

# ProBiS: a web server for detection of structurally similar protein binding sites

Janez Konc<sup>1</sup> and Dušanka Janežič<sup>1,2,\*</sup>

<sup>1</sup>National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana and <sup>2</sup>University of Primorska, Faculty for Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, 6000 Koper, Slovenia

Received February 27, 2010; Revised April 29, 2010; Accepted May 13, 2010

## ABSTRACT

**A web server, ProBiS, freely available at <http://probis.cmm.ki.si>, is presented. This provides access to the program ProBiS (Protein Binding Sites), which detects protein binding sites based on local structural alignments. Detailed instructions and user guidelines for use of ProBiS are available at the server under 'HELP' and selected examples are provided under 'EXAMPLES'.**

## INTRODUCTION

The biochemical function of a protein with known 3D structure can be elucidated by searching for proteins with similar folding patterns and known functions. However, similar folding alone is not a guarantee of similar biochemical function and conversely, the same biochemical function can be performed by differently folded proteins. It now seems clear that the binding sites in a protein, rather than its folding patterns, are a primary determinant of its biochemical function (1–4).

In this article, we introduce ProBiS (protein binding sites), a web server for the recognition of similar surface regions in a database of non-redundant protein structures. The web server ProBiS, is based on the ProBiS algorithm that has been described previously (5–7).

ProBiS first defines the solvent accessible surface by rolling a probe of 1.4 Å radius over the protein atoms represented as van der Waals spheres, then a region ~4 Å below the surface is defined as surface structure for comparison. Therefore, residues that are near the surface, but do not directly contact the ligand, will also be labeled as surface residues. ProBiS then compares this entire surface structure of a query protein with no reference to known binding sites to each of about 24 000 non-redundant protein structures in a database which is updated weekly. Structures with surface regions with geometry and physicochemical properties similar to those in the query structure are retrieved.

The surface structure being compared is searched by the ProBiS algorithm (7) for all possible similarities and the similar regions are identified based on our maximum clique approach. Each maximum clique, i.e. its rotational–translational variation, represents a rigid, local similarity, which is then used to locally superimpose the two compared protein structures. Finally the two compared structures are subject to local alignment of their backbones, which are conserved but have different conformations in the two compared proteins. At this point, the ProBiS can detect conserved structure buried under the protein surface.

Structural conservation scores are calculated for all conserved amino acid residues of the query protein and reveal the extent to which a particular residue appears in the local structural alignments that were found within the protein database. These scores are represented as different colors on the query protein structure.

Given a structure of a protein with unknown binding sites, ProBiS suggests the regions on its surface which may be involved in binding with small ligands, proteins or DNA/RNA. Alternatively, given a protein with an identified binding site, ProBiS finds other proteins with structurally or physicochemically similar binding sites. If used as a pairwise structure alignment program, ProBiS detects and superimposes similar functional sites in a pair of submitted protein structures, even when these do not have similar folds.

A number of web servers for local structural alignment, or for detection of similar 3D structural motifs, have recently become available. These include eF-seek (8), FunClust (9), MegaMotifBase (10), MolLoc (11), MultiBind (12), PAR-3D (13), and PINTS (14). In contrast to these programs, ProBiS performs pairwise local structural alignments of an entire query protein surface to several thousand protein structures in a completely unsupervised mode, and in a reasonable time (<1 h), thus enabling the discovery of previously unknown binding sites in the query protein structure.

\*To whom correspondence should be addressed. Tel: +386 1 476 0321; Fax: +386 1 476 0300; Email: [dusa@cmm.ki.si](mailto:dusa@cmm.ki.si)

## HARDWARE AND SOFTWARE

The ProBiS server consists of 18 computers each with two Intel Xeon 2.26GHz processors. Each processor consists of four cores and each core of two threads. Each submitted job is assigned to the first free computer and 16 ProBiS threads are run in parallel on this computer.

For the most efficient use of ProBiS, Java should be enabled in a browser. The Jmol applet (<http://www.jmol.org>), which is a molecular viewer used to visualize ProBiS results, requires that Java Runtime Environment be installed. If Java is not enabled, ProBiS will still work but with a reduced functionality, in which a static picture, rendered in PyMOL (<http://www.pymol.org>), of the query protein with its residues colored by the grades of similarity will be displayed.

## ProBiS INPUT

ProBiS performs binding site detection on the query structure by multiple pairwise comparison of a query structure with the database of currently some 24 000 protein structures (the number of proteins currently in the database, which is updated weekly, is posted on the input page). ProBiS can also perform pairwise local comparisons of two structures. The data required to use ProBiS are shown in Figure 1.

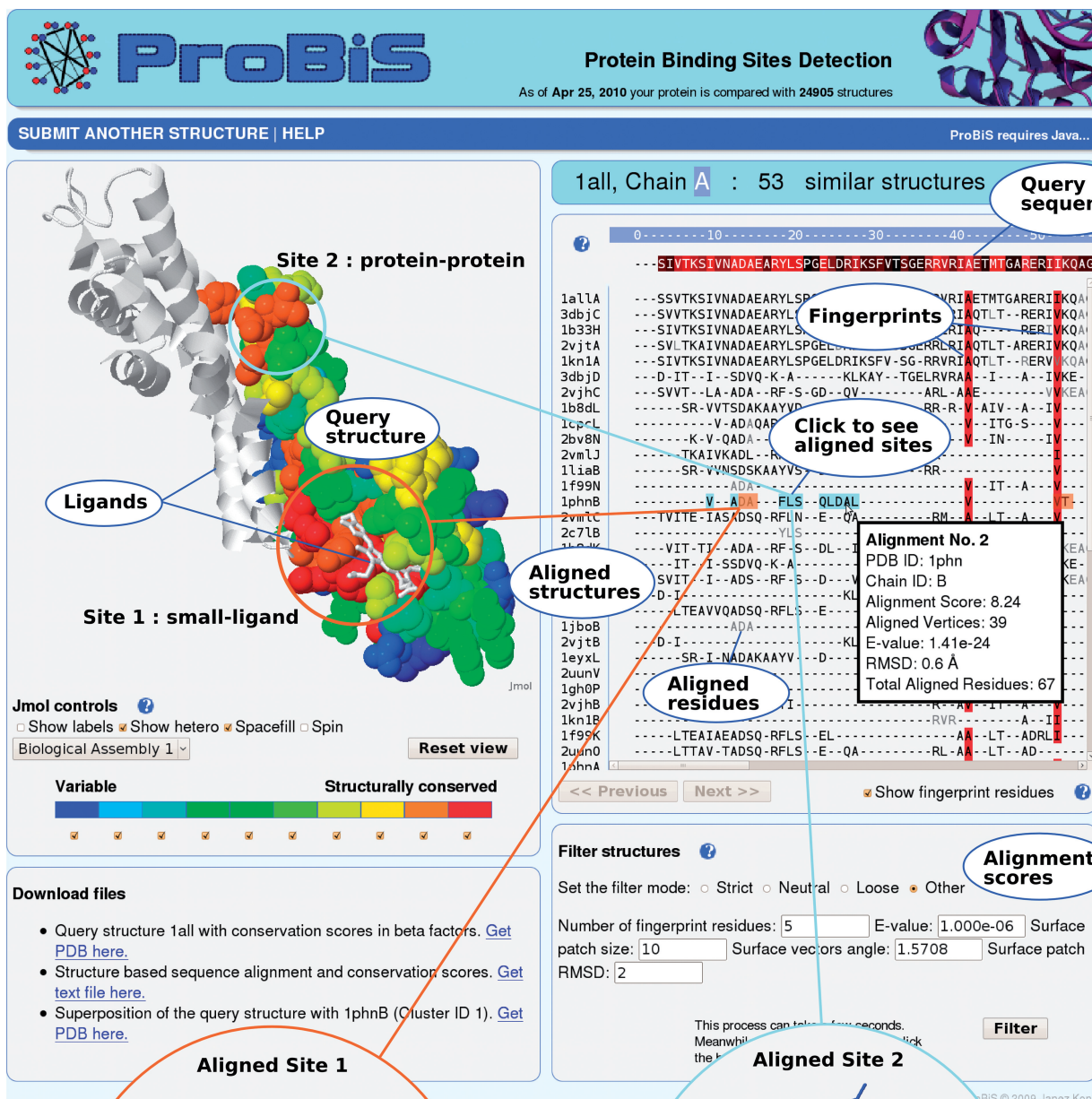
## ProBiS OUTPUT

The ProBiS results page shown in Figure 2 uses an integrated Jmol molecular viewer, in which the input query structure is colored by the discrete structural conservation scores from unconserved (blue) to conserved (red).

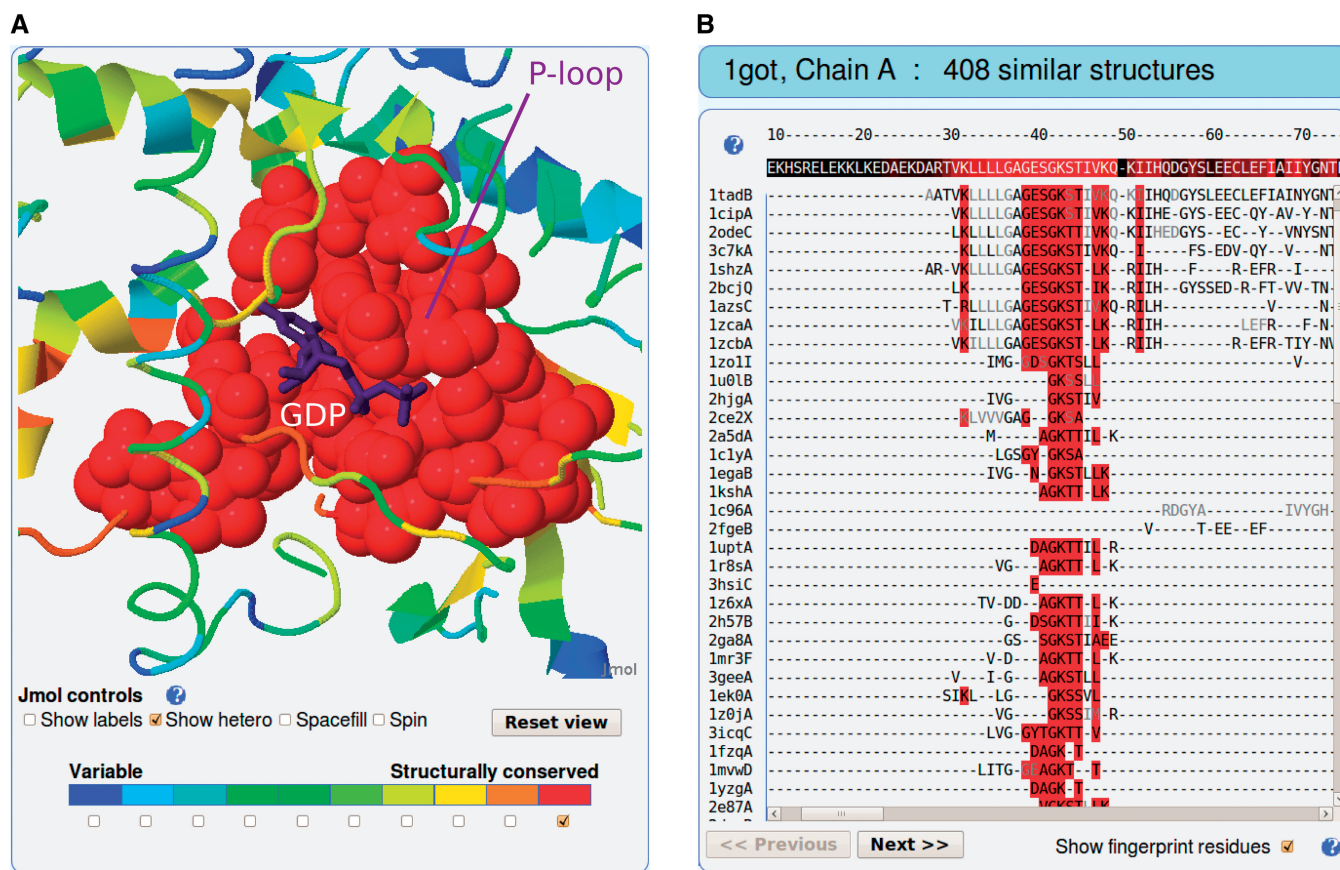
A structure-based sequence alignment (SBSA) box on the right contains local structural alignments of the query with each aligned database protein, represented in tabular form as sequence alignments. At the top of the SBSA box, the query protein's sequence is shown with its amino acids highlighted from unconserved (black) to conserved (red). The panel below contains the first 100 aligned proteins, listed in order of decreasing alignment lengths. Each row of this panel contains the data for one aligned protein, characterized by its PDB code, Chain ID and a tabular list of aligned residues, where each aligned residue is in the same vertical column as its corresponding residue in the query protein; user may use the 'Next' and 'Previous' buttons to navigate between different panels.

Faded, gray colored residues in the SBSA box adopt different conformations in the aligned and in the query protein structures (e.g. see in Figure 2, row with the aligned protein 1jboB and amino acid motif ADA), dark residues are structurally well conserved and define the rotation and translation that superimposes the specific aligned protein onto the query protein. There may be many different local superpositions of an aligned protein

**Figure 1.** ProBiS input page. The PDB ID or an uploaded PDB file; Chain ID(s) (case sensitive); optional email address to receive the results.



**Figure 2.** ProBiS output page. Conserved binding sites and two different local structural alignments (the two insets) of the query allophycocyanin (PDB ID: 1al1) with a retrieved similar protein (PDB ID: 1phn) are shown.



**Figure 3.** Detection of GDP binding site in G-protein. (A) The GDP binding site with the structurally conserved P-loop. Residues with high structural conservation scores are marked red. (B) The P-loop motif GESGKST is designated as the most structurally conserved part of this protein; these residues are detected as the fingerprint residues.

to the query protein and positioning the mouse cursor on an aligned structure's conserved residues, opens a small window with information about this specific alignment, such as alignment number, PDB code and Chain ID of the aligned protein, total alignment length, *E*-value (expectation value) and RMSD between the C- $\alpha$  atoms of the superimposed residues.

Clicking on any part of a retrieved protein sequence, which is highlighted in light-blue color when the cursor is over it, will display the local superposition of the query and the corresponding database protein in the Jmol viewer on the left. When present in the database or the query protein PDB formatted structure file, ligands, designated by the HETATM keyword, are also shown in the Jmol viewer. Ions are displayed as space-filling models, larger ligands are represented as stick models. If Java is not installed, downloading of a PDB formatted file containing the locally superimposed proteins will be initiated. If the 'Show fingerprint residues' checkbox is checked, fingerprint residues, which may be parts of conserved binding sites, are shown as red vertical bars.

Alignment scores determine the number of proteins that are presented in the structure-based sequence alignment. Adjusting the criteria to which each local structural

alignment must conform then pressing the 'Filter' button, leads to recalculation of the web-page with new alignment scores.

### EXAMPLE OF DETECTION OF BINDING SITES

ProBiS was used to detect binding sites in the heterotrimeric G-protein (PDB ID: 1got) which relays hormonal signals from transmembrane receptors to intracellular effectors. This study used the  $\alpha$ -subunit (1got, Chain ID: A) as the input structure for ProBiS. The  $\alpha$ -subunit and the  $\beta$ - $\gamma$  subunit bind to one another and to a GDP molecule. The G-protein contains several possible binding sites, and is thus a useful test of the ability of ProBiS to detect authentic binding sites. In Figure 3A the G-protein, color coded to indicate the degree of structural conservation, is shown. The program retrieved 408 locally similar protein structures and predictions of binding sites were obtained by calculating the structural conservation scores for the query protein residues based on these 408 retrieved structures. The results are at <http://probis.cmm.ki.si/examples.html> and aligned structures are shown in Figure 3B.

## PERFORMANCE

A pairwise alignment of two proteins, each of around 180 amino acid residues takes about 1 s to compute. The time to query the non-redundant protein structures database ranges from 10 min for a query protein with about 180 residues (e.g. 1ytfA), to 50 min for a query protein with 450 residues (e.g. 1bncA).

## CONCLUSION

ProBiS allows local structural alignments of proteins and, with no prior knowledge of binding sites, detects these sites independently of the sequence and the fold of the proteins.

## FUNDING

Ministry of Higher Education, Science and Technology of Slovenia; Slovenian Research Agency (P1-0002). Funding for open access charge: National Institutes of Chemistry, Ljubljana, Slovenia.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Russell,R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol.*, **279**, 1211–1227.
2. Todd,A.E., Orengo,C.A. and Thornton,J.M. (1999) Evolution of protein function, from a structural perspective. *Curr. Opin. Chem. Biol.*, **3**, 548–556.
3. Hegyi,H. and Gerstein,M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.*, **288**, 147–164.
4. Leis,S., Schneider,S. and Zacharias,M. (2010) In silico prediction of binding sites on proteins. *Curr. Med. Chem.*, **17**, 1550–1562.
5. Carl,N., Konc,J. and Janezic,D. (2008) Protein surface conservation in binding sites. *J. Chem. Inf. Model.*, **48**, 1279–1286.
6. Konc,J. and Janezic,D. (2007) Protein-protein binding-sites prediction by protein surface structure conservation. *J. Chem. Inf. Model.*, **47**, 940–944.
7. Konc,J. and Janezic,D. (2010) ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics*, **26**, 1160–1168.
8. Kinoshita,K., Murakami,Y. and Nakamura,H. (2007) eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape. *Nucleic Acids Res.*, **35**, W398–W402.
9. Ausiello,G., Gherardini,P.F., Marcatili,P., Tramontano,A., Via,A. and Helmer-Citterich,M. (2008) FunClust: a web server for the identification of structural motifs in a set of non-homologous protein structures. *BMC Bioinformatics*, **9**, S2.
10. Pugalenti,G., Suganthan,P.N., Sowdhamini,R. and Chakrabarti,S. (2008) MegaMotifBase: a database of structural motifs in protein families and superfamilies. *Nucleic Acids Res.*, **36**, D218–D221.
11. Angaran,S., Bock,M.E., Garutti,C. and Guerra,C. (2009) MolLoc: a web tool for the local structural alignment of molecular surfaces. *Nucleic Acids Res.*, **37**, W565–W570.
12. Shulman-Peleg,A., Shatsky,M., Nussinov,R. and Wolfson,H.J. (2008) MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Res.*, **36**, W260–W264.
13. Goyal,K., Mohanty,D. and Mande,S.C. (2007) PAR-3D: a server to predict protein active site residues. *Nucleic Acids Res.*, **35**, W503–W505.
14. Stark,A. and Russell,R.B. (2003) Annotation in three dimensions. PINTS: patterns in non-homologous tertiary structures. *Nucleic Acids Res.*, **31**, 3341–3344.