Incorporating Sequence and Biochemical Information in TOPS models - for Biologically Significant Pattern Matching and Pattern Discovery in Protein Structures

Mallika Veeramalai[1], David Gilbert[1] and David Westhead[2]

1Bioinformatics Research Centre, Department of Computing Science, University of Glasgow,
Glasgow G12 8QQ, Scotland, UK;
{mallika, drg}@dcs.gla.ac.uk
2 Bioinformatics Group, School of Biochemistry and Molecular Biology, University of Leeds,
Leeds LS2 9JT, England, UK.
D.R.Westhead@leeds.ac.uk

Key words: TOPS, TOPS+, Protein structure comparison, Pattern-Discovery, Pattern-Matching, Protein-ligand interaction, Graph-algorithm

TOPS (Topology of Protein Structure) Database contains 2D abstract spatial representation of secondary structure elements (SSEs) of the protein structures [1]. For example, fig.1 (B) shows the TOPS 2D cartoon of the *methyltransferase* protein 1vpt00 domain. Based on TOPS cartoons TOPS diagrams are developed [2]). Instead of representing spatial positions by elements in a plane, a TOPS diagram contains information about the grouping of beta-strands in beta-sheets (two adjacent elements in a beta-sheet are connected by an H-bond, which can be either parallel or anti-parallel) also information about the orientation of elements (any two SSEs can be connected by either left or right chirality). Figure1.(C) shows the TOPS diagram for 1vpt00. Based on these TOPS diagrams, very fast pattern matching and pattern discovery algorithms for protein topologies were developed [3] and [4]. However, because of its abstract nature, it is possible to loose significant biological information. Incorporation of sequence information (in the form of PSSM/HMM profiles) and biochemical features such as ligand-binding sites, active-sites in the TOPS graph-based representation of the protein structure will increase its biological significance.

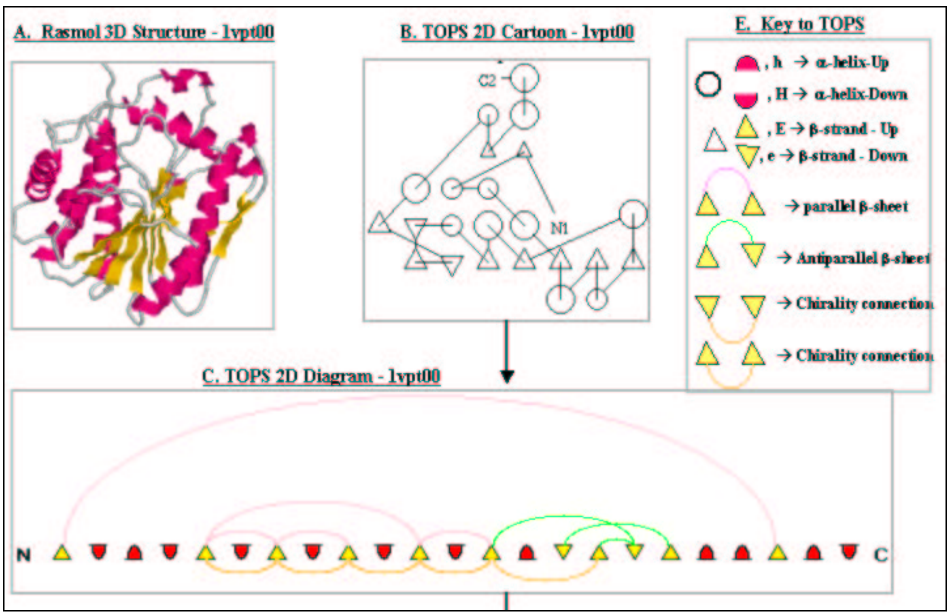

Figure 1. Rasmol and TOPS models for *methyltransferase* protein structure - 1vpt00.

The flowchart of the project is given in the figure 2. Interesting results would be valuable efforts to predict protein structure and function from the sequences, and these problems remain key challenges of direct relevance to projects in structural and functional genomics. TOPS database can be accessible from http://www.tops.leeds.ac.uk

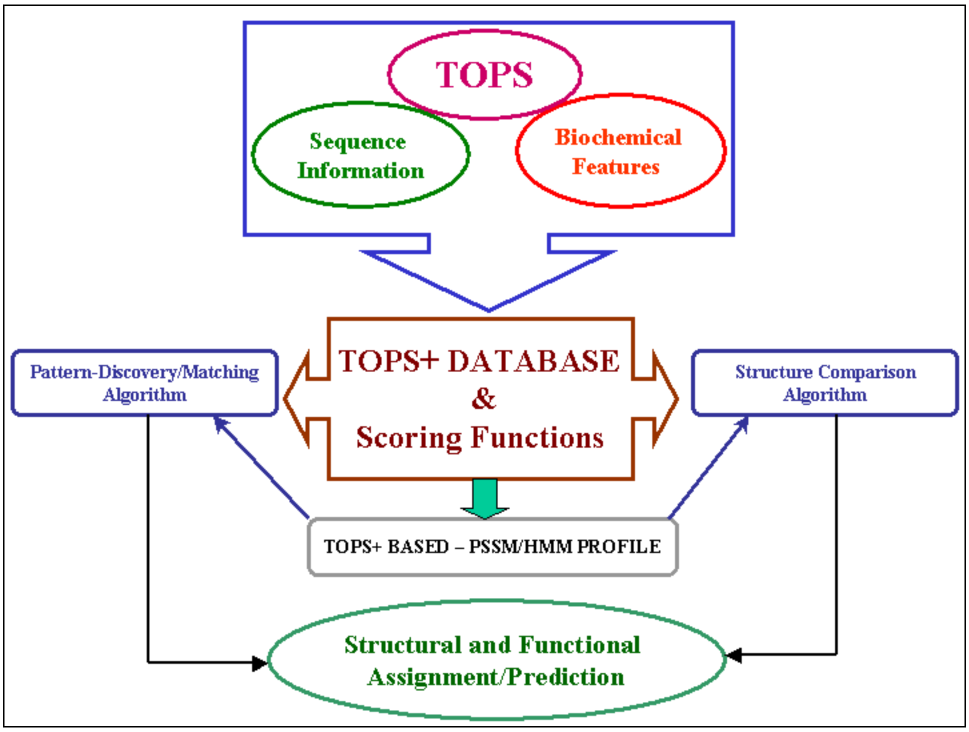

Figure 2. Flowchart.

References

[1] Westhead D, Slidel TW, Flores TP and Thornton JM, Protein structural topology: automated analysis, diagrammatic representation and database searching. Protein Sci 8, 897-904, 1999.

[2] Gilbert D, Westhead DR, Nagano N and Thornton JM, Motif-based searching in TOPS protein topology databases. Bioinformatics, 15:,317-326, 1999.

[3] Gilbert D, Westhead DR, Viksna J and Thornton JM, Topology-based protein structure comparison using a pattern discovery technique, Journal of Computers and Chemistry, 26:1,23-30.3, 2001.

[4] Viksna J and Gilbert D, Pattern matching and pattern discovery algorithms for protein topologies, Algorithms in Bioinformatics: First International Workshop, WABI 2001 Proceedings, LNCS 2149, 98-111, ISBN3-540-42516-0, 2001.