ABSTRACT

Summary: Protein features are often displayed along the linear sequence of amino acids that make up that protein, but in reality these features occupy a position in the folded protein's 3D space. Mapping sequence features to known or predicted protein structures is useful when trying to deduce the function of these features and when evaluating sequence or structural predictions. To facilitate this goal, we developed PDBpaint, a simple tool that displays protein sequence features gathered from bioinformatics resources on top of protein structures, which are displayed in an interactive window (using the Jmol Java viewer). PDBpaint can be used either with existing protein structures or with novel structures provided by the user. The current version of PDBpaint allows the visualization of annotations from Pfam, ARD (detection of HEAT-repeats), UniProt, TMHMM2.0 and SignalP. Users can also add other annotations manually.

Availability and Implementation: PDBpaint is accessible at http://cbdm.mdc-berlin.de/~pdbpaint. Code is available from http://sourceforge.net/projects/pdbpaint. The website was implemented in Perl, with all major browsers supported.

Contact: david.fournier@mdc-berlin.de

Received on April 12, 2011; revised on July 1, 2011; accepted on July 9, 2011

1 INTRODUCTION

A protein’s function is tightly dependant on its 3D structure. In particular, protein features such as catalytic centers, domains and post-translational modifications, have a shape in 3D space that is crucial to understand the function of the protein. However, in protein databases these features are usually defined by their locations in linear protein sequences only. It is actually easier to define and handle protein annotations in a 1D sequence. Second, protein sequence data is much more abundant than protein structure data, which require complex techniques to determine atomic coordinates. As a result, most protein annotations are mapped to protein sequences but not to protein structures (Liu et al., 2008). Nevertheless, tagging available 3D structures with annotations can be extremely informative. This is used in some web tools and databases dealing with protein structures [PDB (Rose et al., 2011), SCOP (Murzin et al., 1995), Dali (Holm and Rosenstrom, 2010)] or with domain predictions [HHpred (Soding et al., 2005), Pfam (Finn et al., 2010)]. Typically, users have to obtain a file with the atomic coordinates of a protein structure (e.g. in PDB format) and then load it into a molecular graphics display program (O’Donoghue et al., 2010). Then, to incorporate annotation information users need to obtain the corresponding protein sequence, have it analyzed by one or more Webservices and then manually transfer the output of these tools into the molecular viewer for 3D representation. This can be time consuming, especially if many sources of annotation are being used on multiple structures.

To facilitate this procedure, we have developed a web tool, PDBpaint, designed specifically to allow loading either existing structures from the PDB, structure models from MODBASE (Pieper et al., 2011), or a PDB file created by the user, for example for a novel structure or from modeling software [e.g. i-TASSER (Roy et al., 2010; Zhang, 2008)], and then tag them according to the user’s preferences.

PDBpaint can compute predicted features from the sequence extracted from the PDB file using external services and locally run methods; protein domains by Pfam (Finn et al., 2010), alpha-solenoid repeats by ARD (Palidwor et al., 2009), signal peptides by SignalP (Emanuelsson et al., 2007) and transmembrane alpha-helices by TMHMM (Krogh et al., 2001). PDBpaint can also represent features from the protein’s corresponding UniProt entry (Magrane and Consortium, 2011) that were experimentally or computationally derived (mapped to the PDB sequence using sequence alignment to account for possible residue numbering differences). Manual annotations can be also input for any set of residues and colors.

The collected sequence features are then mapped and represented in the PDB structure by a script that calls the molecular graphics viewer Jmol [http://www.jmol.org/; (Jmol, 2011)]. The structures to tag are then displayed in an interactive window (Fig. 1).

2 SIMILAR TOOLS

The problem of automating the mapping of features to structures is almost as old as the molecular graphic programs themselves.

In 1994, Saqi and Sayle already presented a script to map protein motifs detected with regular expressions on PDB protein structure (Saqi and Sayle, 1994) using the RasMoL viewer. The PDB website displays secondary structure annotations using Jmol. Standalone graphical molecular viewers like RasMol (Sayle and Bissell, 1992) or PyMOL (PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC) can also do this. Pfam (Finn et al., 2010) allows the display of Pfam domains of a protein in a dynamic protein 3D viewer (Jmol) when the protein’s structure is available from the PDB. Aside from these well-known tools, several small web services allow to display a variety of features on 3D structures. Motif3D is an online tool that focuses on displaying protein motifs from
we believe that PDBpaint will facilitate the use of predicted protein structures when designing experiments.

ACKNOWLEDGEMENTS
We thank Jean-Fred Fontaine, who gave useful advice during the coding process. We also thank warmly Ursula Pieper (MODBASE, University of California San Francisco) and Torsten Schwede (SwissModel, University of Basel) for their great help.

Funding: Collaborative Research Center for Theoretical Biology: Robustness, Modularity and Evolutionary Design of Living Systems [Sonderforschungsbereich 618 (SFB 618)], Humboldt-University of Berlin, Germany.

Conflict of Interest: none declared.

REFERENCES

Fig. 1. Bovine rhodopsin annotated according to the prediction of TMJIMM2.0, which detects transmembrane regions. 1–6: features of PDBpaint: 1: PDB/UniProt ID code input window, with different options (positions to tag, webservice and window size); 2: custom structure upload window, with its options. 3: output of the structure by Jmol: 4: properties (positions to tag, webservice and window size). 5: legend of annotations performed by PDBpaint. In our example, webservice TMHMM2.0 for detection of transmembrane regions has been chosen; 6: help page, to get some tips about PDBpaint.

2606