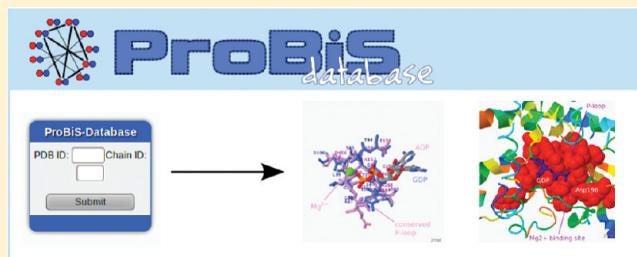


ProBiS-Database: Precalculated Binding Site Similarities and Local Pairwise Alignments of PDB Structures

Janez Konc, Tomo Česnik, Joanna Trykowska Konc, Matej Penca, and Dušanka Janežič*

National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia

ABSTRACT: ProBiS-Database is a searchable repository of precalculated local structural alignments in proteins detected by the ProBiS algorithm in the Protein Data Bank. Identification of functionally important binding regions of the protein is facilitated by structural similarity scores mapped to the query protein structure. PDB structures that have been aligned with a query protein may be rapidly retrieved from the ProBiS-Database, which is thus able to generate hypotheses concerning the roles of uncharacterized proteins. Presented with uncharacterized protein structure, ProBiS-Database can discern relationships between such a query protein and other better known proteins in the PDB. Fast access and a user-friendly graphical interface promote easy exploration of this database of over 420 million local structural alignments. The ProBiS-Database is updated weekly and is freely available online at <http://probis.cmm.ki.si/database>.



INTRODUCTION

Many different questions can be addressed by detection of structural similarities in proteins. These include elucidation of the biochemical functions of newly characterized proteins,^{1,2} prediction of side-effects of known drugs that bind to proteins other than their initial target (off-targets),³ and repositioning of ligands between similar binding sites in different proteins to find a new indication for an old drug.^{4,5} However, comparison of only the folds in proteins fails to shed light on these problems⁶ because the binding sites in a protein rather than its folding patterns control its binding to ligands and hence its biochemical function.^{7–10} Methods for the detection of local structural similarities^{11,12} and computational resources that deal with similar problems^{13,14} have been developed.

Here, we describe ProBiS-Database, a searchable repository of local pairwise alignments of nonredundant protein structures generated by the ProBiS algorithm.^{15,16} ProBiS compares entire protein surfaces in a local manner by searching for similar three-dimensional structural motifs in pairs of proteins without reference to known binding sites or co-crystallized ligands.¹⁵ It retrieves structures that possess surface regions with geometrical and physicochemical properties similar to those in a query protein. The algorithm represents the surfaces of compared proteins as protein graphs, i.e., as 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. It uses a filtering step, which removes nonsimilar protein graph pairs beforehand,¹⁷ and a maximum clique algorithm to compare these protein graphs efficiently.¹⁸ As a consequence, the ProBiS algorithm is able to compare complete protein structures rather than preselected residue motifs, and this facilitates the detection of similar binding sites. Many local alignments between two proteins can be detected, and each

such local alignment is represented by a rotational and translation variation that optimally superimposes a patch of surface residues from each of the proteins. ProBiS has been shown to successfully align binding sites in protein structures with dissimilar folding patterns.¹⁵ Structural similarity scores that are calculated for all amino acid residues in the query protein reveal the frequency of occurrence of a particular residue in the local structural alignments that were found in the protein database. These scores are represented as different colors on the query protein structure.

The initial version of ProBiS-Database was built in 2011 from the PDB of 181,882 protein single chains.⁶ All these single-chain protein structures are clustered with >95% sequence identical structures, and a representative of each cluster is chosen.¹⁵ Surface residues of the selected representative proteins are identified and converted to protein graph representations, which are saved into 29,266 “surface files” enabling faster pairwise comparisons by ProBiS. The ProBiS algorithm is used to complete an “all against all” comparison of these 29,266 nonredundant protein structures that represent the whole PDB. The resulting pairwise local structural alignments that are detected among these nonredundant proteins constitute the ProBiS-Database.

A standard comparison with ProBiS algorithm, available at <http://probis.cmm.ki.si>, of a query protein against the non-redundant PDB (nr-PDB) can require hours, but the precalculated local structural similarity profile for a query protein, which gives essentially the same result, can be obtained in seconds from the ProBiS-Database. ProBiS-Database can be linked by other Web pages, e.g., PDBWiki,¹⁹ which provides users of these Web pages with instant access to local structural

Received: December 2, 2011

Published: January 23, 2012



Figure 1. ProBiS-Database home page.

alignments of PDB protein structures. The ranking of local structural alignments is supported by Z-Score, which provides a statistical measure of protein similarity and is described below. ProBiS-Database can be queried with a protein's PDB/Chain ID to identify regions on the protein's surface that may be involved in binding of various ligands. Alternatively, by querying ProBiS-Database with a protein containing an identified binding site, other proteins can be found with structurally or physicochemically similar binding sites, and superimposition of these functional sites and similar site(s) in the query protein can be achieved. ProBiS-Database holds over 420 million precalculated local structural alignments of complete protein surfaces, which span beyond similar protein folding patterns. This enables the detection of known as well as novel similar binding sites in proteins from PDB, even when these do not have structural homologues.

METHODS

ProBiS-Database Access. Figure 1 shows three means of accessing the ProBiS-Database: (a) the search text box, (b) a ProBiS-Database Widget, and (c) the RESTful Web Service Interface.²⁰

ProBiS-Database Search Text Box. The search text box, centrally located at the top of the ProBiS-Database home page, allows searching of the database with PDB ID as the query. After the *Search* button is clicked, the server returns all protein chains for which there is data in the ProBiS-Database as links under *Search Results*. Selection of such a link, identified by PDB/Chain ID, opens the *Local Structural Similarity Profile* Web page for that protein chain or a similar representative protein chain.

ProBiS-Database Widget. A ProBiS-Database widget, a dynamicWeb element, which can easily be included in one's

Web page, allows access to ProBiS-Database. The widget is a javascript program, which accepts a PDB/Chain ID as a query and directs the user to a *Local Structural Similarity Profile* Web page for the query protein chain. Entry of a nonrepresentative PDB structure prompts redirection to the >95% sequence identical representative of the input protein's corresponding cluster. If a query protein is not in the nonredundant-PDB (nr-PDB) (for definition see below), the user is redirected to the *Local Structural Similarity Profile* Web page for the most similar protein from the nr-PDB. The widget's source code is on the ProBiS-Database server and does not require any installation or programming from the user; a single line of HTML code causes it to be included in the Web page source code. Users can also modify the widget's appearance, such as the size and colors, to tailor it to their own Web page design.

ProBiS-Database RESTful Web Service interface. To allow programmatic access to the ProBiS-Database, it is also available through a RESTful (representational state transfer) Web service interface. The data on our Web server can thus be downloaded by other client applications, e.g., other Web pages, scripts, and on remote computers through HTTP protocol in a fully automated way. The interface is defined by a set of HTTP commands that can be used to retrieve data in JSON, XML, or text/plain formats from the ProBiS-Database. A complete list of commands available is on the ProBiS-Database home page. To download any data from the ProBiS-Database, the user may execute the script in Perl language provided on the ProBiS-Database home page.

ProBiS-Database Construction. The ProBiS Web server¹⁶ enables de novo comparisons of protein structures, while ProBiS-Database provides precalculated structural similarity profiles for all nonredundant PDB entries. The construction of the ProBiS-Database involved the steps described below.

Data Set Reduction. The nr-PDB is built from the PDB protein chains and holds more than 29,000 representative protein structures, covering the current protein structural variability in the PDB.

"All against All" Alignments. Structural comparison of each nr-PDB structure with all other nr-PDB structures using ProBiS algorithm, a total of $(29,000)^2/2 = 420 \times 10^6$ computations, was completed in 18 days using a cluster of 14 high performance computers, and the resulting pairwise local alignments are stored in a searchable ~350 GB MySQL database that is updated on a weekly basis as described below.

Entries in the ProBiS-Database. The ProBiS-Database is composed of a main table and tables containing results and alignments. There are some 420×10^6 entries in the main table, each consisting of the PDB/Chain IDs of two compared proteins. An entry in the main table points to one or more entries in the results table, each consisting of a pair of aligned amino acid residues from the two compared proteins. In the results table, residue–residue correspondences that belong to a particular local pairwise alignment are connected with a single entry in the alignments table, which carries different scores, which all describe the quality of that particular local alignment. This entry also holds a rotational matrix and translational vector, which define the superimposition of the two compared proteins in this local alignment. Efficient indexing of the tables guarantees very fast data retrieval from the ProBiS-Database with PDB/Chain ID queries.

Automatic ProBiS-Database Updates. The ProBiS-Database is updated automatically on a weekly basis. First, a new nr-PDB is built as described above, and then the protein

chains that were absent from the previous week's version of nr-PDB are identified. The new representative protein structural chains, currently some 150 each week, are compared by the ProBiS algorithm to all the structures in the new nr-PDB. Data associated with protein chains that are not in the new nr-PDB are removed from the ProBiS-Database. This automated process performed on a single computer requires ~3 days.

Structural Alignment Scores. *Z-Score*, used to measure the statistical and structural significance of local structural alignments in the ProBiS-Database is calculated as follows. First, the alignment score (al_{score}) is calculated for each local alignment on the basis of the different scores described in ref 15 by the equation

$$al_{score} = \log\left(\frac{n_{vert} \times \log(1 + 1/e_{value})}{rmsd}\right)$$

where $rmsd$ is the root mean square deviation between pairs of superimposed vertices, n_{vert} is the number of aligned vertices, and e_{value} is the alignment expectation value calculated by the Karlin–Altschul equation.²¹

The alignment score is then standardized into *Z-Score* as

$$Z\text{-Score} = \frac{al_{score} - \mu}{\sigma}$$

The population mean (μ) and population standard deviation (σ) were calculated from alignment scores for all 420×10^6 structural alignments, and the values of μ and σ are 2.0 and 2.2, respectively. *Z-Score* indicates how many standard deviations al_{score} differs from the mean, e.g., a pairwise alignment with *Z-Score* of 2.0 is in the top ~2% of all alignments in the ProBiS-Database. Local alignments are ranked by their *Z-Scores*, and only alignments with *Z-Score* > 1.0 are shown in the database user interface.

"Hot" Similar Proteins. Similar proteins that are retrieved but belong to a different protein family than the query protein according to the Protein Family (Pfam) classification,²² are designated as "hot" and are marked with a red star in the ProBiS-Database interface. "Hot" proteins often perform a different biochemical function than the query protein. Pfam accession numbers are used in the ProBiS-Database because Pfam database is updated regularly and promptly and covers most of the PDB structures. The concept of "hot" proteins is introduced into the ProBiS-Database interface to enable users to quickly identify globally dissimilar proteins, sharing only local similarities with the query protein and possible examples of convergent evolution.

Software Requirements. ProBiS-Database requires Sun (Oracle) Java plugin version 6 update 26–29 (<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.

RESULTS

ProBiS-Database, a repository of protein local structural alignments, spans across protein fold space. For a PDB/Chain ID as query, the *Local Structural Similarity Profile* Web page is retrieved in seconds from the ProBiS-Database. This Web page contains (1) structurally similar binding sites, (2) local pairwise alignments of the query protein with the nonredundant PDB protein structures, and (3) "Hot" proteins that are of a different protein family than the query protein

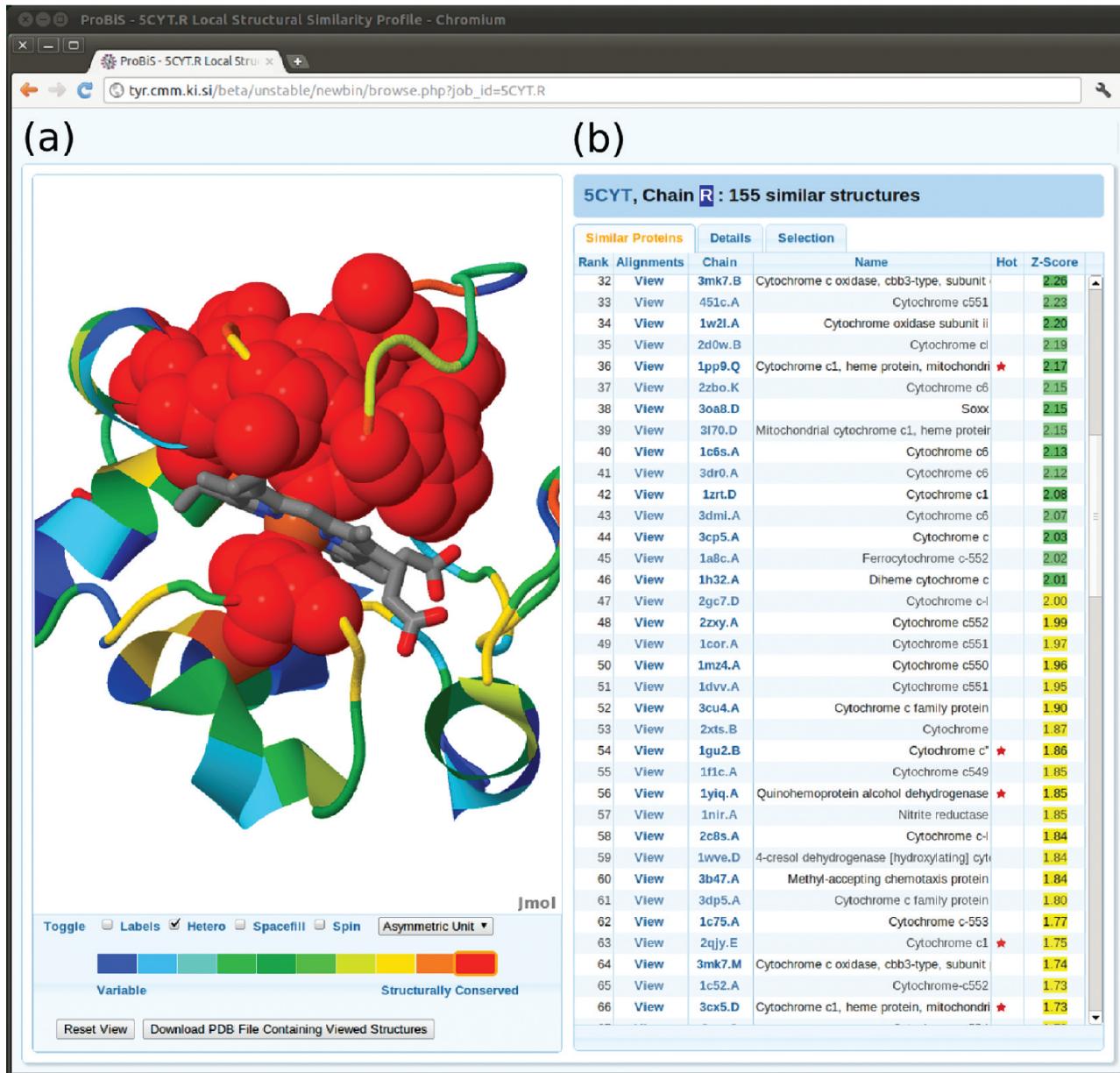


Figure 2. Local structural similarity profile Web page for cytochrome c query protein (PDB/Chain ID: Scyt.R). (a) Similarity scores are mapped onto the three-dimensional cartoon model of the query protein in a Jmol molecular viewer (<http://www.jmol.org>). The heme ligand is shown in the binding site as a wireframe model colored by CPK scheme. A mouse click on the red part of the rainbow-colored band below the Jmol viewer highlights the most structurally similar residues as red colored spheres. (b) An interactive table of similar proteins, which are ranked by their Z-Scores. Click on a View link in the Alignments column opens the Details tab showing all alignments of the query and the similar proteins (see details in Figure 3). Superimposition of the two proteins according to the highest scoring alignments (Alignment No. 1) is shown in the Jmol viewer. Click on a PDB/Chain ID link in the Chain column opens a new Local structural similarity profile Web page for the selected protein chain and thus allows browsing the ProBiS-Database. The Name column presents the names of the similar proteins. A red star in the Hot column indicates that the aligned and query proteins have a different Pfam accession number. Z-Scores > 2.0 are colored green; 1.0 < Z-Score < 2.0 are yellow.

according to Pfam classification but contain a similar surface amino acid motif. The following examples illustrate the various features of the ProBiS-Database. The first three examples describe the technical aspects of the database and the user interface; the latter two deal with the biochemical insights that can be obtained with the ProBiS-Database.

Example 1: Identification of Functionally Important Binding Site Residues. The cytochrome c protein (PDB ID: Scyt) comprises a single polypeptide chain and participates in the electron transport chain by transferring one electron using its heme prosthetic group. With this protein as query, ProBiS-

Database yielded 155 similar protein structures having a similar surface motif, and similarity scores were calculated on the basis of the local alignments of the query protein with all retrieved structures with Z-Score > 2.0. The Local Structural Similarity Profile page for this protein is presented in Figure 2. The three-dimensional model of the query protein is shown on the left in Figure 2, color coded by structural similarity scores, in the Jmol molecular viewer (<http://www.jmol.org>). It is simple to identify functionally important binding site residues, which outline the functional site on this protein, the heme binding site, which is colored red.

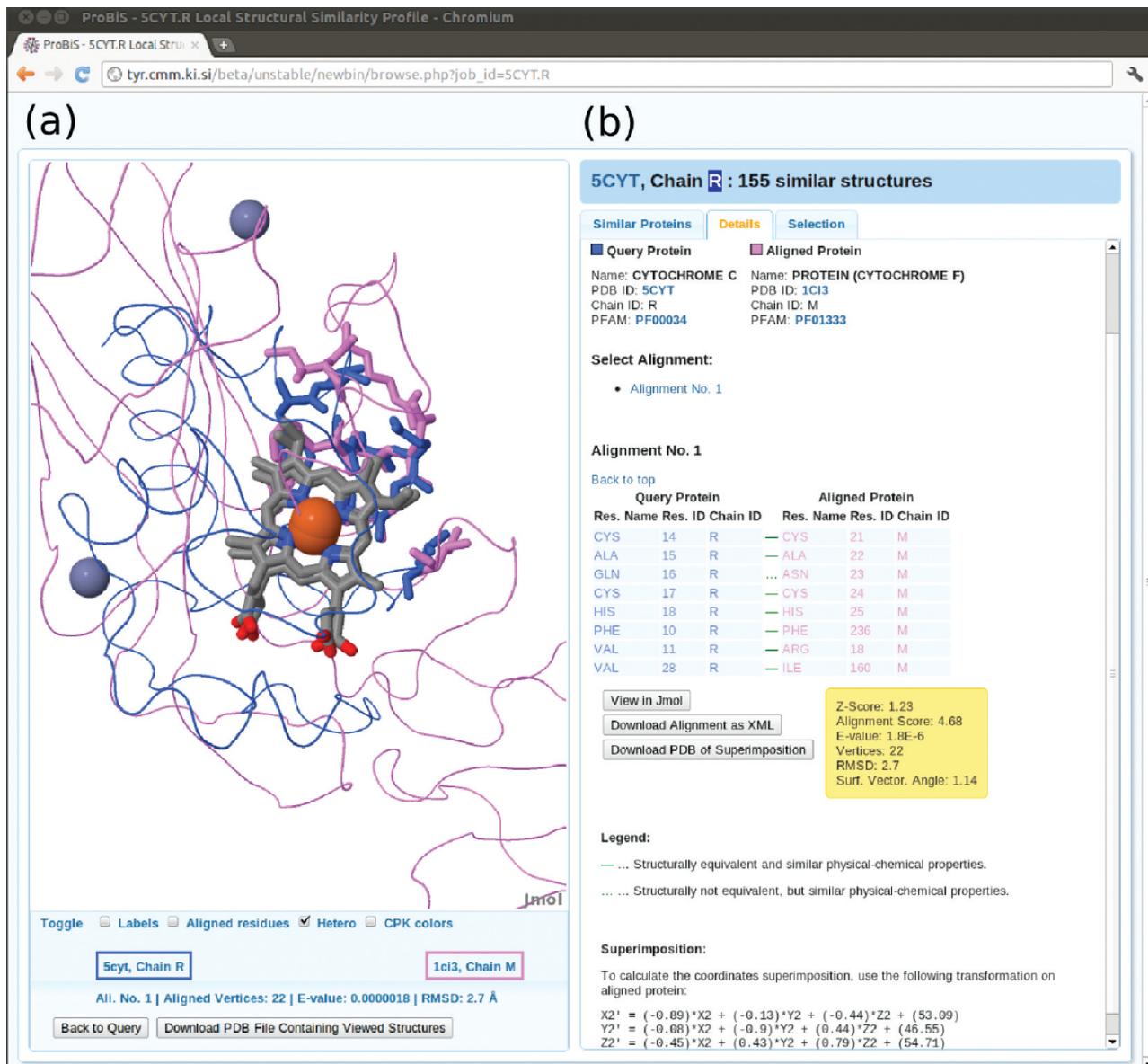


Figure 3. Similar binding sites in protein structures of different Pfam families. To view this example, click the *View* link in the *Alignments* column for protein chain 1ci3.M (rank 116) in the *Similar Proteins* tab on the *Local structural similarity profile* Web page for cytochrome c (Scyt.R). (a) The query protein (Scyt.R) and the aligned protein (1ci3.M) are shown as thin ribbons in blue or violet. Similar binding sites residues are thick wireframe models and are colored according to their respective proteins. The two heme ligands, which are almost perfectly superimposed as a result of the alignment of binding sites residues, are shown in CPK colors as wireframe models; the attached Fe-ions are brown spheres. Thus these two unrelated proteins are linked by ProBiS because they possess a structural motif, the binding site, in common. *Back to Query* button at the bottom resets the Jmol viewer to the original state showing the query protein color coded by structural similarity scores [see Panel (a) in Figure 2]. (b) The *Details* tab for local alignments between the query and similar protein is presented. In general, more than one alignment is possible between the query and the aligned protein structures (in this example only one). The names, PDB IDs, and Pfam accession numbers of the query and the aligned protein are at the top. The PDB and Pfam IDs are clickable links, which take the user to the RCSB PDB Web page or to the Pfam protein annotation database, respectively. Below, the alignments are presented ranked according to their Z-Scores. Each pairwise alignment is shown as a table of residue–residue correspondences between the query and the aligned protein. A continuous green dash connecting a pair of aligned residues indicates a good structural correspondence; 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 shows the alignment in Jmol.

Example 2: Local Pairwise Alignments of PDB Structures. An interactive table of similar proteins appears on the right side of Figure 2. Each of these similar proteins may have many different local pairwise alignments with the query protein; they are ranked by the Z-Score of their highest scoring local pairwise alignment. Similar proteins marked with a red star are “Hot”, which means they are of a different protein family according to the Protein Family (Pfam) classification

system than the query protein.²² In the *Local Structural Similarity Profile* page for cytochrome c in Figure 2, there are 61 “hot” similar proteins; many of these have a fold different from that of the query protein (cytochrome c fold).²³ Among similar proteins are various differently folded proteins, e.g., multiheme cytochrome, cytochrome f, etc. It should be noted that these proteins have no backbone or sequence similarities and thus will not be detected by structural alignment algorithms, which

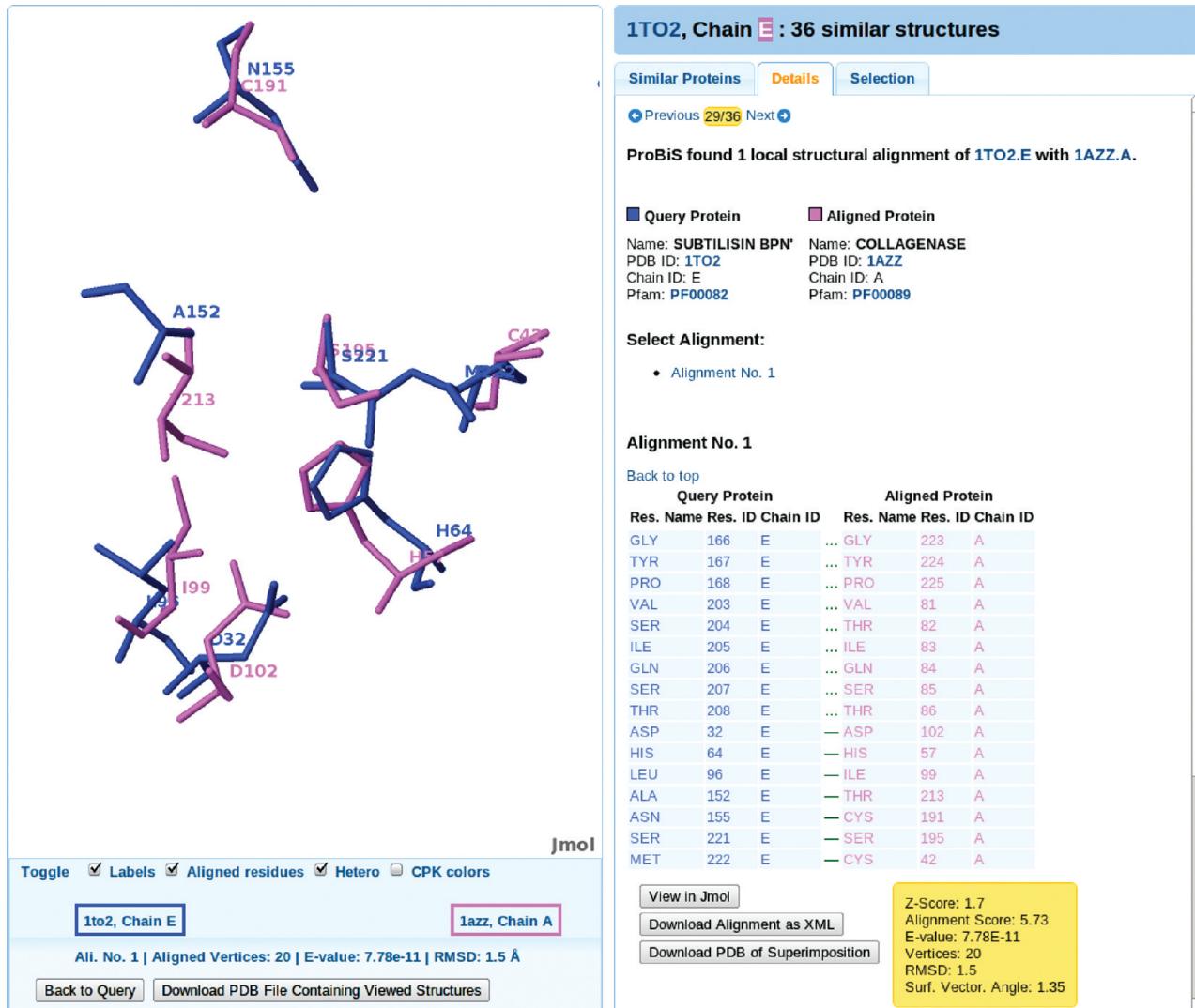


Figure 4. Detection of similarity between convergently evolved binding sites in PDB structures of subtilisin (1to2.E) and trypsin (1azz.A).

compare protein backbones or secondary structure elements.⁶ In the majority of these differently folded proteins, the detected pairwise alignments correspond to amino acids in the heme binding sites of these proteins, and below we present one such example.

Example 3: Similar Binding Sites in Proteins of Different Pfam families. In Figure 3, an example of similar binding sites in “Hot” proteins belonging to different protein families according to Pfam, i.e., cytochrome c (Pfam ID: PF00034) and cytochrome f (PF01333), is presented as provided by the ProBiS-Database.

Example 4: Detection of Convergent Evolution in PDB Structures. ProBiS-Database can also be used to detect weak similarities in proteins with different protein folds. Here, we present a classic example of convergent evolution, i.e., the proteins subtilisin and trypsin, which are evolutionarily unrelated serine proteases with completely different folds but that share the same catalytic mechanism and utilize the same catalytic triad of serine, aspartic acid, and histidine in their binding sites.²⁴ With PDB/Chain ID: 1to2.E (subtilisin fold), we obtain 36 similar proteins, and there are two trypsin-like folds among the “Hot” similar proteins: collagenase (1azz.A) and polyprotein (2fp7.B); an example of the superimposition of

the convergently evolved binding sites in the query subtilisin (1to2.E) and aligned trypsin-like (1azz.A) proteins is shown in Figure 4. The alignment of the catalytic triads in both proteins involves the following residue–residue correspondences: Serine 221–Serine 195, Aspartate 32–Aspartate 102, and Histidine 64–Histidine 57, where the residues in each corresponding pair belong to the query and aligned protein, respectively. These residues are scattered in the sequence of the proteins and thus undetectable by standard sequence or structural alignment algorithms. ProBiS-Database enables the detection of protein similarities in differently folded proteins, which in turn enables functional annotation of proteins that have no structural homologues in the PDB database. To our knowledge, there is no such comprehensive computational approach that would allow discovery of such weak similarities in this automated and intuitive manner.

Example 5: Functional Annotation of PDB Structure from Structural Genomics. Protein ne0167 (PDB/Chain ID: 3k6c.H) is a protein recently deposited in the RCSB PDB by the Midwest Center for Structural genomics.⁶ It is uncharacterized and has no significant sequence similarity to any of the known PDB structures. Using the structural alignment methods in the 3D Similarity tab at the RCSB PDB Web page (<http://>

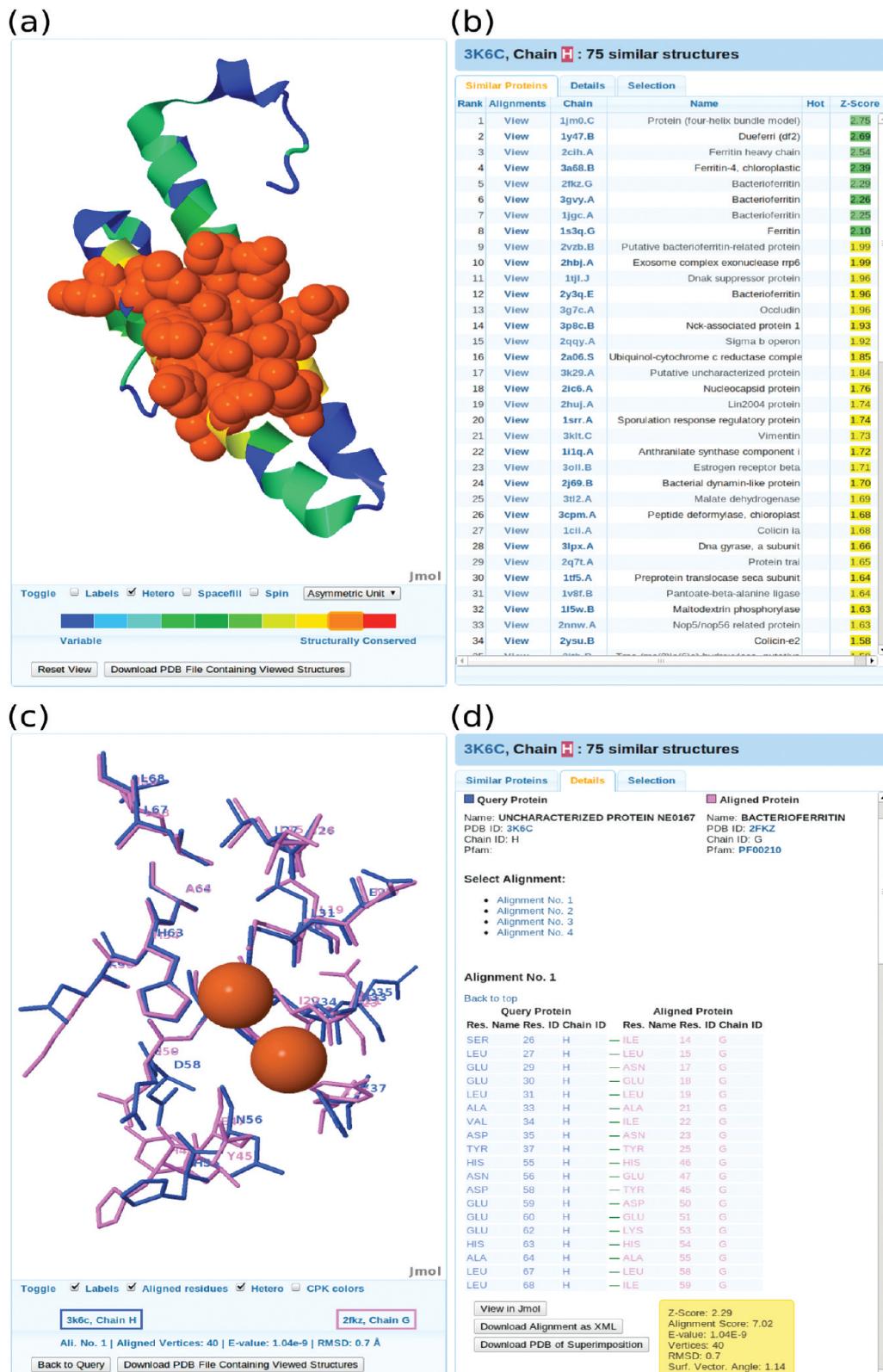


Figure 5. Functional annotation of uncharacterized query protein ne0167 (PDB/Chain ID: 3k6c.H) from a Structural genomics project. (a) Putative binding site is shown as spacefill model and is colored orange. The rest of the query protein is shown as a cartoon model. (b) Among the top similar proteins are ferritin heavy chain (2cih.A), chloroplastic ferritin 4 (3a68.B), and various bacterioferritins (2fkz.G, 3gvy.A, 1jgc.A, and 2vzb.B). The top ranked proteins (1jm0.C and 1y4dt.B) are de novo designed protein structures that bind Fe^{2+} ions. (c) Superimposition of putative binding site in query protein (3k6c.H) and known Fe^{2+} ion binding site in bacterioferritin (2kfz.G). The two Fe^{2+} ions co-crystallized with bacterioferritin are brown spheres. (d) The detailed pairwise alignment of query and bacterioferritin proteins.

www.rcsb.org) provides no unambiguous structural similarities to other PDB protein structures, with the highest scoring alignment (*Golgi to ER traffic protein 1*) having a sequence identity with the query protein of only 6.78%. The similarities obtained at that Web page are too weak to allow a definitive functional annotation of this query protein.

ProBiS-Database provides the following answers about this protein's binding sites and function:

- (1) The similarity scores mapped onto the query protein structure indicate a putative binding site region, which is colored orange in panel (a) of Figure 5.
- (2) Among the most similar proteins found by ProBiS-Database are various iron-binding protein structures, for example, ferritin heavy chain (2cih.A), chloroplastic ferritin 4 (3a68.B), and various bacterioferritins (2fkz.G, 3gvy.A, 1jgc.A, and 2vzb.B), as shown in panel (b) of Figure 5. With Z-Scores > 2.0, these protein structures are significantly similar to the query protein.
- (3) A detailed structural alignment with the bacterioferritin protein (2fkz.G) reveals a significant structural correspondence between amino acid residues in the ferritin Fe²⁺-binding site region and residues of the uncharacterized protein, as shown in panels (c) and (d) of Figure 5. The Fe²⁺ ions, which are co-crystallized in bacterioferritin, are shown in panel (c) of Figure 5, and reveal a probable binding pose of these divalent ions in the query protein (3k6c.H).

Our results reveal that the uncharacterized protein *ne0167* is an iron-binding protein, most likely a previously unknown form of bacterioferritin. Although the global structure of this protein is distantly similar to that in many other proteins, the functional annotation of *ne0167* has to date evaded definition. In such difficult cases functional annotation can only be achieved by finding local similarities with known binding sites. ProBiS-Database is clearly useful in this respect, and it has the potential to become a classic tool for protein functional annotation.

CONCLUSIONS

ProBiS-Database is a repository of local structural similarities between all nonredundant protein structures. It allows detection of similar three-dimensional residue patterns in protein structures irrespective of protein folds and with no prior knowledge of binding sites. The purpose of ProBiS-Database is to generate hypotheses for protein functions, but it can also be used for detection of off-targets and for detection of sites possibly valuable for drug repositioning.

Every new structure may provide new clues as of functions of proteins, and so the weekly updated ProBiS-Database always contains the most recently reported protein structures. In contrast to the ProBiS Web server,¹⁶ the results are precalculated, guaranteeing rapid response to queries.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dusa@cmm.ki.si.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support through Grants P1-0002 and Z1-3666 of the Ministry of Higher Education, Science, and Technology of Slovenia and the Slovenian Research Agency is acknowledged.

REFERENCES

- (1) Jaroszewski, L.; Li, Z.; Krishna, S. S.; Bakolitsa, C.; Wooley, J.; Deacon, A. M.; Wilson, I. A.; Godzik, A. Exploration of Uncharted Regions of the Protein Universe. *PLoS Biol.* **2009**, *7*, e1000205.
- (2) Musiani, F.; Bellucci, M.; Ciurli, S. Model Structures of Helicobacter Pylori UreD(H) Domains: A Putative Molecular Recognition Platform. *J. Chem. Inf. Model.* **2011**, *51*, 1513–1520.
- (3) Xie, L.; Evangelidis, T.; Xie, L.; Bourne, P. E. Drug Discovery Using Chemical Systems Biology: Weak Inhibition of Multiple Kinases May Contribute to the Anti-Cancer Effect of Nelfinavir. *PLoS Comput. Biol.* **2011**, *7*, e1002037.
- (4) Haupt, V. J.; Schroeder, M. Old Friends in New Guise: Repositioning of Known Drugs with Structural Bioinformatics. *Brief. Bioinform.* **2011**, *12*, 312–326.
- (5) Skedelj, V.; Tomasic, T.; Peterlin Masic, L.; Zega, A. ATP-Binding Site of Bacterial Enzymes as a Target for Antibacterial Drug Design. *J. Med. Chem.* **2011**, *54*, 915–929.
- (6) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (7) Russell, R. B. Detection of Protein Three-Dimensional Side-Chain Patterns: New Examples of Convergent Evolution. *J. Mol. Biol.* **1998**, *279*, 1211–1227.
- (8) Kuhn, D.; Weskamp, N.; Schmitt, S.; Hullermeier, E.; Klebe, G. From the Similarity Analysis of Protein Cavities to the Functional Classification of Protein Families Using Cavbase. *J. Mol. Biol.* **2006**, *359*, 1023–1044.
- (9) Martin, J. Beauty Is in the Eye of the Beholder: Proteins Can Recognize Binding Sites of Homologous Proteins in More Than One Way. *PLoS Comput. Biol.* **2010**, *6*, e1000821.
- (10) Shirvanyants, D.; Alexandrova, A. N.; Dokholyan, N. V. Rigid Substructure Search. *Bioinformatics* **2011**, *27*, 1327–1329.
- (11) Xie, L.; Bourne, P. E. Detecting Evolutionary Relationships across Existing Fold Space, Using Sequence Order-Independent Profile–Profile Alignments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5441–5446.
- (12) Jambon, M.; Andrieu, O.; Combet, C.; Deléage, G.; Delfaud, F.; Geourjon, C. The SuMo Server: 3D Search for Protein Functional Sites. *Bioinformatics* **2005**, *21*, 3929–3930.
- (13) Liao, C.; Sitzmann, M.; Pugliese, A.; Nicklaus, M. C. Software and Resources for Computational Medicinal Chemistry. *Future Med. Chem.* **2011**, *3*, 1057–1085.
- (14) Teyra, J.; Samsonov, S. A.; Schreiber, S.; Pisabarro, M. T. SCOWLP Update: 3D Classification of Protein–protein, –Peptide, –Saccharide and –Nucleic Acid Interactions, and Structure-Based Binding Inferences across Folds. *BMC Bioinformatics* **2011**, *12*, 398.
- (15) Konc, J.; Janezic, D. ProBiS Algorithm for Detection of Structurally Similar Protein Binding Sites by Local Structural Alignment. *Bioinformatics* **2010**, *26*, 1160–1168.
- (16) Konc, J.; Janezic, D. ProBiS: A Web Server for Detection of Structurally Similar Protein Binding Sites. *Nucleic Acids Res.* **2010**, *38*, W436–W440.
- (17) Konc, J.; Janezic, D. Protein–protein Binding-Sites Prediction by Protein Surface Structure Conservation. *J. Chem. Inf. Model.* **2007**, *47*, 940–944.
- (18) Konc, J.; Janezic, D. An Improved Branch and Bound Algorithm for the Maximum Clique Problem. *MATCH Commun. Math. Comput. Chem.* **2007**, *58*, 569–590.
- (19) Stehr, H.; Duarte, J. M.; Lappe, M.; Bhak, J.; Bolser, D. M. PDBWiki: Added Value through Community Annotation of the Protein Data Bank. *Database* **2010**, *2010*, DOI:10.1093/database/baq009.
- (20) Fielding, R. T.; Taylor, R. N. Principled Design of the Modern Web Architecture. *ACM Trans. Internet Technol.* **2002**, *2*, 115–150.
- (21) Karlin, S.; Altschul, S. F. Methods for Assessing the Statistical Significance of Molecular Sequence Features by Using General Scoring Schemes. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2264–2268.
- (22) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P.; Sammut, S. J.; Hotz, H.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. L.;

Bateman, A. The Pfam Protein Families Database. *Nucleic Acids Res.* **2008**, *36*, D281–D288.

(23) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* **1995**, *247*, 536–540.

(24) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4523.