High-throughput x-ray crystallography for structure-based drug design

Kenneth E. Goodwill, Michael G. Tennant and Raymond C. Stevens

Knowledge of the 3D structure of proteins can play a key role in both understanding the biochemical function of protein targets, and developing small-molecule drugs that interact with these targets. This review will discuss recent advances in automation and miniaturization, which are making the determination of protein structures faster, more reliable, and more economical than has been possible historically.

Kenneth E. Goodwill* and Michael G. Tennant
Syrrx, 10450 Science Center Drive
San Diego, CA 92121
USA
*tel: +1 858 202 4161
fax: +1 858 678 3984
e-mail: ken.goodwill@syrrx.com

Raymond C. Stevens
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla
CA 92037
USA

▼ With the availability of many fully sequenced genomes, we are now able to focus on the actual targets of most drugs – the proteins. The past few decades have shown that our best understanding of proteins comes from an examination of their 3D structures[1]. Progress in structural biology over the past 20 years has been facilitated by technical advances ranging from molecular biology to synchrotron x-ray sources. Public and private initiatives now underway are promoting the development of large-scale structural biology around the world. As the number of available protein structures increases, this information will play a greater role in guiding drug discovery and understanding drug metabolism. Over time, knowledge of protein structure will help to enable the conversion of genomic information into future medical advances. This review will describe the automation now being applied to protein crystallography and the potential impact of miniaturization on the science of structural biology.

Knowledge of the structure of a protein can serve many purposes once it is available. For proteins of unknown function, the 3D structure can provide insight into the biological role and help guide the development of biochemical assays. Often, the active site can be identified by a combination of sequence patterns and characteristic surface properties[2]. As indicated in Fig. 1, the protein structure can also be used for virtual ligand screening (VLS), where millions of compounds are computationally fitted to the active site surface[3]. This can be combined with HTS, where hundreds of thousands of compounds per week are assayed for activity in small-scale experiments[4]. In practice, the quicker and less expensive method of VLS can be used to filter the compounds used for HTS. As our knowledge of chemical diversity improves, VLS offers the advantage of rapidly screening a very large compound space, generally far larger than a company has in its compound library.

To determine the exact mode of binding, and to find ways to improve lead compounds, co-crystal structures can be determined with the compounds bound. In fact, the protein crystals can be used directly to screen pools of potential ligands, and molecules that have bound to the protein can be visualized in subsequent x-ray structure determinations[5]. The ability to actually see how lead compounds bind to their targets makes drug development faster and more cost efficient[6].

The structure determination process

The path from genome to 3D protein structure is outlined in Fig. 1. The additional steps of compound screening and co-crystallization with lead compounds are shown to illustrate how a complete cycle of lead optimization can be enabled by crystallography. Two crucial elements for the success of protein crystallography projects, both in the past and in the coming era of automation, are diversity and attrition. To obtain the structure of a given protein target, hundreds of versions of

1359-6446/01/$ – see front matter PII: S1471-1931(01)00169-0

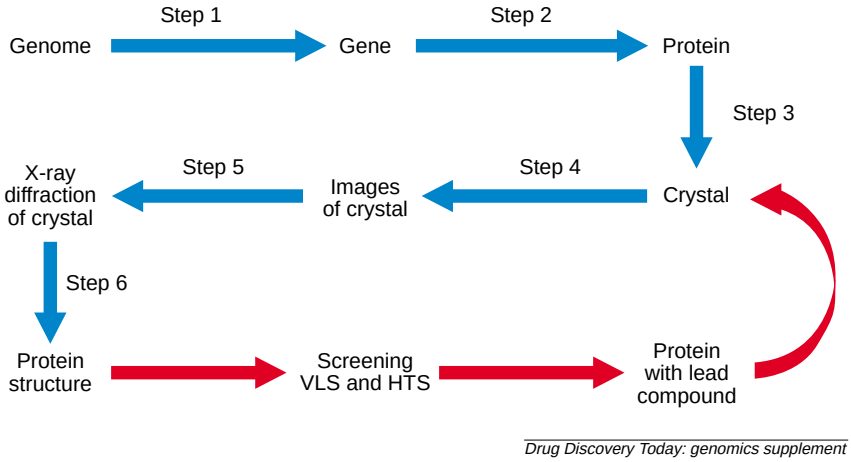

Figure 1. The path from genome to protein structure (blue arrows), and the cycle of structure based drug design. The steps of screening and protein incubation with lead compounds are shown to illustrate the use of crystallography for lead optimization (red arrows). Now, for the first time, all of these steps are being automated. Abbreviation: VLS, virtual ligand screening.

the protein can be expressed and purified, each with thousands of crystallization conditions, leading to hundreds of crystals screened by x-ray diffraction. This process of expansion and selection will be described below for each part of the path in Fig. 1.

Protein purification

In the first two steps of the process, large numbers of genes must be cloned and expressed. Although these steps have often been rate limiting in the past, the availability of standardized protocols has streamlined this part of the structure determination process. Once full length clones are obtained, these are expressed in a standardized prokaryotic host such as *Escherichia coli*, or a eukaryotic host such as baculoviral expression in insect cells[7,8]. To simplify protein purification, an affinity tag is usually added during the cloning step. The most popular combination has been a poly-histidine tag on the protein and metal affinity chromatography. Recently, companies such as Qiagen (Valencia, CA, USA) have adapted robotics designed for PCR and nucleic acid manipulation to the purification of poly-his tagged proteins in a 96-well-plate format.

The amount of protein required for crystallography has typically been between 2 and 20 mg. This amount is sufficient to set up 100–1000 trials, which each use 2 µl of a 10 mg ml^{-1} protein solution. The concentrated protein sample needs to be microscopically homogeneous as measured by mass spectrometry and isoelectric focussing electrophoresis[9]. Also helpful for crystallization is a monodispersed sample with a single oligomeric state, and this can be analysed by size exclusion chromatography and light scattering[10]. High-throughput application of these analytical techniques should help to guide the protein production process.

To be successful, many different versions of the protein are often tested for crystallization. These versions include amino- or carboxy-terminal truncations, point mutations, and orthologs of the protein from different species. Each new variation increases the chances of obtaining well-diffracting crystals. The recent breakthrough crystal structure of a mammalian cytochrome P450 enzyme represents an example of combining a 21 amino acid N-terminal deletion with several point mutations to obtain a protein product that crystallized[11]. The resulting enzyme had activity similar to the unaltered version of the protein.

Crystallization

The next potential bottleneck in the process is the crystallization experiment (Fig. 1; Step 3). Traditionally, this has been a labor intensive procedure where hanging-drop vapor diffusion experiments are manually set-up in 24-well plates[12]. As shown in Table 1, new robotic systems for protein crystal growth have a throughput of 2,500 to 100,000 experiments per day. This automation is expanding beyond the traditional 24-well-plate format, resulting in easier storage and manipulation of large numbers of experiments, and perhaps a ten-fold reduction in material costs[13,14].

Nano-droplets for crystal growth

An important advantage of automated crystallization is the ability to work with smaller protein samples. The increased speed of automation reduces the time that the crystallization droplet is exposed to evaporation before sealing the equilibration chamber. This means that smaller droplets with less protein can be used. In addition, the greater precision of robotics enables the uniform production of small drops. All of the systems in Table 1 are capable of working with protein samples smaller than one microliter.

The reduced amount of protein in these drops eases demands on the step before (protein production) by 10–100-fold. In addition, the smaller drops equilibrate more rapidly[15] leading to a more rapid appearance of crystals. A survey of the time required for crystal formation by four different proteins (Fig. 2) shows that the time elapsed until the appearance of protein crystals can be reduced by 2–10-fold by reducing the drop size. The more rapid appearance of crystals is useful for the iterative process of growth condition optimization, although larger drops with slower growing crystals can also be used later. The topic of collecting x-ray diffraction data from very small crystals will be addressed later in this review.

Table 1. Robotics for protein crystallization

Company	No. of wells per plate	Experiment type	Volume of protein per experiment (μl)	Throughput (experiments per h)
Douglas Instruments (Hungerford, Berkshire, UK)	76	Microbatch under oil	0.20–10.00	200
Gilson Cyberlab C-200 (Middleton, WI, USA)	24	Hanging drop	0.50–3.00	100
Gilson Cyberlab C-240	96	Hanging drop	0.50–1.00	800
Hauptman-Woodward Institute (Buffalo, NY, USA)	1536	Microbatch under oil	0.50	9,200 (for protein addition step only)
LBNL (Berkeley, CA, USA) GNF (San Diego, CA, USA)	48	Hanging drop	0.02–0.10	480
Protein Structure Factory (Berlin, Germany)	96	Hanging drop	0.25–3.00	2,000
Syrrx/GNF (San Diego, CA, USA)	96	Sitting drop	0.02–1.00	4,000

Abbreviations: LBNL, Lawrence Berkeley National Laboratory; GNF, Genomics Institute of the Novartis Research Foundation.

Crystallization imaging

The next key step in the structure determination process is the analysis of the crystallization experiments. The set of 100–1000 conditions per protein must be visually inspected over a time-course of hours to months. The typical protein crystallography lab will have a room (and a cold room) full of crystallization trays, which need to be examined under a microscope on a regular basis. Repeated examinations are required because crystals can grow and disappear, or convert from useable single crystals into unusable multi-nucleated clusters[16]. An example of the growth and demise of a protein crystal as captured by automated imaging is shown in Fig. 3. In this example, the crystal grew to a useful size by Day 3, but was almost completely dissolved by Day 11.

Data collection

The next constriction on the path to structure determination is the collection of x-ray diffraction data. For each project, 10's to 100's of crystals are screened to evaluate how well each one diffracts x-rays. In this step, the crystals are harvested and frozen, then examined by x-ray diffraction. The mounting step remains a manual process for most researchers, although Oceaneering Space Systems (Houston, TX, USA) has built a robotic system for use on the international space station[17]. The collection of diffraction data is carried out either on an in-house x-ray generator or at a synchrotron facility. The synchrotron facility has the advantage of using high-intensity x-rays that reduce the data collection time, but also has the disadvantage of involved safety procedures that limit manual access to the crystal. This limitation is being overcome by efforts at synchrotrons around the world, with the long-term goal of standardized crystal mounting hardware for worldwide use[17,18]. An automatic mounting and alignment system developed by Abbott Laboratories (Abbott Park, IL, USA) has demonstrated 24-hour-a-day unattended data collection from a set of 63 pre-frozen crystals[19].

A key remaining question is the usefulness for diffraction of crystals grown in very small drops. A recent study has found that, in both theory and practice, protein microcrystals are suitable for diffraction analysis[20]. In some cases, smaller crystals perform better than larger ones. When crystals are transferred to liquid nitrogen for use at a synchrotron, smaller and thinner crystals have the advantage of more rapid and uniform freezing rates[21]. A study of various x-ray data sets collected for the transmembrane protein bacteriorhodopsin, showed that the smallest crystals had a lower twinning ratio and a lower mosiacity[22]. That structure was determined to 2.4 Å resolution from x-ray diffraction data collected from a single $30 \times 30 \times 5$ μm^3 crystal[23].

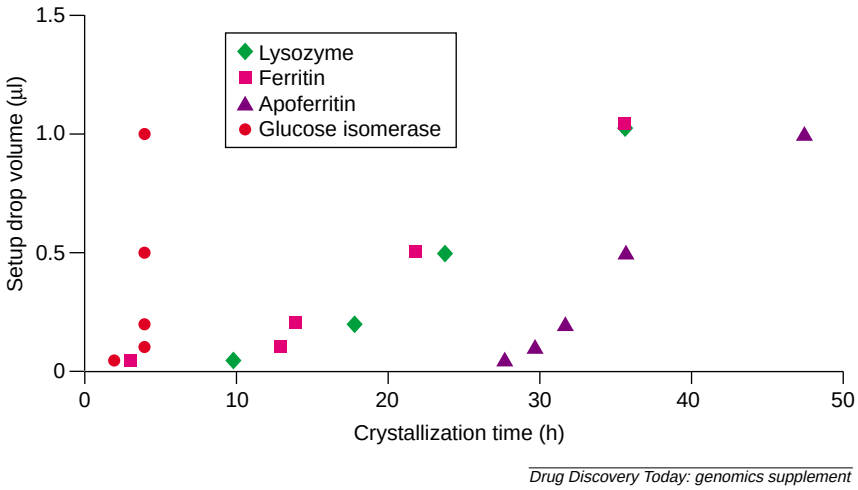

Figure 2. Four different proteins were tested in hanging drop vapor diffusion experiments in drops ranging from 0.02 to 1.00 µl. The elapsed time to the appearance of the first crystals is given for each experiment. Data courtesy of Bernie Santarsiero (Syrrx; San Diego, CA, USA).

In the past decade, significant advances have been made in x-ray data processing and protein structure determination[24] such that the eventual success of a project is usually assured when well-diffracting crystals are available. A large part of this success has come from the widespread adoption of MAD (multiple-wavelength anomalous diffraction) phasing, the method of choice for solving new protein structures[25]. One implementation of this method is to substitute selenomethionine for methionine to provide a source of phase information. This technique can be systematically applied to proteins, enabling the development of high-throughput protocols that can replace the previous standard of soaking crystals with scores of different heavy atom solutions[26].

Structure building

The last part of the structure determination process is the construction of a 3D model of the protein. Rapid progress is being made in automating this task, especially when high-resolution (<2.3 Å) diffraction data are available[27]. These new methods find protein-like features by combining pattern recognition with the growing library of highly accurate structures in the Protein Data Bank (PDB)[28]. Increasingly, the management of the PDB has played a valuable role in checking the accuracy of the finished structures to ensure the quality of this public database.

High-throughput crystallography programs

The promise of structural biology as the logical follow on to our success in genome sequencing has not escaped the notice of governments, venture capitalists, and pharmaceutical companies around the world. In the US, the National Institutes of Health (Bethesda, MD, USA) is funding seven structural genomics centers with US$150 million over five years[29]. All of these projects will develop automation for high-throughput x-ray crystallography and focus much of their efforts on the search for new protein folds. Solving these structures will provide an important opportunity to test the accuracy of fold prediction algorithms, and will also help to fill in the protein-fold dictionary that will eventually be used to look up the structure of every protein sequence[30]. The proteins for the public projects will come from the genome of humans, *Caenorhabditis elegans*, yeast, and bacterial pathogens such as *Mycobacterium tuberculosis*[29].

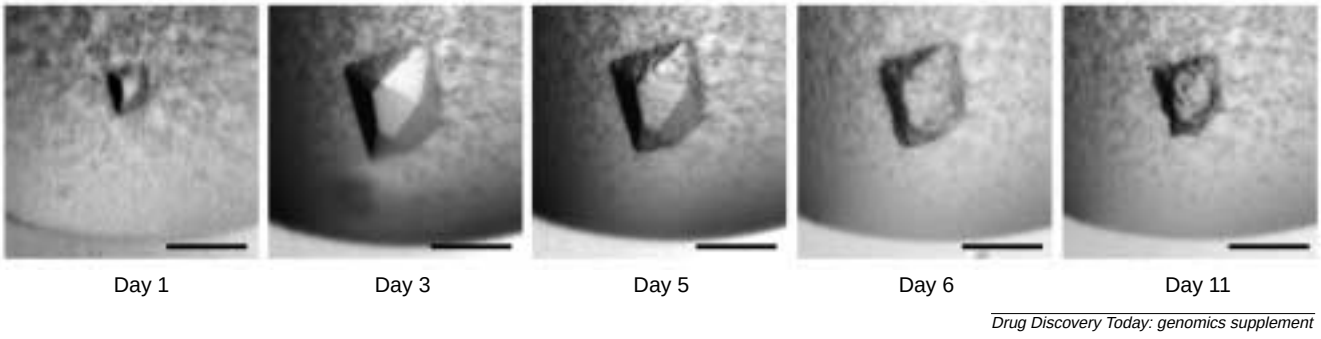

Figure 3. The growth and dissolution of a protein crystal grown in a 50 nl drop. On Day 3, the crystal is greater than 50 µm wide and ready to be harvested for x-ray data collection. By Day 11, the crystal has nearly completely dissolved. Scale bars, 50 µm. Images courtesy of Dan Vaughn (Syrrx; San Diego, CA, USA).

Box 1. Drug design companies

Structure-based drug design
- 3D Pharmaceuticals (Exton, PA, USA)
- Agouron (subsidiary of Pfizer, San Diego, CA, USA)
- BioCryst Pharmaceuticals (Birmingham, AL, USA)
- Emerald Biostructures (subsidiary of Medichem, Bainbridge Island, WA, USA)
- Vertex Pharmaceuticals (Cambridge, MA, USA)

High-throughput structure-based drug design
- Astex (Cambridge, UK)
- Integrative Proteomics (Toronto, Ontario, Canada)
- Structural GenomiX (San Diego, CA, USA)
- Syrrx (San Diego, CA, USA)

In the private sector, structure-based drug design is being pursued by both pharmaceutical companies and biotechnology start-ups. Box 1 lists the larger companies currently involved in these efforts. In the 1980s, structure-based work was embraced by many pharmaceutical companies and also formed the core technology for start-ups such as Agouron and Vertex (see Box 1). In the 1990s, the key new technologies were HTS and combinatorial chemistry. Now, automation and miniaturization are being applied to structure-based drug design by the companies listed in Box 1.

Future impact of HT structure-based drug design

Protein structure is likely to play an increasingly important role in drug discovery. Already, crystallography has facilitated the development of drugs for immunosuppression, and for the treatment of HIV, influenza, glaucoma and cancer[6,31]. Structural information also has the potential to improve the quality of new drugs. As the structures of more proteins involved in toxicity become known (e.g. the P450 enzymes)[11], potential drug candidates can be screened for toxicity through VLS and solving co-crystal structures. Likewise, the specificity of drugs can be improved by screening against close homologs of the intended target.

One of the bottlenecks in drug development is the availability of medicinal chemists. Knowledge of the actual 3D interaction between lead compounds and protein targets should make the work of chemists more directed and efficient. The barriers to successfully bringing a drug to market remain formidable, but the addition of the widespread use of direct structural knowledge of how drugs bind to their targets, will streamline and improve drug discovery.

Acknowledgements

We would like to thank Gordy Hunter of Gilson Cyberlab (Middleton, WI, USA), and Lajos Nyarsik and Holger Eickhoff of the Protein Structure Factory (Berlin, Germany) for providing information for this review. We also acknowledge the contributions to this work of the Genomics Institute of the Novartis Research Foundation (San Diego, CA, USA), especially Peter Schultz, Bob Downs, Mark Weselak, and Andy Meyers. We greatly appreciate the support and suggestions of Ciaran Cronin, Ned David, Mark Knuth, and Wendell Wierenga of Syrrx (San Diego, CA, USA).

References

1 Burley, S.K. *et al.* (1999) Structural genomics: beyond the human genome project. *Nat. Genet.* 23, 151–157
2 Orengo, C.A. *et al.* (1999) From protein structure to function. *Curr. Opin. Struct. Biol.* 9, 374–382
3 Klebe, G. (2000) Recent developments in structure-based drug design. *J. Mol. Med.* 78, 269–281
4 Hertzberg, R.P. and Pope, A.J. (2000) High-throughput screening: new technology for the 21st century. *Curr. Opin. Chem. Biol.* 4, 445–451
5 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using x-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108
6 Bailey, D. and Brown, D. (2001) High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discov. Today* 6, 57–59
7 Edwards, A.M. *et al.* (2000) Protein production: feeding the crystallographers and NMR spectroscopists. *Nat. Struct. Biol.* 7 (Suppl.), 970–972
8 Stevens, R.C. (2000) Design of high-throughput methods of protein production for structural biology. *Structure Fold Des.* 8, R177–R185
9 Lorber, B. and Giegé, R. (1999) Biochemical aspects and handling of macromolecular solutions and crystals. In *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (2nd edn) (Ducruix, A. and Giegé, R., eds), pp. 17–43, Oxford University Press
10 Georgalis, Y. and Saenger, W. (1999) Light scattering studies on supersaturated protein solutions. *Sci. Prog.* 82, 271–294
11 Williams, P.A. *et al.* (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* 5, 121–131
12 McPherson, A., ed. (1998) Practical procedures for macromolecular crystallization. In *Crystallization of Biological Macromolecules*, pp. 159–214, Cold Spring Harbor Laboratory Press
13 Stevens, R.C. (2000) High-throughput protein crystallization. *Curr. Opin. Struct. Biol.* 10, 558–563
14 Mueller, U. *et al.* (2001) Development of a technology for automation and miniaturization of protein crystallization. *J. Biotechnology* 85, 7–14
15 Diller, D.J. and Hol, W.G. (1999) An accurate numerical model for calculating the equilibration rate of a hanging-drop experiment. *Acta Crystallogr. D Biol. Crystallogr.* 55, 656–663

16 Malkin, A.J. *et al*. (1996) Incorporation of microcrystals by growing protein and virus crystals. *Proteins* 24, 247–252
17 Abola, E. *et al*. (2000) Automation of x-ray crystallography. *Nat. Struct. Biol.* 7 (Suppl.), 973–977
18 Service, R.F. (2001) Robots enter the race to analyze proteins. *Science* 292, 187–188
19 Muchmore, S.W. *et al.* (2000) Automated crystal mounting and data collection for protein crystallography. *Structure Fold Des.* 8, R243–246
20 Glaeser, R. *et al.* (2000) Characterization of conditions required for x-ray diffraction experiments with protein microcrystals. *Biophys. J.* 78, 3178–3185
21 David, P.R. (1999) Cryocrystallography: basic theory and methods. In *Practical Protein Crystallography* (2nd edn) (McRee, D., ed.), pp. 409–443, Academic Press
22 Cusack, S. *et al.* (1998) Small is beautiful: protein micro-crystallography. *Nat. Struct. Biol.* 5 (Suppl.), 634–637
23 Pebay-Peyroula, E. *et al.* (1997) X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* 277, 1676–1681
24 Lamzin, V.S. and Perrakis, A. (2000) Current state of automated crystallographic data analysis. *Nat. Struct. Biol.* 7 (Suppl.), 978–981
25 Ealick, S.E. (2000) Advances in multiple wavelength anomalous diffraction crystallography. *Curr. Opin. Chem. Biol.* 4, 495–499
26 Islam, S.A. (1998) HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering. *Acta Crystallogr. D Biol. Crystallogr.* 54, 1199–1206
27 Perrakis, A. *et al*. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458–463
28 Berman, H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
29 Terwilliger, T.C. (2000) Structural genomics in North America. *Nat. Struct. Biol.* 7 (Suppl.), 935–939
30 Mallick, P. *et al*. (2000) Selecting protein targets for structural genomics of *Pyrobaculum aerophilum*: validating automated fold assignment methods by using binary hypothesis testing. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2450–2455
31 Sintchak, M.D. and Nimmesgern, E. (2000) The structure of inosine 5′-monophosphate dehydrogenase and the design of novel inhibitors. *Immunopharmacology* 47, 163–184