nature structural biology
molecular form & function
september 2001
vol. 8 no. 9

Healthy debate of structural analysis

The Comments section of *Nature Structural Biology* has been going strong this year. We publish material in this section for two main reasons: to allow a forum for discussion of results published in our pages and for announcements and commentary of widespread general interest to the structural biology community. The forum is not meant for lengthy, detailed scientific debate, which we believe should be left to the primary literature. Instead, one of the main functions of the Comments section is to raise awareness of certain issues. In this context, it is interesting to note that the past three pairs of Comments discussing work published in *Nature Structural Biology* have all centered on the analysis of structural data.

In the February issue of the journal, Ng and Dickerson[1], and Vargason and Ho[2] discussed the analysis of certain DNA structures. In the August issue, Rupp and Segelke[3], and Hanson and Stevens[4] debated the interpretation of electron density in a particular protein–peptide complex. And, on pages 744–745 of this issue, Paoli[5], and Burley and colleagues[6] discuss the difficulties involved in finding relatedness among protein structures. All of this correspondence suggests that despite the numerous advances in obtaining structural data, the actual production and analysis of structural models is still not completely straightforward.

This fact has serious implications for structural biology in the context of structural genomics. Those in the field of structural genomics expect that thousands of structures will be churned out, almost automatically, and that the final structures will be as accurate as any solved in a 'traditional' structural biology lab. However, many have expressed concern that the structures emerging from the large structural genomics consortia might be less reliable, given how little time might be devoted to each structure determination and the lack of obvious routes for biochemical follow-up of many of the structures. This is indeed a serious issue — after all, who wants the Protein Data Bank (PDB) to be clogged with inaccurate structures? To head off this problem, the structural genomics community has agreed on some fundamental principles of practice that might make this problem surmountable. However, these guidelines will not address the basic need for better programs to analyze and compare structural data.

At the Second International Structural Genomics Meeting in Airlie, Virginia, USA last April[7], it was agreed that all structural data would be deposited in the PDB in a timely manner for release to the public, and that structure factor files and equivalent NMR data would be deposited as well. This should allow any questionable structure to be examined in detail by interested scientists for accuracy. This is also the current policy of *Nature Structural Biology*. However, these procedures could conceivably be hampered by a 'gray zone' regarding when deposition should occur — when the quality of a structure is good enough to be considered 'complete'. Although some numerical guidelines have been established, most believe that a structure's final quality can be reliably assessed only by an experienced crystallographer or NMR spectroscopist. Given these caveats, we will simply have to wait to see how rapid the data deposition from the structural genomics projects becomes, and how the quality of those structures measures up.

It should be noted that even when more complete data sets for a particular structure are available, not everyone is going to agree on how to analyze the data, as the pairs of Comments mentioned above illustrate. While one might think that everything in structural biology should be black and white, reducible to mathematical principles alone, there still seems to be some degree of art and skill to solving structures and comparing them to other structures in the database. There is clearly much work to be done to make structural data analysis more routine and straightforward, so that everyone always arrives at the same answer to a particular question about the data. We suspect our Comment section will continue to provide a forum for discussions about structural analysis and comparison for some time to come, until standards of practice have improved and become so accepted that they are no longer up for debate.

To make our policies on such discussions clear, the general procedures for handling Comment submissions are outlined here. Consideration of any piece for the Comments section is solely at the discretion of the editors. We receive many submissions to this section, and we proceed only with those we believe address the most topical and important issues. Because space in this section of the journal is limited, priority is given to short (less than 500 words) well-written letters, and typically new data are not presented in this section, although they may occasionally be allowed. Comments concerning material previously published in the journal are often sent to the authors of the original piece for a formal reply, and the material is usually peer reviewed to ensure that the discussion is of broad interest and maintains an appropriate tone of scientific discourse. If, after this procedure, we decide to publish the discussion, the Comments are edited for brevity and clarity. Hopefully, these practices will continue to ensure that many interesting topics, including structural data analysis, will be covered in this section of the journal in future issues.

1. Ng, H.-L. & Dickerson, R.E. *Nature Struct. Biol.* **8**, 107 (2001).
2. Vargason, J.M. & Ho, P.S. *Nature Struct. Biol.* **8**, 107–108 (2001).
3. Segelke, B. & Rupp, B. *Nature Struct. Biol.* **8**, 663–664 (2001).
4. Hanson, M. & Stevens, R. *Nature Struct. Biol.* **8**, 664 (2001).
5. Paoli, M. *Nature Struct. Biol.* **8**, 744 (2001).
6. Groft, C.M., Beckmann, R., Sali, A. & Burley, S.K. *Nature Struct. Biol.* **8**, 745 (2001).
7. http://www.nigms.nih.gov/news/meetings/airlie.html