

Bridging the gap

Sir — In response to your recent editorial¹, I would like to point out what we at the Protein Data Bank (PDB), and many members of the community have been actively doing to help bridge the gap that you allude to, between the structure and genome worlds.

As Steve Bryant at the Natural Center for Biotechnology Information (NCBI) observes, "even experienced sequence analysts can sort of get ill when they look at coordinate files". In general, it is quite unnecessary to look at *x,y,z* coordinates; this would be like looking at a DNA gel to determine a protein sequence. But there is more to a PDB entry than just the coordinates: the header of every PDB file contains a considerable series of annotations including links to entries in sequence databases. However, I agree that easy access to these data in a form relevant and understandable to non-structural biologists is essential. Therefore, over two years ago, we established a World Wide Web (WWW) home page (Table 1) with this in mind, and included a user-friendly front end with search capabilities². Molecular biologists are thus able to see 3D structures with public-domain software such as RasMol³. Whenever a particular PDB entry is accessed through the 3DBbrowser (Fig. 1), one obtains an Atlas page (Table 1) that includes PDB data as well as links to other database and information resources. We also provide access to the informative images of 3D structures created by Manuel Peitsch (Glaxo Institute for Molecular Biology in Geneva), which are also available at the Swiss-Prot Web site (Table 1) Each image depicts a key aspect of the structure or experiment that it represents⁴. In parallel, a number of other groups are contributing to making access to 3D structural information easier for molecular biologists. This can be seen

in the work on integrating reference, sequence and structural information in the Entrez system (Table 1) at the NCBI, in the classification and comparison of structural information as seen in SCOP (Table 1) at the Medical Research Council (MRC), Cambridge, CATH (Table 1) at University College, London and DALI (Table 1) at the European Molecular Biology Laboratory (EMBL), in the analysis of DNA and RNA structures at the Nucleic Acid Database (NDB) (Table 1) at Rutgers University, and in many other projects throughout the world. For a comprehensive listing, see Pedro's BioMolecular Research Tools (Table 1).

PDB has from its inception served the role of providing access to structural data to the non-structural community. This is clearly demonstrated by the early decision to store data in orthogonal Angstrom units rather than in fractional crystallographic coordinates which often are non-orthogonal. Current development work, such as the 3DBbase⁵, is being done in close collaboration with prominent members of the genome community, including the Genome Data Base (GDB), Lawrence Berkeley National Laboratory (LBNL), European Bioinformatics Institute (EBI), and the Weizmann Institute of Science Bioinformatics Unit, highlighting the importance that the PDB puts on the interplay between genomic and structural fields.

One of the primary goals of the PDB is, as you quote me, "validation of the quality of the all structural data [sic]"¹, which is a service of fundamental importance to the community, as it makes it possible for users of this data to have a high level of confidence that the information archived in the PDB is reliable and has been carefully validated both at the PDB and in consultation with the depositors.

Although in the past people in the

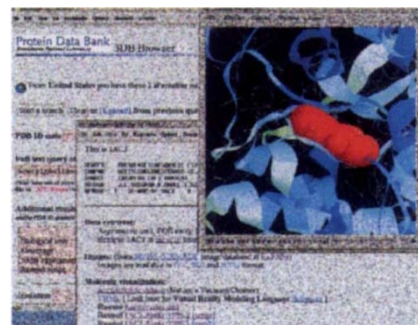


Fig. 1 The 3DBbrowser in action. In the background is the browser screen, with windows to enter search strings in the appropriate categories. At the bottom is the Atlas page for the entry 1ACJ (acetylcholinesterase complexed with the drug tacrine¹⁰). In the upper right, the structure of this complex is seen through the molecular viewer RasMol³. The enzyme is shown as a ribbon diagram with the drug, tacrine, in its active site, represented as a space filling model. The 3DBbrowser was developed by Jaime Prilusky (Weizmann Institute of Science) and Enrique Abola (PDB).

genome field experienced some difficulty in accessing and using structural information, this problem has now been almost completely overcome. In this regard, the PDB together with the Swiss-Prot database, organized an international symposium (Table 1), on 'Bioinformatics and Structure' in Jerusalem, at the end of last year. The meeting was devoted entirely to cross-fertilization between the sequence and structural communities. There was clearly a common language between the two communities with a real excitement in the interchange of ideas. Among others, Steve Bryant, Janet Thornton and Graham Cameron were invited and made excellent presentations (see their published abstracts⁶⁻⁸, and a review of the meeting⁹).

Joel L. Sussman

Head, Brookhaven Protein Data Bank, Professor of Structural Biology, Weizmann Institute of Science
Correspondence should be addressed to J.L.S
email: jls@bnl.gov

Table 1 Web sites mentioned in this letter

Description	URL
PDB Homepage	http://www.pdb.bnl.gov/
3DBbrowser	http://www.pdb.bnl.gov/cgi-bin/pdbmain
Swiss-Prot database	http://www.expasy.ch/
Entrez system	http://www3.ncbi.nlm.nih.gov/Entrez/
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/
DALI	http://roma.ebi.ac.uk/dali/
Nucleic Acid Database	http://ndbserver.rutgers.edu:80/
Pedro's BioMolec Research Tools	http://www.public.iastate.edu/~pedro/research_tools.html
Bioinformatics & Struct Conference	http://www.pdb.bnl.gov/pdb25p10/

1. *Nature Struct. Biol.* **4**, 329-330 (1997).
2. Stampf, D.R., Felder, C.E. & Sussman, J.L. *Nature* **374**, 572-574 (1995).
3. Sayle, R.A. & Milner-White, E.J. *TIBS* **20**, 374-376 (1995).
4. Peitsch, M.C., Stampf, D.R., Wells, T.N.C. & Sussman, J.L. *TIBS* **20**, 82-84 (1995).
5. Sussman, J.L., Abola, E.E., Manning, N.O. & Prilusky, J. in *Crystallogr. Computing 7* (Eds. Bourne, P.E. & Watenpaugh, K.D.), Oxford University Press in the press.
6. Madej, T., Gibrat, J.-F., Hogue, C., Ohkawa, H. & Bryant, S. *Folding & Design 1* (Supplement), S39-S40 (1996).
7. Thornton, J.M. et al. *Folding & Design 1* (Supplement), S63 (1996).
8. Cameron, G. *Folding & Design 1* (Supplement), S14 (1996).
9. Adman, E. *PDB Quart. Newsletter* **79**, 9-11 (1997).
10. Harel, M. et al. *Proc. Natl. Acad. Sci. USA* **90**, 9031-9035 (1993).