



3MOTIF: visualizing conserved protein sequence motifs in the protein structure database

Steven P. Bennett¹, Craig G. Nevill-Manning^{2,†} and Douglas L. Brutlag^{1,*}

¹Department of Biochemistry, B400 Beckman Center, Stanford University, CA 94305-5307, USA and ²Computer Science, Rutgers University, Piscataway, NJ 08854, USA

Received on March 6, 2002; revised on September 6, 2002; accepted on September 10, 2002

ABSTRACT

Summary: 3MOTIF is a web application that visually maps conserved sequence motifs onto three-dimensional protein structures in the Protein Data Bank (PDB; Berman *et al.*, *Nucleic Acids Res.*, **28**, 235–242, 2000). Important properties of motifs such as conservation strength and solvent accessible surface area at each position are visually represented on the structure using a variety of color shading schemes. Users can manipulate the displayed motifs using the freely available Chime plugin.

Availability: <http://motif.stanford.edu/3motif/>.

Contact: brutlag@stanford.edu

Supplementary information: <http://motif.stanford.edu/3motif/supplementary/>

INTRODUCTION

Discrete protein sequence motifs are widely used to describe homology between proteins and establish relationships between well-known and new protein sequences. More specifically, discrete motifs identify amino acids sharing important properties conserved in evolution. Further, they are often able to identify structurally or functionally important regions within protein families, such as active sites and protein–protein interaction sites.

In addition to identifying these regions, biologists would often like to determine the specific interactions or mechanisms of activity these conserved regions represent. Since activity and interactions depend heavily on three-dimensional relationships between amino acids, it can be difficult to determine specific roles of conserved amino acids strictly from sequence motifs. Our goal is to bridge this gap by visually mapping conserved sequence motifs to structural examples of those motifs. Previous efforts in structural representations of conserved sequence data have included the JOY software

(Mizuguchi *et al.*, 1998), in which the font in a multiple sequence alignment diagram is modulated according to structural properties. More recent approaches include the PDBsum database (Laskowski, 2001), in which conserved patterns are color-coded by conservation in PDB structures, and COMBOSA3D (Stothard, 2001), which is a world wide web based display of multiple sequence alignment information mapped onto protein structures. Here, we present 3MOTIF, a web application that provides three-dimensional visualization of conserved residues in discrete sequence motifs. The benefits of 3MOTIF are 2-fold: first, the structural representation provides clues as to why certain positions are conserved in protein families. Second, knowing the structural environments of these conserved residues allows biologists to better target them for further experimentation.

3MOTIF OVERVIEW

Search options

3MOTIF provides a number of ways to visualize discrete sequence conservation data. Three of the most common representations for protein sequence conservation are PROSITE patterns (Falquet *et al.*, 2002), eMOTIFs (Huang and Brutlag, 2001) and BLOCK multiple sequence alignments (Henikoff *et al.*, 1999). 3MOTIF can be accessed using any of these, or any regular expression a user may have from another motif-building method. 3MOTIF then displays the first PDB structure found that contains the query, with the option to view all other structures that have it as well (Fig. 1). In addition to searching by motif, the user can also search using a PDB structure or multiple sequence alignment accession number.

3MOTIF is also designed to integrate with other bioinformatics resources on the Internet. For example, the eMOTIF-SEARCH component of the eMOTIF software suite (<http://motif.stanford.edu/emotif/>) has been enhanced such that when a user submits a protein sequence, a 3MOTIF hyperlink appears next to each resulting

*To whom correspondence should be addressed.

† Present address: Google, Inc., 2400 Bayshore Parkway, Mountain View, CA 94043, USA

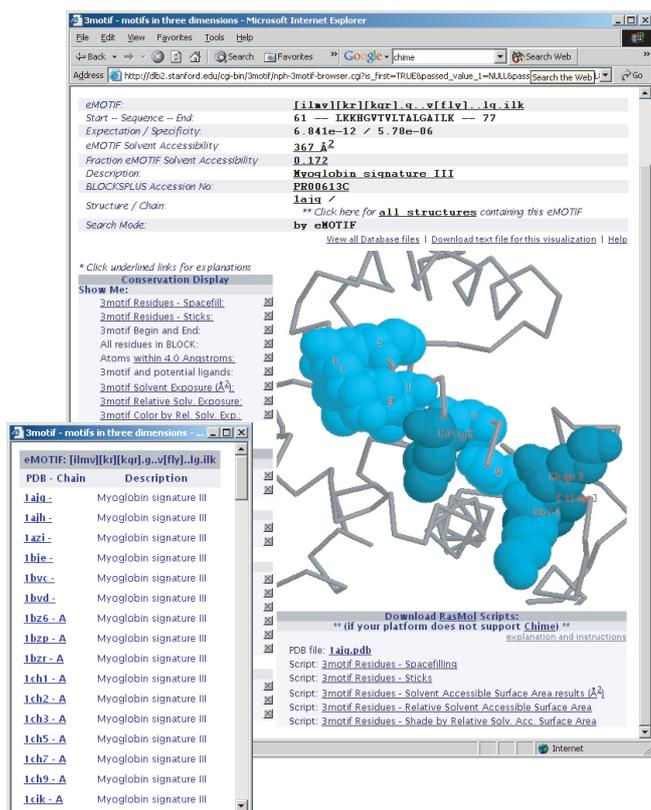


Fig. 1. 3MOTIF results for the eMOTIF, [ilmv][kr][kqr].g.v[fly]..lg.ilkk. The top of the main page provides basic information about the eMOTIF currently displayed in this structure (1AJG, a sperm whale myoglobin), such as the location of the eMOTIF, the solvent accessible surface area of the conserved amino acids, and the PRINTS accession number denoting the multiple sequence alignment from which the eMOTIF was derived. The series of panels to the left of the structure display area provide options for displaying different atom representations, and different shading schemes, such as shading by solvent accessible surface area or by amino acid conservation strength (the blue color scheme shown here). The smaller window (foreground) contains a list of all structures containing this eMOTIF and is generated if the user selects the link to view 'all structures containing this eMOTIF' at the top of the main page. Selecting any structure in this smaller window loads it into the main viewing window and highlights the eMOTIF.

eMOTIF if the eMOTIF has a structural example. In this way, the user can seamlessly move from the eMOTIF suite of sequence analysis tools to the structural information displayed in 3MOTIF. Any similar resource can easily link to 3MOTIF in the same way.

VISUALIZATION

Visualizations in 3MOTIF are displayed using the freely available Chime plugin (<http://www.mdlchime.com/>

chime/). For users who cannot run Chime in their browsers, every 3MOTIF visualization page provides the option to download visually equivalent RasMol scripts.

3MOTIF encodes multiple types of information in the visualization of conserved motif residues. We discuss two of these encodings here—the visualization of conservation 'strength' at conserved positions and the visualization of the chemical environments of amino acids at those positions. By conservation strength, we refer to the degree of amino acid variability allowed at a given sequence position in a motif. In 3MOTIF, all motif residues in the displayed structure are colored according to positional variability (see **Supplementary information** for details). For example, positions specified by a single amino acid appear as a brighter blue than positions described by an amino acid substitution group. In this way, 3MOTIF provides a visual cue for assessing which residues of a motif are more strongly conserved (Fig. 1).

As mentioned above, an important feature of 3MOTIF is the encoding of chemical environments of motif residues. One way this is done in 3MOTIF is through the calculation and display of solvent accessible surface area. In any 3MOTIF visualization, the top of the page displays the motif's overall solvent accessible surface area in \AA^2 , as well as the average relative solvent accessibility of the amino acids in the motif (see **Supplementary information** for details). Solvent accessible surface area data for individual motif amino acids can also be mapped directly onto the structure. This can be done either quantitatively, displaying the numerical values as labels attached to the residues in the structure, or chromatically as a green color gradient similar in concept to that of the conservation strength shading scheme discussed above and shown in Figure 1.

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.
- Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.*, **30**, 235–238.
- Henikoff, S., Henikoff, J.G. and Pietrokovski, S. (1999) Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics*, **15**, 471–479.
- Huang, J.Y. and Brutlag, D.L. (2001) The eMOTIF database. *Nucleic Acids Res.*, **29**, 202–204.
- Laskowski, R.A. (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.*, **29**, 221–222.
- Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S. and Overington, J.P. (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**, 617–623.
- Stothard, P.M. (2001) COMBOSA3D: combining sequence alignments with three-dimensional structures. *Bioinformatics*, **17**, 198–199.