Computer Program Abstracts

The category Computer Program Abstracts provides a rapid means of communicating up-to-date information concerning both new programs or systems and significant updates to existing ones. Following normal submission, a Computer Program Abstract will be reviewed by one or two members of the IUCr Commission on Crystallographic Computing. It should not exceed 500 words in length and should follow the standard format given on page 189 of the June 1985 issue of the Journal [J. Appl. Cryst. (1985). ***18****, 189–190] and on the World Wide Web at http://www.iucr.org/journals/jac/software/.*

J. Appl. Cryst. (1999). **32**, 1190

GETENTRY: a simple Unix script for accessing the Protein Data Bank's anonymous ftp server

JOHN P. ROSE,[a]* MICHAEL LIBESON,[b] BI-CHENG WANG[a] AND ENRIQUE E. ABOLA[b]

[a]*Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA, and* [b]*Protein Data Bank, Department of Biology, Brookhaven National Laboratory, Upton, NY 11973, USA*

(*Received 11 January 1999; accepted 18 May 1999*)

The crystallographic problem: The Protein Data Bank's (Bernstein *et al.*, 1977; Abola *et al.*, 1987) archival computer database of three-dimensional structures of biological macromolecules is a unique resource serving the international scientific community. The database contains atomic coordinates, bibliographic citations, and primary sequence and secondary structure information, as well as crystallographic structure factors and two-dimensional-NMR experimental data. Information is available on protein, DNA, RNA, virus and carbohydrate structures. The current release of the Protein Data Bank (PDB) contains atomic coordinate entries for over 8800 structures with about 170–200 new coordinate sets being added to the archive monthly.

In 1992, the PDB established an anonymous ftp server (pdb.pdb.bnl.gov, 130.199.146.1) giving users access to the current archive *via* the Internet. Network access to the archive has several advantages: user access to the archive is free, new Layer 1 coordinate entries become available to the user immediately upon approval by the depositor, and disk and other hardware requirements are minimized. However, as the size of the PDB has grown, finding a particular entry among the thousands of entries in the PDB has become increasingly difficult. The PDB World Wide Web (WWW) browsers, 3DB-Browser and PDB-Lite, have proven to be powerful tools that allow most users to search the archive using various text fields. However, there remain a number of users requiring a more basic interface to the PDB files, particularly *via* VT-type terminals.

Method of solution: A simple user-friendly command-line Unix shell script, *GETENTRY*, has been developed to search the PDB ftp archive and retrieve coordinate entries. It has been in use at the PDB since early 1993 and unlike the WWW-based browsers (Peitsch *et al.*, 1995; Stampf *et al.*, 1995; Biggs *et al.*, 1996), *GETENTRY* can be run on any terminal or window connected to a Unix host.

The following options are available.

Entry retrieval by ID code (`getentry` *ID code*).

Author search (`getentry -a` *author name*).

Compound search (`getentry -c` *compound name*).

Source search (`getentry -s` *source name*).

Resolution search (`getentry -r` [`<, <=, ==, >=, >`] *resolution*).

Download the list of pending or on hold entries (`getentry -p`).

Search for crystallographer's e-mail address (`getentry -x` *name*).

Retrieve BIOMOL entries (`getentry -b` *ID code*).

Help text (`getentry -h`).

Software environment: *GETENTRY* is command-line driven and uses standard Unix commands such as `echo`, `ftp`, `pipe`, `awk`, `more` and `egrep`. The script uses the PDB author, compound, source and resolution indices as well as the PDB-maintained e-mail list of Martha Teeter in its pattern searches. Pattern searches are performed using full regular expressions (*i.e.* the full set of alphanumeric and special characters). *GETENTRY* is fast, with search/transfer times ranging from 10 to 60 s (depending on the server location and Internet load) for most X-ray coordinate entries. *GETENTRY* requires that the user has a `.netrc` file with the correct protection (700) resident in his/her home directory and that the terminal, computer or workstation has access to the Internet. In addition, the script can be easily tailored to use local PDB file servers.

Hardware environment: Unlike WWW-based browsers, hardware and installation requirements are minimal since *GETENTRY* will run on almost any terminal or PC connected to a Unix workstation.

Availability: The script is available at no charge from the PDB anonymous ftp server or from the PDB WWW server (http://www.pdb.bnl.gov/pub-docs/software.html).

Keywords: Protein Data Bank; ftp; Unix script; data retrieval.

References

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzel, T. F. & Weng, J. (1987). *Protein Data Bank* in *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107–132. Bonn: Data Commission of the International Union of Crystallography.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Biggs, J., Pu, C., Groeninger, A. & Bourne, P. E. (1996). *J. Appl. Cryst.* **29**, 484–490.

Peitsch, M. C., Wells, T. N. C., Stampf, D. R. & Sussman, J. L. (1995). *TIBES*, **20**, 82–84.

Stampf, D. R., Felder, C. E. & Sussman, J. L. (1995). *Nature (London)*, **374**, 572–574.

J. Appl. Cryst. (1999). **32**, 1190

Crystallographers

The American Crystallographic Association (ACA) has presented its A. L. Patterson Award to **Gérard Bricogne**, Howard Hughes International Research Scholar at MRC Cambridge, in England, and Director of Research at LURE in Orsay, France, for 'his fundamental analysis of structure factor statistics and implementation of optimal computational algorithms to update, sample, and evaluate accurate joint probability distributions of structure factors . . . '.

The ACA has presented its 1999 Elizabeth Wood Science Writing Award to **Robert A. Weinberg**, the Daniel K. Ludwig Professor for Cancer Research at MIT, for his writing for a general audience.

The International Centre for Diffraction Data (ICDD) has elected Professor Dr **Walter Eysel** to its list of Distinguished Fellows, for his 'sustained, outstanding work in the field of powder diffraction'.