J Comput Aided Mol Des (2010) 24:887–890
DOI 10.1007/s10822-010-9382-0

WARR'S PIECE

Collaboration, competition, validation and plans for the future

An interview with Gerard Kleywegt, Head of the Protein Data Bank in Europe

Wendy A. Warr

Received: 8 August 2010 / Accepted: 31 August 2010 / Published online: 5 October 2010

Gerard Kleywegt obtained his PhD in 1991 from the University of Utrecht. After a brief spell with a software company he joined Alwyn Jones's protein crystallography laboratory in Uppsala. He stayed there for 17 years and developed many tools that are widely used by structural biologists worldwide. From 1996 to 2009 Gerard was involved in running the Swedish Structural Biology Network (SBNet), first as network coordinator and later as program director. From 2002 to 2007 he was a Research Fellow of the Royal Swedish Academy of Sciences. In 2009, he was appointed Professor of Structural Molecular Biology at Uppsala University. Since July 2009 Gerard has been head of the Protein Data Bank in Europe (PDBe) [1, 2] at the European Bioinformatics Institute (EBI) near Cambridge in the United Kingdom. He has an extensive publication record and has served on the PDBe Scientific Advisory Board for some years. He has also been one of the European representatives on the Worldwide PDB (wwPDB) [3] advisory committee.

W. A. Warr (✉)
Wendy Warr & Associates, Holmes Chapel, UK
e-mail: wendy@warr.com

Interview

WAW: Almost all your career has been spent at Uppsala. Why did you move after 17 years?

GK: It wasn't a sudden move and it was certainly not an easy decision. Even after I accepted the job I stayed in Uppsala for a year to fulfill my teaching and supervision duties. I had been on the PDBe [1, 2] scientific advisory committee, I knew that the job was coming up, I knew Kim Henrick, my predecessor, quite well, and I was encouraged to apply by Janet Thornton. I have worked with Janet and her team on validation efforts since the mid-1990s. In the end, the overriding consideration was the fact that as head of PDBe I could have a significant impact on the way in which biomacromolecular structures are archived and served back to the entire biomedical community.

WAW: Do you miss Sweden? Do you go back to visit?

GK: I had a good job in Sweden and I miss living there. I still have my house in Sweden and I like to go back, to meet old friends and to enjoy the fresh air and nature.

WAW: What is your job title now? What do you *do*?

GK: In EMBL (European Molecular Biology Laboratory) parlance I am a "Senior Team Leader" and head of PDBe [1, 2]. The last year has been a learning exercise, as part of which I have talked to each individual in the team, trying to figure out what they do on a day to day basis and where they came from, and so on and so forth. I also got to know a large number of new colleagues in the EBI and EMBL and in our wwPDB [3] and Electron Microscopy Data Bank (EMDB) [4] partner organizations. EMBL rules insist on staff turnover and therefore almost nobody has tenure. As a consequence, in my first 18 months at the EBI, PDBe will lose staff whose total experience adds up to more than 100 years. However, this opens up opportunities as well. PDBe is an important project and attracts good people. Already I have been able to hire about a dozen talented new staff members, but this has taken up a lot of my time of course. The PDB is a global project and EMBL is an international organization, which means that I must travel once or twice a month as well, attending meetings related to the wwPDB and EMDB, attending EMBL faculty meetings, teaching courses, lecturing at conferences and so on. I have been really busy in my first year.

WAW: Where does this fit in with the rest of EBI?

GK: There are 400 people at EBI; 70 are in research; almost all the rest are in services (in bioinformatics). There are also outreach and training, an industry program and admin. So there are research groups (usually small) and service teams (usually large) such as UniProt and PDBe (not as large as UniProt). My role does include some research as well, but so far my responsibility for PDBe has left no time for that. There are about 25 people in the PDBe team. Tom Oldfield runs the team that is responsible for the technical side of the databases and services, including weekly updates. There are five people in the deposition and annotation team (the people who process the structures that come in), and three in scientific content and integration, who are concerned with getting us to talk to other databases in EBI and developing new ways to access structural data. If I had to swap jobs, that is the team I would want to work in. Then there are three people in the electron microscopy (EM) project and two in NMR. The EMDB was initiated here by Kim Henrick but is now operated jointly by PDBe, RCSB (the Research Collaboratory for Structural Bioinformatics) [5] and a group at Baylor College of Medicine.

WAW: Do you have time to do any research as well as running PDBe? What are your research interests?

GK: My first year has been too busy but I will be doing similar research to that I did in Uppsala: things that could tie in with PDBe services or things that could become services. In Uppsala I wrote a number of programs where we took the PDB and derived a database from it. This allowed people to search, for example, 3D fragments of protein structures. Maybe I will develop structural bioinformatics tools for nucleic acids as there aren't many of those at the moment. Eventually such tools could become PDBe services.

WAW: I am told that the cooperation between the RCSB and PDBe is a truly substantial international collaboration. Tell me about the synergies between RCSB and PDBe. How do you work together?

GK: The annotators talk to each other almost every day when they find problem structures or unusual ligands, for example. Helen Berman [6] and I have weekly Skype sessions, and there are e-mails, phone calls and video teleconferences all the time. The four wwPDB partners are working on a common software tool for deposition and annotation of X-ray, NMR and EM data and structure models. At the moment PDBe and RCSB have different software but eventually everything will be consistent for the depositor. It is an enormous project. We have been working on it for 2 years; the anticipated delivery date is the end of 2011. Biomacromolecular structure data are very complex: there are so many aspects that have to be handled and we have to talk to other databases and do cross-referencing. X-ray, EM and NMR all have specific requirements. The software has to handle not just deposition but also annotation by wwPDB curators. We need secure access for example to run sequence searches against the UniProt database. We also want to link structures to lots of other databases: the SIFTS initiative (Structure Integration with Function, Taxonomy and Sequence) [7] aims to integrate PDBe, UniProt, sequence-family databases such as Pfam [8], function-annotation databases such as GO [9] and Interpro [10], and the structure-classification databases SCOP [11] and CATH [12].

PDBj in Osaka [13] is also involved in the work to produce a common wwPDB deposition and annotation tool (they currently use the same software as RCSB), and the BMRB (BioMagResBank at the University of Wisconsin) [14] is collaborating too. The wwPDB partnership and collaboration is all about what we call "data *in*", which includes the deposition and annotation software and procedures, standards, validation, formats and common libraries.

WAW: I noticed that you have separate web pages with different tools for searching and analysis.

GK: Collaboration on "data in" is crucial to ensure a single, uniform global archive. Serving the data back to users, on the other hand, allows for diversity. The wwPDB partners respect each other's efforts to produce the best possible services in a mildly competitive way, which ultimately offers the greatest benefit to science. PDBe, RCSB

and PDBj therefore all produce their own set of unique, value-added services, tools and activities, based on the local environment, strengths and interests. For instance, the fact that PDBe is part of the EBI puts us in a perfect position to work on integration of structural and other biological data resources.

WAW: How does cooperation benefit the user?

GK: The new high quality software for deposition and annotation will benefit the user and all of us will benefit from load balancing. The user need not know where he or she is depositing a structure. There will be only one piece of software to maintain, the dictionaries will be the same and the data coming out will be better.

WAW: You've been running PDBe for a year now. Have you had any surprises? Is your vision still the same?

GK: When you are on the outside it's easy to say to the PDB: "W*hy can't you just change this*?" In academia you can change software, databases, websites and so on without too many constraints or ramifications. Once you get here you realize that managing the PDB is an industrial scale project, full of complexities and dependencies. It has taken me a year just to wrap my head around everything and start to formulate my 5–10 year plan for PDBe.

WAW: Do you foresee any substantial changes in the way PDBe is run over the next 5 years?

GK: A key issue, in my opinion, will be whether or not we should stay an historic archive. Coordinates from a structure deposited in 1976 are still the same today, even though we could automatically improve a 1976 structure into a state-of-the-art model, which might be more useful to biologists. Personally, I think that, while the historic archive must be preserved, we must build from it a user-oriented resource for biomedicine. Crystallographers won't necessarily all agree that this is the way forward: many of them worry about *"my coordinates"*. Crystallographers also don't always like quality indicators about their structures, but a non-expert user wants to know "W*hich of these two kinase structures is the better*?" Scientific and funding constraints mean that we need to serve new user communities: medicinal chemists, geneticists and clinicians, for example, and not just crystallographers and NMR spectroscopists. Cell biologists think in terms of pathways or complexes, not PDB codes. Thus we need to provide new ways for them to access structural information, using terminology and classifications that they are familiar with. The community should discuss these issues and PDBe will certainly work on it. The PDB celebrates its 40th anniversary next year [15], so hopefully we will see new and wider uses of structural data by the time the PDB turns 50.

WAW: How about funding?

GK: We have core funding from EMBL and the Wellcome Trust for about 15 people to do curation and keep the databases and web services up and running. Other funds come from NIH (National Institutes of Health), EU (the European Union) and BBSRC (the UK Biotechnology and Biological Sciences Research Council). NIH funds EMDB in a joint project with RCSB and Baylor College of Medicine. EMBL has a five-year plan so we are not so sensitive to reductions in research funding in the short term. The Wellcome Trust grant runs for 5 years from January 2010. In the future, new developments in scientific content and integration are going to be a major focus of our work. For example, we developed the Electron-Density Server (EDS) in Uppsala [16]: it makes electron-density maps available in a user-friendly way so that modelers and medicinal chemists can visually assess how well a ligand fits the crystallographic data without knowing anything about calculating such maps. It is complete as an academic project and needs to go into production now at PDBe. I also want to put emphasis on validation (you can see my publications on this); it means assessing the quality and reliability of models, and identifying problematic structures, such as ligands with no density.

WAW: I suppose everyone asks you about "StructureGate" [17]. Could it happen again?

GK: First and foremost, we should keep things in perspective. The vast majority of scientists have, in the forty-year lifespan of the PDB, made an honest attempt to deposit accurate data and structure models that are their best possible interpretation of that data. However, we have battled with "honest incompetence" and the best way to combat this is to provide good tools for data interpretation and validation as well as to educate every new generation of students that enters the field of structural biology. Unfortunately, there is now a case under investigation by the ORI (Office of Research Integrity) [18] that appears to involve scientific fraud. It may be possible to be fraudulent and not be found out, but we as database centers have to assume that people are honest.

In the past I have made friends and enemies with my views on this, but nowadays everyone seems interested in proper validation. wwPDB has convened expert validation task forces for X-ray, NMR and EM to advise on the most suitable criteria to use for validating structure entries when they are deposited [19]. The results will be summarized in validation reports that can be submitted alongside a manuscript so that the editors and referees can assess the reliability of the data, the model and the inferences drawn from them. Such reports are already mandatory for structure papers submitted to IUCr (International Union of Crystallography) journals (*Acta Crystallographica D* and *F*) and we expect that a number of high-profile journals will follow soon.

Many of your publications concern the imprecision of crystal structures and in particular the quality of

ligand structure and conformation. What should drug designers look out for as they use crystal complexes of drug-protein interactions? How much should they trust them? How can they judge their reliability?

I think the lessons for modelers described in our DDT paper [20] sum up my findings and recommendations very well. Researchers don't always appreciate the uncertainties introduced during the process of deriving an atomic model from the experimentally observed electron density. Our paper highlights some of those uncertainties with examples from the literature, and it has snippets of advice for medicinal chemists. Here's some of that advice.

Don't assume that the protein structure is necessarily correct. Use global quality indicators such as those provided by WhatCheck [21] or MolProbity [22] to assess the quality of the model as a whole. You might also want to have a look at my validation tutorial [23]. In the future (hopefully from 2012), the wwPDB sites are expected to provide detailed quality statistics for every crystal structure in the PDB (following the recommendations of the wwPDB X-ray Validation Task Force).

Don't assume that the presence, location, orientation and conformation of the ligand and its interactions with the protein are necessarily unambiguously defined or supported by the experimental data. Check if the electron density map is available from the Uppsala Electron-Density Server [16]. If it is, view the ligand and the binding site with the electron density superimposed; if it all fits snugly, it's probably okay. In the future, this type of information may also become available from the wwPDB sites.

WAW: What do you like to do when you are not working in structural bioinformatics?

GK: Apart from the usual suspects, reading and movies, I play football once or twice a week, and a Swedish sport called floorball [24], which is a bit like indoor hockey. However, my first year here has been mostly work; I don't even have a television. I like to visit Sweden and to go on road trips in Scandinavia [25] (or New Zealand), and also to spend time with my precocious 7-year-old (who, among other things, tends to beat me at most card and board games).

WAW: Gerard, thank you for being such a good sport about this interview. I'm sure that our readers are going to find it interesting.

References

1. The Protein Data Bank in Europe (PDBe). http://pdbe.org/. Accessed Aug 2010
2. Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Sousa Da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, Krissinel EB, Newman R, Oldfield T, Pajon A, Penkett CJ, Pineda-Castillo J, Sahni G, Sen S, Slowley R, Suarez-Uruena A, Swaminathan J, van Ginkel G, Vranken WF, Henrick K, Kleywegt GJ (2010) PDBe: Protein Data Bank in Europe. Nucleic Acids Res 38(Database issue):D308–D317
3. Worldwide PDB. http://www.wwpdb.org. Accessed Aug 2010
4. Electron Microscopy Databank. http://emdatabank.org. Accessed Aug 2010
5. Research Collaboratory for Structural Bioinformatics (RCSB). http://www.rcsb.org/. Accessed Aug 2010
6. Warr WA (2008) Future structural genomics initiatives: an interview with Helen Berman, director of the Protein Data Bank. J Comput-Aided Mol Des 22(10):707–710
7. Structure integration with function, taxonomy and sequence (SIFTS) initiative. http://www.ebi.ac.uk/pdbe/docs/sifts/. Accessed Aug 2010
8. Pfam. http://pfam.sanger.ac.uk. Accessed Aug 2010
9. Gene ontology, GO. http://www.geneontology.org/. Accessed Aug 2010
10. Interpro. http://www.ebi.ac.uk/interpro/. Accessed Aug 2010
11. SCOP. http://scop.mrc-lmb.cam.ac.uk/scop/. Accessed Aug 2010
12. CATH. http://www.cathdb.info/. Accessed Aug 2010
13. Japan PDB (PDBj). http://www.pdbj.org/. Accessed Aug 2010
14. BMRB, BioMagResBank, Wisconsin. http://www.bmrb.wisc.edu/. Accessed Aug 2010
15. Berman HM (2008) The Protein Data Bank: a historical perspective. Acta Crystallogr Sect A Found Crystallogr A64(1):88–95
16. Electron-Density Server at Uppsala University (EDS). http://eds.bmc.uu.se/. Accessed Aug 2010
17. StructureGate. http://bytesizebio.net/index.php/2009/12/10/structuregate/. Accessed Aug 2010
18. Office of Research Integrity, ORI. http://ori.dhhs.gov/. Accessed Aug 2010
19. wwPDB statement on retraction of PDB Entries. http://www.wwpdb.org/UAB.html. Accessed Aug 2010
20. Davis AM, St-Gallay SA, Kleywegt GJ (2008) Limitations and lessons in the use of X-ray structural information in drug design. Drug Discov Today 13(19–20):831–841
21. WhatCheck. http://swift.cmbi.ru.nl/gv/whatcheck/. Accessed August 2010
22. MolProbity. http://molprobity.biochem.duke.edu/. Accessed Aug 2010
23. Model validation tutorial. http://xray.bmc.uu.se/embo2001/modval/. Accessed Aug 2010
24. Floorball. http://en.wikipedia.org/wiki/Floorball. Accessed Aug 2010
25. http://xray.bmc.uu.se/smultron/. Accessed Aug 2010