Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

SSEP-2.0: Secondary Structural Elements of Proteins

B. Balamurugan,[a] K. Samaya Mohan,[a] J. Ramesh,[a] M. N. A. Md. Roshan,[a] K. Sumathi[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

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

Received 10 December 2004
Accepted 23 February 2005

The Secondary Structural Elements of Proteins (SSEP) database is an integrated and comprehensive knowledge base for accessing information related to all the secondary-structural elements present in non-redundant (25 and 90%) protein chains. The new version 2.0 of the database contains 2485 and 8595 protein chains from the 25 and 90% non-redundant data sets, respectively. The necessary web interfaces have been developed that enable users to visualize the three-dimensional structure of the secondary-structural element in the client machine using the free molecular-visualization program *RASMOL*. This source is updated at regular intervals and can be accessed through the bioinformatics web server at the URL http://cluster.physics.iisc.ernet.in/ssep or http://144.16.71.148/ssep/.

1. Introduction

The sequence and structural properties of various secondary-structural elements are of considerable interest in view of the occurrence of these fragments in different regions in the protein molecules available in the PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000). Identity or similarity in sequences is used commonly as a basis for assigning putative function to newly solved protein structure(s) or identified gene products. Thus, analysis of a specific sequence or a structural motif is greatly enhanced by determining the sequence, structure and functional relationships between the individual proteins. To the best of the knowledge of the authors, there is no efficient computing server (including the earlier version of SSEP; Shanthi *et al.*, 2003) to search for and superpose a particular secondary-structural element from the same or different non-redundant (25 and 90%) protein chains (Hobohm & Sander, 1994). In addition, there is no mechanism available to delineate different types of β- and γ-turns available in the non-redundant protein chains. The updated search engine addresses these issues in detail and offers several useful utilities (see §3 for details) for the user's convenience. Hence, in this paper we will present only the enhanced facilities (compared with the earlier version of SSEP; Shanthi *et al.*, 2003) in detail.

2. Materials and methods

To maintain data quality and uniformity, the widely used program *PROMOTIF* (Hutchinson & Thornton, 1996) was used to identify all secondary-structural elements present in non-redundant protein chains. The programs *STAMP* (Russell & Barton, 1992) and *ProFit* (A. C. R. Martin; http://www.bioinf.org.uk/software/profit) are interfaced with the search engine to superpose the secondary-structural elements retrieved by the search engine. The details of the various secondary-structural elements available in the knowledge base are given in Table 1.

3. Data structure and new functionalities

The primary aim of this project is to support structural bioinformatics research by maintaining a high-quality knowledge base containing secondary-structural elements. Additionally, SSEP provides an effi-

cient interactive web interface with the following options: (i) search for α-helix, (ii) search for 3_{10}-helix, (iii) search for β-strand, (iv) search for β-turns, (v) search for γ-turns, (vi) search for hairpin loops, (vii) sequence pattern matching, (viii) advanced search facility, (ix) available non-redundant protein chains and (x) secondary-structural elements distorted by residues. The following useful functionalities have been newly added in order to understand better the behavior of a particular secondary-structural element.

(i) A particular type of secondary-structural element (see §4) identified in various non-redundant protein chains can be superposed. For superposition calculation, the programs *STAMP* and *ProFit* are interfaced with the search engine. Additionally, the superposed fragments can be visualized in different colours on the client machine using the molecular-visualization tool *RASMOL* (Sayle & Milner-White, 1995) by clicking the button 'View in RASMOL'. A maximum of 20 fragments at a time can be superimposed.

(ii) The updated search engine fetches different types (Table 1) of β- and γ-turns and hairpin loops available in the knowledge base.

(iii) A particular type of secondary-structural element (for example, all α-helices or β-strands) can be visualized in a given protein structure.

(iv) A helical wheel plot can be constructed. If a particular α-helix exceeds the theoretical limit of 18 residues, an option is provided in the plot (forward direction button available at the top right corner) to see the remaining residues.

(v) The distribution of the isotropic temperature factors of the residues present in a queried secondary-structural element can be visualized.

(vi) Pattern matching (identical) is carried out using the new fast and efficient pattern-matching algorithm *SSABS* (Sheik *et al.*, 2004).

4. Case study

A sample output of a typical search for a helix containing more than 20 residues in all the structures (resolution better than 1.0 Å and crystallographic R factor better than 15%) solved using X-ray crystallography available in the 25% non-redundant protein chains is shown in Fig. 1. This simple search aided in the recognition of seven hits from five different protein structures. Fig. 1 shows the helical wheel plot (top right), the location of the 24-residue-long helix (bottom right window), shown as ribbons, and the distribution of temperature factor of the residues in the queried helix (top left). Users can compare the corresponding information available in the PDB by clicking the link 'View Headers' provided in the output frame (Fig. 1). Additionally, to find a similar occurrence in the gene sequences, the user needs to

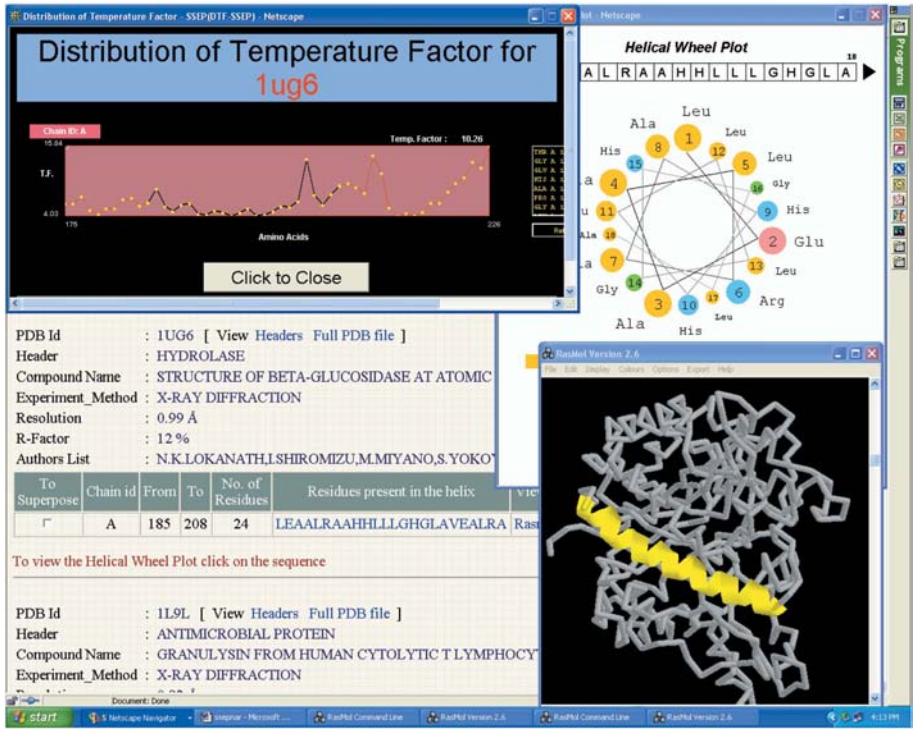

Figure 1
A screen shot of the output frame displaying some of the structural properties of an α-helix with 24 residues present in a β-glucosidase (resolution, 0.99 Å; crystallographic R factor, 12%).

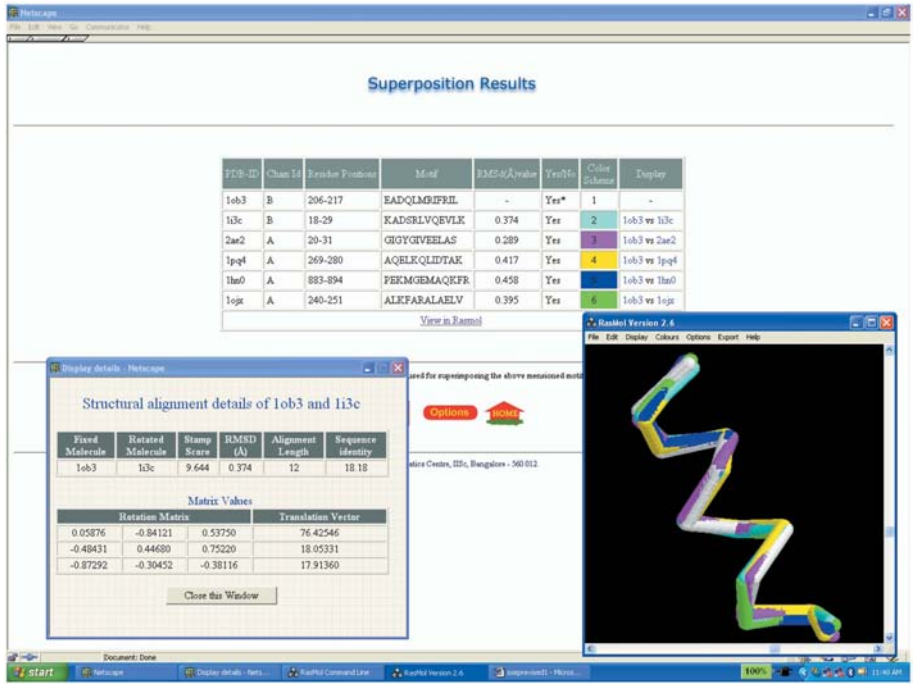

Figure 2
An output frame showing the detailed superposition results output for five α-helical segments of 12 residues (exact length) as queried by the search engine from 90% non-redundant protein chains.

Table 1
Various secondary-structural elements and their number of occurrences in 25 and 90% non-redundant protein chains.

Non-redundant protein chains (%)	25	90
No. of protein structures	2379	7836
No. of chains	2485	8595
Secondary-structural elements present in the database		
No. of α-helices	10850	53546
No. of 3_{10}-helices	3916	21359
No. of β-strands	14873	77249
No. of β-turns and their various types		
Type I	8985	46271
Type I′	980	5159
Type II	2791	15664
Type II′	587	3088
Type IV	14454	62583
Type VIa	176	956
Type VIb	243	1314
Type VII	2313	12416
Total	30529	147451
No. of γ-turns and their various types		
Classic type	621	2224
Inverse type	4221	18391
Total	4842	20615
No. of hairpin loops	5178	25269

click the link 'Occurrence in genome database' (not shown). The centre maintains the anonymous Genome Sequence FTP server and the current version contains all genome sequences available in the NCBI (National Centre for Biotechnology Information) site. Fig. 2 depicts the superposition results of five α-helical segments (12 residues in length) queried from 90% non-redundant protein chains (resolution better than 2.0 Å and crystallographic R factor better than 20.0%). The output frame shows 60 hits from 42 protein chains. The bottom right panel shows the graphical representation of the superposed α-helical segments (colouring schemes as shown in the top panel) and the bottom left panel shows the detailed alignment results of the two fragments from PDB codes 1ob3 (fixed fragment) and 1i3c (rotated fragment). This information is obtained by clicking '1ob3 vs 1i3c' available in the last column of the middle panel. The described search engine *SSEP* is easy to use and has been tested through the popular web browsers Netscape (v.4.7) and Mozilla (the built-in browser in the Linux operating system). The user needs to interface the graphics program *RASMOL* with the web browser (for the first use only; see http://144.16.71.148/ssep/downloadrasmol.htm for instructions).

We are grateful for the individual project support from the Department of Biotechnology (DBT), Government of India. Part of this work was supported by the Institute-wide Computational Genomics Program of the Department of Biotechnology, Government of India. Finally, KS thanks Dr Geoff Barton and Dr Andrew Martin for permitting the use of their superposition programs in the search engine. The authors thank Ms P. Mridula for careful reading of the manuscript.

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.
Hobohm, U. & Sander, C. (1994). *Protein Sci.* **3**, 522–524.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
Russell, R. B. & Barton, G. J. (1992). *Proteins Struct. Funct. Genet.* **14**, 309–323.
Sayle, R. A. & Milner-White, E. J. (1995). *Trends. Biochem. Sci.* **20**, 374–382.
Shanthi, V., Selvarani, P., Kiran Kumar, C., Mohire, C. S. & Sekar, K. (2003). *Nucleic Acids Res.* **31**, 3404–3405.
Sheik, S. S., Aggarwal, S. K., Poddar, A., Balakrishnan, N. & Sekar, K. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 1251–1256.