

Personal remarks on the future of protein crystallography and structural biology*

Mariusz Jaskolski[✉]

Department of Crystallography, Faculty of Chemistry, Adam Mickiewicz University, and Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

Protein crystallography, the main experimental method of structural biology, has undergone in the recent past three revolutionary changes leading to its unexpected renaissance. They were connected with (i) the introduction of synchrotron radiation sources for X-ray diffraction experiments, (ii) implementation of Se-Met multiwavelength anomalous diffraction (MAD) for phasing, and (iii) initiation of structural genomics (SG) programs. It can be foreseen that in the next 10–15 years protein crystallography will continue to be in this revolutionary phase. We can expect not only an avalanche of protein crystal structures from SG centers, but also attacking of more demanding projects, such as the structure of membrane proteins and of very large macromolecular complexes. On the technological front, the introduction of X-ray radiation from free-electron lasers will revolutionize the experimental possibilities, making feasible even the imaging of single molecules and of intact biological cells.

Keywords: high-throughput crystallography, single-particle imaging, structural genomics, synchrotron radiation, X-ray free-electron laser

Received: 09 July, 2010; **revised:** 20 July, 2010; **accepted:** 15 August, 2010; **available on-line:** 19 August, 2010

INTRODUCTION

Since it is very difficult to make credible predictions, especially regarding the future, I will begin my remarks by noting some revolutionary changes that have taken place in the field of biological crystallography, or in popular language — protein crystallography, in the recent past. And as protein crystallography has been and continues to be the principal method of structural biology, all my remarks will be applicable to a significant degree also to structural biology in general.

In the late 1970s and early 1980s, it appeared to many, including crystallographers themselves, that protein crystallography had already said its most important word and reached a level of post-maturity saturation, with a constant flow (trickle?) of a dozen or two structures per year. And then, in the following decades, three revolutions took place which have changed protein crystallography entirely, leading to its renaissance, with an explosion of productivity, now reaching 7000 structures per annum and culminating at a staggering 60 000 macromolecular X-ray structures in the PDB (Protein Data Bank) (Berman *et al.*, 2000) today.

THREE RECENT REVOLUTIONS

Introduction of synchrotron radiation

The first revolution took place in the 1980s and was connected with the introduction of synchrotron radiation (SR) for routine X-ray diffraction experiments (Dauter *et al.*, 2010). At first regarded with cautious curiosity, SR has had a spectacular effect on the way the X-ray experiment is done in protein crystallography. From a very slow process taking usually many days, the measurement has been shortened to merely a few hours, and today, with the most powerful third-generation beamlines, even seconds are enough. In addition, the data quality has increased dramatically, despite the ever shrinking size of acceptable crystals (today merely 10 μm). A number of other important innovations have been fostered by the use of SR, the routine use of cryoconditions being perhaps the most successful one. Full automation of the diffraction experiment is another consequence of the SR-related revolution. Today, remote SR diffraction experiments ("Fedex crystallography") are quickly gaining popularity (Gilski, 2008).

MAD phasing

The second revolution occurred in the early 1990s and was inspired by the elaboration of the theoretical principles of MAD (Multiwavelength Anomalous Diffraction) phasing (Karle, 1980) and particularly by its development into a practical algorithm (Hendrickson, 1991), most notably with selenomethionine (Se-Met) as the phasing "workhorse". Today, practical algorithms are also available for phasing using the much simpler single-wavelength SAD approach (Dauter *et al.*, 2002). With Se-Met, the formidable phase problem could become a matter of routine, but another breakthrough was necessary.

Impact of genetic engineering and genomics

It came about in full splendor with the new Millennium. The brave successes of the genome sequencing

[✉]e-mail: mariuszj@amu.edu.pl

*Based on remarks presented at a panel discussion "Future trends of experimental crystallography" organized by the Polish Crystallographic Society, Wrocław, Poland, 26 June, 2010.

Abbreviations: HT, high throughput; MAD, multiwavelength anomalous diffraction; PDB, Protein Data Bank; SAD, single-wavelength anomalous diffraction; SG, structural genomics; SR, synchrotron radiation; XFEL, X-ray free-electron laser

projects, culminating with the sequencing of the human genome in 2001 (International Human Genome Consortium, 2001; Venter *et al.*, 2001) (many hundreds of genomes have been sequenced by now) have created an embarrassing aftertaste: scientists had to admit that they have very little idea about the information written in the ATGC alphabet of the book of the genome. It became apparent that to be understood, the book has to be translated into the language of protein structure. A method was needed that would quickly and in a high-throughput (HT) fashion furnish the missing protein structures. This is how protein crystallography was reborn, boldly aspiring to fulfill this expectation. It is today the best, cheapest, fastest, and most reliable method for the job. In combination with genetic engineering (and protein engineering), protein crystallography has become a tremendously powerful tool. Structural genomics (SG) centers, with a motto to characterize structurally entire recombinant proteomes, have been thus mushrooming throughout the world. Although they have been the driving force for an incredible flow of technological advances, the promised flood of structures has not quite occurred yet. But...

FUTURE PERSPECTIVES

Changing my hat of a “historian” to that of a “prophet”, I would venture a statement that HT-SG will come of age in the near future, delivering not only the low-hanging fruit, but also more ambitious target structures, and in quickly increasing numbers. We have about 30 000 genes in our genome, and most of the proteins encoded by them are just blunt question marks. In addition, many more proteins are made owing to alternative splicing and other RNA-specific mechanisms. And what with the proteins of all the other fellow creatures on our planet... From this point of view, protein crystallography will be needed, and busy, for quite some years to come.

Membrane proteins

But numbers are not everything. Protein crystallography can be expected to attack proteins that have remained a taboo for a long time, because of the near-impossibility to crystallize them: integral membrane proteins, which probably dominate our protein make-up. Increasing numbers of such proteins are already landing in the PDB but I would predict that the near future will see a breakthrough in crystallographic characterization of membrane proteins and that this will change our understanding of the protein world, so far practically limited to soluble globular proteins.

Macromolecular complexes

Another direction where a revolution can be expected is structural characterization of huge macromolecular complexes. We already know the structure of hundreds of viruses, and the recent Nobel Prize in Chemistry (2009) awarded to Ramakrishnan, Steitz and Yonath has crowned the efforts to map the structure of the macromolecular factory, the ribosome, which is responsible for the biosynthesis of all proteins in all living cells (Ban *et al.*, 2000; Schluenzen *et al.*, 2000; Wimberly *et al.*, 2000; Selmer *et al.*, 2006). The ribosome is a huge edifice, with roughly 200 000 non-hydrogen atoms, whose positions in the three-dimensional structure are now precisely known. But it is a very stable macromolecular complex. What I

foresee is that protein crystallography will be attacking large macromolecular assemblies that are not cemented forever, but have a transient nature, forming and reforming, according to the rhythms dictated by the processes of life. Molecular recognition is the keyword here and the trick will be to map and understand those tenuous interactions.

X-Ray free-electron lasers

Challenges presented to us by biology will be, however, only one of the engines propelling the continuing revolution in protein crystallography that I predict for the next one-two decades. Technological progress will be the second engine. We can already discern the looming shape in this area: novel astronomically powerful sources of X-ray radiation. If you thought that with third-generation synchrotrons we have reached the edge of the X-ray universe, think again... The dream of an X-ray laser is finally becoming true. Such devices do not operate in an oscillating mode (as do classical lasers) but have an “open” construction, extending even for a number of kilometers. The lasing action is based on relativistic electrons propagating in a linear fashion, thus the name X-ray Free-Electron Laser (XFEL). The XFELs will provide X-ray light more than 10 orders of magnitude brighter than from existing synchrotron sources. We are not sure yet how to harness such powerful radiation for structural studies; imagination (or imaging!) is the best source of inspiration.

Mapping of macromolecular dynamics

One possibility, already being exploited with traditional synchrotron sources, is to map protein crystal structures with incredible speed (Bourgeois & Royant, 2005). It is already possible to record the complete diffraction pattern in about 1 ns, but 1 ps or less may soon be a real possibility. Thus, by taking a succession of “structure photographs” with such a fast “shutter speed”, we should be able to map the trajectory of a chemical or physical process that occurs in our crystal. Even if the process was very fast by the standards of our senses (say 1 ms or even 1 μ s), we would still be able to take many snapshots as it progresses from beginning to end. Thus, not only static protein structure but also protein dynamics and kinetics of protein transformations can be studied by protein crystallography.

Single-molecule diffraction

But with such powerful X-rays, do we need a crystal at all? In crystallography, we form a periodic crystal from our molecules because we need an amplifier that would concentrate the weak scattering signals only in discrete Bragg peaks. With a beam many billions of times brighter, the amplifier is not necessary, and a single molecule should do. This bold proposition has been investigated theoretically (Hajdu, 2000) and appears to be a viable approach. We would thus be able to map the structure of a single molecule injected into the X-ray beam of an XFEL laser. It does not matter that the molecule would disintegrate very quickly. If we are able to capture the scattering pattern fast enough, the structure is ours! And there would be no phase problem because the Fourier transform of a non-periodic structure is continuous, providing plenty of data to recover the missing phases. There is an intellectual beauty in this method: starting from a crystallographic concept, we are able to leave the

constraints of the crystal behind, and go forward using a mathematical apparatus that is still crystallographic in its essence.

Imaging of large biological structures

But why limit our non-periodic object to just one molecule? We could have a huge macromolecular assembly, perhaps even the entire living cell, and still try to map its structure by X-ray scattering. Such pilot experiments have actually been done already and the first results are very promising. It is extremely gratifying that these breakthrough approaches are spearheaded by a pioneer of the creation of direct methods in crystallography, David Sayre (2007).

CONCLUDING REMARKS

When I said that XFEL lasers will become available soon, I was not quite correct because the first XFEL installation is already functioning at Stanford University and, indeed, one can apply for experimental time at a laser X-ray beam already! Another XFEL is under construction at DESY (Hamburg, Germany) and should start operation in the X-ray regime, hopefully, before long. Several other devices of this type are planned in Europe, for example at the DIAMOND (UK) and ALBA (Spain) synchrotron centers. The linac injector of the new MAX IV synchrotron (Lund, Sweden) will be converted into a laser accelerator in the future. So, if we keep looking at synchrotrons as the X-ray sources of the future, we might be actually looking into the past...

To summarize what I have said above, protein crystallography should continue to experience the “revolutionary phase” for some time to come, greatly benefiting and advancing the whole area of structural biology. This will be connected with dramatic technological developments in the methods of experimentation, coming mostly from physics, and with bold and visionary challenges posed by biological sciences. In this sense, protein crystallography will retain its role as a focal point at the intersection of the principal natural and life sciences. But in its niche, it should strongly hold to its chemical legacy. The strength of protein crystallography has been, and no doubt will continue to be, in penetrating chemical interpretations of the structure and function of the matter of life. I could paraphrase by saying, “he that understandth structure of matter, he hath the power”. And this person is, without doubt, the protein crystallographer.

Two other aspects seem to me important for the future of protein crystallography: one is the quality of protein crystallographic research, and the other the question of the upbringing of the next generation of protein crystallographers. With regard to the quality of the crystallographic product, i.e. of a crystal structure determination, it is gratifying to note that thanks to the various technological and methodological advances, the quality of protein crystal structures has been improving systematically, a fact that must be emphasized if we remember that the complexity of the structures tackled has been increasing as well. Generally, protein structures are being determined with better statistical parameters (also visible in the statistics of the raw data) and at better resolution. There is now a wealth of protein structural data available, with quality equal to or surpassing that characterizing standard small-molecule crystallography (Wang *et al.*, 2007). There are also exciting cases, still quite rare at present, where it was possible to refine multipole models

of proteins, which essentially describe the distribution of bonding electrons and the deformation of atomic charge distribution from the normally assumed spherical approximation (Jelsch *et al.*, 2000; Guillot *et al.*, 2008). One hears sometimes the opinion that such accurate studies of protein structure are unnecessary because the basic questions can be answered at a more crude level. I could not disagree more with such an opinion. First, as scientists, we have the obligation to search for the objective truth in the best way we can. Second, our strength lies in the ability to provide not only rather nebulous structural models, but to push the resolution with the ultimate goal of mapping the structure atom-for-atom, without ambiguity or approximation. One can never know what to expect from even the most “commonplace” structure. With poor data and model quality, many of the exciting, unexpected features could be simply overlooked!

All the prognostication about the future of protein crystallography will be of no use if there are no protein crystallographers in the future. As ridiculous as it might sound, it is in fact a very real danger. With crystallography quickly disappearing from university curricula around the globe, we are facing a situation where there would be no specialists to carry on and carry out the research, let alone to advance the discipline (Wlodawer *et al.*, 2008). If by unwise (but trendy) policies we lose a generation, there may be no one in the next 15–20 years to implement the wonders about which I have been pondering above, or even to interpret the results of quite routine studies. This would really mean cutting a thriving branch of the tree of human knowledge and culture. Let us not only hope, but, indeed, *do* all we can to eradicate such a bleak scenario!

REFERENCES

- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905–920.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242.
- Bourgeois D, Royant A (2005) Advances in kinetic protein crystallography. *Curr Opin Struct Biol* **15**: 538–547.
- Dauter Z, Dauter M, Dodson E (2002) Jolly SAD. *Acta Crystallogr D* **58**: 494–506.
- Dauter Z, Jaskolski M, Wlodawer A (2010) Impact of synchrotron radiation on macromolecular crystallography: A personal view. *J Synchrotron Rad* **17**: 433–444.
- Gilski M (2008) Automation and remote synchrotron data collection. *Acta Phys Polon A* **114**: 331–338.
- Guillot B, Jelsch C, Podjarny A, Lecomte C (2008) Charge-density analysis of a protein structure at subatomic resolution: the human aldose reductase case. *Acta Crystallogr D* **64**: 567–588.
- Hajdu J (2000) Single-molecule X-ray diffraction. *Curr Opin Struct Biol* **10**: 569–573.
- Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51–58.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Jelsch C, Teeter MM, Lamzin V, Pichon-Pesme V, Blessing RH, Lecomte C (2000) Accurate protein crystallography at ultra-high resolution: Valence electron distribution in crambin. *Proc Natl Acad Sci U.S.A* **97**: 3171–3176.
- Karle J (1980) Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int J Quantum Chem* **S7**: 357–367.
- Sayre D (2007) Single-particle imaging. In: *Evolving Methods for Macromolecular Crystallography*. Read RJ, Sussman JL, eds, pp 181–191. Springer, Dordrecht.
- Schlutzenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A (2000)

- Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**: 615–623.
- Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**: 1935–1942.
- Venter JC, Adams MD, Myers EW (2001) The sequence of the human genome. *Science* **291**: 1304–1351.
- Wang J, Dauter M, Alkire R, Joachimiak A, Dauter Z (2007) Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr D* **63**: 1254–1268.
- Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vornrhein C, Hartsch T, Ramakrishnan V (2000) Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.
- Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **275**: 1–21.