

W0001

Quality Assurance in Structural Bioinformatics. Philip Bourne, RCSB Protein Data Bank, San Diego Supercomputer Center, University of California-San Diego, 9500 Gilman Dr., MC 0537, La Jolla, CA 92093-0537 USA.

The primary source of all publicly accessible macromolecular structure data is the Protein Data Bank (PDB). A series of secondary resources are available that further process these data and make them available to the community. This processing includes, but is not limited to, classifying structures according to family and fold, the study of protein-protein interactions, use in the prediction of structure from primary sequence and functional classification. This processing combined with the representation and storage of these data as well as visualization comprises a significant part of the young field of structural bioinformatics [1].

This field offers many algorithms and software for their application, but often without rigorous testing and comparison of methods. This testing and comparison is important to crystallographers trying to determine which of a variety of methods best meets their specific needs. We will illustrate the need for such quality assurance using two commonly applied automated methods, structure comparison and alignment and domain recognition from 3-dimensional structure. Structure comparison methods are evaluated by comparison to an accurate hand alignment made over a number of months [2].

Domain analysis is examined by looking for consensus within automated versus manual methods. We find that in 80% of cases there is agreement between the numbers of domains across all methods, while 20% of cases are evenly distributed such that in 10% of cases the experts agree and the automated methods disagree. However, in 10% of cases the automated methods agree and the experts disagree [3]. Domain boundaries present an even greater challenge as will be shown.

[1] P.E. Bourne and H. Weissig (eds) Structural Bioinformatics (2003) John Wiley and Sons NY.

[2] E. Scheeff,(2003) unpublished

[3] S. Verentik, P.E. Bourne, N.N. Alexandrov, I.N. Shindyalov (2003) Towards consistent assignment of structural domains in proteins. Journal of Molecular Biology, submitted.