W214–W221 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* *Published online 16 May 2012*
doi:10.1093/nar/gks435

ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins

Janez Konc and Dušanka Janežič*

National Institute of Chemistry, Ljubljana, Slovenia

Received February 18, 2012; Revised April 17, 2012; Accepted April 25, 2012

ABSTRACT

The ProBiS web server is a web server for detection of structurally similar binding sites in the PDB and for local pairwise alignment of protein structures. In this article, we present a new version of the ProBiS web server that is 10 times faster than earlier versions, due to the efficient parallelization of the ProBiS algorithm, which now allows significantly faster comparison of a protein query against the PDB and reduces the calculation time for scanning the entire PDB from hours to minutes. It also features new web services, and an improved user interface. In addition, the new web server is united with the ProBiS-Database and thus provides instant access to pre-calculated protein similarity profiles for over 29 000 non-redundant protein structures. The ProBiS web server is particularly adept at detection of secondary binding sites in proteins. It is freely available at http://probis.cmm.ki.si/old-version, and the new ProBiS web server is at http://probis.cmm.ki.si.

INTRODUCTION

Detection of structural similarities in proteins can be applied to many open questions. These include elucidation of the biochemical functions of newly characterized proteins (1–4), identification of novel indications for existing drugs—drug repositioning (5–7), and prediction of interactions of known drugs with secondary targets, or off-targets that may lead to undesirable side effects (8,9). However, comparison of only the sequence and folding of proteins fails to address these problems because protein binding sites, rather than protein folding, control interaction with ligands and hence biochemical function (10). Web servers that allow the detection of local similarities in proteins have been developed (11–21), and there is an increasing number of approaches that deal with drug repositioning and off-target prediction from different perspectives (22–24).

Two programs developed at the National Institute of Chemistry in Ljubljana, ProBiS web server (25,26) and ProBiS-Database (27) enable the detection of structurally similar protein binding sites and local pairwise alignment of crystallographically or NMR determined protein structures from the PDB. ProBiS-Database holds pre-calculated structural similarity profiles for over 29 000 non-redundant proteins in the PDB, and allows access to these results in seconds. The ProBiS web server allows *de novo* similarity calculations, using the ProBiS algorithm, for any protein in the PDB. ProBiS is different from most other structural alignment algorithms, in that it can align proteins having different folds, if they share similar binding sites.

ProBiS conducts searches for similar three-dimensional structural regions in proteins without reference to known binding sites or co-crystallized ligands, and takes into account entire protein surfaces. It accepts a protein structure as a query, and compares it against each protein in the non-redundant PDB (nr-PDB). The nr-PDB, updated weekly, is derived from the entire PDB which currently has ~182 000 protein single chain structures (28). All these single chain protein structures are clustered with >95% sequence identical structures and a representative of each cluster is chosen. These 29 000 representatives identified in this way, constitute the nr-PDB; each is identified with a PDB number and chain, as in 2q4u.A. The ProBiS algorithm represents the surfaces of compared proteins as protein graphs, i.e. structures of vertices and edges, the vertices corresponding to functional groups of surface amino acid residues and the edges determined by distances between pairs of adjacent vertices. This representation captures both geometric as well as physicochemical characteristics of protein surfaces. ProBiS compares the query protein to each of the database proteins, using the maximum clique algorithm (29), which allows it to efficiently detect the largest similar subgraphs of compared

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

protein graphs. After the comparison of the query protein to every nr-PDB protein structure is complete, the degrees of structural conservation are calculated for all amino acid residues in the query protein. These are analogous to degrees of sequence conservation, represented, e.g. in sequence logos with amino acid letters of different sizes, and reveal the frequency of occurrence of a particular residue in the local structural alignments that were found in the nr-PDB. These degrees of structural conservation are represented as different colors on the query protein structure as in Figure 3A, and often indicate the position of binding or other functionally important sites.

The latest ProBiS web server shown in Figure 1 has a number of powerful features. It has an order of magnitude faster computation time, pre-calculated results and an improved user interface. All the functions of the ProBiS web server can now be accessed fully automatically from user scripts through the new ProBiS web services based on RESTful (Representational State Transfer) technology (30). The new additions to the ProBiS web server invite users to explore similarities among binding sites in

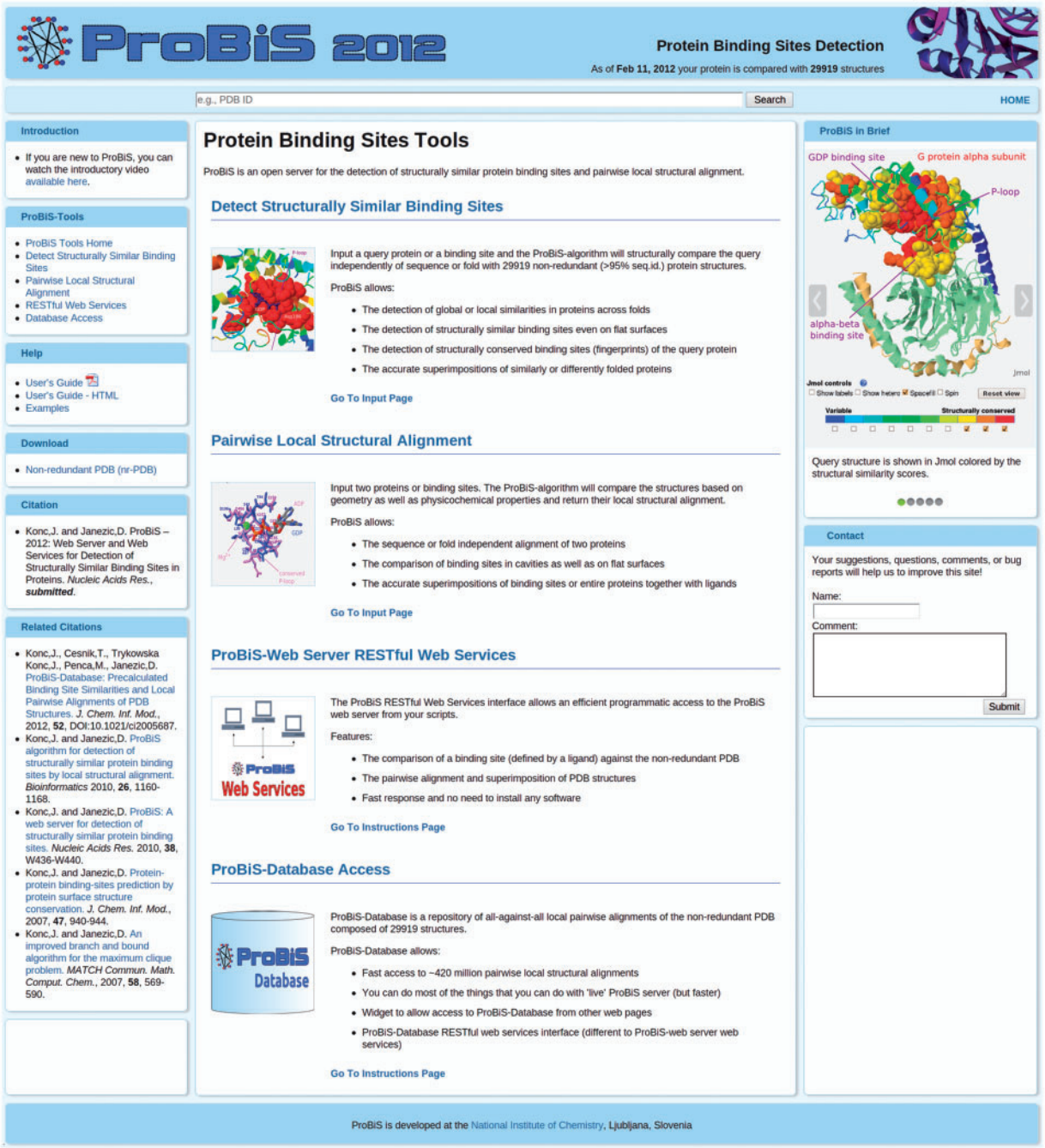

Figure 1. ProBiS web server page provides access to the Protein Binding Sites Tools.

proteins, and provide increased annotation of uncharacterized protein structures, drug repositioning and off-target predictions. Details regarding the new ProBiS are available at http://probis.cmm.ki.si.

ProBiS WEB SERVER

Input

An example of the input to ProBiS is shown in Figure 2. The user first defines the query protein by either providing PDB/Chain ID(s) or uploading a PDB file. In the first case, a yellow pop-up window appears, containing a link to the 'Local Structural Similarity' web page with the pre-calculated results for a homologous member of the nr-PDB. Following this link, the results are taken directly from the ProBiS-Database, and no computations will be performed; in the other case, one continues with *de novo* search for structural similarities. This can optionally be limited to a specific region on the query protein surface, for example, a binding site, by clicking the 'Select Motif' button, shown in Figure 2. This opens a new browser window with the 3D query protein model. A binding site, or any other part of the protein, can then be selected by clicking on the query protein 3D model; the surface atoms within a radius of 15 Å around this selected point are highlighted in yellow. This default value of 15 Å can be changed in the text box above the 'Select & Close Window' button. When the window is closed, the selected, highlighted surface is converted to a text format in the 'Residue Motif' text box, and can be immediately used as the input to the ProBiS web server. This enables selection of binding sites to be used as queries. The comparison database can also be changed in the new ProBiS web server and the query can now be compared against either the non-redundant PDB (default) or a custom list of any protein structures. The latter are represented by the PDB/Chain IDs, e.g. 1all.A, 3dbj.C, 2vjt.A, which can be entered into the text box that opens when the user selects the 'List of PDB/Chain IDs' option in the 'Comparison Database' drop-down list.

Access to the pre-calculated similarity profiles in the 'ProBiS-Database' is provided through the search text box located on the top of the ProBiS web server page as shown in Figure 2, and through ProBiS-Database widget or ProBiS-Database web services, which have been described previously (27).

Output

The ProBiS output page shown in Figure 3 contains the query protein cartoon model 2q4u.A colored according to degrees of structural conservation from unconserved (blue) to structurally conserved (red) and visualized in an integrated Jmol molecular viewer (see panel A). The structurally conserved residues are shown as red spheres and indicate the location of the putative binding sites.

There is also an interactive table of available similar proteins (see panel B). Clicking on the 'View' link in the 'Alignments' column shows the superimposition of the query (2q4u.A) and the similar protein (2ehb.A), and opens the 'Details' tab of residue–residue correspondences of all alignments between the query and the similar database protein as shown in panels C and D. Clicking on any protein chain link in the 'Chain' column opens a new output page with the pre-calculated structural similarity profile for a homologous member of the nr-PDB. The 'Name' column presents the names of the similar proteins; the 'Pfam', 'SCOP' and 'UniProt' columns provide links, where available (31), to the corresponding external protein annotation resources. The similar

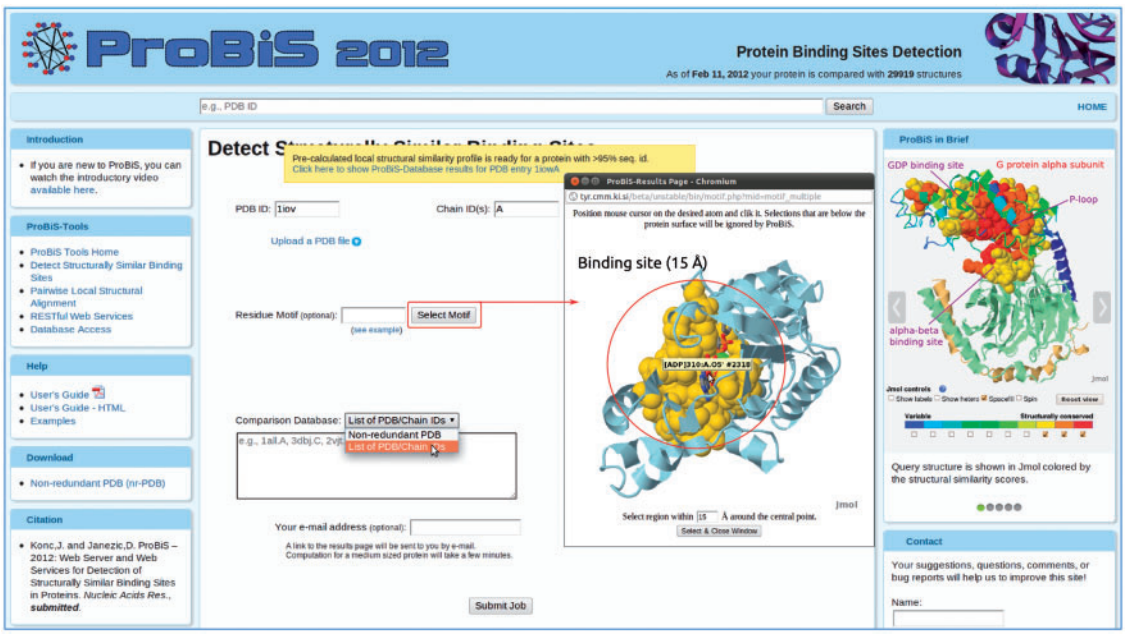

Figure 2. Detection of Structurally Similar Binding Sites. ProBiS input page. Detailed instructions are provided in the User's Guide at http://probis.cmm.ki.si.

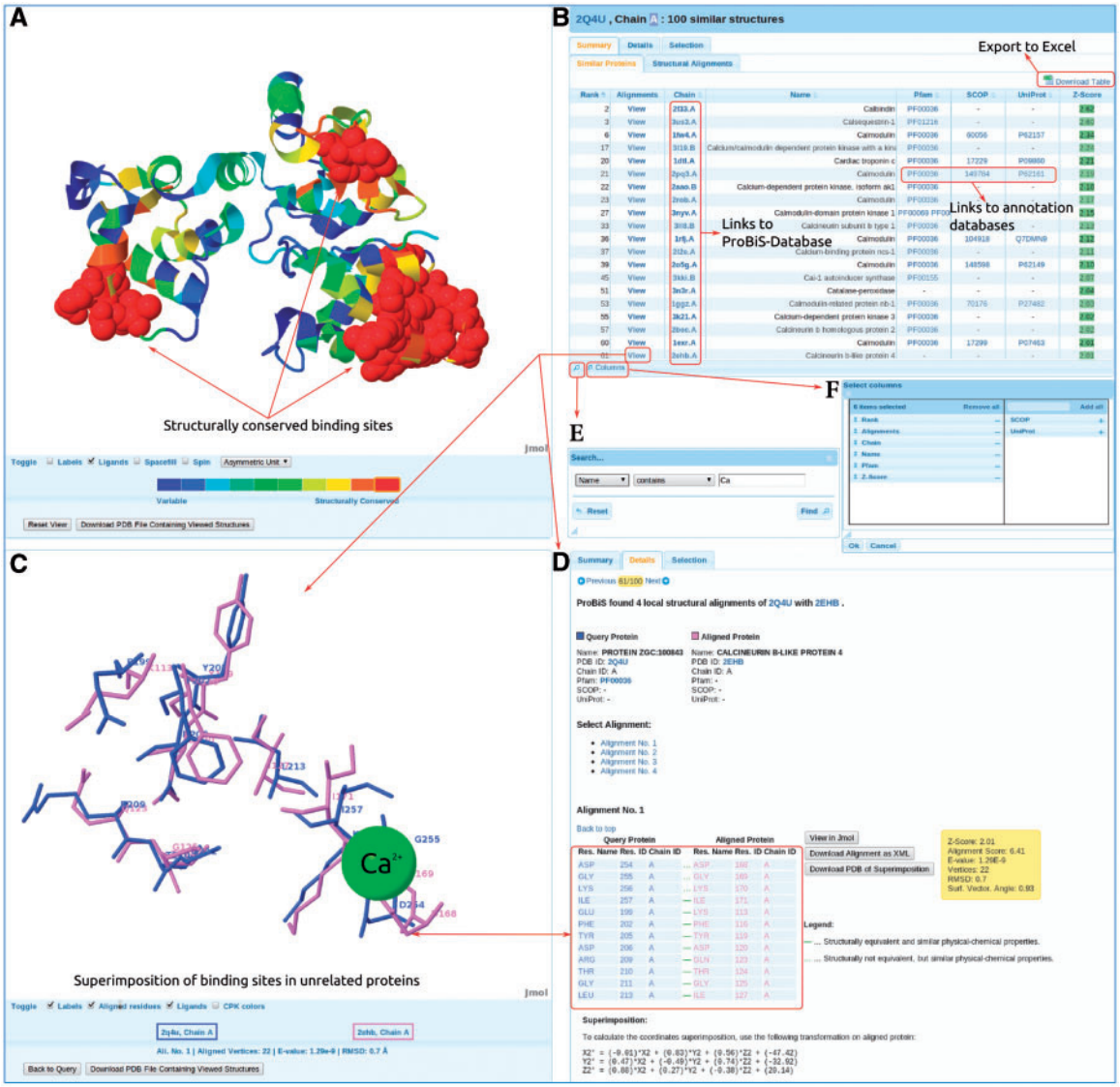

Figure 3. ProBiS output page for query protein (PDB/Chain ID: 2q4u.A) of unknown function. (**A**) Structurally conserved binding sites. (**B**) An interactive table of similar proteins, filtered to show only calcium binding proteins. (**C**) Local structural superimposition of the blue query (2q4u.A) and the violet similar protein (2ehb.A) that contains a co-crystallized Ca^{2+} ion shown as green sphere. (**D**) The alignments presented as tables of residue-residue correspondences. (**E**) Integrated search tool. (**F**) Columns reordering tool.

proteins in the interactive table are ranked by the standard *Z*-scores; *Z*-Scores >2.0 are colored green and those between 1.0 and 2.0 are yellow. The table can be downloaded in a CSV format, and thus directly exported to a spreadsheet program, such as Excel.

Local structural superimposition of a query protein, 2q4u.A in blue, and a similar protein 2ehb.A in violet, is shown in Figure 3C. The calcium ion was co-crystallized in 2ehb.A. Alignments of 2q4u.A with 2ehb.A are also shown in tables of residue-residue correspondences in panel D; the alignment 1 is that seen in panel C. Where available (31), the PDB, Pfam, SCOP and UniProt accession numbers are at the top, and are links to external databases for structural and functional protein annotation. Below, the pairwise local alignments are presented in tabular form, ranked according to their *Z*-Scores. A continuous green dash connecting a pair of aligned residues indicates a good structural correspondence; an interrupted green dash indicates a poorer correspondence between the residues. The 'Download' buttons allow downloading the alignment in various formats, and the 'View in Jmol' button loads the alignment as shown in Figure 3C. Alignment scores for each pairwise alignment are shown in a yellow box, and are explained in detail in references (26,27).

Filtering of the table by different search conditions is accomplished by the integrated search tool, shown in panel E. Here, the table was filtered, so that the protein names in 'Name' column must contain a 'Ca' keyword, which filters out all but calcium-related proteins.

The table columns can also be reordered, shown or hidden, using the 'Reorder Columns' tool in panel F.

ProBiS web services

The new ProBiS web server uses RESTful web services to provide ready access from user scripts to the binding site similarities and local pairwise alignments for any PDB protein structure (30). Specification of the web services interface input data, a full set of commands and useful examples can be found at the 'ProBiS-Web server RESTful Web Services' instructions page at http://probis.cmm.ki.si/?what=webservices. The results of the web services calculations are returned in XML, Json or PDB formats, which are well supported in modern programming languages.

NEW FEATURES IN THE 2012 ProBiS WEB SERVER

Faster calculation

In 2010, a batch script ran the non-parallel version of the ProBiS algorithm on 16 processors of a single computer (25). In 2012, the ProBiS algorithm (26) was parallelized using the Open-MPI library, and now runs on ~250 processors of the ProBiS web server, which has shortened search time from hours to minutes.

Integration with the ProBiS-Database

The new ProBiS web server features integrated access to the ProBiS-Database (27), which is a searchable repository of local pairwise alignments of non-redundant protein structures (nr-PDB) generated by the ProBiS algorithm. This database consists of ~420 million pre-calculated pair-wise alignments and presents a faster alternative to the standard *de novo* protein similarity detection used by the ProBiS web server: structural similarity results are obtained in seconds from the ProBiS-Database. However, the database holds results only for the ~29 000 nr-PDB proteins; for non-nr-PDB proteins, results are for the closest homologue in the nr-PDB.

Improved user interface

New features in ProBiS include: (i) submission of a binding site as a query—previously only complete proteins could be used as queries; (ii) comparison of the query protein against a user-provided list of PDB/Chain IDs. Previously, this was possible only against proteins in the nr-PDB; (iii) links to other protein annotation resources, such as Pfam, SCOP or UniProt; (iv) extended download options—results can be downloaded as CSV, XML, Json

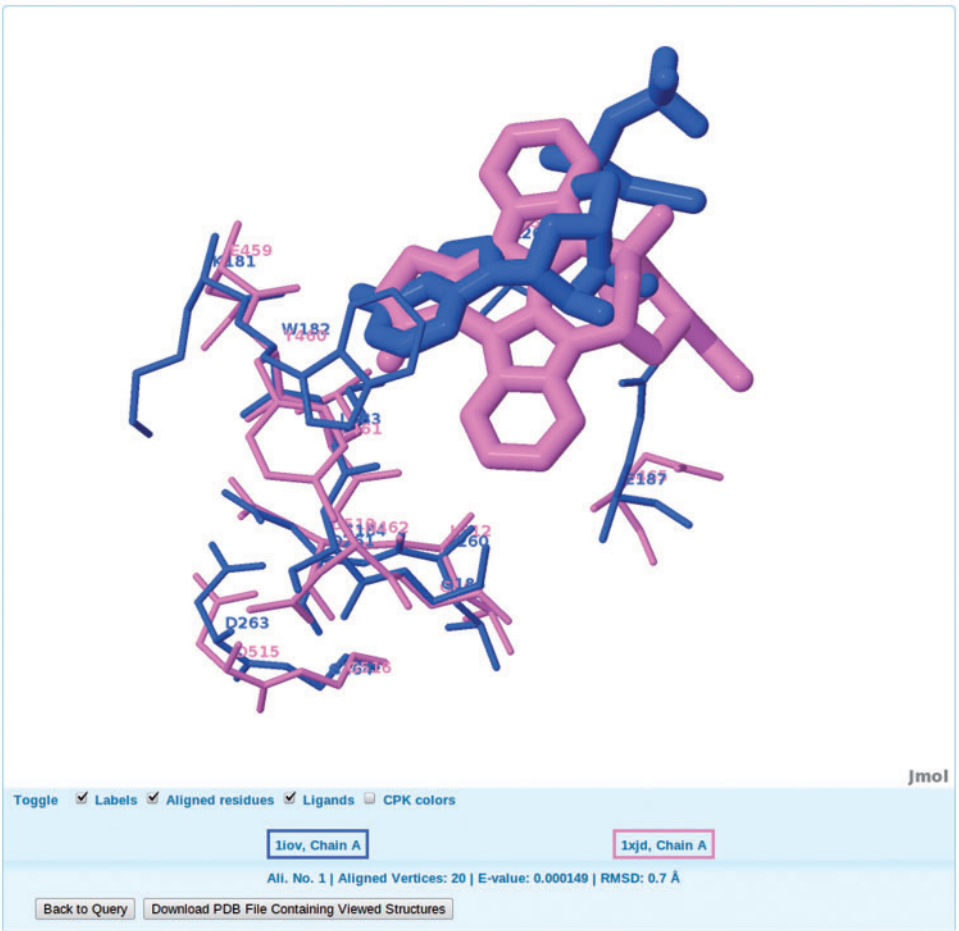

Figure 4. Superimposition of similar binding sites in Ddl (1iov.A; blue) and PKC (1xjd.A; violet). The superimposed residues from Ddl and protein kinase C are shown as thin wireframe models. The co-crystallized ligands, ADP (blue) and staurosporine (violet), are shown as thick wireframe models.

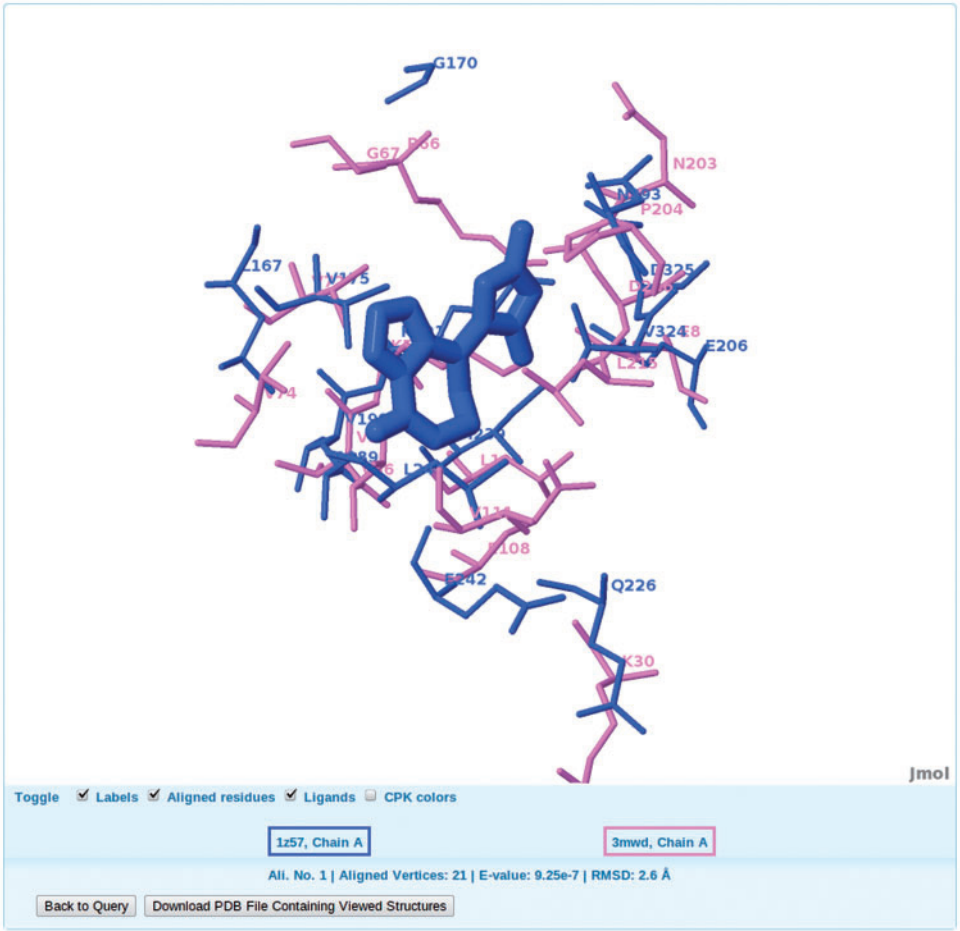

Figure 5. Superimposition of similar binding sites in dual-specificity kinase (1z57.A; blue) and ATP-citrate synthase (3mwd.A; violet). The superimposed residues are thin wireframe models. The inhibitor debrohymenialdisine is colored blue and is a thick wireframe model.

or PDB files; and (v) searching within the table of similar proteins. A complete list of new features is available in the User's Guide at http://probis.cmm.ki.si.

ProBiS web services

RESTful web services have been implemented to allow access to the ProBiS web server by user written programs or scripts. A list of commands available is on the ProBiS web server home page.

Methodological improvements

Statistically meaningful ranking of the identified locally similar proteins or binding sites is achieved through the use of standard scores (*Z*-Scores) (27), which replace the various alignment scores described previously (26).

EXAMPLES

Binding sites detection

For the query protein (PDB/Chain ID: 2q4u.A), ProBiS accurately detects three conserved binding sites as shown in Figure 3A. The Ca^{2+} ion, which is co-crystallized in the similar protein calcineurin B (2ehb.A), is shown in Figure 3C, and reveals a probable binding pose of the Ca^{2+} in the query protein. Proteins 2q4u.A and 2ehb.A have similar binding sites, despite their low amino acid identity, i.e. between 2 and 10%, as judged by different pair-wise structural alignment methods in the '3D Similarity tab' at the RCSB PDB Web page (28). ProBiS also predicts binding of 5 other Ca^{2+} ions to the query protein, which is not shown in Figure 3; these can be seen by viewing alignments 1–4 in Jmol for the similar protein calmodulin (1fw4.A).

Drug repositioning

ProBiS can detect weak binding site similarities in proteins with different protein folds. Here, we present an example of drug repositioning. Protein kinase inhibitors were considered for inhibition of bacterial enzyme D-alanine–D-alanine ligase (Ddl), based on the similarities between ATP binding sites in protein kinases and in Ddl (32,33). However, repositioning was only done for a small proportion of all available protein kinases, which manifested this similarity. Using the ATP binding site of Ddl, 1iov.A, shown in Figure 2, as the query, ProBiS detects a previously unrecognized similarity between this binding site

and an ATP binding site in protein kinase C (PKC; 1xjd.A); the superimposition of these two binding sites is shown in Figure 4. Here, the ADP ligand of Ddl is superimposed upon the PKC inhibitor, staurosporine, as a consequence of the alignment of the two similar binding sites. This supports the suggestion that staurosporine may bind to Ddl, and that it should be experimentally tested for inhibition of Ddl.

Off-target prediction

Many inhibitors have been developed to compete for the ATP binding sites of protein kinases, but their poor selectivity usually eliminates them from consideration as clinical agents (34). The kinase 1z57.A is important in the control of alternative splicing and selective inhibitors targeting its ATP binding site have been developed. Debrohymenialdisine, for example, co-crystallized with human dual-specificity kinase, has been reported as 1z57.A (35), and we found that the ATP binding site of this protein kinase is similar to an ATP binding site in human ATP-citrate synthase (3mwd.A) as shown in Figure 5. Sequence identity of 1z57.A and 3mwd.A proteins is between 4 and 8% as computed by the FATCAT (36) and DaliLite (37) structural alignment algorithms. However, these two algorithms are not able to detect the binding site similarity. This result thus suggests that debrohymenialdisine, inhibiting the ATP binding site in this human dual-specificity protein kinase, would probably have undesirable side effects in human patients, due to off-target binding to ATP-citrate synthase.

SOFTWARE REQUIREMENTS

The ProBiS web server requires Sun Java plugin Version 6 Update 26 or higher (http://www.java.com), and has been shown to function correctly with Firefox, IE8, Chrome 14.0, Safari 5.1 and Opera 11.5 web browsers. It also works with OpenJDK (IcedTea-Web 1.1.1) plugin on Firefox.

CONCLUSION

ProBiS is a web server for detection of local structural similarities in proteins. It allows detection of similar three-dimensional patterns of residues in protein structures irrespective of protein folds and with no prior knowledge of binding sites. ProBiS enables the detection of similar binding sites in differently folded proteins, and can suggest protein targets amenable to drug repositioning. It can also be used to generate hypotheses for protein functions and for the prediction of off-target effects. To our knowledge, there is no such comprehensive, freely available web server that would allow these functions in this automated and intuitive manner. ProBiS can provide useful insights to experimentalists, and can directly suggest molecules that have a potential value in pharmaceutical applications.

FUNDING

Ministry of Higher Education, Science and Technology of Slovenia; Slovenian Research Agency [P1-0002, Z1-3666]. Funding for open access charge: National Institute of Chemistry, Ljubljana, Slovenia.

Conflict of interest statement. None declared.

REFERENCES

1. Jaroszewski,L., Li,Z., Krishna,S.S., Bakolitsa,C., Wooley,J., Deacon,A.M., Wilson,I.A. and Godzik,A. (2009) Exploration of uncharted regions of the protein universe. *PLoS Biol.*, **7**, e1000205.
2. Musiani,F., Bellucci,M. and Ciurli,S. (2011) Model structures of Helicobacter pylori UreD(H) domains: a putative molecular recognition platform. *J. Chem. Inf. Model.*, **51**, 1513–1520.
3. Wong,M.T., Choi,S.B., Kuan,C.S., Chua,S.L., Chang,C.H., Normi,Y.M., Too,W.C., Wahab,H.A. and Few,L.L. (2012) Structural modeling and biochemical characterization of recombinant KPN_02809, a zinc-dependent metalloprotease from Klebsiella pneumoniae MGH 78578. *Int. J. Mol. Sci.*, **13**, 901–917.
4. Kar,G., Keskin,O., Nussinov,R. and Gursoy,A. (2012) Human proteome-scale structural modeling of E2-E3 interactions exploiting interface motifs. *J. Proteome Res.*, **11**, 1196–1207.
5. Ashburn,T.T. and Thor,K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug. Discov.*, **3**, 673–683.
6. Haupt,V.J. and Schroeder,M. (2011) Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief. Bioinform.*, **12**, 312–326.
7. Kinnings,S.L., Liu,N., Buchmeier,N., Tonge,P.J., Xie,L. and Bourne,P.E. (2009) Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.*, **5**, e1000423.
8. Xie,L., Evangelidis,T. and Bourne,P.E. (2011) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput. Biol.*, **7**, e1002037.
9. Defranchi,E., Schalon,C., Messa,M., Onofri,F., Benfenati,F. and Rognan,D. (2010) Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One*, **5**, e12214.
10. Russell,R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol.*, **279**, 1211–1227.
11. 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.
12. Ito,J., Tabei,Y., Shimizu,K., Tsuda,K. and Tomii,K. (2012) PoSSuM: a database of similar protein-ligand binding and putative pockets. *Nucleic Acids Res.*, **40**, D541–D548.
13. 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.
14. 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.
15. Parca,L., Mangone,I., Gherardini,P.F., Ausiello,G. and Helmer-Citterich,M. (2011) Phosfinder: a web server for the identification of phosphate-binding sites on protein structures. *Nucleic Acids Res.*, **39**, W278–W282.
16. Regad,L., Saladin,A., Maupetit,J., Geneix,C. and Camproux,A.C. (2011) SA-Mot: a web server for the identification of motifs of

interest extracted from protein loops. *Nucleic Acids Res.*, **39**, W203–W209.
17. Shirvanyants,D., Alexandrova,A.N. and Dokholyan,N.V. (2011) Rigid substructure search. *Bioinformatics*, **27**, 1327–1329.
18. Jambon,M., Andrieu,O., Combet,C., Deleage,G., Delfaud,F. and Geourjon,C. (2005) The SuMo server: 3D search for protein functional sites. *Bioinformatics*, **21**, 3929–3930.
19. Stark,A. and Russell,R.B. (2003) Annotation in three dimensions. PINTS: patterns in non-homologous tertiary structures. *Nucleic Acids Res.*, **31**, 3341–3344.
20. 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.
21. 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.
22. Ren,J., Xie,L., Li,W.W. and Bourne,P.E. (2010) SMAP-WS: a parallel web service for structural proteome-wide ligand-binding site comparison. *Nucleic Acids Res.*, **38**, W441–W444.
23. von Eichborn,J., Murgueitio,M.S., Dunkel,M., Koerner,S., Bourne,P.E. and Preissner,R. (2011) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res.*, **39**, D1060–D1066.
24. Luo,H., Chen,J., Shi,L., Mikailov,M., Zhu,H., Wang,K., He,L. and Yang,L. (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. *Nucleic Acids Res.*, **39**, W492–W498.
25. Konc,J. and Janezic,D. (2010) ProBiS: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.*, **38**, W436–W440.
26. Konc,J. and Janezic,D. (2010) ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics*, **26**, 1160–1168.
27. Konc,J., Cesnik,T., Trykowska Konc,J., Penca,M. and Janezic,D. (2012) ProBiS-Database: precalculated binding site similarities and local pairwise alignments of PDB structures. *J. Chem. Inf. Model.*, **52**, 604–612.
28. 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.
29. Konc,J. and Janezic,D. (2007) An improved branch and bound algorithm for the maximum clique problem. *MATCH Commun. Math. Comput. Chem.*, **58**, 569–590.
30. Fielding,R.T. and Taylor,R.N. (2002) Principled design of the modern web architecture. *ACM Trans Internet. Technol.*, **2**, 115–150.
31. Velankar,S., McNeil,P., Mittard-Runte,V., Suarez,A., Barrell,D., Apweiler,R. and Henrick,K. (2005) E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Res.*, **33**, D262–D265.
32. Skedelj,V., Tomasic,T., Masic,L.P. and Zega,A. (2011) ATP-binding site of bacterial enzymes as a target for antibacterial drug design. *J. Med. Chem.*, **54**, 915–929.
33. Triola,G., Wetzel,S., Ellinger,B., Koch,M.A., Hubel,K., Rauh,D. and Waldmann,H. (2009) ATP competitive inhibitors of D-alanine-D-alanine ligase based on protein kinase inhibitor scaffolds. *Bioorg. Med. Chem.*, **17**, 1079–1087.
34. Huang,D., Zhou,T., Lafleur,K., Nevado,C. and Caflisch,A. (2010) Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. *Bioinformatics*, **26**, 198–204.
35. Fedorov,O., Huber,K., Eisenreich,A., Filippakopoulos,P., King,O., Bullock,A.N., Szklarczyk,D., Jensen,L.J., Fabbro,D., Trappe,J. *et al.* (2011) Specific CLK inhibitors from a novel chemotype for regulation of alternative splicing. *Chem. Biol.*, **18**, 67–76.
36. Ye,Y. and Godzik,A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, **19(Suppl. 2)**, 246–255.
37. Holm,L. and Park,J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics*, **16**, 566–567.