

---

# ProBiS Web Server

2012

## User's Guide

September 07, 2012

Laboratory for Molecular Modeling  
National Institute of Chemistry  
Hajdrihova 19  
1000 Ljubljana, Slovenia  
[www.sicmm.org](http://www.sicmm.org)  
Support: [konc@cmm.ki.si](mailto:konc@cmm.ki.si)  
Collaborations: [dusa@cmm.ki.si](mailto:dusa@cmm.ki.si)

# Table Of Contents

Background.....	3
Software Requirement.....	4
Linux - Ubuntu.....	4
Sun Java (Version 6 Update 30).....	4
OpenJDK (IcedTea Version 1.1.1).....	4
Windows 7.....	4
Sun Java (Version 6 Update 30).....	4
New User Interface Features.....	4
Submit a Binding Site as a Query.....	4
Change the Default Comparison Database.....	5
Tabs.....	5
Similar Proteins Tab.....	5
Structural Alignments Tab.....	5
Search Tool.....	5
Details Tab.....	5
Selection Tab.....	5
Progress Indicator.....	6
ProBiS-Database Widget.....	6
Jmol Help.....	6
Protein Binding Sites Tools.....	7
Detect Structurally Similar Binding Sites.....	8
Input.....	9
Simple Query.....	9
Precalculated Results.....	10
Advanced Query.....	10
Output.....	12
Pairwise Local Structural Alignment.....	20
Input.....	21
Output.....	22
ProBiS Web Server RESTful Web Services .....	22
Input.....	22
Pairwise Alignment of PDB Structures .....	22
Detect Similar Proteins or Binding Sites in the Non-redundant PDB.....	23
Perl Example Script.....	23
Output.....	24
ProBiS-Database Access.....	26
Input.....	26
The ProBiS-Database Widget.....	27
The ProBiS-Database RESTful Web Service Interface.....	27
Perl Example Script.....	28
Output.....	28
FAQ.....	29

Can I upload a protein model?.....	29
Can I upload a PDB file with only backbone atoms C $\alpha$ ?.....	29
How is the protein surface defined?.....	29
Why upload of a PDB file failed?.....	29
Why are the results wrong for an uploaded PDB?.....	29
Can I use more than one Chain ID as query?.....	29
Can I upload an NMR structure?.....	29
Can I use a binding site as query?.....	30
What happens with co-crystallized ligands in query protein structure?.....	30
Why ProBiS finds only a few similar proteins, when I know that there should be hundreds?..	30
What is the non-redundant PDB?.....	30

# Background

Binding sites are often conserved in evolution of proteins. ProBiS can detect such conserved binding sites regions on a protein structure given as query. ProBiS performs a local, surface oriented structural comparison of a query structure to a database of non-redundant protein structures (nr-PDB), and finds proteins that are locally similar to the query protein. Then it calculates degrees of structural conservation for query amino acid residues, which measure frequency of occurrence of a particular residue in the local structural alignments that were found. The degrees of structural conservation are represented as colors on the query protein from blue (unconserved) to red (conserved). In contrast to most structural alignment algorithms, ProBiS can detect similar binding sites, even when proteins are of different folds.

# Software Requirement

ProBiS web server requires Java due to the use of Jmol applet. You can get Java at <http://java.com>. ProBiS has been shown to work on the following platforms.

## Linux - Ubuntu

### Sun Java (Version 6 Update 30)

- Firefox
- Chrome
- Opera

### OpenJDK (IcedTea Version 1.1.1)

- Firefox

## Windows 7

### Sun Java (Version 6 Update 30)

- Firefox
- Chrome
- Opera
- Safari (version 5.1)
- Internet Explorer (version 8)

# New User Interface Features

## Submit a Binding Site as a Query

The new Jmol based tool for selection of binding sites, which is available on the *Detection of Structurally Similar Binding Sites* and *Pairwise Local Structural Alignment* input pages, now allows one to easily define a binding site, and submit it as a query to the ProBiS web server. This focuses the search for structural similarities to the interesting part of a protein.

## Change the Default Comparison Database

The *Detection of Structurally Similar Binding Sites* page now provides the option to change the default Comparison Database, the non-redundant PDB, for a user-provided list of protein chains. A search for similar binding sites can thus be narrowed to a subset of the PDB, e.g., proteins of the same fold as query, or even to complete PDB of currently ~180 thousand chains.

## Tabs

The *ProBiS output page* has a new streamlined tab layout, which allows better organization of the larger content.

## Similar Proteins Tab

The *ProBiS output page* now supports a tabular view of the similar proteins, to aid in structural and functional annotation of the query protein. Relevant information, i.e., links to the Pfam, SCOP, UniProt, or ProBiS databases, are displayed for each similar protein, which also allows identification of unexpected binding sites similarities across protein folds. In addition, the table can be downloaded in Column Separated Values (CSV) format, which is supported in popular software, e.g., Excel.

## Structural Alignments Tab

The structural alignments of proteins can now be downloaded in the CLUSTAL format as multiple sequence alignments.

## Search Tool

The table of similar proteins on the *ProBiS output page* has an integrated keyword search tool. For example, all similar proteins with the same Pfam accession number can be identified.

## Details Tab

The *ProBiS output page* also provides detailed alignment view for each pairwise local structural alignment. This view allows examination of correspondences between the aligned residues in tabular format, which aids in the detection of mutations in binding sites. A pairwise alignment can be loaded into the Jmol to view it in 3D, or can be downloaded as PDB or XML file.

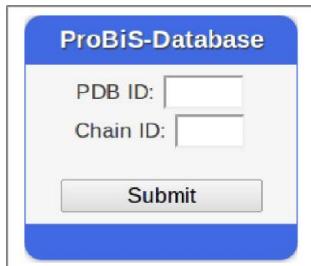
## Selection Tab

The selection tab on the *ProBiS output page* shows a list of molecules that are in the Jmol viewer, and enables hiding/showing the molecules.

## Progress Indicator

A prominent progress indicator allows precise tracking of the status of a ProBiS job. It is colored like a semaphore, i.e., red - job was cancelled, yellow - job is waiting in queue, and green - the calculation is in progress.

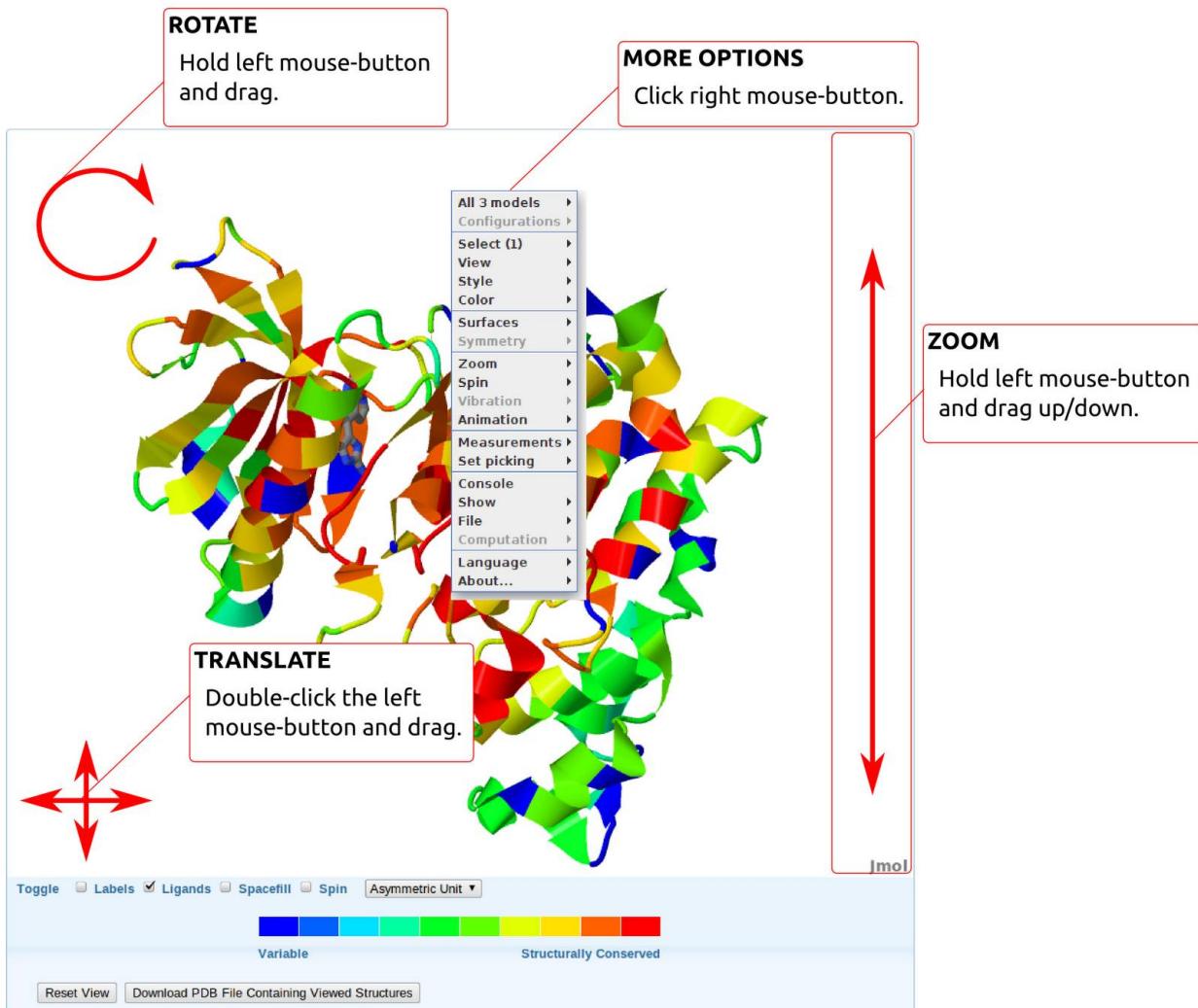
## ProBiS-Database Widget



The ProBiS-Database Widget provides access to local structural similarity profiles available in the ProBiS-Database. It can be included in any web page by adding a single line of code.

# Jmol Help

Some key tips on how the user can manipulate the 3D protein model in Jmol are presented in Figure 1.



**Figure 1.** Mouse operations in Jmol.

# Protein Binding Sites Tools

ProBiS web server is used for detection of structurally similar binding sites in PDB and local pairwise alignment of protein structures. The main page of the ProBiS is shown in Figure 2.

The screenshot shows the ProBiS web server interface. At the top, there's a navigation bar with links for "Introduction", "ProBiS-Tools", "Help", "Citation", "Related Citations", "ProBiS-Web Server RESTful Web Services", "ProBiS-Database Access", and "Contact". The main content area has several sections:

- Protein Binding Sites Tools**: Describes the server as open for detecting structurally similar protein binding sites and pairwise local structural alignment. It includes a "Go To Input Page" link.
- Detect Structurally Similar Binding Sites**: Shows a 3D molecular surface representation of a protein. Text describes the tool's purpose and features, including detection of global/local similarities, structurally similar binding sites, conserved binding sites, and accurate superimpositions. It also includes a "Go To Input Page" link.
- Pairwise Local Structural Alignment**: Shows a 3D molecular surface representation of two proteins. Text describes the tool's purpose and features, including sequence/fold-independent alignment, comparison of binding sites in cavities and flat surfaces, and accurate superimpositions. It includes a "Go To Input Page" link.
- ProBiS-Web Server RESTful Web Services**: Shows a diagram of two computers connected to a central "ProBiS Web Services" box. Text describes the RESTful interface and its features, including comparison of binding sites against non-redundant PDB, pairwise alignment, and fast response. It includes a "Go To Instructions Page" link.
- ProBiS-Database Access**: Shows a "ProBiS Database" icon. Text describes the database as a repository of pairwise alignments and its features, including fast access, local structural alignments, and a RESTful web services interface. It includes a "Go To Instructions Page" link.
- ProBiS In Brief**: Displays a 3D Jmol visualization of a G protein alpha subunit with colored regions representing structural similarity scores. Labels include "GDP binding site", "G protein alpha subunit", "P-loop", "alpha-beta binding site", and "Jmol". It includes "Jmol controls" like "Show labels", "Show helixes", "Spacefill", "Spin", and "Reset view". A color scale at the bottom indicates "Variable" and "Structurally conserved".
- Contact**: A form for users to provide suggestions, questions, comments, or bug reports. It includes fields for "Name", "Comment", and a "Submit" button.

At the bottom of the page, a footer states: "ProBiS is developed at the National Institute of Chemistry, Ljubljana, Slovenia".

Figure 2. The ProBiS web server main page provides access to the Protein Binding Sites Tools.

# Detect Structurally Similar Binding Sites

To start using this tool the user should follow instructions described in Figure 3.

**Introduction**

- If you are new to ProBiS, you can watch the introductory video available here.

**ProBiS-Tools**

- Home
- Detect Structurally Similar Binding Sites
- Pairwise Local Structural Alignment
- RESTful Web Services
- Database Access

**Help**

- Website Help
- Pre-calculated Examples
- Website FAQ
- Browser Compatibility

## Protein Binding Sites Tools

ProBiS is an open server for the detection of structurally similar protein binding sites and pairwise local structural alignment.

### Detect Structurally Similar Binding Sites

[Go To Input Page](#)

**C**  **A**

Input a query protein or a binding site and the ProBiS-algorithm will structurally compare the query independently of sequence or fold with 29412 non-redundant (>95% seq.id.) protein structures.

ProBiS allows:

- The detection of global or local similarities in proteins across folds
- The detection of structurally similar binding sites even on flat surfaces
- The detection of structurally conserved binding sites (fingerprints) of the query protein
- The accurate superimpositions of similarly or differently folded proteins

**D** [Go To Input Page](#)

**Figure 3.** ProBiS main allows access to the input page: (A) by clicking Detect Structurally Similar Binding Sites, (B) by clicking a link located in the bar on the left-hand side of the ProBiS main page, (C) by clicking the picture or (D) by clicking Go To Input Page.

## Input

### Simple Query

The form in Figure 4 allows the comparison of a query protein against the non-redundant PDB proteins.

### Detect Structurally Similar Binding Sites

**A**  
PDB ID:

**B**  
Chain ID(s):

**C**  
Upload a PDB file [?](#)

Select Binding Site (optional) [?](#)

Local Alignments Only      Cutoff Z-Score:

Proteins to Compare Against:

Your e-mail address (optional):   
A link to the results page will be sent to you by e-mail.  
Computation for a medium sized protein will take a few minutes.

**Submit Job**

**Figure 4.** Simple query. Enter (A) PDB ID; (B) Chain ID; (C) alternatively to (A) upload a PDB file (D) optional email address. Then click the Submit Job button.

### Precalculated Results

Alternatively, the user can get instant results from the ProBiS-Database. If the user enters a PDB ID and chain ID, ProBiS will display a pop-up window, presented in Figure 5, which has a link to precalculated results (see also ProBiS-Database Access).

Pre-calculated local structural similarity profile is ready for a protein with >95% seq. id.  
[Click here to show ProBiS results for PDB entry 1gotA](#)

**Figure 5.** Pop-up window that allows to obtain pre-calculated results from the ProBiS-Database.

## Advanced Query

Figure 6 shows, how to submit a binding site (or any other site) as a query, or compare the query protein against a custom list of proteins.

### Detect Structurally Similar Binding Sites

Pre-calculated local structural similarity profile is ready for a protein with >95% seq. id.  
[Click here to show ProBiS-Database results for PDB entry 1vpqA](#)

PDB ID:  Chain ID(s):

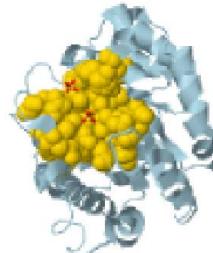
Upload a PDB file

**A** Select Binding Site (optional)

Select Input

**B**  Binding Sites  
 SO4.262.A  
 SO4.263.A  
 SO4.264.A  
Distance (Å)

**C**  Custom Selection  
e.g., :A and (12-15,18) or :B and (C)



Jmol

**D**  Local Alignments Only      Cutoff Z-Score:

**E**

Proteins to Compare Against:

**F**

e.g., 1all.A, 3dbj.C, 2vjt.A

Your e-mail address (optional):

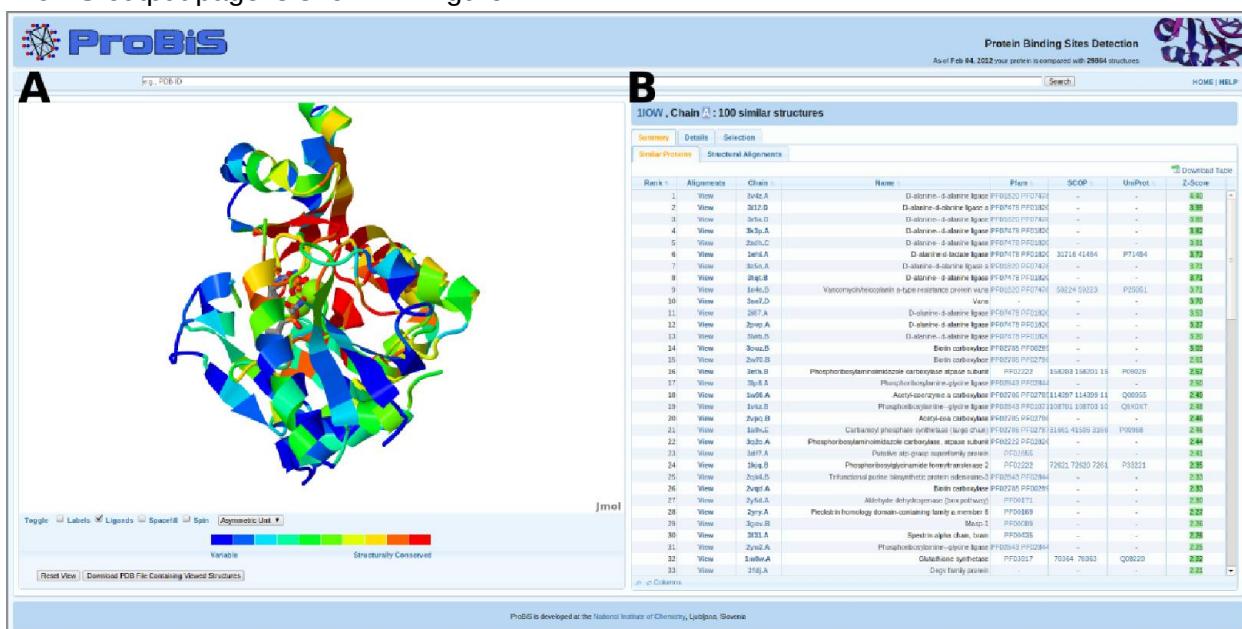
A link to the results page will be sent to you by e-mail.  
Computation for a medium sized protein will take a few minutes.

**Figure 6.** Advanced query. Clicking on (A) Select Binding Site link will open the form (red rectangle), where you can define part of the surface on the query protein that will be taken as input. Input can be

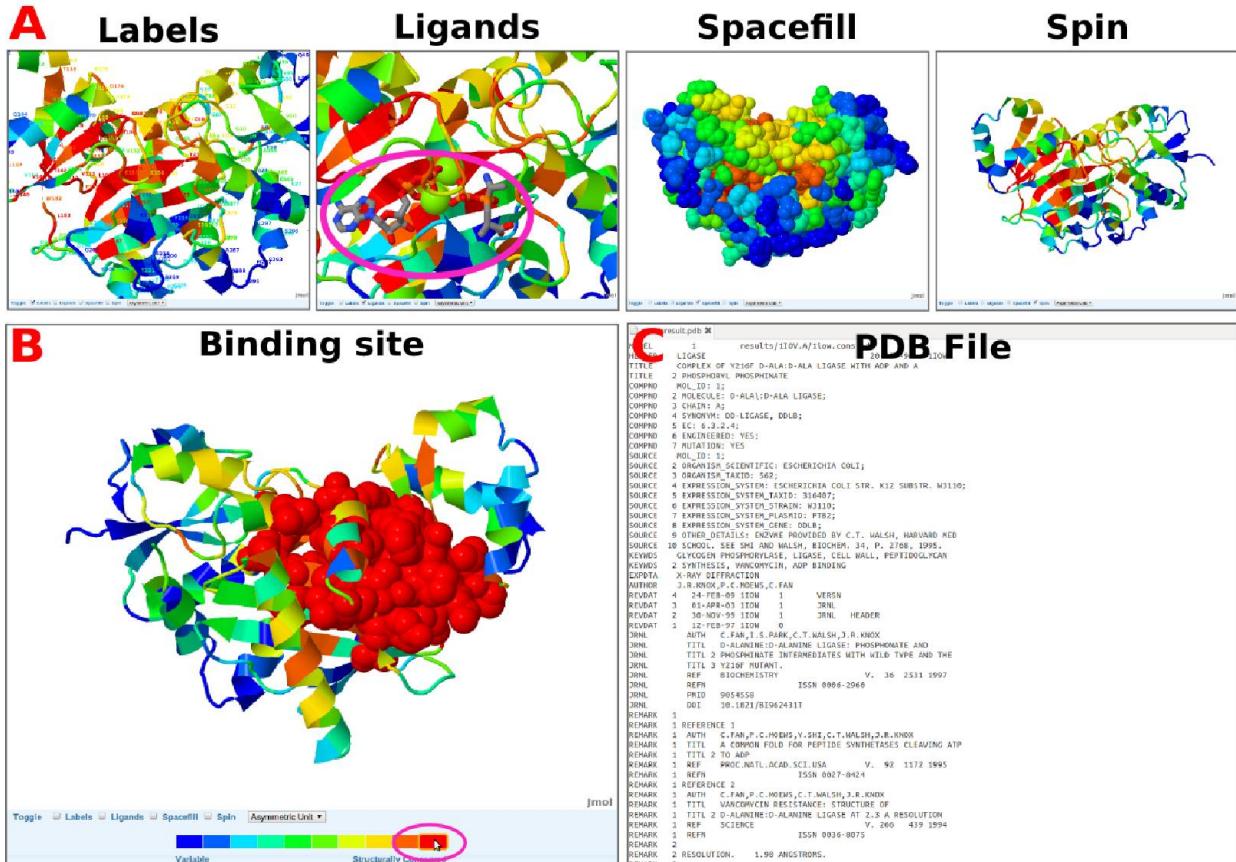
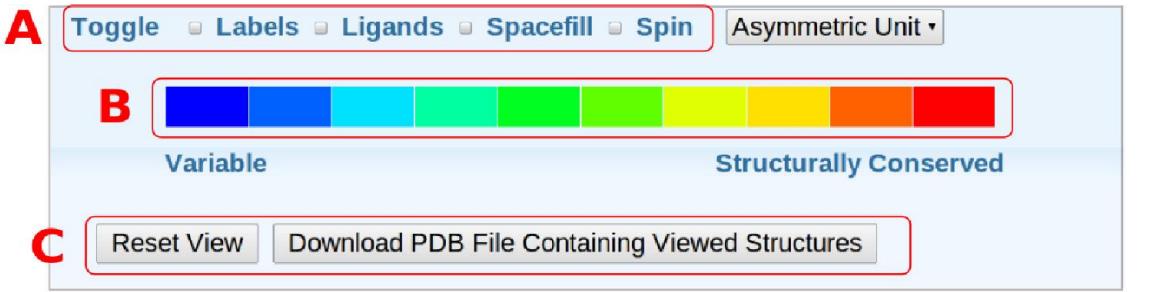
(B) a binding site, which you select by a ligand (e.g., SO4.262.A) and distance: all residues in a radius of 3.0 Å within SO4 ion will be considered as a query binding site; (C) a custom selection of residues given by chain identifier and residue numbers, e.g., ":A and (12-15,18)". (To select residues numbered 12,13,14,15,18 on chain A and 64,67,89,90 on chain B of a query protein, enter :A and (12-15,18) or :B and (64,67,89-90).) The selected surface region appears as yellow spheres in the Jmol viewer. (D) If enabled, the alignments will only be searched for in the selected surface region and the local alignments found will not be extended along compared proteins' backbones. (E) Regulated filtering of the alignments. Lower Z-Score means that more alignments will be displayed among results. (F) The drop-down list enables to enter a custom list of proteins against which to compare the query protein, e.g., 1all.A, 3dbj.C, 2vjt.A. Default is to compare against the non-redundant PDB.

## Output

ProBiS output page is shown in Figure 7.



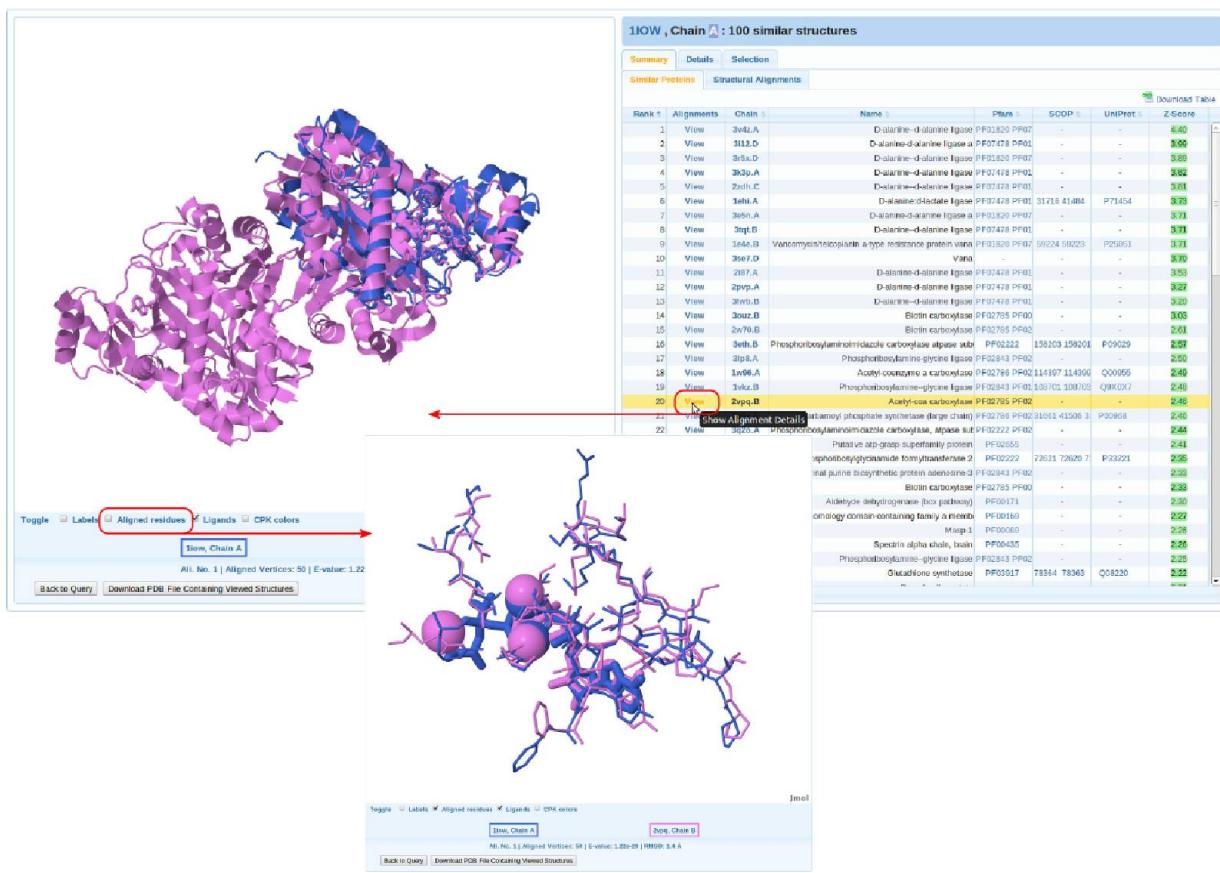
**Figure 7.** ProBiS output page: (A) 3D query protein colored by degrees of structural conservation from blue (unconserved) to red (structurally conserved) in Jmol viewer; (B) table of similar proteins.



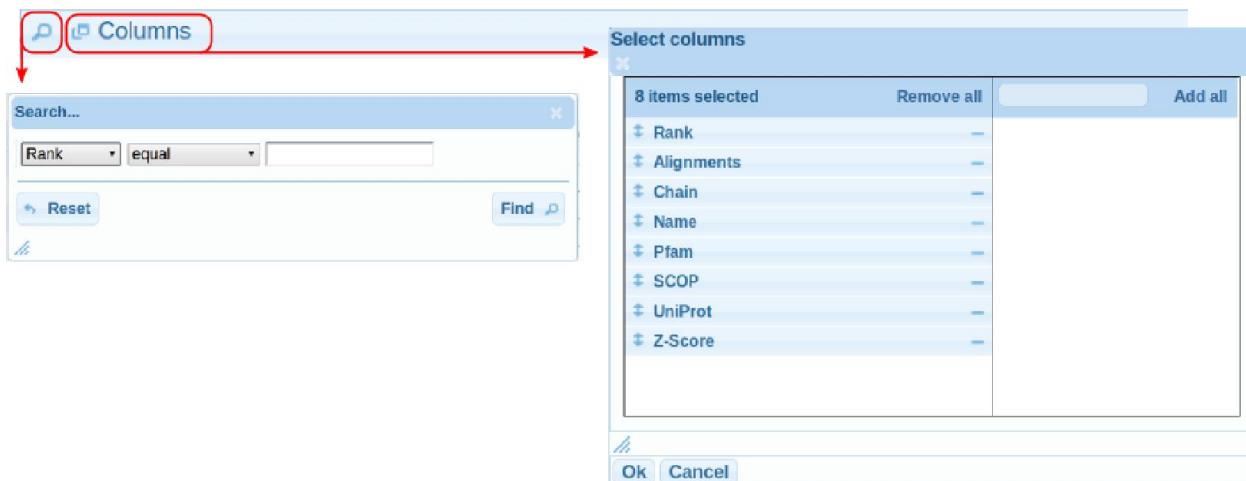
**Figure 8.** Jmol options. (A) Toggle options; (B) Click on a color, shows residues of this color in the 3D query protein model. Each color represents a conservation grade from 0 to 9. (C) The downloaded PDB file of query protein will have conservation grades from 0.0 to 0.9 in beta factor fields.

1IOW , Chain A : 100 similar structures							
Summary		Details		Selection			
Similar Proteins		Structural Alignments					
Rank	Alignments	Chain	Name	Pfam	SCOP	UniProt	Z-Score
1	View	3v4z.A	D-alanine-d-alanine ligase	PF01820 PF07	-	-	4.40
2	View	3I12.D	D-alanine-d-alanine ligase a	PF07478 PF01	-	-	3.99
3	View	3r5x.D	D-alanine-d-alanine ligase	PF01820 PF07	-	-	3.89
4	View	3k3p.A	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.82
5	View	2zdh.C	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.81
6	View	1ehl.A	D-alanine:d-lactate ligase	PF07478 PF01	31716 41484	P71454	3.73
7	View	3e5n.A	D-alanine-d-alanine ligase a	PF01820 PF07	-	-	3.71
8	View	3tqt.B	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.71
9	View	1e4e.B	Vancomycin/telcoplanin a-type resistance protein vanA	PF01820 PF07	59224 59223	P25051	3.71
10	View	3se7.D	Vana	-	-	-	3.70
11	View	2I87.A	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.53
12	View	2pvp.A	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.27
13	View	3lwB.B	D-alanine-d-alanine ligase	PF07478 PF01	-	-	3.20
14	View	3ouz.B	Biotin carboxylase	PF02785 PF00	-	-	3.03
15	View	2w70.B	Biotin carboxylase	PF02785 PF02	-	-	2.61
16	View	3eth.B	Phosphoribosylaminoimidazole carboxylase atpase sub	PF02222	158203 158201	P09029	2.57
17	View	3lp8.A	Phosphoribosylamine-glycine ligase	PF02843 PF02	-	-	2.50
18	View	1w96.A	Acetyl-coenzyme a carboxylase	PF02786 PF02	114397 114399	Q00955	2.49
19	View	1vkz.B	Phosphoribosylamine-glycine ligase	PF02843 PF01	108701 108703	Q9X0X7	2.48
20	View	2vpq.B	Acetyl-coa carboxylase	PF02785 PF02	-	-	2.46
21	View	1a9x.E	Carbamoyl phosphate synthetase (large chain)	PF02786 PF02	31661 41506 3:	P00968	2.46
22	View	3q2o.A	Phosphoribosylaminoimidazole carboxylase, atpase sut	PF02222 PF02	-	-	2.44
23	View	3df7.A	Putative atp-grasp superfamily protein	PF02655	-	-	2.41
24	View	1kjq.B	Phosphoribosylglycinamide formyltransferase 2	PF02222	72621 72620 7:	P33221	2.35
25	View	2qk4.B	Trifunctional purine biosynthetic protein adenosine-3'	PF02843 PF02	-	-	2.33
26	View	2vqd.A	Biotin carboxylase	PF02785 PF00	-	-	2.33
27	View	2y5d.A	Aldehyde dehydrogenase (box pathway)	PF00171	-	-	2.30
28	View	2yry.A	Pleckstrin homology domain-containing family a memb	PF00169	-	-	2.27
29	View	3gov.B	Masp-1	PF00089	-	-	2.26
30	View	3f31.A	Spectrin alpha chain, brain	PF00435	-	-	2.26
31	View	2yw2.A	Phosphoribosylamine-glycine ligase	PF02843 PF02	-	-	2.25
32	View	1m0w.A	Glutathione synthetase	PF03917	78364 78363	Q08220	2.22

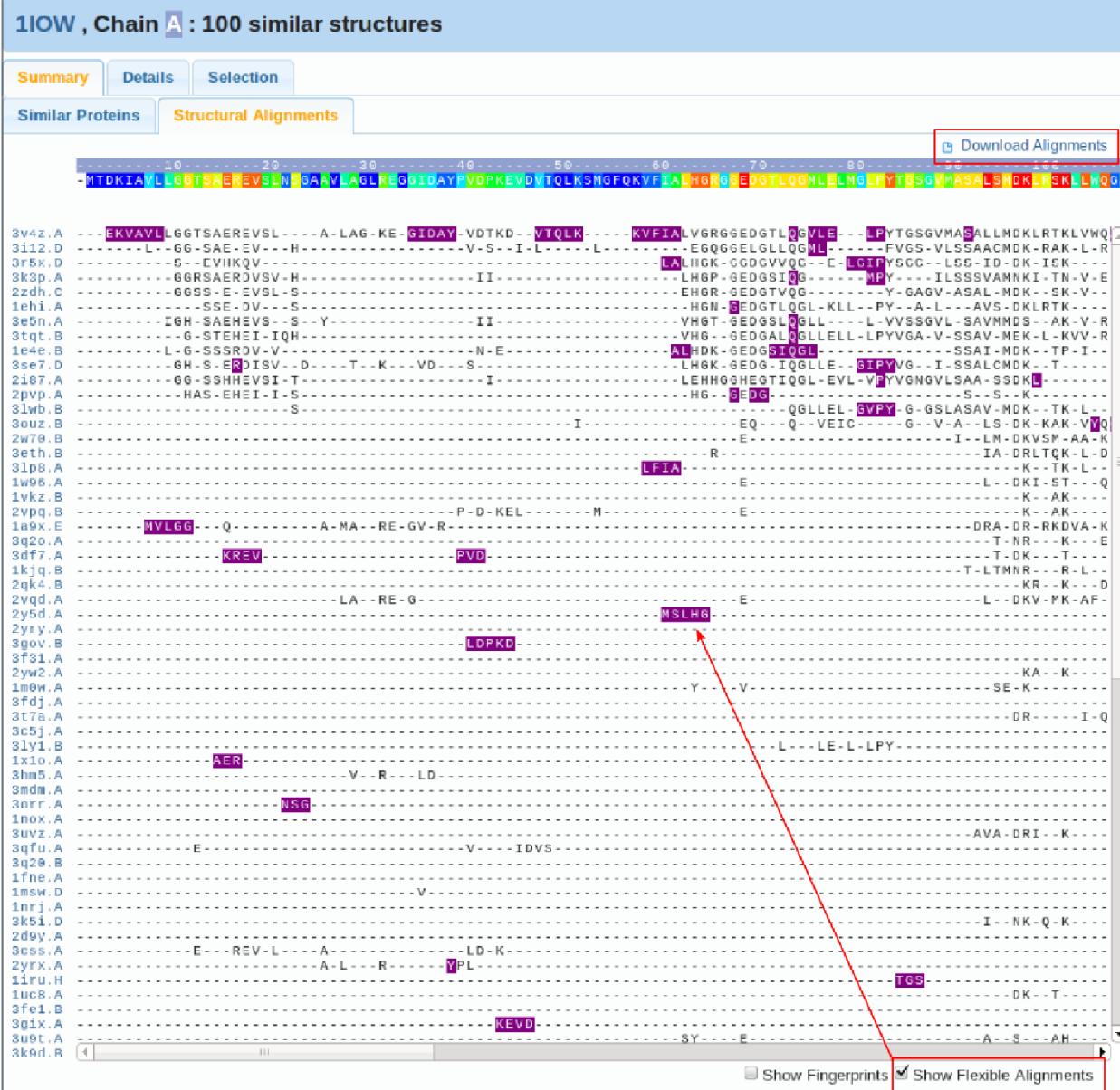
**Figure 9.** Table of similar proteins. Columns specifications are: Alignments - click on View shows local structural superimposition of the query and the similar protein in Jmol viewer and opens the Details tab, Chain - click on a link opens a new web page with precalculated results for the clicked chain from ProBiS-Database, Name - protein name, Pfam, SCOP, UniProt - links to corresponding proteins annotation databases, Z-Score - indicate how many standard deviations each alignment differs from the mean, e.g., a pairwise alignment with Z-Score of 2.0 is in the top ~2% of all alignments.



**Figure 10.** Click on a View link in the Alignments column loads the local structural alignment between query protein and similar protein into the Jmol viewer. To see only the aligned residues, click the respective checkbox under the Jmol window.



**Figure 11.** Search & Reorder columns tools. They are located below the table of similar proteins.



**Figure 12.** Structure-based sequence alignments. On the top is the query protein's sequence colored by degrees of structural conservation. The ruler above this sequence is colored by the Chain ID(s) of the query protein. Below are aligned amino acid residues of similar proteins. The flexible alignments, i.e., residues of similar proteins that cannot be aligned to a RMSD<2 Å, are highlighted in purple. Click on aligned residues loads the alignment between the similar protein and the query protein into the Jmol viewer. Click on a PDB/Chain ID of an aligned protein structure, will open a new tab with the information about this structure in the RCSB Protein Data Bank. Click on the Download Alignments downloads a file with the aligned sequences in the CLUSTAL format.

Summary		Details		Selection			
Similar Proteins		Structural Alignments					
<a href="#">Download Alignments</a>							
10	20	30	40	50	60		
-MEAIAKHDFSATADDLSFRKG-----DSNWYRAELDGKEG--P-NYIEMK--	-YDQ-S-EVT-KKGDLIL-LN----NKDWKK-EV--RQGFVP--YVK--	-YDQ-SS-DEVT-KKGDLIL-LN----NKDWKK-EV--RQGFVP--YVK--	-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--	-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--	-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--		
2a37.A -MEAIAKHDFSATADDLSFRKG-----DSNWYRAELDGKEG--P-NYIEMK--	1neg.A -YDQ-S-EVT-KKGDLIL-LN----NKDWKK-EV--RQGFVP--YVK--	3ig1.B -A-YDFE---EL-FK-GDII-----DENWYE-L---GF-----NYVEI--	2dix.C -VA-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--	3ulr.A -IA-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--	31q9.A -A-YDQ-QAA-DDEISF---DII-----W-----G-----NYVEI--		
3m0s.A -YDQ-Q-SS-DEVT-KKGDLIL-LN----NKDWKK-EV--RQGFVP--YVK--	2hda.A -A-YDFEAT---EDLSFKKG-----WWEA-----KNGY-----NYV--	1x69.A -VA-YDQ-QAA-DDEISF---DII-----WWR-----G-----P-NYVEI--	2dlm.A -A-KDFQAS-S---ELT-QKGDIV----DKNW-----NYVEV--	1sem.A -A-FDF-----ELAFKRGDVI-----NWE-----NYV--	3rea.B -VA-YDQ-S-SSDLSFKKG-----VL-E-G-WWKA-----EGY-P-NYV--		
1ujo.A -A-YDFEA---NELTFK-GELI-----D-NWWQ-----G-----NFV--	1qcf.A -A-YDFEAH---EDLSFKKG-----WWKA-----KEGY-P-NYV--	ibui.F -VA-YDQ-EAH---EDLSFKKG-----WL-----G-----NYV--	1oeb.A -A-YDFEA---EDEL-FR-GEVV-----W-L-K-G-----NYV--	2d13.A -AKDFPKA-----QKGDIV-----DQNWYE-----G-----P-YIEL--	1zlm.A -A-YTFEAE-DEL-FE-GDII-----DNNWWK-----G-----NYV--		
2drm.B -A-YDQ-QAT-DGELTFK-GD_I-----WWE-ELNGK-G-P-NYVQ-	3m0r.A -YDQ-S-DEVT-KKGDLIL-----NKDWKK-----GFVP--YVK--	1j08.A -AE-YDQ-AA-DNELTFF---D-----D-DW---L-----G-----NYV--	3ngp.A -YDQ-Q-SS-ELT-KKGDLIL-----NKDWKK-----GFVP--YVK--	3nhn.A -VA-YDQ-EA---SFQKGDM-----WWKA-----GY-----NYV--	3h0h.A -A-YDQETEEDLSF-KGE-----DWW-----GY-----NYV--		
3uf4.A -A-YDFEAR-ETDLSF-KGE-----N-E-----DWW-----TGY-----NYV--	2vrf.A -A-FDF-----L-FRRGD-----NWWK-----G-----PRNYV--	1ituc.A -LA-YDQ-S-T-KKGDLIL-----NKDWKK-----GFVP--YVK--	2x3w.D -A-YDQ-SS-DELSFKAGD-----EQGW-----P-NYVE--	2vir.A -A-YDF-P-E-EVA-KKGDLM-----DD-----DWWK-----GY-----NYIEI--	1cka.A -A-FDF---EDL-FKKGDLIRI-----W-----G-----P-YVE--		
2vge.A -Y-A-DELSFR-----R-----E-DWW-----LHGQEGYYVRNY--	2a28.C -MEAI-YEQADDEIS---GDII-----W-E-----G-----P-NYI--	1uhc.A -FKAR-NELS-----LKILE-----W-AEVNGK-GY-P-NYI--	3qwq.X -V-YDFE-----D-FEQGE-L-----NQDWWEA-----G-----P-NYVQIQ--	3rb3.D -I-VA-YDQ-S-SS-DELSFKKG-----VL-E-GEEWWKA-----GY-P-NYV--	2kbt.A -ARYDF-A-S-K-GDIIKILNKK-----QGWWR-I-----NYVE--		
2dm1.A -VARYNF-AR-----ELS-R-GD-----DQGWWK-----G-----P-YVE--	2eq1.A -YDQKA-DELTFR-GL-----WWK-----G-----P-NYVE--	1h07.A -A-FDF-----FKKGDLIRI-----G-----P-YVE--	2ak5.A -AKFNF-T-DELSF-KGDOI-----W-W-----GW-----NYV--	1ark.A -A-YDQ-A-DEVSFK-GD-----W-----G-----P-NYVE--	1uti.A -YDFEAL-EDEL-F-----N-SWW-----L-K-G-----NYV--		
2dbm.A -A-YDFE-----FK-GDII-----DENWYE-----NYVEI--	1g2b.A -YDQ-Q-S-EVT-KKGDLIL-LN----NKDWKK-EV--RQGFVP--YVK--	1ycs.B -A-WDQYEPQ-DDEL-K-GD-----D-E-WWW-----KEGYVRN--	2yup.A -VA-YDFEP-T-DELSFR-E-I-----NENWYE-----SYVQ--	1yn8.B -VA-YDFEP---DNEI-----GDIV-----QGW-----G-----P-FV--	2yup.A -VA-YDFEP-T-DELSFR-E-I-----NENWYE-----SYVQ--		
2z33.A -FEY-N-EDEL-K-GDII-IIINEE-W-W-----LN-----G-----P-NFVK--	2j6f.A -I-YDQ-AV-DDELT-RV-EI-----W-----ELNFR-G-----NFVK--	2gncc.A -IAK-YDQ-TA-ELSFKKG-L-----DW-----P-YI--	3sa0.B -VA-YNFO-S-QLS-Q-GDGV-----DWR-----G-----KSFI--	2d18.A -IAK-YDQ-TA-ELSFKKG-L-----DW-----DG-----P-YI--	2cre.A -ARA-YDQ-DELAFR-RGDIL-----WWK-----L-----LO--		
2ed0.A -A-YDQ-EDELSFKQ-G-GI-----GWYE-----P-NYVE--	315r.A -A-YDQ-EEDID-GDIL-----W-----G-----P-YVE--	3ie6g.A -V-YDQ-Q-SS-ELT-KKGDLIL-----WKDWKK-V-----GFI-----YLV--					

**Figure 13.** Fingerprint residues are highlighted as red vertical stripes, and often correspond to conserved active site residues. They are not always available, because they cannot always be determined, e.g., when there are no similar proteins with Z-Score>3.0.

**1IOW , Chain A : 100 similar structures**

**Summary** **Details** **Selection**

Previous 1/100 Next

ProBiS found 4 local structural alignments of **1IOW** with **3V4Z**.

**Query Protein**      **Aligned Protein**

Name: D-ALAI:D-ALA LIGASE	Name: D-ALANINE-D-ALANINE LIGASE
PDB ID: <b>1IOW</b>	PDB ID: <b>3V4Z</b>
Chain ID: A	Chain ID: A
Pfam: <b>PF01820 PF07478</b>	Pfam: <b>PF01820 PF07478</b>
SCOP: <b>41481_31713</b>	SCOP: -
UniProt: <b>P07882</b>	UniProt: -

**Select Alignment:**

- Alignment No. 1
- Alignment No. 2
- Alignment No. 3
- Alignment No. 4

**Alignment No. 1**

Back to top

Query Protein			Aligned Protein				
Res.	Name	Res. ID	Chain ID	Res.	Name	Res. ID	Chain ID
ASP	3	A	...	GLU	3	A	
LYS	4	A	...	LYS	4	A	
ILE	5	A	...	VAL	5	A	
ALA	6	A	...	ALA	6	A	
VAL	7	A	...	VAL	7	A	
LEU	8	A	...	LEU	8	A	
GLY	34	A	...	GLY	34	A	
ILE	35	A	...	ILE	35	A	
ASP	36	A	...	ASP	36	A	
ALA	37	A	...	ALA	37	A	
TYR	38	A	...	TYR	38	A	
VAL	47	A	...	VAL	47	A	
THR	48	A	...	THR	48	A	
GLN	49	A	...	GLN	49	A	
ARG	288	A	...	ARG	288	A	
MET	292	A	...	LEU	292	A	
SER	293	A	...	SER	293	A	
PHE	294	A	...	PHE	294	A	
SER	295	A	...	SER	295	A	
LEU	297	A	...	LEU	297	A	
ILE	301	A	...	ILE	301	A	
LEU	302	A	...	LEU	302	A	
ILE	304	A	...	LEU	304	A	
ALA	305	A	...	ALA	305	A	
ASP	306	A	...	ASP	306	A	

**Superimposition:**

To calculate the coordinates superimposition, use the following transformation on aligned protein:

$$X2' = (0.99)*X2 + (0.14)*Y2 + (0.07)*Z2 + (-38.77)$$

$$Y2' = (0.15)*X2 + (-0.78)*Y2 + (-0.61)*Z2 + (24.74)$$

$$Z2' = (-0.03)*X2 + (0.62)*Y2 + (-0.79)*Z2 + (35.58)$$

**View in Jmol** **Download Alignment as XML** **Download PDB of Superimposition**

Z-Score: 4.4  
 Alignment Score: 11.7  
 E-value: 1.18E-243  
 Vertices: 574  
 RMSD: 0.7  
 Surf. Vector. Angle: 0.52

**Legend:**

- ... Structurally equivalent and similar physical-chemical properties.
- ... Structurally not equivalent, but similar physical-chemical properties.

**Alignment No. 2**

Back to top

Query Protein			Aligned Protein				
Res.	Name	Res. ID	Chain ID	Res.	Name	Res. ID	Chain ID
LYS	4	A	...	LYS	4	A	
ILE	5	A	...	VAL	5	A	
ALA	6	A	...	ALA	6	A	
VAL	7	A	...	VAL	7	A	

**View In Jmol** **Download Alignment as XML** **Download PDB of Superimposition**

Z-Score: 4.14  
 Alignment Score: 11.13  
 E-value: 1.13E-226  
 Vertices: 216  
 RMSD: 0.4  
 Surf. Vector. Angle: 0.07

**Figure 14.** Detailed pairwise alignment. ProBiS found in this case four different alignments. Click on the Next at the top, opens the detailed alignments for the next similar protein.

**1IOW , Chain A : 100 similar structures**

Summary   Details   **Selection**

	Query protein structure	<b>1</b>	N/A	N/A	
	Structural superimpositions	<b>2</b>	N/A		
PDB ID1	Chain ID1	PDB ID2	Chain ID2	Ali. No.	View
1iow	A	2vpq	B	1	<input type="checkbox"/>
1iow	A	3q2o	A	1	<input checked="" type="checkbox"/>

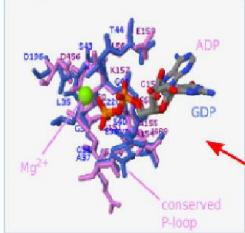
**Figure 15.** The Selection tab holds a list of structures that are loaded in Jmol viewer (but may be also hidden from view). Box with the Query protein structure is always present, and expands on click, showing the Asymmetric Units and Biological Units of the query protein. Box with the Structural superimpositions expands to show alignments that are already in the Jmol viewer. You can use the View checkbox to view or hide the alignments or query protein.

# Pairwise Local Structural Alignment

The access to this tool from the ProBiS main page is shown in Figure 16.

**A**

### Pairwise Local Structural Alignment



Input two proteins or binding sites. The ProBiS-algorithm will compare the structures based on geometry as well as physicochemical properties and return their local structural alignment.

ProBiS allows:

- The sequence or fold independent alignment of two proteins
- The comparison of binding sites in cavities as well as on flat surfaces
- The accurate superimpositions of binding sites or entire proteins together with ligands

[Go To Input Page](#) **click**

**B**

**ProBiS-Tools**

- Home
- Detect Structurally Similar Binding Sites
- **Pairwise Local Structural Alignment** 
- RESTful Web Services
- Database Access

**Figure 16.** The access to the Pairwise Local Structural Alignment tool is from the ProBiS main page, as shown in panels (A) and (B).

## Input

### Pairwise Local Structural Alignment

A

First PDB ID (e.g., 1all)  Chain ID(s) (e.g., A, AB)

Upload a PDB file

Residue Motif (optional):  C   
(see example)

---

B

Second PDB ID (e.g., 3dbj)  Chain ID(s) (e.g., C, CD)

Upload a PDB file

Residue Motif (optional):  D   
(see example)

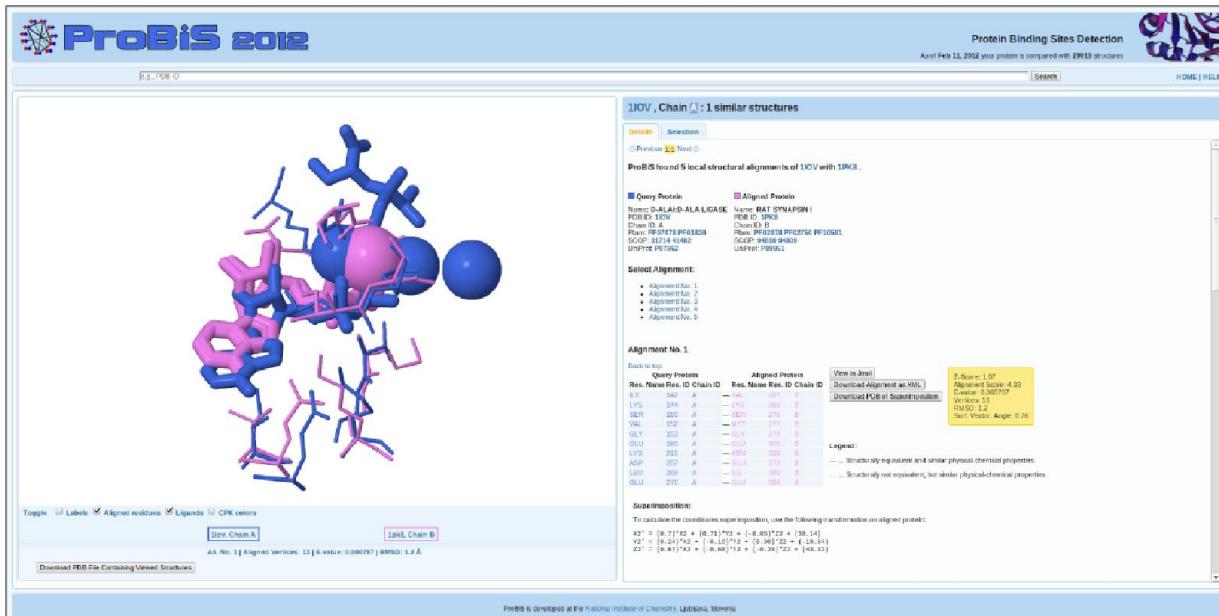
Your e-mail address (optional):  E

A link to the results page will be sent to you by e-mail. Alignment of two proteins will take seconds.

F

**Figure 17.** Pairwise Local Structural Alignment input page. To align a pair of complete proteins the user should only fill in (A) and (B) fields, (E) optionally provide email address, and (F) submit the query. To compare only selected structural motifs of one or both proteins, e.g., binding sites, the (C) and (D) buttons should be used in the same way as shown in Figure 6.

## Output



**Figure 18.** Pairwise Local Structural Alignment output page. This page shown only the structural alignment of the two inputted proteins.

## ProBiS Web Server RESTful Web Services

The ProBiS web server features [RESTful](#) (Representational State Transfer) web services to make the binding site similarities and local pairwise alignments for any PDB protein structure easily accessible from your scripts. Full set of commands and useful examples are listed below.

## Input

All commands below are working URLs, but in the text they do not appear as full URLs. To use the command “/rest/align?structure\_id1=1all.A&structure\_id2=3dbj.C”, you should precede it with “<http://probis.cmm.ki.si/update2012>”, e.g., the full command is “[http://probis.cmm.ki.si/update2012/rest/align?structure\\_id1=1all.A&structure\\_id2=3dbj.C](http://probis.cmm.ki.si/update2012/rest/align?structure_id1=1all.A&structure_id2=3dbj.C)”.

## Pairwise Alignment of PDB Structures

### Examples:

- Align two entire PDB structures: [/rest/align?structure\\_id1=1all.A&structure\\_id2=3dbj.C](/rest/align?structure_id1=1all.A&structure_id2=3dbj.C)
- Align two binding sites (defined as residues in 7 Å radius of ligands): [/rest/align?structure\\_id1=1all.A&bsite1=CYC.175.A.7&structure\\_id2=3dbj.C&bsite2=CYC.201.C.7](/rest/align?structure_id1=1all.A&bsite1=CYC.175.A.7&structure_id2=3dbj.C&bsite2=CYC.201.C.7)
- Align two entire PDB structures and return superimposition PDB file: [/rest/align?structure\\_id1=1all.A&structure\\_id2=3dbj.C&return=pdb](/rest/align?structure_id1=1all.A&structure_id2=3dbj.C&return=pdb)

### Specification:

- **structure\_id1, structure\_id2** - Specifies the PDB/Chain ID(s) of PDB structures to compare (ex. 1all.A, 3dbj.C). For two chains use 1all.AB, 3dbj.CD.
- **bsite1, bsite2** (optional) - Specifies the binding sites to compare (ex. CYC.175.A.7, CYC.201.C.7). E.g., CYC - ligand name, 175 - ligand number, A - chain id, 7 - radius in Angstroms to select residues around ligand. If omitted, entire PDB structures are aligned
- **return** (optional) - Specifies the return type, which can be pdb, json, or xml. The default is xml.

## Detect Similar Proteins or Binding Sites in the Non-redundant PDB

*Caution:* Examples in this section require all CPUs of the ProBiS-web server, and can load several minutes depending on the server work load.

### Examples:

- Search the non-redundant PDB (nr-PDB) with a query protein: [/rest/scan?structure\\_id=5cyt.R](/rest/scan?structure_id=5cyt.R)
- Search the nr-PDB with a binding site (defined as residues in a 5 Å radius of a ligand): [/rest/scan?structure\\_id=5cyt.R&bsite=HEM.105.R.5](/rest/scan?structure_id=5cyt.R&bsite=HEM.105.R.5)
- Search the nr-PDB with a query protein and return alignments with Z-Score>2.0 in json format: [/rest/scan?structure\\_id=5cyt.R&z\\_score=2.0&return=json](/rest/scan?structure_id=5cyt.R&z_score=2.0&return=json)

### Specification:

- **structure\_id** - Specifies the PDB/Chain ID(s) of the query PDB structure to be compared with all proteins in the nr-PDB (ex. 5cyt.R).
- **bsite** (optional) - Specifies the query binding site to search against nr-PDB (ex. HEM.105.R.5). E.g., HEM - ligand name, 105 - ligand number, R - chain id, 5 - radius in Angstroms to select residues around ligand. If omitted, entire PDB structure is considered as a query.
- **z\_score** (optional) - Specifies the cutoff Z-Score; alignments with Z-Score>cutoff will be returned. The default is 1.0.
- **return** (optional) - Specifies the return type, which can be json or xml. The default is xml.

## Perl Example Script

```
use strict;
use LWP::Simple qw( $ua );

# Make a request command (uncomment lines below if you want something else)
```

```

my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/align?
structure_id1=1all.A&structure_id2=3dbj.C');
#my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/align?
structure_id1=1all.A&bsite1=CYC.175.A.7&structure_id2=3dbj.C&bsite2=CYC.201.C.7');
#my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/align?
structure_id1=1all.A&structure_id2=3dbj.C&return=pdb');
#my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/scan?
structure_id=5cyt.R');
#my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/scan?
structure_id=5cyt.R&bsite=HEM.105.R.5');
#my $request = HTTP::Request->new( GET => 'http://probis.cmm.ki.si/update2012/rest/scan?
structure_id=5cyt.R&z_score=2.0&return=json');

# Decide about the content type you want to get in return (default is XML) (applies to
get_alignments and get_representative; other two commands return "text/plain")
$request->header(Accept => "application/json");
#$request->content_type( 'application/xml' );

# Send the HTTP request
my $response = $ua->request( $request );

# Check to see if there is an error
unless( $response->is_success ) {
print "\n Error: ", $response->status_line, "\n";
}

# Output response
print "ProBiS returned:\n", $response->content;

```

## Output

The diagram shows an XML document structure. On the left, the full XML code is displayed. A red box highlights a specific node under the 'alignment' section, which is further expanded into five separate alignment details on the right. A red arrow points from this expanded section to the text '5 Alignments'. The XML code is as follows:

```
<aligned_structures>
  <node>
    <pdb_id>3dbj</pdb_id>
    <chain_id>C</chain_id>
    <nfp>0</nfp>
    <protein_name>ALLOPHYCOCYANIN</protein_name>
  </node>
  <alignment>
    <node>
      +<scores></scores>
      +<rotation_matrix></rotation_matrix>
      +<translation_vector></translation_vector>
      +<aligned_residues></aligned_residues>
    </node>
    +<node></node>
    +<node></node>
    +<node></node>
    +<node></node>
  </alignment>
</node>
</aligned_structures>
```

→ 5 Alignments

```
- <scores>
  <alignment_no>0</alignment_no>
  <aligned_vertices>371</aligned_vertices>
  <e_value>5.99E-166</e_value>
  <rmsd>0.4</rmsd>
  <sva>0.51</sva>
  <z_score>4.43</z_score>
  <alignment_score>11.77</alignment_score>
</scores>
+<rotation_matrix></rotation_matrix>
+<translation_vector></translation_vector>
+<aligned_residues></aligned_residues>
```

**Figure 19.** The web services output for alignment of two proteins 1all.A and 3dbj.C. The command is [http://probis.cmm.ki.si/update2012/rest/align?structure\\_id1=1all.A&structure\\_id2=3dbj.C](http://probis.cmm.ki.si/update2012/rest/align?structure_id1=1all.A&structure_id2=3dbj.C). Local structural alignments are in XML format. Alternatively, they can be returned in Json format.

# ProBiS-Database Access

ProBiS-Database can be accessed through the pop-up window in the *Detect Structurally Similar Binding Sites* tool as shown in Figures 5 and 6. Here, we present the other means of access: through the search text box shown in Figure 20, through the ProBiS-Database widget in Figure 22, and from user scripts by ProBiS-Database web services.

## Input

The screenshot shows the ProBiS 2012 homepage. At the top, there is a logo and the text "ProBiS 2012" next to "Protein Binding Sites Detection". Below this, a message says "As of Feb 11, 2012 your protein is compared with 29919 structures". A search bar contains the placeholder "e.g., PDB ID" and a "Search" button. To the right of the search bar is a "HOME" link. On the left, there are two columns: "Introduction" and "ProBiS-Tools". The "Introduction" column has a link to an introductory video. The "ProBiS-Tools" column links to "ProBiS Tools Home" and "Detect Structurally Similar Binding". On the right, there is a "ProBiS in Brief" section featuring a 3D molecular model of a G protein alpha subunit with various binding sites labeled: "GDP binding site", "G protein alpha subunit", and "P-loop".

**Figure 20.** The ProBiS-Database access from the search text box, which is marked with red. This is the most common way of access, since it is available on all ProBiS web pages.

The screenshot shows the search results for the query "1GOT". The search bar at the top has "1GOT" entered. The results are displayed in a table:

	Representative Chain:	Title:
1GOT.A	1gotA	GT-ALPHA/GI-ALPHA CHIMERA
1GOT.B	1gotB	GT-BETA
1GOT.G	1gotG	GT-GAMMA

To the right of the table is a 3D molecular model of the protein 1GOT, showing the "alpha-beta binding site" and "GDP binding site". The "ProBiS in Brief" section is also visible on the right side of the page.

**Figure 21.** The search results for query protein "1got". Three Chain IDs, A, B, and G are found in this protein. Click on the red encircled link opens precalculated results page for 1got.A.

## The ProBiS-Database Widget

### The ProBiS-Database Widget

The ProBiS-Database Widget provides access to the local structural similarity profile for a protein chain. You can try the fully functioning example below!

**Widget:**

ProBiS-Database

PDB ID:

Chain ID:

**Usage:**

To embed in your own website, copy/paste the following line to your HTML code.

```
<script type="text/javascript" src="http://tyr.cmm.ki.si/beta/unstable/bin/pw/?width=150px&title-color=%23faa&body-color=whitesmoke"></script>
```

**Options:**

The ProBiS-Database Widget is highly customizable to best fit within your site. There are a number of options you specify to customize the widget. You can omit any of these options, and the widget will use default values.

- **width** - Specify the width of the widget (ex. 150px). The default width is 280px.
- **title-color** - Specify the color of the title and borders in html hex form (ex. #faa). The default is orange.
- **body-color** - Specify the color of the body in html hex form (ex. whitesmoke). The default is whitesmoke.

**Figure 22.** ProBiS-Database widget can be included in any web page to provide access to the ProBiS-Database.

## The ProBiS-Database RESTful Web Service Interface

The ProBiS-Database features [RESTful](#) (REpresentational State Transfer) Web Services to make our data easily accessible from your scripts. ProBiS-Database contains data only for non-redundant PDB chains. This means that you have to use representative non-redundant PDB & Chain IDs as queries (see examples below).

### Get Representative

Translate your PDB & Chain ID to > 95% sequence identical representative:

[/rest/get\\_representative?structure\\_id=1ytf.A](/rest/get_representative?structure_id=1ytf.A)

### Get Local Structural Alignments

Get local structural alignments with Z-Scores>2.0 for a representative: [/rest/get\\_alignments?structure\\_id=1ytb.B&z\\_score=2.0](/rest/get_alignments?structure_id=1ytb.B&z_score=2.0)

### Get Superimposition

Get the PDB file of two superimposed representative structures: [/rest/get\\_superimposition?structure\\_id1=1all.A&structure\\_id2=3nmd.B&alignment\\_no=1](/rest/get_superimposition?structure_id1=1all.A&structure_id2=3nmd.B&alignment_no=1)

### Get Similarity Scores

Get representative with similarity scores in beta-factors at Z-Score>2.0:

[/rest/get\\_beta\\_marked\\_PDB?structure\\_id=1all.A&z\\_score=2.0](http://probis.cmm.ki.si/update2012/rest/get_beta_marked_PDB?structure_id=1all.A&z_score=2.0)

## Perl Example Script

```
use strict;
use LWP::Simple qw( $ua );

# Make a request command (uncomment lines below if you want something else)
my $request = HTTP::Request->new( GET =>
'http://probis.cmm.ki.si/update2012/rest/get_alignments?structure_id=1all.A&z_score=2.0');
#my $request = HTTP::Request->new( GET =>
'http://probis.cmm.ki.si/update2012/rest/get_representative?structure_id=1all.A');
#my $request = HTTP::Request->new( GET =>
'http://probis.cmm.ki.si/update2012/rest/get_beta_marked_PDB?
structure_id=1all.A&z_score=2.0');
#my $request = HTTP::Request->new( GET =>
'http://probis.cmm.ki.si/update2012/rest/get_superimposition?
structure_id1=1all.A&structure_id2=3nmd.B&alignment_no=1');

# Decide about the content type you want to get in return (default is XML) (applies to
get_alignments and get_representative; other two commands return "text/plain")
$request->header(Accept => "application/json");
#$request->content_type( 'application/xml' );

# Send the HTTP request
my $response = $ua->request( $request );

# Check to see if there is an error
unless( $response->is_success ) {
print "\n Error: ", $response->status_line, "\n";
}

# Output response
print "ProBiS-Database returned:\n", $response->content;
```

## Output

Refer to figure 19.

# FAQ

## Can I upload a protein model?

Yes. A protein model in a standard [PDB file format](#). The ATOM records must be listed as usual on [the RCSB web page](#), i.e., for THR residue: N, CA, C, O, CB, OG1, CG2. If the modelling program that you used reordered the ATOM records in any other way, the results you will get, will be wrong.

## Can I upload a PDB file with only backbone atoms C $\alpha$ ?

No. PDB file must be a full-atom representation of a protein. Hydrogens are ignored.

## How is the protein surface defined?

The surface of a protein is defined by rolling a 1.4 Å radius sphere over its atoms. Cavities that have no connection to the outside are not considered, i.e., only the outer surface is used.

## Why upload of a PDB file failed?

PDB models from various programs, such as CHARMM or MODELLER, could be in non standard formats. The ordering of ATOM records may be different, side chain ATOM records may be missing, ENDMDL records between different models in NMR structures may not be there - these are necessary, because ProBiS does not tolerate restarting of residue numbers. If the PDB file has no chain ID record then the whole structure will be taken as chain A - if there were to be two models in such file, not separated by ENDMDL records, then upload will fail.

## Why are the results wrong for an uploaded PDB?

The ATOM records must be listed in the same order as in PDBs at [the RCSB web page](#), i.e., for THR residue: N, CA, C, O, CB, OG1, CG2. If the modelling program that you used (such as MODELLER) reordered the ATOM records in any other way, the results you get will be wrong.

## Can I use more than one Chain ID as query?

Yes. Up to 3 different chain IDs within the same PDB record can be considered, e.g., if you would like to input the complex of chains A and B, just write AB in the Chain ID(s) input box.

## Can I upload an NMR structure?

You can upload an NMR structure as long as each model is in own MODEL/ENDMDL record.

## **Can I use a binding site as query?**

Yes. Use the *Select binding site* button on the input page.

## **What happens with co-crystallized ligands in query protein structure?**

Ligands in a PDB file, i.e., all HETATM records, are ignored.

## **Why ProBiS finds only a few similar proteins, when I know that there should be hundreds?**

Similar proteins found by ProBiS are members of the non-redundant PDB which is a subset of the entire PDB. This means that each is a representative of a cluster of homologous proteins (>95 seq.id.) in the PDB. A cluster may have hundreds of members, e.g., as in protein kinases, but ProBiS will represent all these proteins by only one structure.

## **What is the non-redundant PDB?**

The non-redundant PDB (nr-PDB) is obtained from the entire PDB by clustering the protein chains in the PDB. Clusters with >95% sequence identical proteins are generated. Then, a representative of each cluster is chosen, which is preferably an X-ray structure with lowest Resolution. These representatives constitute the nr-PDB. The nr-PDB is updated each week.