Full length article

Structural biology: Past, present, and future

Peter B. Moore

Department of Chemistry, Yale University, PO Box 208107, New Haven, CT, USA

ARTICLE INFO

Article history:
Received 9 June 2016
Accepted 19 July 2016
Available online xxx

ABSTRACT

In March 2014, a symposium was held at Aarhus University to commemorate the 40th anniversary of the founding of its Division of Biostuctural Chemistry by Professor Brian Clark. This article is a lightly edited version of the talk the author gave to open that event. It consists of some reflections about the discipline known as structural biology, and some comments about the author’s interactions with Professor Clark and the Division of Biostuctural Chemistry over the years.

© 2016 Elsevier B.V. All rights reserved.

Forty years is a long time in the life of any academic unit, especially one like the Division of Biostuctural Chemistry at Aarhus, which is devoted to a single kind of scientific activity. Fashions change. What has kept this particular unit healthy for so long is on the one hand, the success its leaders have enjoyed in recruiting and retaining good scientists, and on the other hand, its focus on a topic of undying interest to biologist: structure.

Biologists have been interested in structure for a very long time. De Humani Corporis Fabrica, the book that marked a revolution in our understanding of the structure of the human body, was published by Andreas Vesalius in 1543. Somewhat more than a century later, using primitive microscopes, Anton van Leeuwenhoek (1632–1723) and Robert Hooke (1635–1702) discovered the existence of an entire universe of organisms all too small to be seen with the naked eye, and demonstrated that organisms large enough to see are made of cells.

While these groundbreaking discoveries all had to do with the structure of biological systems, they do not correspond to what biologist mean today by the phrase “structural biology”. Structural biology is the determination of the three dimensional structures of biological macromolecules at atomic resolution, and the use of those structures to explain the biological properties of those molecules in chemical terms.

As scientific disciplines go, structural biology has a strange history. It is the only one I know of about which one can state without fear of contradiction that its first triumph was the greatest it ever will have. The triumph I refer to, of course, is embodied in the proposal for the structure of DNA that was published by Watson and Crick in 1953 [1].

There are two reasons Watson and Crick’s cartoon-like structure (Fig. 1) had the impact it did. First, because it spoke directly to an issue at the heart of biology, namely the chemical basis of heredity, it had the biological relevance granting agencies profess to love at level unlikely ever to be surpassed. Its second virtue was that the relationship of this structure to the biological function it performs was/is transparently obvious. In the words of its authors,

“It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.”

J.D.Watson, & Francis H.C. Crick [1].

Nothing so transformational has happened in structural biology since. Fig. 2, which portrays the next two landmarks in structural biology, may help the reader understand the problem structural biologists have been contending with ever since DNA. The structure on the left is myoglobin [2], and the one on the right is lysozyme [3]. As everyone knows, myoglobin binds molecular oxygen reversibly, and lysozyme hydrolyzes polysaccharides, but it is not at all obvious from Fig. 2 that this should be so, and more detailed, less schematic renderings of these two proteins would not help. Molecules like these two, which are typical of those on which structural biologists have concentrated for the last 60 years, hide their secrets well. Decades of experience have shown that atomic resolution structures are a prerequisite for arriving at a chemical understanding of molecular function, but are seldom sufficient by themselves. It usual takes work, often a lot of work, to make the connection between structure and function.

Structural biologists have another problem, which most would rather not talk about. Most of the time most biologists do not care about the structural details. Because function has already been
explained by structure for so many macromolecules, biologists are confident that a chemical explanation will ultimately be found for the functional properties of the macromolecules they care about, if they have not been already. Hence many are content to explain biochemical processes using “jelly bean” diagrams in which entire protein molecules are represented using little colored circles, squares or ellipses. It is only when the atomic details are likely to be enabling that their ears will prick up.

The event that first stirred my own interest in macromolecular structure was a pair of seminars given by Max Perutz at the Harvard Medical School in the spring of 1963. He was the Dunham Lecturer that year, and the Medical School rented a huge auditorium in a nearby high school to accommodate the crowd it anticipated would want to attend. They were right to do so; the place was packed. After a few minutes of introduction, Max asked the projectionist to show the first stereo slide. We all donned the cardboard stereo glasses that we had picked up as we entered the room, and after the projectionist had adjusted the alignment and focus of the projector, a highly audible “oooh” erupted from the audience as a three-dimensional image of myoglobin, which was about 10 m across, emerged from the screen right over Max’s head. More than 99% of the people in the audience, including me, had never seen an atomic resolution structure of a protein in three-dimensions before; it was a revelation.

Four years later I had the good fortune to become a postdoc at Max’s home institution, the MRC Laboratory of Molecular Biology at Cambridge. There I shared an office with three other postdocs: David DeRosier, one of the pioneers in the field of EM image reconstruction, about which more below, Jean Witz, a French scientist interested in virus structure, and Brian F.C. Clark. What I can tell you about that office is that if had been any larger, there would have been five desks in it, not four.

Unlike me, Brian was seldom to be found in that office. Most of the time he was upstairs doing experiments on transfer RNAs with the other molecular biologists at the MRC, of whom there were many. Sometime in 1968 rumors started to circulate to the effect that a young German scientist named Hasko Paradies had obtained crystals of a tRNA, an astonishing development, if true. Suffice it say that in retrospect, it appears not to have been true [4]. Nevertheless, I had the distinct impression at the time that this rumor stimulated groups all around the world, including at the MRC, to get serious about tRNA crystallization. Brian was deeply involved in this enterprise, and in 1974, the year Brian moved to Aarhus to establish the unit we are celebrating, the first tRNA structures were published, one of them from the MRC [5,6] (Fig. 3).
**Fig. 2.** The structures of myoglobin and lysozyme. These ribbon representations of the structures of myoglobin and lysozyme were obtained from the Protein Data Bank (PDB), which is operated by the Research Collaboratory for Structural Bioinformatics. They are the images the PDB uses to represent the structures identified as 1MBN (myoglobin) and 1LYZ (lysozyme). 1MBN was derived by refinement from the 2 Å resolution structure of myoglobin reported in 1960 by Kendrew and coworkers [2]. 1LYZ is a refined version of the original 2 Å resolution structure of lysozyme [3,18].

**Fig. 3.** The structure of yeast phenylalanyl tRNA. The image shown represents 1EHZ [8], but any of the other structures that have been obtained for this molecule would be all but indistinguishable in presentations like this one. This is the iconic image of this particular molecule used by the Protein Data Bank. The red “wire” in the figure depicts the trajectory of the molecule’s backbone. Bases are color-coded as follows: adenine (red), cytosine (yellow), guanine (green), and uracil (blue).
Brian recruited some excellent people to help him get the new enterprise started. The two I knew the best in the first wave of hires are Roger Garret and Jens Nyborg, who, sadly, passed away several years ago. A decade or so later, Morton Kjeldgaard, who had spent a few years as a postdoctoral fellow with me, also joined the staff.

For the next 20 years, the tRNA structures published in the mid-1970s constituted the lion’s share of what was known about the structure of RNA at atomic resolution, and in the late 1990s, for reasons irrelevant here, we decided that it might be worth finding out what would emerge if data were collected from similar crystals using synchrotron radiation, which was not available in the early 1970s, and their structures resolved using 1990s techniques. Unknown to us, Daniella Rhodes, who had been a member of the team that produced the MRC structure of tRNA in 1974, had the same idea at the same time, and the two groups published their results in 2000 almost simultaneously [78]. As both groups had hoped, the resolutions of the data sets collected from these crystals were significantly higher than those of the data sets collected from similar crystals 25 years previously: ~2.5 Å vs. ~2 Å. However, save for a not inconceivable improvement in coordinate accuracy, little new was learned about tRNA structure.

In 1974, the year Brian opened up shop in Aarhus, the number of atomic resolution macromolecular structures available stood at about 60. All of them had been solved using X-ray crystallography because at the time there was no alternative. It was not until 1984 that the first NMR-derived protein structures were published, and while by 1974 there already existed a small cabal of subversives, the members of which thought it possible to obtain atomic resolution structures for macromolecules by image reconstruction starting with electron microscopic images, even the most optimistic of them was willing to admit that it might be decades before that ship came in. It followed that if there was to be any structural biology at Aarhus, it would have to be X-ray crystallography, and that is the activity that Brian got organized.

In the 1970s, X-ray crystallography was not for the weak of heart. The best X-ray source available was the Eliot rotating anode, and the preferred method of data collection was diffractometry, one reflection at a time. It could take months to collect a single data set, and the data had to be analyzed on expensive mainframe computers using programs you either wrote yourself or borrowed from a colleague. It was decades before things improved, but improve they did. The construction of the first synchrotron beam lines designed for the study of biological samples in ~1976 was a harbinger of things to come. However, the number of crystal structures solved each year did not increase dramatically until the mid-1990s, when the first third-generation synchrotrons came on line, as well as X-ray detectors of improved design and powerful, user-friendly data processing programs. As of this writing (May 31, 2016), the number of structures deposited in the Protein Data Bank stands at 119,137, roughly 90% of which have been produced by X-ray diffraction. (Most of the rest have been solved by NMR.)

I first visited Aarhus in September 1994. On that occasion I gave a seminar on the structure of EF-G, which had just been solved both at Lund and at Yale [9,10]. After the talk, a student in the back of the room raised his hand and said that he thought that the structure of EF-G resembled the structure of the ternary complex that EF-Tu forms with tRNA and GTP, which had just been solved at Aarhus by the team led by Jens Nyborg [11]. After the question period was over, we went to a workstation in another room, and in a remarkable display of competence, I managed to upload the coordinates for EF-G from Yale, and the comparison was done. The student in question, Poul Nissen, turned out to be right, and thus the idea of molecular mimicry idea was born (Fig. 4) [12]. I do not know who thought up the phrase “molecular mimicry”, but it was inspired. If you want your ideas to be remembered, you must give them memorable names.

I am still not sure quite why these two molecules look so much alike: function? evolution? Professor Harry Noller (UC Santa Cruz) and I agreed at the time that the most important thing we had learned from the discovery of molecular mimicry was that God has a sense of humor. It seemed obvious that he (or she) thought it would be amusing to watch people scurry around trying to concoct proposals for the mechanism of protein synthesis based on this enigmatic, but mesmerizing discovery.

After he finished up at Aarhus, Poul Nissen spent 2 years or so at Yale as a postdoctoral fellow with Professor Thomas Steitz. There he got involved in the then on-going effort to obtain a crystal structure for the large ribosomal subunit from H. marismortui, which involved my group as well as Steitz’s. In the end, that enterprise worked out well too, e.g. [13,14], and in 2005, Steitz won an all expenses paid trip to Stockholm.

Poul returned to Aarhus in 2000, and has become one of the leaders of the Division of Biostructural Chemistry. He transformed himself into a membrane protein crystallographer. This burgeoning field is well on its way to filling one of the most obvious and longstanding gaps in the PDB, namely the relative paucity of structures for members the very large and very important class of proteins that associate with membranes. Structural Biology is alive and well at Aarhus.

Where is the structural biology enterprise headed? Of what will it consist in the future, say, 40 years from today? As the noted baseball player and philosopher Yogi Berra is once alleged to have said, “It is hard to make predictions, especially about the future.” There are, however, few observations one can make without going too far out on a limb.

First, when it comes to generating new structures for biological macromolecules, the day of the NMR spectroscopist has come and gone. It can be done, but the molecular weight limitations inherent in the method place severe restrictions on what can be accomplished. It makes much more sense for NMR spectroscopists to concentrate on the dynamics of macromolecules, rather than their (average) structures because they can provide important information in that arena that no one else can access. This is what many of them are now doing.

Second, it could happen that the bio-molecular simulation enterprise will advance to the point that macromolecular structures no longer have to be solved experimentally; all that will be needed are sequences. In fact, crystallographers have been busy undermining their own long-term employment prospects by flooding the databases with new structures that computational biologists can use to inform their modeling efforts. This is a kind of cheating, of course, but crystallographers can hardly complain. They do the same thing every time they solve a structure by molecular replacement. That said, the experimentalists are not out of a job yet. If a computational biologist were to tell me tomorrow he had obtained a structure for a protein I cared about, I would ask him to prove it. The only means he would have for doing so are crystallography, NMR, or, possibly, EM. But if that is what it takes to be sure a structure is correct, why bother with computation at all?

Third, those who have been paying attention will know about the excitement surrounding the use of the radiation produced by free electron lasers (FEL) to solve macromolecular crystal structures. These devices produce very short pulses of X-rays that are many orders of magnitude brighter than the pulses of radiation generated by 3rd generation synchrotrons. When you are collecting data this way you do not have to freeze the crystals you use because the energy deposited in a crystal by a single pulse of radiation is so great that it instantly destroys the impacted region. Happily the photons diffracted by these crystals emerge
Fig. 4. The structure of elongation factor G compared to that of the ternary complex formed by elongation factor Tu, an aminosacyl tRNA, and GTP. Both structures are shown as ribbon diagrams. This image is taken from [12].

from them before the irradiated material disintegrates. This technology can be used to collect diffraction data from crystals that are orders of magnitude smaller than those required for conventional data collection. In fact, it may someday become possible someday to collect atomic resolution FEL data from single macromolecules. In addition, FEL data should represent macromolecular crystals that have suffered little or no radiation damage, and FELs could be used to perform experiments that require exceptional time resolution. All of the above notwithstanding, I doubt that FELs will ever become as important for biological scientists as 3rd generation light sources. The number of such beams available is never going to be large because they are so expensive, and currently, at least, the method is very clumsy.

Fourth, as far as the crystallographic community is concerned, it is the electron microscopists who pose a clear and present “danger”. In 1997, David DeRosier whom I mentioned earlier, published an article in Nature that had the cheeky title, “Who Needs Crystals Anyway?” [15]. As he was not so subtly pointing out, no crystals are required for structure determinations by EM. In addition, by comparison with crystallography, the amounts of material needed are miniscule; femtomoles will do. In addition, EM is comparatively insensitive to sample heterogeneity, and can in fact take advantage of it to map out the conformational landscapes. Finally, EM is extremely size-tolerant; the larger the object whose three-dimensional structure is required, the easier it is to deal with.

Fig. 5 shows portions of a 3.3 Å resolution density map that was obtained for an icosahedral virus in 2010, which is representative of the best that electron microscopists could do in early 2014, and it is very good indeed. The fitting of sequences into these maps was unambiguous. However, it is hard to imagine a specimen more amenable to structure determination by EM than a small icosahedral virus. Small viruses are comparatively rigid, which reduces conformational heterogeneity, and the quality of the maps obtained could be (and were) greatly improved by symmetry averaging. For less rigid, asymmetric objects like the ribosome, the resolutions then attainable were substantially less, e.g. 5–6 Å [16], and chemical interpretation of maps correspondingly more difficult.

Historical Note: At almost exactly the same time the talk was given on which this article is based, the first reports appeared of EM structures solved using data collected using direct electron detectors. It had been realized for some time that EM specimens move around in unpredictable ways (“creep”) when they are exposed to the electron beam, and that creep was degrading image resolution. The speed of this new class of detectors is so high that one can obtain what amounts to freeze-frame images with them, and this has resulted in a revolution in the efficiency with which

Please cite this article in press as: P.B. Moore, Structural biology: Past, present, and future. New Biotechnol. (2016), http://dx.doi.org/10.1016/j.nbt.2016.07.010
structures can be solved by EM. For example, ribosome density maps are being obtained by EM today that have resolutions in the 3 Å range, e.g. [17], and consequently are just as interpretable as the virus maps shown in Fig. 4. In short, there is now a new game in town.

I am confident that the structural biologists at Aarhus will take advantage of these developments, which hold so much promise for their field. That said, the future prosperity of this unit will be determined by the choices its members make about which problems to pursue, rather than by the methods they select to address them. The day has long past when so little was known about macromolecular structure and data collection was so difficult that it made sense to selected molecules for structural determination based on ease of preparation and/or crystallization. Now some facet of biology must be illuminated by the structures one seeks to solve, and the more important the biological implications of a structure, the better. Here “important” can mean important in a pure science sense, but it can also mean important in an applied, medical sense. If the scientists in this unit make choices in the future as well as they have made choices in the past, there is every reason to be confident that it will survive to celebrate an 80th anniversary.

References


Please cite this article in press as: P.B. Moore, Structural biology: Past, present, and future, New Biotechnol. (2016), http://dx.doi.org/10.1016/j. nbt.2016.07.010