BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 15 2004, pages 2460–2462
doi:10.1093/bioinformatics/bth248

POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins

*Aleksey A. Porollo, Rafal Adamczak and Jaroslaw Meller**

Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

Received on November 10, 2003; revised and accepted on February 24, 2004
Advance Access publication April 8, 2004

ABSTRACT

Summary: The POLYVIEW visualization server can be used to generate protein sequence annotations, including secondary structures, relative solvent accessibilities, functional motifs and polymorphic sites. Two-dimensional graphical representations in a customizable format may be generated for both known protein structures and predictions obtained using protein structure prediction servers. POLYVIEW may be used for automated generation of pictures with structural and functional annotations for publications and proteomic on-line resources.
Availability: http://polyview.cchmc.org
Contact: jmeller@cchmc.org

The rapid growth of the amount of structural data derived from experimental studies and computational modeling requires parallel development of new tools for fast and convenient parsing, viewing and analyzing these data. In particular, tools and servers for protein structure visualization need to address new challenges arising in the context of functional annotations, studies on protein–protein interactions and protein pathways.

There are many public domain programs for the visualization and three-dimensional (3D) rendering of protein structures, e.g. RASMOL (Sayle and Milner-White, 1995) or WebMol (Walther, 1997). Such programs are powerful tools for the 3D analysis and may be used to analyze polymorphisms and structural motifs. On the other hand, however, 3D representation may be cumbersome when the overall view of the structure with multiple annotations is required. For example, some information is inevitably lost while making static 2D projections of 3D rendering for paper media publications or on-line resources. Therefore, 2D representations of protein structure, specifically designed to highlight and enable analysis of multiple attributes and annotations are often used instead.

PDBsum (Laskowski, 2001) is one of the most informative programs for the 2D protein structure visualization. Based on the Protein Data Bank (PDB) (Berman *et al*., 2000) information, it generates a comprehensive representation of protein secondary structures (SS) and interactions with ligands. However, the PDBsum lacks information about the residue relative solvent accessibility (RSA) and other attributes, such as hydrophobicity profile, for example. Moreover, PDBsum server provides static and precomputed (and therefore often unavailable for new structures) representations of structures deposited in the PDB.

We have developed a web server called POLYVIEW that addresses these limitations and offers a flexible annotation and 2D visualization tool for both experimental or modeled 3D structures and predicted 1D structural profiles, such as SS or RSA. For example, the POLYVIEW annotations allow one to identify putative globular soluble as well as membrane domains, and thus make preliminary conclusions about domain structure of a protein. Important functional motifs and polymorphic residues can be analyzed in terms of their structural environments.

The POLYVIEW server deals with three types of input data: (1) a file with a structure in the PDB format or a four-letter PDB code if the protein of interest can be found in the PDB; (2) results from protein structure prediction servers with predicted protein SS and RSA in the standard CASP (CASP5, 2002, http://predictioncenter.llnl.gov/casp5/doc/casp5-format.html) format [results from our own protein prediction server (Adamczak *et al*., 2004, http://sable.cchmc.org) may be submitted in the original format] and (3) arbitrary protein sequence and 1D sequence profiles to be submitted using copy and paste technique. In order to obtain the SS and RSA for structures submitted in the PDB format, the POLYVIEW server runs the DSSP program (Kabsch and Sander, 1983). Solvent accessibilities (SA) are then normalized using values of SA for the corresponding amino acid residues in tri-peptides (Chothia, 1976).

Pictures generated by POLYVIEW are highly customizable, allowing one to include different combinations of relevant attributes. For example, one may optionally include residue numeration, amino acid sequence, graphical representation and/or three-state codes for SS (H, E, C), confidence level for SS prediction, gray scale bar and/or numeric data for RSA,

*To whom correspondence should be addressed.

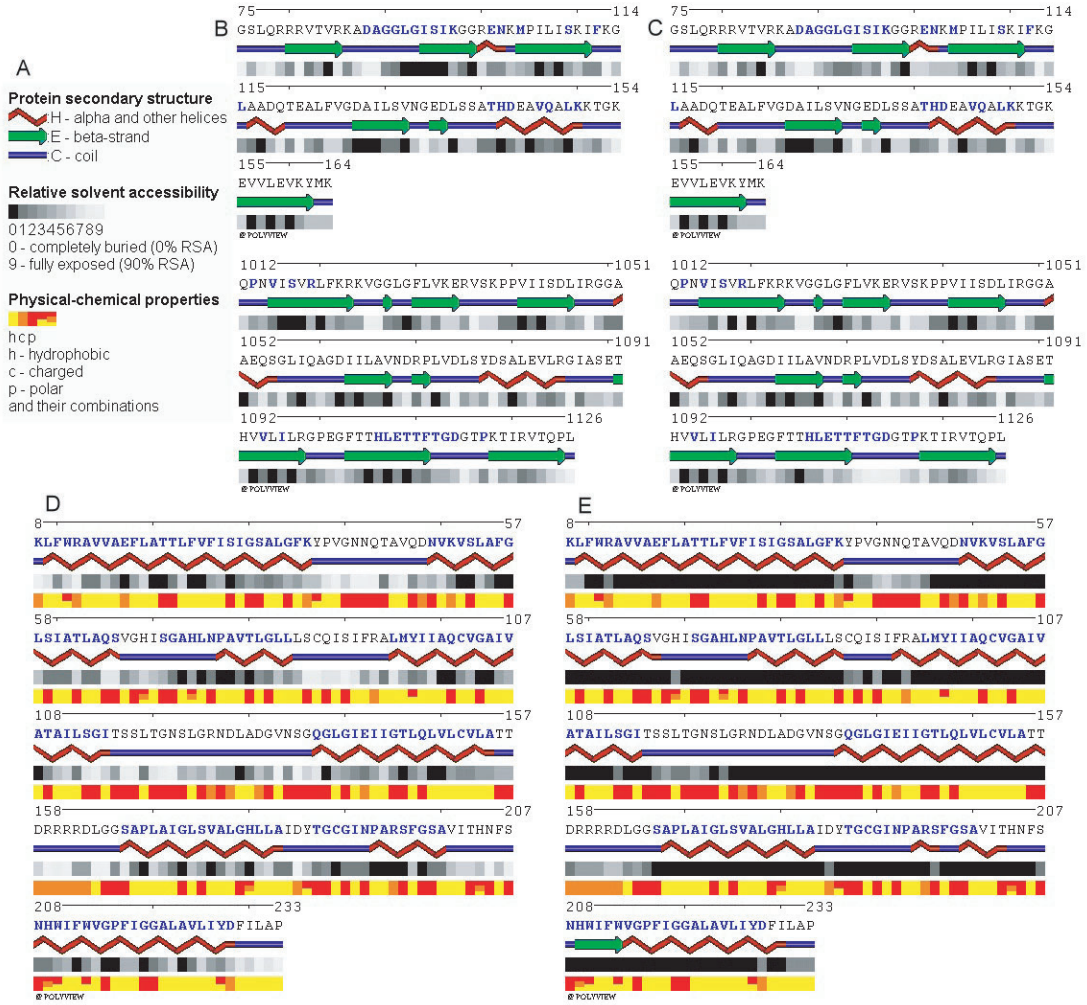

Fig. 1. Examples of protein sequence annotations by the POLYVIEW server: (**A**) legend; (**B**) two chains of the protein 1qav as complex; (**C**) the same sequences but as isolated chains (highlighted residues have different RSA in complex relative to isolated chains); (**D**) aquaporin (PDB code is 1fqy) structure annotation; (**E**) annotation of the same protein using SABLE prediction (highlighted residues show trans-membrane regions).

as well as hydrophobicity, polarity and charge profiles. One may also highlight specific residues by color and font style. It allows users to emphasize the sites of interest, e.g. crucial structural motifs or polymorphic residues. In the case of a protein complex, POLYVIEW may be used to automatically identify those residues that are located at the interaction interface, based on observed changes in RSA in the complex relatively to isolated chains.

Two examples of how the POLYVIEW server can be used for functional and structural annotations are included in Figure 1. The top panels (B and C) show the example of protein sequences within a complex (PDB code is *1qav*) and as isolated chains. Changes in RSA reveal residues that form the interaction interface and are highlighted in blue color and bold font style. The two bottom panels represent, in turn, an annotation of a membrane protein aquaporin (PDB code is *1fqy*). The actual PDB structure as analyzed by DSSP and as predicted by the SABLE server is shown in panels D and E, respectively. Residues that are located in trans-membrane and membrane-embedded regions (according to Swiss-Prot entry P29972) are highlighted in blue. As can be seen from panel E, most of the residues in the trans-membrane regions are predicted as fully or nearly fully buried (i.e. as having very low water accessible area). Prediction and visualization of such patterns may serve as a tool for the determination of trans-membrane regions.

POLYVIEW can be used in both interactive and batch modes. Using scripts available from the server's web site, one may automatically generate pictures for further annotations, publications and on-line resources. For example,

POLYVIEW was used to create graphical representations and annotations of polymorphisms included in the PolyDom database of Human coding SNPs (Jegga *et al*., http://polydoms.cchmc.org).

REFERENCES

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

CASP5 (2002) Fifth Community Wide Experiment on the critical assessment of techniques for protein structure prediction—CASP5: formats for prediction submission.

Chothia,C. (1976) The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.*, **105**, 1–14.

Jegga,A.G., Gowrisankar,S., Chen,J. and Aronow,B.J. (2004) PolyDom: mapping of human coding SNPs onto protein domains, manuscript in preparation.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Laskowski,R.A. (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.*, **29**, 221–222.

Adamczak,R., Porollo,A. and Meller,J. (2004) Accurate prediction of solvent accessibility using Neural Networks based regression. *Proteins: Struct. Funct. and Bioinformatics* (in press).

Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374–376.

Walther,D. (1997) WebMol—a Java-based PDB viewer. *Trends Biochem. Sci.*, **22**, 274–275.