Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# LySDB – Lysozyme Structural DataBase

**K. Samaya Mohan,[a]‡ Soma Das,[a]‡ C. Chockalingam,[a] V. Shanthi[a] and K. Sekar[a,b]***

[a]Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India, and [b]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

Received 15 September 2003
Accepted 18 December 2003

LySDB (Lysozyme Structural DataBase) is an integrated database containing 740 three-dimensional structures of lysozyme available in the Protein Data Bank. The database can be used to visualize the three-dimensional structure of the entire protein model or the substructures in which the user is interested (for example, insertions and deletions of amino acids) using the three-dimensional atomic coordinates. The database is provided with a search engine with several useful built-in facilities. The public domain graphics program *RASMOL* has been deployed for visualization. The three-dimensional structures used to create the database are updated at regular intervals and hence the users are provided with the current information available in the literature. The database LySDB is available over the World Wide Web and can be accessed at the URL http://iris.physics.iisc.ernet.in/lysdb/ or http://144.16.71.2/lysdb/.

## 1. Introduction

Lysozyme is considered as the model enzyme in biology and is found in a variety of organisms ranging from viruses to higher organisms. The enzyme lysozyme is an $\alpha + \beta$ protein with helical $\beta$-sheet domains and is found in mammalian secretions and secretory organs. There are different types of lysozymes designated A to G and they belong to the lactalbumin/lysozyme family (Bachali *et al.*, 2002). In addition, various types of lysozyme structures have been reported from different organisms, *e.g.* lysozyme c (chicken), g (goose) and i (invertebrates) and those from phages, bacteria, plants *etc.* (McKenzie, 1996). Plant lysozyme, which is found in ficus and papaya latex, is chemically distinct from the egg-white enzyme.

Lysozyme or muramidase (peptidoglycan *N*-acetylmuramoylhydrolase) possesses antibacterial activity against a number of bacteria. It preferentially hydrolyzes the $\beta$-1,4-glucosidic linkages between *N*-acetylmuramic acid and 2-acetamido-2-deoxy-D-glucose residues in peptidoglycan heteropolymers (mucopolysaccharide or muropeptide) of the bacterial cell wall, suggesting that these enzymes play a major role in the body's defence mechanisms. Lysozyme has also been demonstrated to be a mediator in the anti-tumour function of macrophages. The antibacterial specificity of lysozyme is directed against certain Gram-positive bacteria and to a lesser extent against Gram-negative bacteria by disrupting their cell walls (Pabo, 1987).

It is interesting to note that lysozyme was the first enzyme for which the three-dimensional structure was determined at a molecular level (Blake *et al.*, 1965). This enzyme has a folded structure typical of bioactive proteins. All lysozymes perform a similar enzymatic function and have an overall similarity in their three-dimensional structures.

To the best of our knowledge, there is no separate value-added web-based database unique to lysozyme structures. Thus, to improve comparison and analysis, we have created the LySDB by assimilating all the known three-dimensional structures of lysozymes available in the Protein Data Bank (PDB; Bernstein *et al.*, 1977; Berman *et al.*, 2000). The main aim of this database is to collate all the lysozyme structures in one place and to provide links to various available easy-to-use interfaces for structure analysis. In addition, this software is also useful for a wide range of tasks required by those working in the area of structural bioinformatics (see below for details).

## 2. Utilities of LySDB

The data in the database are organized in a format appropriate for quick searching. The major options provided in the search engine associated with the database are (i) 'Protein details' and (ii) 'Search'. The option 'Protein Details' lists various structures with a brief description and the number of available structures associated with the protein. The output frame shows the experimental details, refinement details, space group, unit-cell parameters and amino-acid sequence for the protein structure in which the user is interested in a convenient tabular format. The output frame also displays the secondary-structural

elements (as available in the corresponding PDB file) present in the protein structure.

Furthermore, with the links provided, the user can visualize the three-dimensional structure of the entire molecule or the region of interest (regions in which deletions or insertions of amino-acid residues are made) or the secondary-structural elements of the protein molecule in order to better understand and to unravel the role of individual amino acids at the molecular level. This can be invoked by clicking on the button 'View in Rasmol'. The resultant pop-up window provides several options to the users; the free graphics program *RASMOL* (Sayle & Milner-White, 1995) is used for visualization. In addition to viewing the molecule, users can also extract the three-dimensional atomic coordinates on the client machine. To invoke the *RASMOL* graphics display on the client machine, the user needs to interface the graphics program with the Netscape browser (only during initial use; see instructions at http://144.16.71.2/lysdb/rasmol.html).


**Figure 1**
A display panel showing a sample output frame of a typical search for hydrolase (O-glycosyl; PDB 1l64).

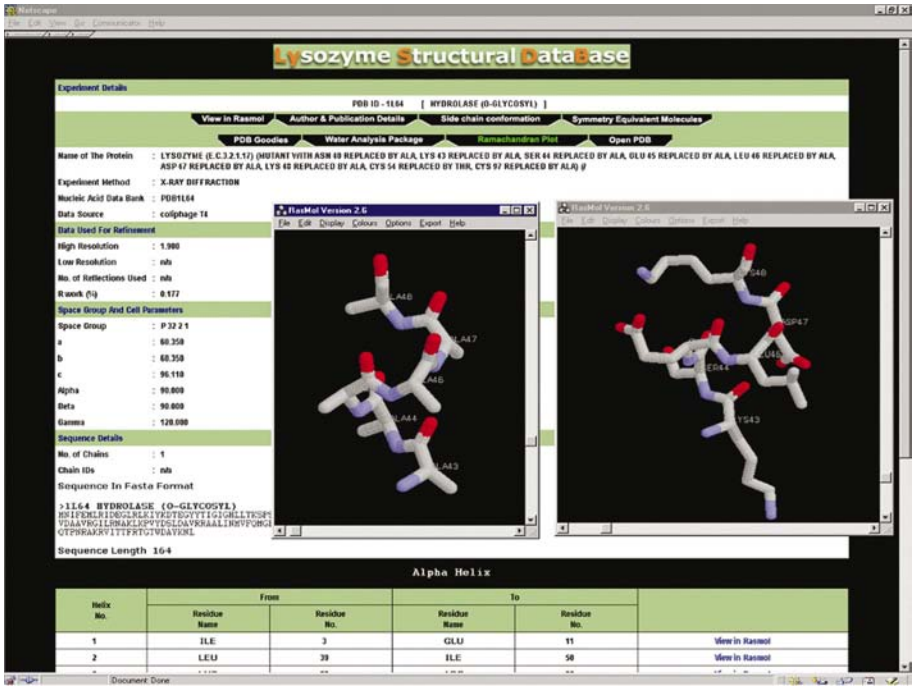


**Figure 2**
A screen shot of the output page, showing the three-dimensional structure of the mutated region (left-hand *RASMOL* panel), residues 43–48, in the mutated protein molecule (PDB code1l64) and the corresponding region of the native protein (right-hand *RASMOL* panel) (PDB code 3lzm).

To enhance the utility of the database, the following user-friendly structure-analysis and validation packages have been linked to the proposed database. The Ramachandran plot package (*RP*) is a World Wide Web-based graphics package that displays the main-chain torsion angles for a particular amino acid or all amino acids present in the protein structure (Sheik *et al.*, 2002). The side-chain conformation angles package (*CAP*) displays all the side-chain conformation angles for a particular amino-acid residue in a selected protein structure (Sheik *et al.*, 2003). The graphics packages described above (*RP* and *CAP*) have several built-in utilities for user convenience. *PDB Goodies* (Hussain *et al.*, 2002) is a web-based software package that manipulates and cuts the three-dimensional atomic coordinates of the protein molecule. The internet computing package *WAP* (Shanthi *et al.*, 2003) is used to calculate the distances and angles between the water O atoms and the polar and non-polar atoms of the protein molecule. The package *SEM* (Symmetry Equivalent Molecules) has been incorporated. This is used to generate symmetry-equivalent molecules so that users can visualize the unit-cell packing (Hussain *et al.*, 2003). The above packages provide detailed information about the protein and its environment in order to unravel the role of a potential amino acid in the structure, particularly in regions where insertions and deletions are made. In addition, the database has several search facilities available to users. Searching can be performed using any of the following keywords: PDB code, name of the protein, protein sequence, experimental method, author list, title of the article, journal name, space group and data source. These options are straightforward and easy to use. In addition, the user can search for a particular identical pattern using *RASMOL*. Using the dedicated pop-up window, the user can visualize the location of the selected pattern in the entire molecule (Selvarani *et al.*, 2004).

The search engine is written using CGI/PERL scripts and runs on an Intel Solaris (3.06 GHz Pentium IV processor, 1 Gb of

random access memory). The front-end input data part of this tool is written in HTML and Dynamic HTML. The user-friendly web forms were tested using Java script. The search engine is very easy to use and was tested on Windows 95/98/2000, Windows NT, Linux and Silicon Graphics (SGI) platforms through the most popular web browser Netscape. The described database is freely available over the World Wide Web at http://iris.physics.iisc.ernet.in/lysdb/ or http://144.16.71.2/lysdb/.

Overall, the LySDB database provides a centralized resource for those interested in studies related to the three-dimensional lysozyme structures available in the Protein Data Bank. The proposed database and the associated search engine permits the users to analyze the three-dimensional structures of lysozyme from various species. In addition, the database assures a quick search based on the user choice using the WWW interface. The lysozyme database will be updated from the primary Protein Data Bank (Research Collaboratory for Structural Bioinformatics, RCSB) server at regular intervals and hence the results obtained correspond to the most recent information. In trial runs, the output page appeared in about 5–10 s; however, the speed depends on the nature of the query and the network traffic.


**Figure 3**
A screen shot of the browser output page, displaying the conformation angles (side chain and main chain) of the mutant protein (PDB code 1l64). The highlighted residue in the left-hand panel (side-chain conformation angles plot) corresponds to Ser136. The main-chain conformation angle ($\varphi$, $\psi$) of this residue is highlighted in the right-hand panel (Ramachandran plot).

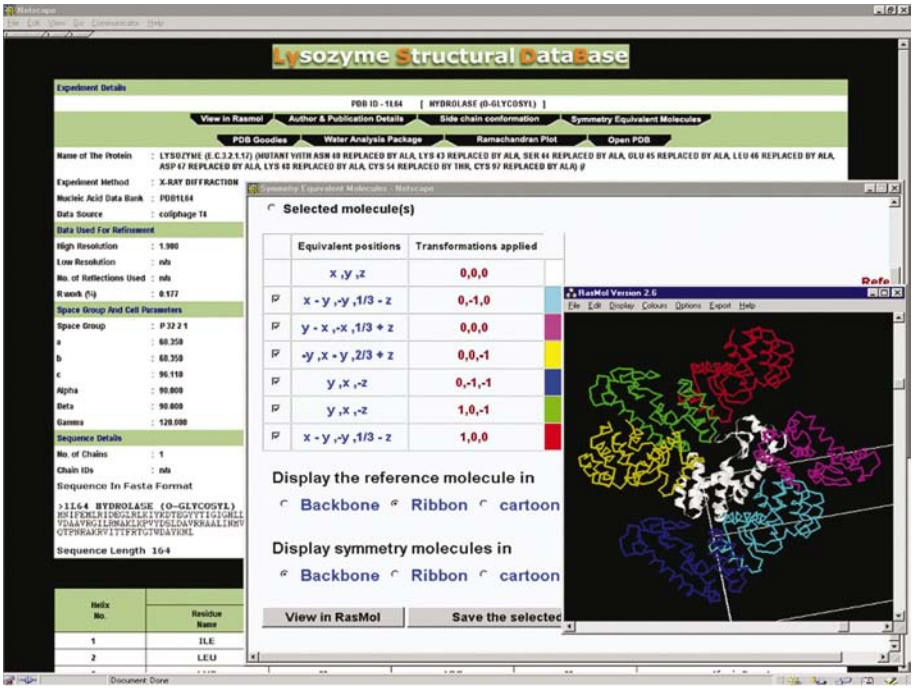


**Figure 4**
A screen shot of the page displaying the symmetry-equivalent molecules along with the reference molecule of the mutated protein (PDB code 1l64). The white-coloured molecule corresponds to the reference molecule and the corresponding symmetry-related molecules are shown in different colours.

## 3. Case study

A sample output (only part of the frame is shown) of the result of a typical search for a lysozyme, hydrolase (O-glycosyl; PDB code 1l64), is shown in Fig. 1. As can be seen in Fig. 1, the frame has links to various packages. As stated above, various structural features can be seen by clicking the appropriate option provided in the page (see the top panel of Fig. 1 for details). The graphics display of the mutated protein (PDB code 1l64) along with the native protein is shown in Fig. 2. The left-hand *RASMOL* graphics panel in Fig. 2 displays the structure of the mutated region (residues 43–48); the corresponding region in the native protein is shown in the right-hand *RASMOL* graphics panel. The Newman projection on the left in Fig. 3 shows the side-chain conformation angles of all serine residues present in the mutated protein and the Ramachandran plot on the right shows the main-chain conformation angles of all the residues of the mutated protein. The highlighted conformation angles in the graphics panel of Fig. 3 correspond to Ser136. The *RASMOL* graphics panel in Fig. 4 displays the arrangement of the symmetry-equivalent molecules of the structure (PDB code 1l64) along with the reference molecule.

The authors gratefully acknowledge the use of the Bioinformatics Centre (DIC), Interactive Graphics Based Molecular Modelling (IGBMM) facility and the Supercomputer Education and Research Centre (SERC). The DIC and IGBMM facilities are supported by the Department

of Biotechnology (DBT), Government of India. KS thanks the DBT for financial support in the form of an individual project.

## References

Bachali, S., Jager, M., Hassanin, A., Schoentgen, F., Jollès, P., Fiala-Medioni, A. & Deutsch, J. S. (2002). *J. Mol. Evol.* **54**, 652–664.

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

Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–761.

Hussain, A. S. Z., Kiran Kumar, Ch., Rajesh, C. K., Sheik, S. S. & Sekar, K. (2003). *Nucleic Acids Res.* **31**, 3404–3405.

Hussain, A. S. Z., Shanthi, V., Sheik, S. S., Jeyakanthan, J., Selvarani, P. & Sekar, K. (2002). *Acta Cryst.* D**58**, 1385–1386.

McKenzie, H. A. (1996). *Lysozymes: Model Enzymes in Biochemistry and Biology*, edited by P. Jollès, pp. 365–409. Basel: Birkhauser.

Pabo, C. O. (1987). *Nature (London)*, **327**, 467.

Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374–382.

Selvarani, P., Shanthi, V., Rajesh, C. K., Saravanan, S. & Sekar, K. (2004). Submitted.

Shanthi, V., Rajesh, C. K., Jayalaksmi, J., Vijay, V. G. & Sekar, K. (2003). *J. Appl. Cryst.* **36**, 167–168.

Sheik, S. S., Sundararajan, P., Hussain, A. S. Z. & Sekar, K. (2002). *Bioinformatics*, **18**, 1548–1549.

Sheik, S. S., Sundararajan, P., Shanthi, V. & Sekar, K. (2003). *Bioinformatics*, **19**, 1043–1044.