

# *PDB Goodies* – a web-based GUI to manipulate the Protein Data Bank file

A. S. Z. Hussain,† V. Shanthi,†  
S. S. Sheik, J. Jeyakanthan,  
P. Selvarani and K. Sekar\*

Bioinformatics Centre, Supercomputer  
Education and Research Centre, Indian Institute  
of Science, Bangalore 560 012, India

† These authors made equal contributions to  
this work.

Correspondence e-mail:  
sekar@physics.iisc.ernet.in

*PDB Goodies* is a web-based graphical user interface (GUI) to manipulate the Protein Data Bank file containing the three-dimensional atomic coordinates of protein structures. The program also allows users to save the manipulated three-dimensional atomic coordinate file on their local client system. These fragments are used in various stages of structure elucidation and analysis. This software is incorporated with all the three-dimensional protein structures available in the Protein Data Bank, which presently holds approximately 18 000 structures. In addition, this program works on a three-dimensional atomic coordinate file (Protein Data Bank format) uploaded from the client machine. The program is written using CGI/PERL scripts and is platform independent. The program *PDB Goodies* can be accessed over the World Wide Web at <http://144.16.71.11/pdbgoodies/>.

Received 27 March 2002  
Accepted 31 May 2002

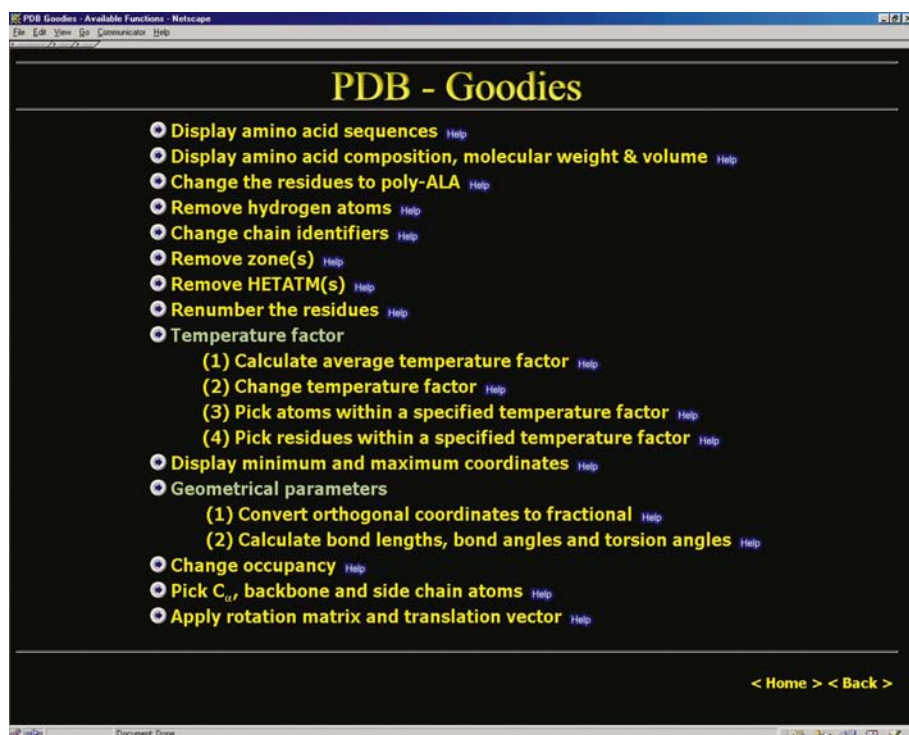
## 1. Introduction

In the present structural genomics era, advanced tools are essential to facilitate the retrieval of useful information from the wealth of data available in the Protein Data Bank public repository. The Protein Data Bank (PDB; Bernstein *et al.*, 1977; Berman *et al.*, 2000) is an archive of three-dimensional biological macromolecular structures determined experimentally using well known techniques such as X-ray crystallography and NMR spectroscopy (mostly by the former). At present, the Protein Data Bank is maintained by the Research Collaboratory for Structural Bioinformatics (RCSB). The three-dimensional structures in the Protein Data Bank are constantly being used by many researchers all over the world and provide a wealth of information towards better understanding of underlying structure–function relationships. Towards this effort, researchers need to use relevant information from the protein structures available in the PDB. These modified segments are used at various stages; for example, as search models for molecular-replacement calculations, in comparison of various secondary-structural elements in the protein model, interactions of the motifs at the interface, protein–protein interactions *etc.* To the best of our knowledge, there is no general program available to manipulate various parts of a PDB file containing the three-dimensional atomic coordinates. In order to perform the

above with less human intervention, a user-friendly graphics user interface (GUI), a World Wide Web-based package *PDB Goodies*, has been developed. On the whole, the proposed package acts as a World Wide Web dissemination of the collated data available in the public domain Protein Data Bank and provides a hardware- and software-independent front end to the database, making the database virtually independent of the client accessing system.

## 2. Discussion

This software is fully incorporated with all the available protein structures solved using X-ray crystallography and NMR spectroscopy. All the three-dimensional structures in the Protein Data Bank (Bernstein *et al.*, 1977; Berman *et al.*, 2000) are also available from the Bioinformatics anonymous PDB FTP server (Indian Institute of Science, Bangalore, India). For the structures available in the PDB, the user needs to provide only the four-character PDB code. In addition, the user has the option of uploading atomic coordinates of interest from the client machine through the World Wide Web browser by clicking the button 'browse' available in the package. For structures determined using NMR spectroscopy, the software has an additional option to choose the model number to perform the required function. The model number can be selected from



**Figure 1**

The web page of *PDB Goodies* showing the available functions. Clicking on the appropriate function activates the search engine.

the pull-down menu available on the input page. The software has the following functions (Fig. 1) to manipulate the three-dimensional atomic coordinate files as required by chemists, biologists or crystallographers.

- (i) Display amino-acid sequences.
- (ii) Display amino-acid composition, molecular weight and volume.
- (iii) Change the residues to poly-Ala.
- (iv) Remove hydrogen atoms.
- (v) Change chain identifiers.
- (vi) Remove zone(s).
- (vii) Remove HETATM(s).
- (viii) Renumber the residues.
- (ix) Temperature factor:
  - (1) Calculate average temperature factor.
  - (2) Change temperature factor.
  - (3) Pick atoms within a specified temperature factor.
  - (4) Pick residues within a specified temperature factor.
- (x) Display minimum and maximum

coordinates.

- (xi) Geometrical parameters.
  - (1) Convert orthogonal coordinates to fractional.
  - (2) Calculate bond lengths, bond angles and torsion angles.
- (xii) Change occupancy.
- (xiii) Pick  $C_{\alpha}$  backbone and side-chain atoms.
- (xiv) Apply rotation matrix and translation vector.

For most of the functions, the software firstly culls the user-inputted PDB file and displays the chain(s) information along with the number of residues involved in each chain. The user has the further option to choose the entire file, a particular chain or a part of a chain to perform the task of interest. Users can save the displayed output onto their local system by clicking the button 'save to disk' provided on the page. The three-dimensional atomic coordinates are updated every week and hence users can

access all structures available in the Protein Data Bank. In trial runs, results appear in about 5–10 s depending upon the nature of the function selected and the network speed. Additional functions will be made available in accordance with progress and based on requests from the scientific and user communities.

### 3. Program implementation

The WWW-based program *PDB Goodies* is written in PERL programming language (CGI/PERL). The front-end input data part and the pull-down menu of this program are written in HTML and JavaScript. The program runs on our Linux server (a 1.5 GHz Pentium IV processor; 1 Gb of main memory). The program *PDB Goodies* is self-explanatory, easy to use and runs on Windows 95/98/2000, Windows NT, Linux and Silicon graphics (SGI) platforms through the browser Netscape. The program is completely general and will work for any protein structures available in the PDB. Help information is provided for all the utilities available in this package. The program is freely accessible without restriction for academic research users over the World Wide Web at the URL <http://144.16.71.11/pdbgoodies/>. Please send any comments and suggestions for the inclusion of additional options to Dr K. Sekar ([sekar@physics.iisc.ernet.in](mailto:sekar@physics.iisc.ernet.in)).

The authors gratefully acknowledge the use of the Bioinformatics Centre (DIC), the Interactive Graphics Based Molecular Modelling (IGBMM) and the Super-computer Education and Research Centre (SERC). The facilities DIC and IGBMM are supported by the Department of Biotechnology (DBT), Government of India.

### References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids. Res.* **28**, 235–242.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. J. (1977). *J. Mol. Biol.* **112**, 535–542.