

# 3dSS: 3D structural superposition

K. Sumathi<sup>1</sup>, P. Ananthalakshmi<sup>1</sup>, M. N. A. Md. Roshan<sup>1</sup> and K. Sekar<sup>1,2,\*</sup>

<sup>1</sup>Bioinformatics Centre and <sup>2</sup>Supercomputer Education and Research Centre,  
Indian Institute of Science, Bangalore 560 012, India

Received October 6, 2005; Revised and Accepted November 28, 2005

## ABSTRACT

**3dSS is a web-based interactive computing server, primarily designed to aid researchers, to superpose two or several 3D protein structures. In addition, the server can be effectively used to find the invariant and common water molecules present in the superposed homologous protein structures. The molecular visualization tool RASMOL is interfaced with the server to visualize the superposed 3D structures with the water molecules (invariant or common) in the client machine. Furthermore, an option is provided to save the superposed 3D atomic coordinates in the client machine. To perform the above, users need to enter Protein Data Bank (PDB)-id(s) or upload the atomic coordinates in PDB format. This server uses a locally maintained PDB anonymous FTP server that is being updated weekly. This program can be accessed through our Bioinformatics web server at the URL <http://cluster.physics.iisc.ernet.in/3dss/> or <http://10.188.1.15/3dss/>.**

## INTRODUCTION

In the post-genome era, the structural and conformational properties of the 3D protein molecules are of considerable interest owing to its importance in various biological processes. Owing to the recent technological advances like high power tunable synchrotron radiation, powerful number crunching computers and due to ambitious structural genomics programs in different parts of the world, there has been a tremendous increase in the number of protein structures in the Protein Data Bank (PDB) (1). Now there are ~34 000 3D structures available in this entity. Analysis of the 3D structure of protein molecules is greatly enhanced by understanding the relationship between the individual protein molecules. Furthermore, knowledge of the 3D structural relationship between different protein molecules is a key issue in

understanding the structure and function. In order to find the common structural region, one need to lay one molecule over the other by appropriate rotation and translation and this process is termed as superposition of the 3D structures. Several programs are available in the literature (2–9) for this purpose. Most of these programs are stand-alone versions and have their own merits and demerits. Two most recent ones are web-based servers, namely, SSM (8) and SuperPose (9). The program SSM uses the procedure of matching graphs generated using the secondary structural elements followed by the alignment of C $\alpha$  atoms of the protein molecule. Using one of the programs (9), SuperPose, we experienced problems while trying to superpose multiple structures as well as portions of molecules. In fact, it was difficult to superpose different subunits available in multi-subunit protein structures. In addition, most of the existing programs use only the first model of the ensemble in the case of structures solved using NMR technique and there is no provision for the users to superpose all the models in the ensemble.

It is well known that water molecules play a vital role in protein structures, aiding in stabilizing the protein fold and in ligand design (10–14). In addition, investigations on the invariant water molecules in several well studied homologous protein structures shed light on the specific roles of water molecules such as catalytic, structural and functional (15–18). Thus, it is necessary to find the invariant and common water molecules (for definition see below) in homologous protein structures, for which 3D structural superposition step is crucial. But the existing programs do not have provisions for the users to identify the invariant and common water molecules. Hence, we created a unique computing server to superpose the three-dimensional structures and to find the invariant and common water molecules in homologous protein structures.

## BACKGROUND

The water molecules present in two highly similar (the best example is native structure and its mutant structure) or highly

\*To whom correspondence should be addressed. Tel: +91 080 23601409; Fax: +91 080 23600085; Email: [sekar@physics.iisc.ernet.in](mailto:sekar@physics.iisc.ernet.in), [sekar@serc.iisc.ernet.in](mailto:sekar@serc.iisc.ernet.in)

This work is dedicated to the late professor M. Sundaralingam

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact [journals.permissions@oxfordjournals.org](mailto:journals.permissions@oxfordjournals.org)

homologous structures (the inhibitor free and inhibitor bound structure) are known as invariant water molecules. Further, such situation is also possible in multi-subunit protein structures. For example, if a molecule has four identical subunits, the water molecules that interact with the residues in the same position in different subunits (e.g. subunit A and B) can be considered as invariant water molecules. On the other hand, common water molecules are those, which lie at the interface and interact with the selected subunits.

In the computing server, two widely recognized programs STAMP (19) and ProFit (A. C. R. Martin, <http://www.bioinf.org.uk/software/profit/>) are deployed for superposition purposes. The program STAMP uses multiple sequence alignment using the amino acid sequence information followed by an initial superposition of structures. In contrast, the program ProFit uses the McLachlan fitting algorithm, essentially a steepest descent minimization (3). The user-friendly molecular visualization tool RASMOL (20) is interfaced to view the superposed molecules in the client machine. This server is developed using PERL, HTML and JAVASCRIPT. Ploticus [Copy right 1998–2002, Stephan C. Grugg (scg@jax.org)], a data display engine is used for generating plots to display root mean square deviation (r.m.s.d.) graphically.

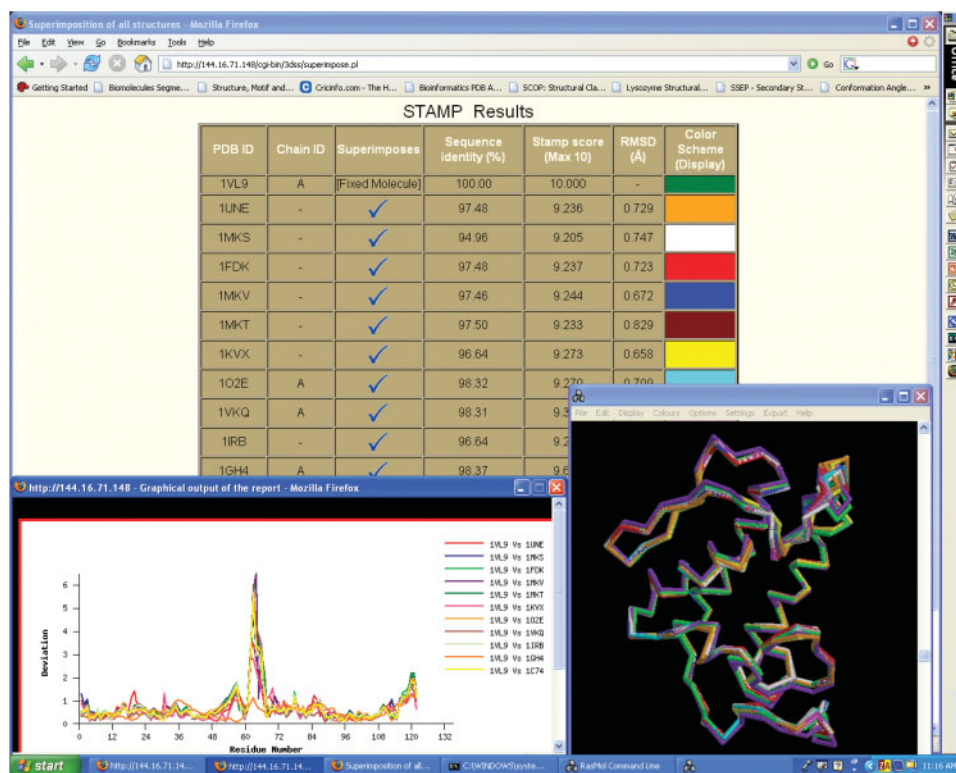
## DATA PRESENTATION AND AVAILABILITY

The software is developed and optimized for Intel based Solaris (Version 10.0) and is driven by 3.0 GHz pentium IV

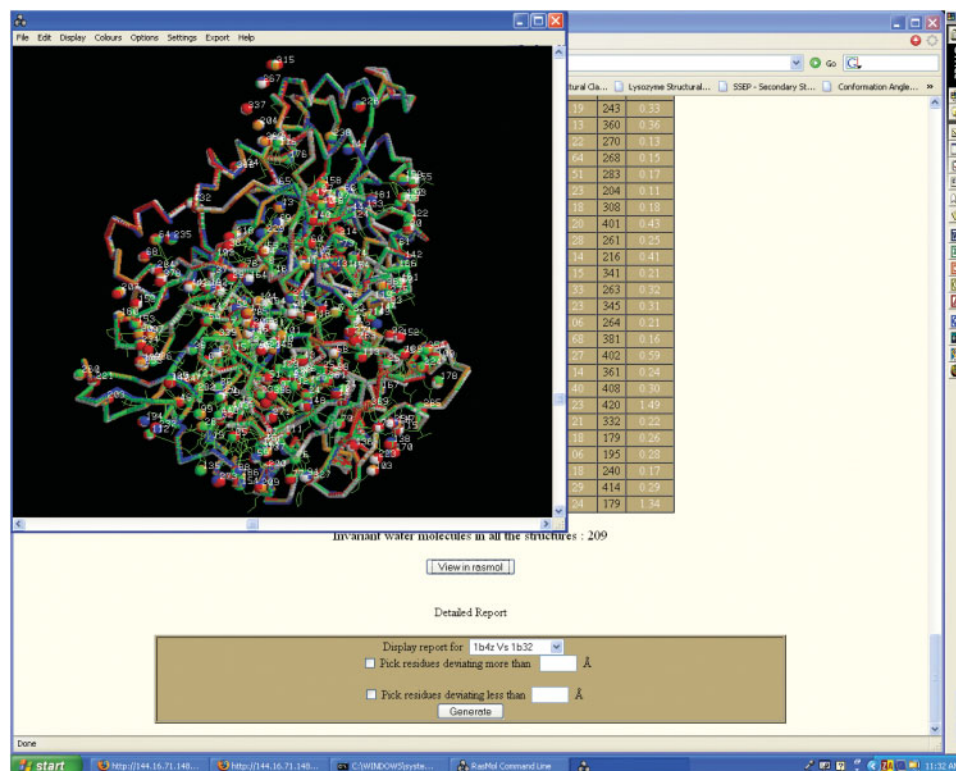
processor equipped with 2 GB RD RAM. This operating system is chosen for better security, scalability and reliability. The software and its functionalities are well tested on Windows 95/98/2000, Linux and SGI platforms. During validation of the software, we realized that two web browsers, namely, NETSCAPE (version 4.7 and 7.2) and MOZILLA behaved well. To visualize the superposed 3D structures in the client machine, user needs to interface the molecular visualization tool RASMOL (only for the first time usage of the software) and the necessary instructions are provided in the link (<http://cluster.physics.iisc.ernet.in/3dss/rasmol.html>). The following are the four major options provided in the proposed computing server.

- Superpose only two structures,
- Superpose several structures,
- Superpose subunits within a structure, and
- Superpose different models present in NMR ensemble.

All the above options, allow users to select the structures available in the PDB by providing its unique PDB-id or by up-loading the 3D atomic coordinates (PDB format) from the local hard disk of the client machine. Once the file is uploaded, the program automatically culls the input PDB file and displays all the chain details of the structure in a convenient form. Using the check box, users can select the entire file, a particular chain or a portion of the chain(s) for superposition. For the option (b), firstly the user needs to provide the number of molecules to be superposed on the fixed molecule.



**Figure 1.** The screen snapshot shows the superposition of 12 structures of recombinant phospholipase A<sub>2</sub>. The top panel shows the status of superposition and the right RASMOL graphics panel displays the superposition in different colors (see the last column of the top panel for coloring scheme). The bottom left panel shows the graphical display of the r.m.s.d. values of the 12 structures and is generated using the data display engine, Ploticus. It is clear from the plot that the region 60–70 is having large deviations compared with the remaining portion of the molecule.

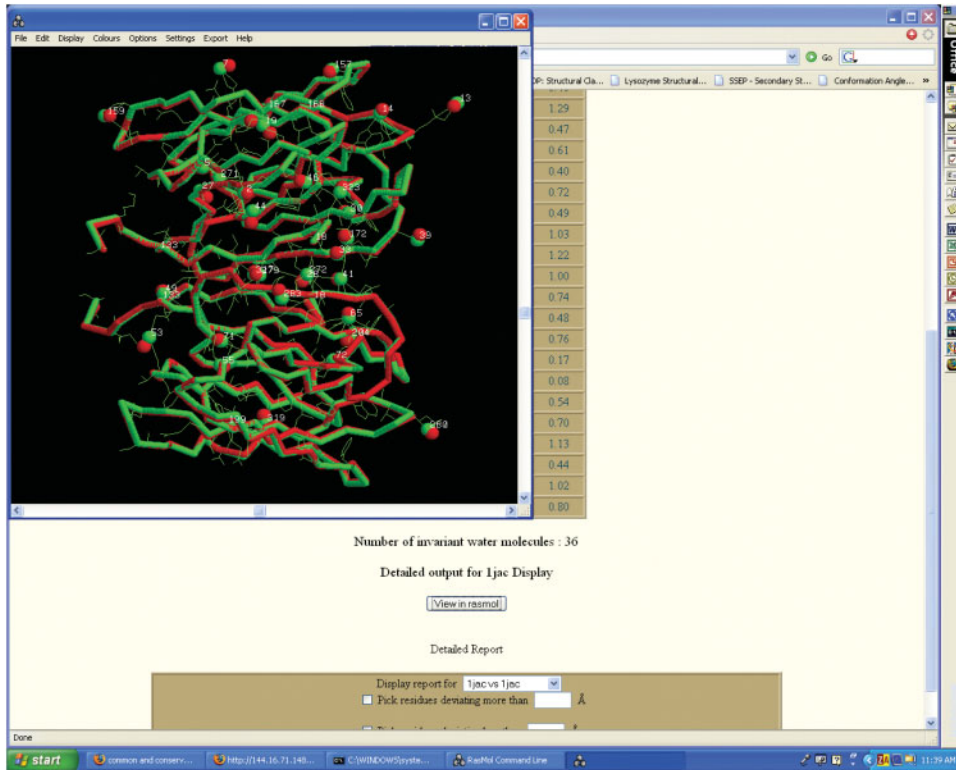


**Figure 2.** The screen shot displays the superposition of six OppA along with 209 invariant water molecules. This is carried out using the option (b), ‘Superpose several structures’ and ‘Superpose and identify invariant water molecules’.

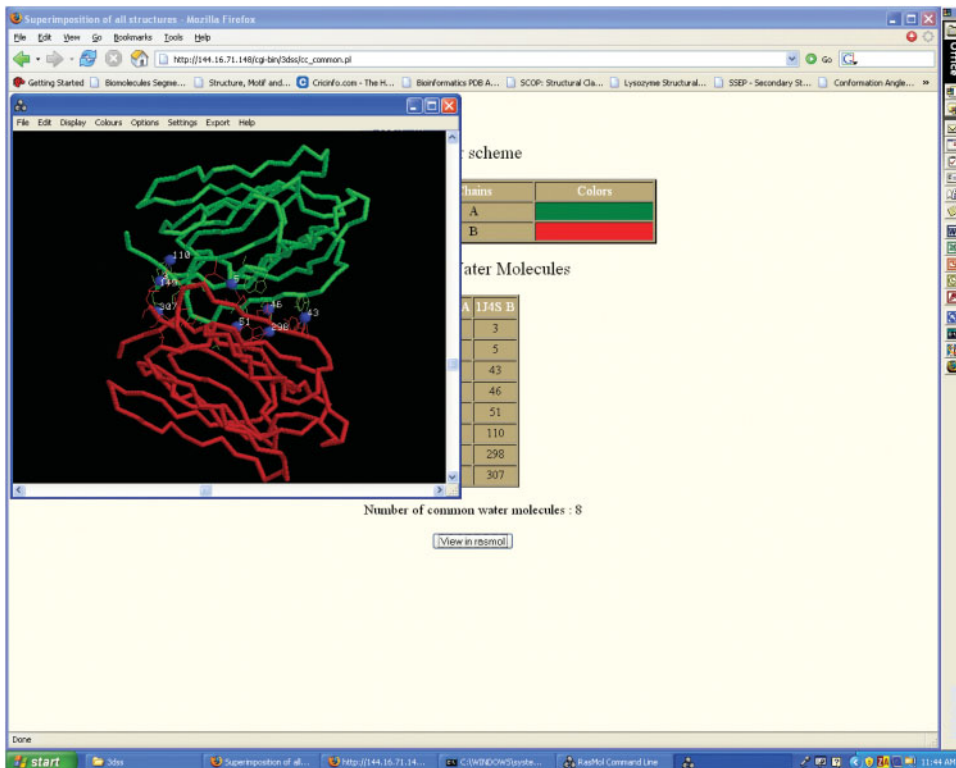
Based on this number, provisions will be available to the user to either supply the PDB-id’s or upload the 3D atomic coordinates from the client machine. By default, the server produces only the structural superposition output. It is worth mentioning that necessary check boxes are provided in the options (a) and (b) to find the invariant water molecules present in the structures. Owing to computational complexity, the number of structures to be superposed on the fixed molecule is limited to 20 at any given time. For option (c), the molecule needs to contain more than one copy of the same polypeptide chain. Using this option, the users can perform three different calculations: (i) superpose different subunits present in a selected structure, (ii) superpose and identify the invariant water molecules and (iii) identify the common water molecules. The option (d) performs structural superposition of various models present in a NMR ensemble and the user can select the models of interest. Here again, the number of mobile molecules is limited to 20 for superposition. In the first three major options, the server displays all models of NMR structure so that the users can select any particular model using the pull-down menu. As mentioned above, two superposition programs (STAMP and ProFit) are deployed for structural superposition and the user has the freedom to choose a program of interest. A detailed output containing r.m.s.d. values, sequence identity, rotation matrix, translation vector and so on will be displayed. Most importantly, users can save the superposed atomic coordinates in the local client machine for further analysis. The users of the program are requested to cite this article and the URL address in their research proceedings.

## CASE STUDY

The output of a typical superposition of 12 (native, mutants and inhibitor complexes) structures of recombinant phospholipase A<sub>2</sub> (21–23) (1VL9, 1UNE, 1MKS, 1FDK, 1MKV, 1MKT, 1KVX, 1O2E, 1VKQ, 1IRB, 1GH4 and 1C74 solved using X-ray crystallography is shown in Figure 1. The PDB-id 1VL9 is used as a fixed molecule and the remaining 11 structures are treated as mobile molecules (molecules to be superposed on the fixed molecule). The program STAMP is used for superposition. The top panel shows a detailed output like status of superposition, sequence identity, stamp score and r.m.s.d. values. The RASMOL graphics panel on the right shows the superposition of all the structures in different colors. Figure 2 displays the invariant water molecules in six different crystal structures of Oligo-peptide binding proteins (OppA) (24). The structure (1B4Z {457}) is used as fixed molecule and the remaining five (1B32 {437}, 1B3F {455}, 1B3G {356}, 1B46 {374} and 1B51 {433}) are treated as mobile molecules. The number within braces represents the number of water molecules present in the 3D structures. The server reports 209 invariant water molecules in all the structures. It is interesting to note that 58.7% (209/356) of the water molecules is invariant. The invariant water molecules are identified after superposition within a distance of 1.8 Å (between the water molecules). Figure 3 shows the invariant water molecules between different subunits of a tetramer. The PDB-id used here is 1JAC (25) and it has eight different chains [four heavy chains (A, C, E, G) and four light chains (B, D, F, H)]. The superposition of different chains A, B, C, D (green) and E, F, G, H (red) along with 36 invariant water



**Figure 3.** The output panel depicts the superposition of eight different chains along with 36 invariant water molecules in PDB-id: 1JAC. The chains A, B, C, D (fixed) are colored green and the color red is used for the chains E, F, G, H (mobile). The invariant water molecules are having the same color as the corresponding subunits.



**Figure 4.** The output shows the common water molecules between the subunits A and B. The RASMOl panel shows eight common water molecules (blue color). This is carried out using the option (c) 'Superpose subunits within a structure and 'identify common water molecules'.

molecules and their interactions with the subunits are shown. The calculation is performed using the options 'Superpose subunits within a structure' and 'identify invariant water molecules'. The common water molecules between two different subunits (only subunit A and B are used) of a tetrameric protein [PDB-id 1J4S (26)] are shown in Figure 4. Here, the options (c), 'Superpose subunits within a structure' and 'Identify common water molecules' are used. The subunits A and B are shown in green and red colors, respectively. There are eight water molecules (blue), which are common between the chains A and B.

## CONCLUSIONS

At the outset, 3dSS is created to better serve the research community working in the area of structural bioinformatics. This computing server is very useful to superpose either complete or partial structures. Furthermore, the server can effectively be used to identify the invariant and common water molecules. The knowledge base (PDB) used by the server is up-to-date and hence the user will be able to access the latest information available in the PDB. As described, it is tempting to conclude that the software will certainly be beneficial for many macromolecular crystallographers and the undergraduate/graduate students working in the area of structural bioinformatics.

## ACKNOWLEDGEMENTS

The corresponding author (K.S.) thanks Dr Geoffrey Barton and Dr Andrew Martin for permitting to use their superposition programs in the computing server. The authors thank Ch. Kiran Kumar for his help at the initial stages of this work. One of the authors (K.S.) thanks Ms P. Mridula for critical manuscript reading. The proposed search engine is developed and maintained at the Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India. All the contributing authors acknowledge the use of the facilities: the Interactive Graphics Based Molecular Modeling, Bioinformatics centre and the Supercomputer Education and Research Centre. The first two facilities are supported by the Department of Biotechnology (DBT), Government of India. We are grateful for the individual (K.S.) project support from DBT. A part of this work is supported by the Institute wide Computational Genomics Program. The Open Access publication charges for this article were waived by Oxford University Press.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.Jr, Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M.J. (1977) The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.
- Kabsch, W.A. (1978) Discussion of solution for best rotation of two vectors. *Acta Crystallogr.*, **A34**, 827–828.
- MacLachlan, A.D. (1982) Rapid comparison of protein structures. *Acta Crystallogr.*, **A38**, 871–873.
- Kearsley, S.K. (1990) An algorithm for the simultaneous superposition of a structural series. *J. Comput. Chem.*, **11**, 1187–1192.
- Diamond, R. (1992) On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.*, **1**, 1279–1287.
- Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 29–32.
- Kaplan, W. and Littlejohn, T.G. (2001) SWISS-PDB Viewer (Deep View). *Brief. Bioinformatics*, **2**, 195–197.
- Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr.*, **D60**, 2256–2268.
- Maiti, R., Domselaar, G.H.V., Zhang, H. and Wishart, D.S. (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.*, **32**, W590–W594.
- Kuntz, I.D. and Kauzmann, W. (1974) Hydration of proteins and polypeptides. *Adv. Protein Chem.*, **28**, 239–345.
- Eisenberg, D. and McLachlan, A.D. (1986) Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.
- Sanschagrin, P.A. and Kuhn, L.A. (1998) Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. *Protein Sci.*, **7**, 2054–2064.
- Sundaralingam, M. and Shekarudu, Y.C. (1989) Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333–1337.
- Sessions, R.B., Thomas, G.L. and Parker, M.J. (2004) Water as a conformational editor in protein folding. *J. Mol. Biol.*, **343**, 1125–1133.
- Biswal, B.K., Sukumar, N. and Vijayan, M. (2000) Hydration, mobility and accessibility of lysozyme: Structures of a pH 6.5 orthorhombic form and its low humidity variant and a comparative study involving 20 crystallographically independent molecules. *Acta Crystallogr.*, **D56**, 1110–1119.
- Zhang, X.-J. and Matthews, B.W. (1994) Conservation of solvent-binding sites in 10 crystal forms of T4 Lysozyme. *Protein Sci.*, **3**, 1031–1039.
- Prasad, B.V.L.S. and Suguna, K. (2002) Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr.*, **D58**, 250–259.
- Kishan, K.V.R., Chandra, N.R., Sudarsanakumar, C., Suguna, K. and Vijayan, M. (1995) Water-dependent domain motion and flexibility in its ribonuclease A and the invariant features in its hydration shell. An X-ray study of two low-humidity crystal forms of the enzyme. *Acta Crystallogr.*, **D51**, 703–710.
- Russell, R.B. and Barton, G.J. (1992) STAMP: multiple protein sequence alignment from tertiary structure comparison. *Proteins*, **14**, 309–323.
- Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374–382.
- Sekar, K. and Sundaralingam, M. (1999) High resolution refinement of the orthorhombic form of bovine pancreatic phospholipase A<sub>2</sub>. *Acta Crystallogr.*, **D55**, 46–50.
- Sekar, K., Rajakannan, V., Gayathri, D., Velmurugan, D., Poi, M.-J., Dauter, M., Dauter, Z. and Tsai, M.-D. (2005) Atomic resolution (0.97 Å) structure of the triple mutant (K53,56,121M) of bovine pancreatic phospholipase A<sub>2</sub>. *Acta Crystallogr.*, **F61**, 3–7.
- Sekar, K., Eswaramoorthy, S., Jain, M.K. and Sundaralingam, M. (1997) Crystal structure of the complex of bovine pancreatic phospholipase A<sub>2</sub> with the inhibitor 1-hexadecyl-3-(trifluoroethyl)-sn-glycero-2-phosphomethanol. *Biochemistry*, **36**, 14186–14191.
- Tame, J.R., Murshudov, G.N., Dodson, E.J., Neil, T.K., Dodson, G.G., Higgins, C.F. and Wilkinson, A.J. (1994) The structural basis of sequence-independent peptide by OppA protein. *Science*, **264**, 1578–1581.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M. (1996) A novel mode of carbohydrate recognition in Jacalin, a *moracea* plant lectin with a beta-prism fold. *Nature Struct. Biol.*, **3**, 596–602.
- Pratap, J.V., Prakash, A.A., Rani, P.G., Sekar, K., Surolia, A. and Vijayan, M. (2002) Crystal structures of Artocarpin, a *moracea* Lectin with mannose specificity and its complex with methyl-alpha-D-Mannose: implications to the generation of carbohydrate specificity. *J. Mol. Biol.*, **317**, 237–247.