75% of the proteins used for crystallization studies are expressed as fusion constructs [7], using a number of small proteins and tags for affinity purification as well as for improving solubility. A wide range of systems for protein expression in different hosts is available. The fusion partners such as the hexaHis tag [8, 9], glutathione-*S*-transferase (GST) [10], and maltose-binding protein (MBP) [11] are particularly popular among structural biologists. Some of the less commonly used include thioredoxin [12], Z-domain from protein A [13], NusA [14] GB1-domain from protein G [15, 16]. None of the tags is universally superior, and often only parallel expression experiments can determine which the best strategy is. Families of vectors allowing for parallel assays of this kind have been developed [17-20].

Many proteins contain highly flexible or even completely unfolded fragments that dramatically interfere with crystallization. This is particularly true for large multidomain signaling proteins, in which the unstructured linker regions often account for more than 50% of the molecule. This can be accomplished in three ways: elimination of flexible N- and/or C-terminal polypeptides, removal of internal flexible W-loops, or extraction of individual domains from a multidomain protein. Truncations and deletions constitute the most frequently used protein engineering tool [21], with numerous examples of this approach available in the literature [4].

Protein heterogeneity and its propensity to aggregate are among the reasons why purified proteins do not crystallize or crystallize poorly. Even in cases where the protein sequence is homogeneous, mixed populations of proteins may coexist due to their inherent flexibility, different folding states and stability. Since proteins are not rigid bodies, a protein molecule can display different shapes and, in consequence, assemble poorly in a crystalline lattice. The heterogeneity in the amino acid sequence renders molecules

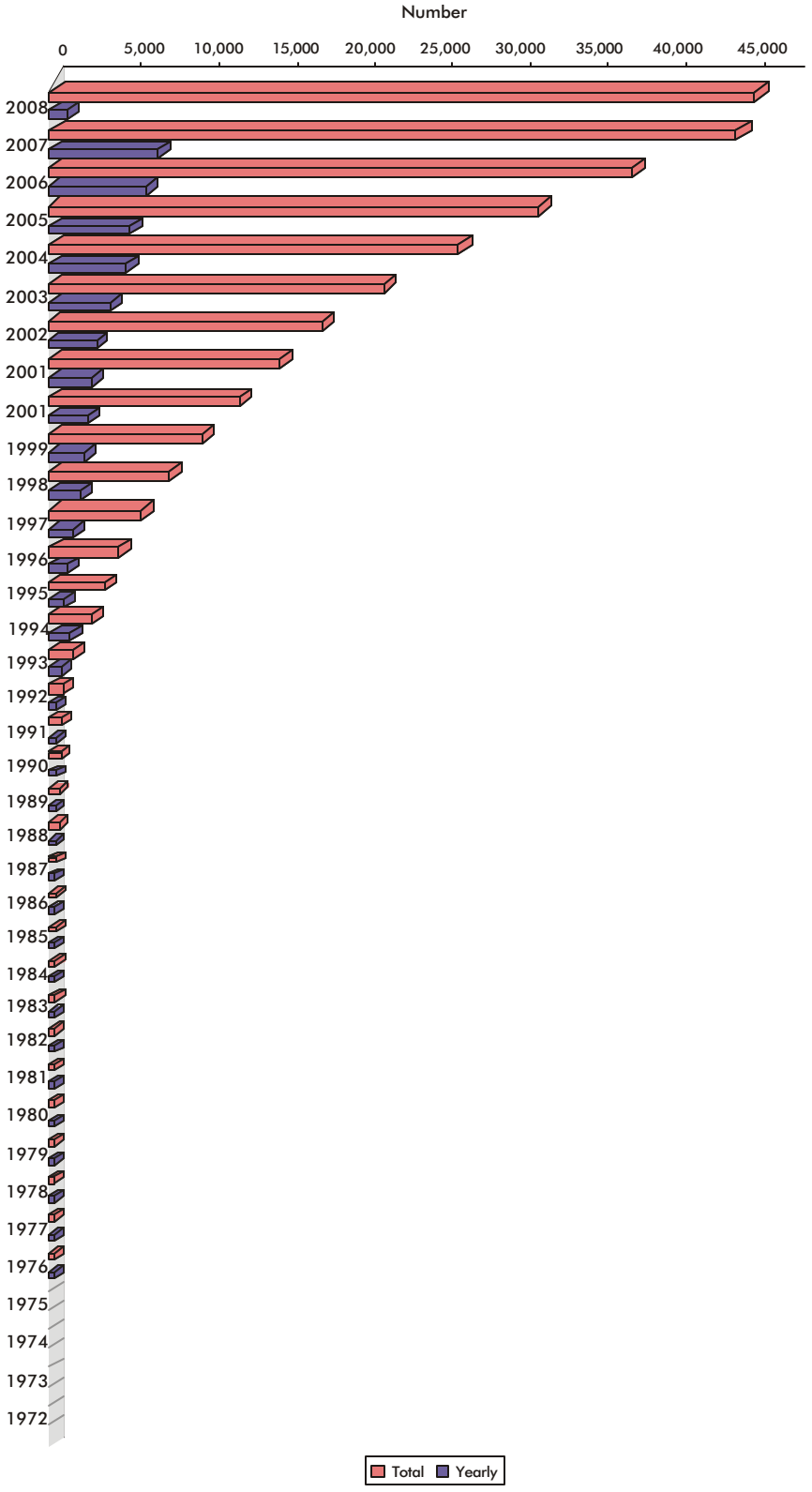

Figure 1. Number of searchable protein structures per year.

of different size and shape. So, the protein of interest is contaminated with variants of itself or with other

types of proteins or foreign materials that may hinder crystal nucleation or growth [22].

Once a biologically active recombinant protein is eluted from the last chromatography step, it can suffer from denaturation, proteolysis, oxidation or other decaying processes, leading to lose of biological activity or structure. A protein molecule changes constantly while samples are waiting for crystallization setups or kept in storage. Therefore, no single protein crystallization trial will be exactly alike. Common treatments against conformational and chemical heterogeneity consist of mixing inhibitors and adding reducing agents to protein samples, changing expression systems and altering pH, temperatures or salt concentrations [23]. However, these actions would require an individual case consideration and, in consequence, are impractical in high-throughput formats [22].

Poor solubility of the protein is another common problem in crystallization experiments. Even if the protein is expressed predominantly in the soluble fraction, particularly if it is expressed as a fusion protein, it may precipitate at concentrations required for crystallization after its cleavage from the carrier protein. Solubility is a function of surface hydrophobicity [24, 25] and can be advantageously altered by mutational modification of selected surface residues. Provided that the protein is stable and properly folded, an alternative and rarely used approach is to introduce a solubilizing motif [26-28]. Preparing a protein for crystallization experiments necessarily includes a previous step of high-speed centrifugation or filtration to remove and minimize insoluble protein species that may lead to heterogeneity and aggregation.

Protein expression, solubility and purification procedures may be completely automated. Several biotechnology companies have developed integrated solutions for large scale recombinant protein purification. Syrrx (San Diego, CA, USA) employs a system that purifies sufficient protein for crystallization, combining robotic centrifugation and sonication with a parallel column chromatography system capable of purifying 96-192 proteins per day [29]. In this system, other steps, including desalting and concentration, are carried out offline before the samples are ready for crystallization. Affinium Pharmaceuticals (Toronto, Canada) developed the ProteoMax system to process cell extracts, purify and concentrate protein samples suitable for analysis. This system clarifies the cell lysate, performs the column chromatography, desalts and concentrates the purified material. In optimal cases, the protein is ready for structural studies after these steps. By obviating centrifugation, and integrating all steps into an automated system, ProteoMax completely performs high-throughput purification [30].

Crystallization

Crystallization techniques

The automation of cloning, expression and purification processes has had a great impact in obtaining a huge amount of pure proteins for several purposes. To keep up with this demand, it is critical to employ automated crystallization, crystal visualization and optimization techniques [30].

Current commercial screening kits and computer algorithms for designing arrays of potential conditions are readily accessible, therefore it is no longer a problem to conduct trials automatically [31, 32]. The automatic generation of high-throughput screening crystallization trials is under way [32-34] as well as automated follow-up and analysis of the results.

The first semi-high-throughput experiments for both screening and optimization were reported in 1990 as microbatch trials under oil [35]. Many proteins have been successfully crystallized using the microbatch method [36-39].

A recent tendency is to perform the initial crystallization trials in a myriad of conditions with small amounts of biological material, using droplets of very small volume, of the order of tens of nanolitres. Preparation of thousands of small droplets of varying, but strictly prescribed composition can be very effectively performed by robots. Indeed, crystallization was one of the first steps of X-ray structure analysis where automation was successfully introduced [40] and currently many academic and industrial laboratories are equipped with various crystallization robots.

Conventional methods used in protein crystallization include batch [41-46], liquid-liquid (free interface) diffusion [47-49], vapor diffusion (using either sitting or hanging drops) [50, 51], capillaries [52, 53] and gels [54]. Currently, the two methods chosen for high-throughput crystallization development are microbatch [35, 55] and sitting or hanging drop vapor diffusion. Both techniques are capable of working with very small volumes of protein solution (nL) and exhibit comparable performances.

For microbatch, the mixed droplet is placed in a well or depression and covered with oil to prevent or reduce evaporation. The gas permeability of the oil layer can be adjusted to allow for the slow evaporation of the drop resulting in a similar concentration effect. In some microbatch approaches, there can be difficulties in getting the reagents through the oil and in forcing them to coalesce in a single droplet. All automated crystallization approaches require automated liquid dispensing methods, capable of accurately dispensing very small volumes of liquids (mL) with a wide variety of surface properties and viscosity [56].

For hanging drop, a mixed droplet is placed on a coverslide which is inverted, and sealed above a well containing the undiluted crystallization solution. For sitting drop, the droplet is placed directly on a shelf adjacent to a well containing the crystallization solution and the entire tray is sealed to prevent evaporation.

In both methods, the mixed droplet is allowed to equilibrate with the well solution through vapor diffusion. This results in a gradual concentration of the ingredients which will hopefully lead to supersaturation and crystallization.

Each of these methods has strengths and weaknesses irrespective of their amenability to automation. The hanging drop has been the most commonly used crystallization method in recent years in the traditional, non-automated setting [56]. A benefit of hanging drop is the individual access to each crystallization experiment in a given tray. Since each well is individually

3. Chayen NE. Tackling the bottleneck of protein crystallization in the post-genomic era. Trends Biotechnol (2002);20:98.

4. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. Methods (2004); 34:354-63.

5. Lundstrom K. Structural genomics and drug discovery. J Cell Mol Med (2007); 11: 224-38

6. Goulding CW, Perry LJ. Protein production in *Escherichia coli* for structural studies by X-ray crystallography. J Struct Biol (2003); 142:133-43.

7. Uhlen M, Forsberg G, Moks T, Hartmanis M, Nilsson B. Fusion proteins in biotechnology. Curr Opin Biotechnol (1992);3: 363-9.

8. Murphy MB, Doyle SA. High-throughput purification of hexahistidine-tagged proteins expressed in *E. coli*. Methods Mol Biol (2005); 310:123-30.

9. Smith MC, Furman TC, Ingolia TD, Pidgeon C. Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. J Biol Chem (1988); 263:7211-5.

10. Smith DB, Johnson KS. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene (1988); 67:31-40.

11. di GC, Li P, Riggs PD, Inouye H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene (1988); 67:21-30.

12. LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Biotechnology (NY) (1993); 11:187-93.

13. Nilsson B, Moks T, Jansson B, Abrahmsen L, Elmblad A, Holmgren E *et al*. A synthetic IgG-binding domain based on staphylococcal protein A. Protein Eng (1987); 1:107-13.

14. Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol Bioeng (1999); 65:382-8.

15. Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Sci (1997); 6:2359-64.

16. Balbas P. Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. Mol Biotechnol (2001); 19:251-67.

17. Betton JM. High throughput cloning and expression strategies for protein production. Biochimie (2004); 86:601-5.

18. Chambers SP. High-throughput protein expression for the post-genomic era. Drug Discov Today (2002); 7:759-65.

19. Ding HT, Ren H, Chen Q, Fang G, Li LF, Li R *et al*. Parallel cloning, expression, purification and crystallization of human proteins for structural genomics. Acta Crystallogr D Biol Crystallogr (2002); 58:2102-8.

sealed, crystals can be harvested from one well without disturbing others. In addition, it is easier to harvest crystals from a hanging drop, since they are likely to grow at the bottom of the drop where they do not interact with the surface.

On the other hand, sitting drop advantageously maintains the vapor-diffusion nature of a hanging drop experiment while greatly simplifying the setup procedure. Even without robotics, many sitting drop experiments can be set up simultaneously using multichannel hand pipettors. However, recovery of crystals from sitting drop trays can be more difficult. It has also been difficult to design sitting drop trays with optical properties that are optimized for automated imaging, although some improvements have been made in this sense.

Crystallization robots

The first automated crystallization robot was designed by Cox and Weber in 1987 [57]. Regardless of the crystallization methodology, the primary task of a crystallization robot is to create mixed droplets of protein and crystallization solutions at the desired volume and place them in a crystallization tray. Each method presents unique difficulties in effectively carrying out this seemingly simple task. For hanging drop, the process of presenting coverslides to the pipetting mechanism and then sealing them onto the wells requires special attention. For sitting drop, the difficulties often involve ensuring that the droplet remains in the middle of its shelf rather than being drawn to the walls by static electricity making subsequent observation very difficult. Since all of the wells are sealed simultaneously once the tray is set up, sitting drops must be dispensed quickly or in a humidified environment so that the drops do not evaporate before they are sealed.

Although the automated production of macromolecular crystals is not necessarily the only issue of concern, it could integrate solutions to other problems like protein availability. Rupp *et al* [58] pointed out that protein availability is the true limitation given the current status of automation, and we therefore must, at the very least, focus on the use of the smallest possible quantities of protein to obtain crystals. This restriction on protein quantity does not only pertains to the volume of crystallization drops but also to the equipment and processes used to set them up either through automation or by hand [22].

Over the last several years, a number of specialized crystallization robots have appeared on the market. Many of them use innovative approaches to address these difficulties and some have achieved admirable success [34, 58-70]. Nanovolume crystallization services are now commercially available from Syrrx, who offer the setup of 1000 crystallization trials using only 100 ml of solution. The dominant characteristics of these systems are that they are generally based upon use of higher-density 96-, 384-, or even 1536-well plates and that they use 1 mL or less of protein solution per experiment. Several systems have been described that use only 100 nL or less of macromolecule solution per crystallization [32, 61, 62, 71]. Recently, a low-cost manual approach to rapidly setting up drops as small as 25 nL has been described which is suitable for laboratory use [72]. However, the high prices of automated systems difficult their massive application.

The reduction of volumes to nanoliter scale has significantly reduced the amount of protein required and, together with the high throughput crystallization in 96 well or higher microplate format, it has permitted the simultaneous analysis of numerous crystallization parameters and conditions [73]. In this context, variables such as pH, ionic strength, temperature and concentration of salts and detergents can be screened and up to 100, 000 crystallization trials conducted per day [5].

Crystals grown from sub-microliter-sized drops are suitable for X-ray diffraction studies. There is evidence that smaller crystals may actually diffract to higher resolution [72, 74]. Empirical observations suggest that crystals nucleate and grow faster and precipitate less, when the reactions are carried our in smaller volumes [75]. Additionally, more rapid vapor equilibration kinetics has been theoretically shown with smaller drop size [67, 76].

A revolutionary approach to crystallization screening has come from the use of microfluidics technology [77]. Hansen *et al*. [75, 78] have described a microfluidics-based liquid-liquid diffusion apparatus. There are 144 parallel reactions executed in each chip. The volume of the crystallization chambers is 25 nL and the volumetric ratios of protein and precipitant solutions are pre-set by design with three ratios of 1:4, 1:1 or 4:1 for each protein-precipitant pair. The device is claimed to have picoliter accuracy and can be linearly scaled up to larger numbers of fixed conditions. This technology has now been commercialized by Fluidigm Corporation with a device that can screen 96 different crystallization conditions using only a few nanograms of material [79].

Another recently example, the automation and miniaturization of high-throughput protein crystallization is SPINE (Structural Proteomics, Europe) [80], a project created to drive the development and hardware uptake for robotic handling of nanolitre quantities of protein, to trial new strategies for crystallization and to automate the processes involved in crystal imaging and recognition [81].

Crystal optimization

Crystal optimization aims to turn poor quality crystals into diffraction-quality crystals that can be used for structure determination. There are a variety of methods that can be used to improve crystal quality including crystal seeding [82], which has been shown to be very effective to grow diffraction-quality crystals of proteins where little or no nucleation is normally observed. Microseed matrix seeding is an extension of conventional seeding techniques in which microseeds from the nucleation step are transferred into new conditions where all drop components vary to screen the growth of well ordered crystals. This technique has been successfully applied to improve the diffraction quality of crystals of the yeast cytosine deaminase [83]. Recently, this method has been automated so that crystals grown from one set of conditions can be seeded into a secondary screen of 96 solutions by using a crystallization robot [84].

20. Sheffield P, Garrard S and Derewenda Z. Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. Protein Expr Purif (1999); 15:34-9.

21. Porter RR. Structural studies of immunoglobulins. Science (1973); 180:713-6.

22. Pusey ML, Liu ZJ, Tempel W, Praissman J, Lin D, Wang BC *et al*. Life in the fast lane for protein crystallization and X-ray crystallography. Prog Biophys Mol Biol (2005); 88:359-86.

23. Lorber B, Giegé R. Preparation and handling of biological macromolecules for crystallization. 1992;1st:19-45.

24. Wagner JR, Sorgentini DA, Anon MC. Relation between Solubility and Surface Hydrophobicity as an Indicator of Modifications during Preparation Processes of Commercial and Laboratory-Prepared Soy Protein Isolates. J Agric Food Chem (2000); 48:3159-65.

25. Schein CH. Solubility as a function of protein structure and solvent components. Biotechnology (1990); 8:308-17.

26. Bianchi E, Venturini S, Pessi A, Tramontano A, Sollazzo M. High level expression and rational mutagenesis of a designed protein, the minibody. From an insoluble to a soluble molecule. J Mol Biol (1994); 236:649-59.

27. Brewer SJ, Haymore BL, Hopp TP, Sassenfeld HM. Engineering proteins to enable their isolation in a biologically active form. Bioprocess Technol (1991); 12:239-66.

28. Smith JC, Derbyshire RB, Cook E, Dunthorne L, Viney J, Brewer SJ *et al*. Chemical synthesis and cloning of a poly (arginine)-coding gene fragment designed to aid polypeptide purification. Gene (1984); 32:321-7.

29. Lesley SA. High-throughput proteomics: protein expression and purification in the postgenomic world. Protein Expr Purif (2001); 22:159-64.

30. Hui R, Edwards A. High-throughput protein crystallization. J Struct Biol (2003); 142:154-61.

31. Chayen NE, Shaw Stewart PD, Baldock P. New developments of the IMPAX small-volume automated crystallization system. Acta Crystallogr D Biol Crystallogr (1994); 50:456-8.

32. Stevens RC. High-throughput protein crystallization. Curr Opin Struct Biol (2000); 10:558-63.

33. Abola E, Kuhn P, Earnest T, Stevens RC. Automation of X-ray crystallography. Nat Struct Biol (2000); 7 Suppl:973-7.

34. Mueller U, Nyarsik L, Horn M, Rauth H, Przewieslik T, Saenger W *et al*. Development of a technology for automation and miniaturization of protein crystallization. J Biotechnol (2001);85:7-14.

35. Chayen NE, Shaw Stewart PD, Maeder DL, Blow DM. An automated system for micro-batch protein crystallization and screening. Journal of Applied Crystallography (1990); 23:297-302.

36. Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J *et al*. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. Cell (1998); 92:117-29.

Images

Crystallization robots are often linked to devices that automatically inspect all crystallization setups through an optical microscope and are able to classify the individual drops for the presence of various features, such as amorphous precipitate, microcrystals, sizable crystals etc. Several efforts are currently in progress towards the reliable automatic identification of crystals in the drops. The final judgment of the preclassified crystals is ultimately performed by the human eye because it is currently not possible to automate the process of fishing out the selected crystals from the drop. This may change in the not too distant future. One proposed way of circumventing this problem is based on growing crystals in special 'matrixes' and inserting all setups with drops into the X-ray beam for diffraction experiments [85-87].

Automated image capture offers various advantages over human inspection, namely ergonomics and throughput. With an image capture system, most protein drops can be examined on a computer, but it is 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 [30].

Pusey *et al.* recently reviewed several automated crystallization plate reading systems that have become commercially available [22]. These systems are focused on rapid acquisition and storage methods for subsequent human analysis. Rapidity of image acquisition, color, illumination and the use of bright- or dark-field polarization and image resolution are major selling points along with the ability of carrying out the scoring from ones desktop computer, not over a microscope. The scoring typically goes directly into a database, which may be subsequently analyzed for future optimization or other crystallization experiments [22].

Diversified Scientific provides image analysis software with their instrumentation [88]. The software then resolves whether a crystal is present or not. If a crystal is found, it is further scored for quality on the basis of straightness of edges, defects, fractures, etc. However, intermediate outcome scoring, for example phase changes, precipitates and spherulites, is apparently not carried out.

The Hough transform has been used as the basis for a crystallization drop analysis system [89]. This approach starts by defining the drop edges to set the boundaries for subsequent analytical operations. The primary attributes are whether the edges are straight or irregular, and there are five levels of classification: crystals, aggregated crystals, interesting objects, skins or junk. The software is able to correctly identify 86% of human-identified crystals and 77% of unfavorable objects.

Both visual and automated analysis methods become more difficult as the size of the crystallization drop is decreased. Higher magnifications mean a lower depth of field and greater difficulty in resolving features. Cumbaa *et al.* [90] described a technique for automated evaluation of 400 nL batch crystallization drops set up under oil in 1536-well plates. The technique was found to be 85% accurate when compared to human-scored results. Features that generally gave false positives included skinned over drops and speckled precipitate. Microcrystals, as well as dendrites, also gave false negative results. Other methods have been developed to discriminate between crystals and other microparticles. Asanov *et al.* [91] found that intrinsic protein fluorescence was a potentially rapid and efficient method of screening existing protein crystals for their likely X-ray diffraction quality. Hampton Research commercially markets the dye methylene blue under the trade name Izit for distinguishing between protein and salt crystals. The dye can diffuse through the solvent channels of protein crystals, not present in those of salt or other small molecules, and bind to the protein, giving a blue colored crystal and a clear solution (www.hamptonresearch.com/products).

Conclusion

X-ray crystallography is one of the most common techniques for protein structure analysis. The advance in technologies for protein crystallization and the implementation of automated tools have clearly reduced the amount of time required to setup a series of crystallization experiments. The automation of nearly every stage of the crystallization process leads this methodology to be applied in structural genomics projects, increasing technical capabilities for academic and industrial scientists. However, high throughput protein expression and crystallization still represent the main challenges to the process of automated structure determination, which requires novel experimental techniques and computational tools.

37. Chayen NE. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. Acta Crystallogr D Biol Crystallogr (1998); 54:8-15.

38. Stock D, Leslie AG, Walker JE. Molecular architecture of the rotary motor in ATP synthase. Science (1999; 286:1700-5.

39. Zhang X, Roe SM, Pearl LH, Danpure CJ. Crystallization and preliminary crystallographic analysis of human alanine:glyoxylate aminotransferase and its polymorphic variants. Acta Crystallogr D Biol Crystallogr (2001); 57:1936-7.

40. Weselak M, Patch MG, Selby TL, Knebel G, Stevens RC. Robotics for automated crystal formation and analysis. Methods Enzymol (2003); 368:45-76.

41. Alderton G, Lewis JC, Fevold HL. The relationship of lysozyme, biotin and avidin. Science (1945); 101:151-2.

42. Carbone MN, Etzel MR. Seeded isothermal batch crystallization of lysozyme. Biotechnol Bioeng (2006); 93:1221-4.

43. Chernov AA. Protein crystals and their growth. J Struct Biol (2003); 142:3-21.

44. Northrop JH. Crystalline pepsin. Science (1929); 69:580.

45. Rayment I. Small-scale batch crystallization of proteins revisited: an underutilized way to grow large protein crystals. Structure (2002); 10:147-51.

46. Sumner JB, Dounce AL. Crystalline catalase. Science (1937); 85:366-7.

47. Fu S, Kang H, Liu S, Zhou J, Jiang P. A new protein crystallization system with large number of cells. Conf Proc IEEE Eng Med Biol Soc (2005); 5:4500.

48. Sygusch J, Coulombe R, Cassanto JM, Sportiello MG, Todd P. Protein crystallization in low gravity by step gradient diffusion method. J Cryst Growth (1996); 162:167-72.

49. Cudney R, Patel S, McPherson A. Crystallization of macromolecules in silica gels. Acta Crystallogr D Biol Crystallogr (1994); 50: 479-83.

50. Davies DR, Segal DM. Protein crystallization. Micro techniques involving vapor diffusion. Methods Enzymol (1971); 22:266-9.

51. Saridakis E and Chayen NE. Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. Protein Sci (2000); 9:755-7.

52. Garcia-Ruiz JM. Counterdiffusion methods for macromolecular crystallization. Methods Enzymol (2003); 368:130-54.

53. Ng JD, Gavira JA, Garcia-Ruiz JM. Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. J Struct Biol (2003);142:218-31.

54. Garcia-Ruiz JM, Morena A. Investigations on protein crystal growth by the gel acupuncture method. Acta Crystallogr D Biol Crystallogr (1994); 50:484-90.

55. Chayen NE. The role of oil in macromolecular crystallization. Structure (1997); 5: 1269-74.

56. Bard J, Ercolani K, Svenson K, Olland A , Somers W. Automated systems for protein crystallization. Methods (2004); 34:329-47.

57. Cox MJ, Weber PC. Experiments with automated protein crystallization. Journal of Applied Crystallography (1987); 20:366-73.

58. Rupp B, Segelke BW, Krupka HI, Lekin T, Schafer J, Zemla A *et al*. The TB structural genomics consortium crystallization facility: towards automation from protein to electron density. Acta Crystallogr D Biol Crystallogr (2002); 58:1514-8.

59. Brown J, Walter TS, Carter L, Abrescia NGA, Aricescu AR, Aricescu AR *et al*. A procedure for setting up high-throughput nanolitre crystallization experiments. II. Crystallization results. J Appl Crystallogr (2003); 36:315-8.

60. DeLucas LJ, Hamrick D, Cosenza L, Nagy L, McCombs D, Bray T *et al*. Protein crystallization: virtual screening and optimization. Prog Biophys Mol Biol (2005); 88:285-309.

61. Hosfield D, Palan J, Hilgers M, Scheibe D, McRee DE, Stevens RC. A fully integrated protein crystallization platform for small-molecule drug discovery. J Struct Biol (2003); 142:207-17.

62. Krupka HI, Rupp B, Segelke BW, Lekin TP, Wright D, Wu HC *et al*. The high-speed Hydra-Plus-One system for automated high-throughput protein crystallography. Acta Crystallogr D Biol Crystallogr (2002); 58: 1523-6.

63. Luft JR, Wolfley J, Collins R, Bianca M, Weeks D, Jurisica I *et al*. Gearing up for structural genomics: the challenge of hundreds of proteins and hundreds of thousands of crystallization experiments per year. Acta Crystallogr; 2000.

64. Luft JR, Wolfley J, Jurisica I, Glasgow J, Fortier S, DeTitta GT. Macromolecular crystallization in a high throughput laboratory-the search phase. J Cryst Growth (2001); 232:591-5.

65. Oldfield TJ, Ceska TA, Brady RL. A flexible approach to automated protein crytallization. J Appl Crystallogr (1991); 24:255-60.

66. Sadaoui N, Janin J, Lewit-Bentley A. TAOS: an automatic system for protein crystallization. J Appl Crystallogr (1994); 27:622-6.

67. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D *et al*. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr (2002); 35:278-81.

68. Soriano TMB and Fontecilla-Camps JC. ASTEC: an automated system for sitting drop crystallization. J Appl Crystallogr (1993); 26: 558-62.

69. Sulzenbacher G, Gruez A, Roig-Zamboni V, Spinelli S, Valencia C, Pagot F *et al*. A medium-throughput crystallization approach. Acta Crystallogr D Biol Crystallogr (2002); 58:2109-15.

70. Walter TS, Diprose J, Brown J, Pickford M, Owens RJ, Stuart DI *et al*. A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. J Appl Crystallogr (2003); 36:308-14.

71. DeLucas LJ, Bray TL, Nagy L, McCombs D, Chernov N, Hamrick D *et al*. Efficient protein crystallization. J Struct Biol (2003); 142: 188-206.

72. Yeh JI. A manual nanoscale method for protein crystallization. Acta Crystallogr D Biol Crystallogr (2003); 59:1408-13.

73. Tickle I, Sharff A, Vinkovic M, Yon J, Jhoti H. High-throughput protein crystallography and drug discovery. Chem Soc Rev (2004); 33:558-65.

74. Bodenstaff ER, Hoedemaeker FJ, Kuil ME, de Vrind HP, Abrahams JP. The prospects of protein nanocrystallography. Acta Crystallogr D Biol Crystallogr (2002); 58:1901-6.

75. Hansen CL, Skordalakes E, Berger JM, Quake SR. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. Proc Natl Acad Sci USA (2002); 99:16531-6.

76. Diller DJ, Hol WGJ. An accurate numerical model for calculating the equilibration rate of a hanging-drop experiment. Acta Crystallogr (1999);D 55:656-63.

77. van der WM, Ferree D, Pusey M. The promise of macromolecular crystallization in microfluidic chips. J Struct Biol (2003); 142:180-7.

78. Hansen C, Quake SR. Microfluidics in structural biology: smaller, faster em leader better. Curr Opin Struct Biol (2003); 13: 538-44.

79. Self K, Hansen CL, Berger JM, Quake SR, Ng S, Godley S *et al*. A microfluidic method for protein crystallization. ACA meeting abstracts (2003); P115:132.

80. Bahar M, Ballard C, Cohen SX, Cowtan KD, Dodson EJ, Emsley P *et al*. SPINE workshop on automated X-ray analysis: a progress report. Acta Crystallogr D Biol Crystallogr (2006); 62:1170-83.

81. Berry IM, Dym O, Esnouf RM, Harlos K, Meged R, Perrakis A *et al*. SPINE high-throughput crystallization, crystal imaging and recognition techniques: current state, performance analysis, new technologies and future aspects. Acta Crystallogr D Biol Crystallogr (2006); 62:1137-49.

82. D'Arcy A, Mac SA, Haber A. Using natural seeding material to generate nucleation in protein crystallization experiments. Acta Crystallogr D Biol Crystallogr (2003); 59:1343-6.

83. Ireton GC, Stoddard BL. Microseed matrix screening to improve crystals of yeast cytosine deaminase. Acta Crystallogr D Biol Crystallogr (2004); 60:601-5.

84. D'Arcy A, Villard F, Marsh M. An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr D Biol Crystallogr (2007); 63:550-4.

85. Watanabe N, Murai H, Tanaka I. Semi-automatic protein crystallization system that allows in situ observation of X-ray diffraction from crystals in the drop. Acta Crystallogr D Biol Crystallogr (2002); 58:1527-30.

86. Jacquamet L, Ohana J, Joly J, Borel F, Pirocchi M, Charrault P *et al*. Automated analysis of vapor diffusion crystallization drops with an X-ray beam. Structure (2004); 12: 1219-25.

87. Dauter Z. Current state and prospects of macromolecular crystallography. Acta Crystallogr D Biol Crystallogr (2006); 62:1-11.

88. Gester TE, Rosenblum WM, Christopher GK, Hamrick DT, DeLucas LJ, Tillotson B. Method for acquiring, storing, and analyzing crystal images. US Patent US 6,529, 612 B1; 2003.

89. Wilson WW. Monitoring crystallization experiments using dynamic light scattering: assaying and monitoring protein crystallization in solution. Methods (1990); 1:110-7.

90. Cumbaa CA, Lauricella A, Fehrman N, Veatch C, Collins R, Luft J *et al*. Automatic classification of sub-microlitre protein-crystallization trials in 1536-well plates. Acta Crystallogr (2003); D 59:1619.

91. Asanov AN, McDonald HM, Oldham PB, Jedrzejas MJ, Wilson WW. Intrinsic fluorescence as a potential rapid scoring tool for protein crystals. J Cryst Growth (2001); 232: 603-9.

Received in September, 2007. Accepted for publication in December, 2007.