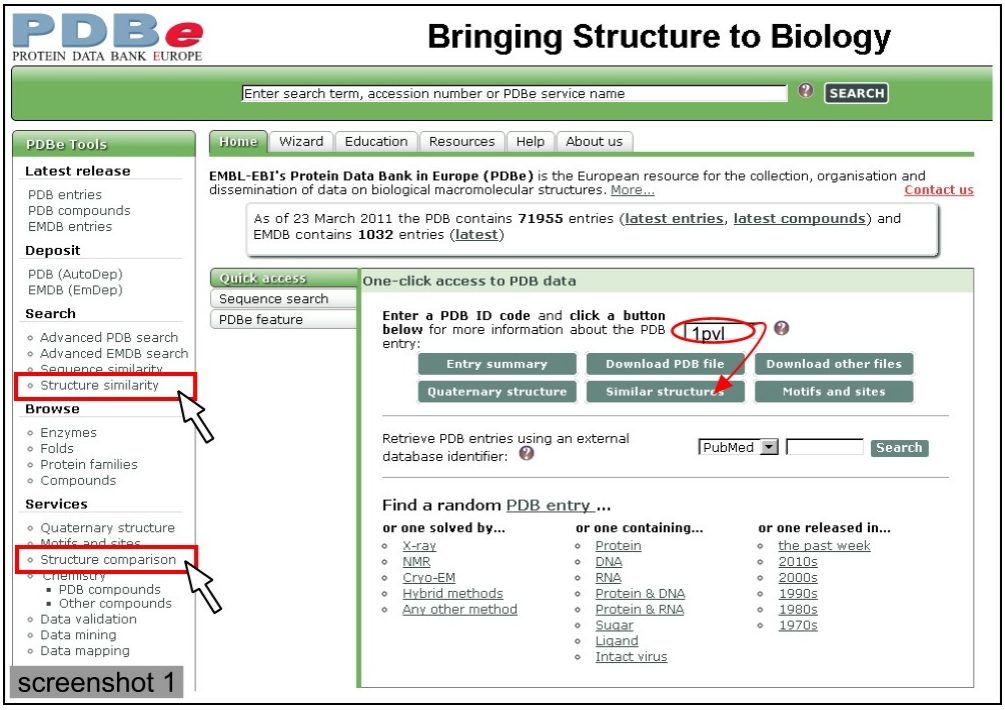
PDBeFold for structural comparison in the Leukotoxin family

This mini-tutorial is a walk-through of the **PDBeFold** server. This service equips you with a method for identifying similarities between protein structures in the PDB archive. You can run a live **PDBeFold** session in another browser window alongside this tutorial to look for structural similarities in the **Leukotoxin** family. **PDBeFold** searches and matches proteins by considering the three-dimensional arrangement of secondary structure elements - β-strands and α-helices.

Let's start with the **Panton Valentine Leukotoxin F** component **1pvl** which was described in the accompanying **PDBeQuips** article.

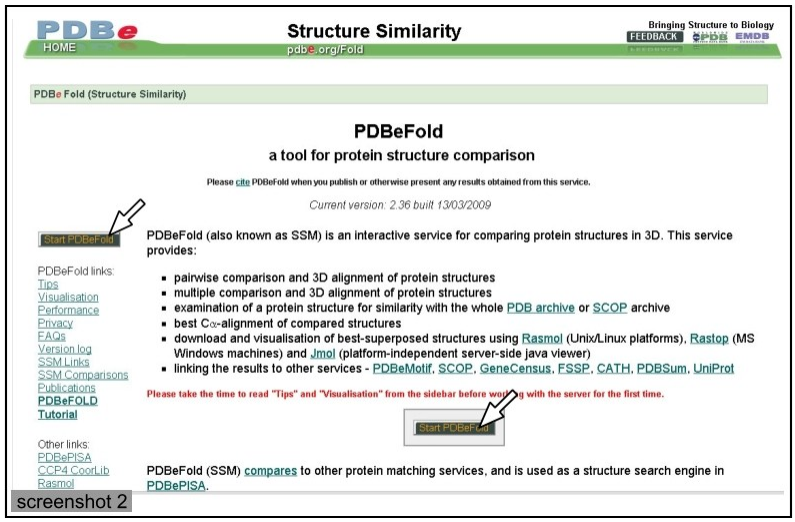
Accessing **PDBeFold**.
Either enter the PDB entry id in the One-click access field (circled in red) and click on Similar structures (red arrow).
Or click on either of the arrowed, red boxed links **Structure similarity** or **Structure comparison**.



Starting up PDBeFold from the PDBe website

If you know which PDB entry you want to analyse then, you can enter it directly in the '**One click access**' field (circled in red on screenshot 1) and hit the '**Similar structures**' button (indicated by the arrow). Alternatively you can get to the **PDBeFold** start page from either of the arrowed links '**Structure similarity**' or '**Structure comparison**' (boxed in red, screenshot 1) at the left of the **PDBe** home page. These take you to the **PDBeFold** start page.

PDBeFold start page.
Click on either of the **'Start PDBeFold'** buttons (mouse arrows) to get a submission form (see screenshot 3).

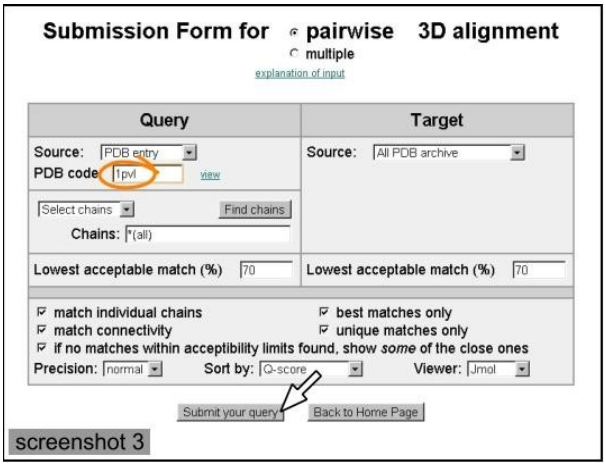


screenshot 2

The introductory explanation tells you that **PDBeFold** expects you to supply a starting structure that will be analysed first and then used as a query to search the PDB archive. After the search **PDBeFold** will report any structures with a similar arrangement of secondary structure elements (**SSE**s). It will rank these by how many elements it believes are significantly similar to those in your query structure, and it will then present the matches for you to look at and assess.

Click on either **'Start PDBeFold'** buttons (marked by arrows in screenshot 2) on the **PDBeFold** start page to get a search submission form.

PDBeFold submission form.
1pvl entered in the entry field (circled in orange) will be used to search the entire PDB archive indicated as **Target**.
Clicking on the **submit your query** button (mouse arrow) to see results as in screenshot 4.



screenshot 3

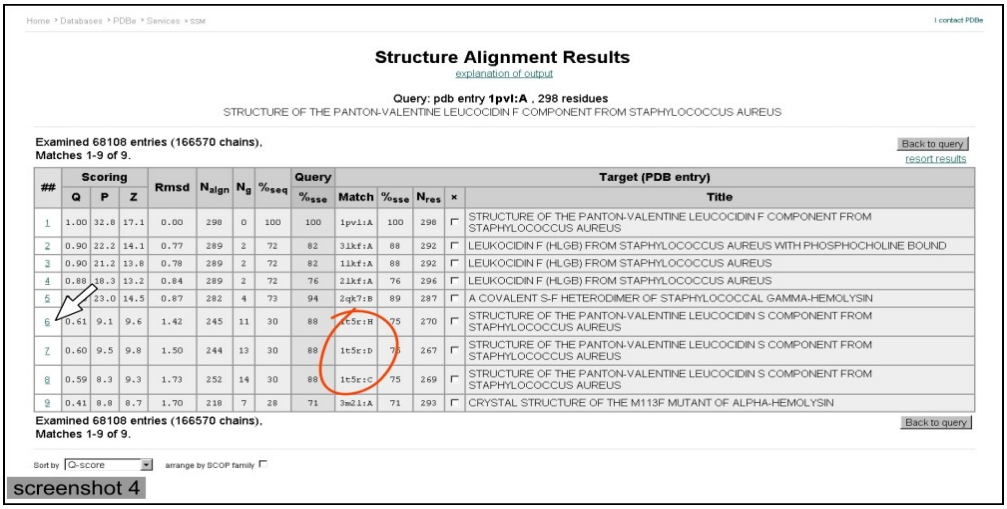
Search results for PDB entry 1pvl

Type in **1pvl** as the search query in the box labelled '**PDB code**' (circled in screenshot 3). The submission form also allows you to upload your own coordinate file, or multiple files - if for example you want to submit a family of proteins that share a particular fold. The submission form can be used to select a subset of the chains available in the query PDB entries (or just a fragment of a chain) and also to alter any program parameters before running the search.

For now just accept the defaults and hit the '**Submit your query**' button (arrowed in

screenshot 3). You should then get a message that your '**Matching is now in progress...**'.

PDBeFold search results. Each line in the table is a hit in the archive. But some entries appear more than once owing to their having multiple chains. Hits to the PVL **S** component can be identified from their titles, the entry id and chains for these are circled in orange.
Click on top **S** component hit (mouse arrow) to see details of match (shown in screenshot 5).



Home > Databases > PDBe > Services > SSM

I contact PDBe

Structure Alignment Results

explanation of output

Query: pdb entry 1pvl:A , 298 residues

STRUCTURE OF THE PANTON-VALENTINE LEUCOCIDIN F COMPONENT FROM STAPHYLOCOCCUS AUREUS

Examined 68108 entries (166570 chains),
Matches 1-9 of 9.

Back to query

resort results

##	Scoring			Rmsd	N_{align}	N_g	$\%_{seq}$	Query	Target (PDB entry)				
	Q	P	Z					$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	32.8	17.1	0.00	298	0	100	100	1pvl:A	100	298	☐	STRUCTURE OF THE PANTON-VALENTINE LEUCOCIDIN F COMPONENT FROM STAPHYLOCOCCUS AUREUS
2	0.90	22.2	14.1	0.77	289	2	72	82	3lkf:A	88	292	☐	LEUKOCIDIN F (HLGB) FROM STAPHYLOCOCCUS AUREUS WITH PHOSPHOCHOLINE BOUND
3	0.90	21.2	13.8	0.78	289	2	72	82	1lkf:A	88	292	☐	LEUKOCIDIN F (HLGB) FROM STAPHYLOCOCCUS AUREUS
4	0.88	18.3	13.2	0.84	289	2	72	76	2lkf:A	76	296	☐	LEUKOCIDIN F (HLGB) FROM STAPHYLOCOCCUS AUREUS
5		23.0	14.5	0.87	282	4	73	94	2qk7:B	89	287	☐	A COVALENT S-F HETERODIMER OF STAPHYLOCOCCAL GAMMA-HEMOLYSIN
6	0.61	9.1	9.6	1.42	245	11	30	88	1t5c:H	75	270	☐	STRUCTURE OF THE PANTON-VALENTINE LEUCOCIDIN S COMPONENT FROM STAPHYLOCOCCUS AUREUS
7	0.60	9.5	9.8	1.50	244	13	30	88	1t5c:D	75	267	☐	STRUCTURE OF THE PANTON-VALENTINE LEUCOCIDIN S COMPONENT FROM STAPHYLOCOCCUS AUREUS
8	0.59	8.3	9.3	1.73	252	14	30	88	1t5c:C	75	269	☐	STRUCTURE OF THE PANTON-VALENTINE LEUCOCIDIN S COMPONENT FROM STAPHYLOCOCCUS AUREUS
9	0.41	8.8	8.7	1.70	218	7	28	71	3m2l:A	71	293	☐	CRYSTAL STRUCTURE OF THE M113F MUTANT OF ALPHA-HEMOLYSIN

Examined 68108 entries (166570 chains),
Matches 1-9 of 9.

Back to query

Sort by Q-score arrange by SCOP family ☐

screenshot 4

Displaying the results

After the search has ended, matching structures ('hits') in the PDB archive are presented in a table which may continue over several pages. By default the results are ordered by the **PDBeFold** program's estimate of the quality of each match (**Q-score**). The **Q-score** takes into account the number of residues in the matched **SSE**s and their positions in space. High **Q-scores** are obtained for structures where a large number of residues in equivalent structural elements superimpose well in three-dimensional space. For the **1pvl** search run here, the **PDBeFold** results table gives a quite a number of structures that match well enough to be presented to the you, the user. The actual number of presented structures can vary depending on the search parameters and the actual content of the PDB archive when the search is run. In your own work you may want to go back and alter the initial search settings to suit your particular query.

The top hit in our search here (labelled 1 in the first column of the table) can be found in the '**Match**' column. Unsurprisingly it is **1pvl** with 100% of the **SSE**s matched.

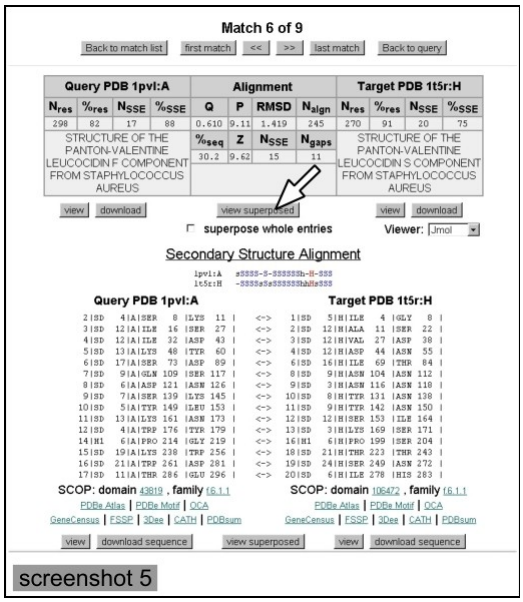
Although it may seem a bit odd to see our query returned from the search, **PDBeFold** is searching the whole PDB archive for us, and so, as long as **1pvl** is in the archive, it should be found and be a perfect hit!

In fact several other hits appear in the top half of the table that are related structures of the **Panton Valentine Leukotoxin F** component. This often happens as the PDB will typically contain multiple entries of the same protein from different experiments, crystal forms, or with different small molecules bound. (By the way, from its title you might guess that **2qk7** also has the **S** component in it although interestingly **PDBeFold** has in fact only matched the **F** component which is chain B in this entry!).

Although the top hits are not very interesting, our search for structural similarity between **Panton Valentine Leukotoxin** components has in fact succeeded in finding non-trivial hits since three chains from **1t5r** are returned in the bottom half of the table (circled in orange on screenshot 4). You may remember from the associated **PDBeQuips** article that this structure was the **Panton-Valentine Leukotoxin S** component.

This is a general rule for **PDBeFold** searches - the most interesting hits are generally on the bottom of the table on the very last page of the results! The interface provides you with a '**last page**' button to go there quickly. Another good tip is to change the '**Sort by**' pull-down (at the bottom of the table) to '**Seq %**' instead of '**Q score**'. The last entries on the last page will then have hits with little sequence identity to your query but with structural similarity - these may be homologues that you may have not found with a simple sequence search.

PDBeFold 1pvl to **1t5r** match. These details are produced for every hit in the table (screenshot 4). Click on **'view superposed'** (mouse arrow). The superimposed coordinates can also be downloaded from this page.



screenshot 5

Getting more detail about your results

To look in detail at any returned structural match just click on the underlined number in the first column (arrowed in screenshot 4). For instance, screenshot 5 above is the summary for the top **1t5r** hit in our search (number 6).

This shows our Panton Valentine Leukotoxin **F** component **1pvl** query on the left and the **S** component **1t5r** match on the right. In the middle are the details of the superimposition from the structural alignment that **PDBeFold** has done. A total of 15 secondary structural elements (N_{SSE}) could be matched satisfactorily between the two structures corresponding to 245 residues (N_{align}) and with an RMSD of 1.419Å.

Below the table is a quick summary of the secondary structure alignment with **S** for aligned strands and **H** for aligned helices (**s** and **h** are for elements that don't align well, or which correspond to a loop in one of the structures).

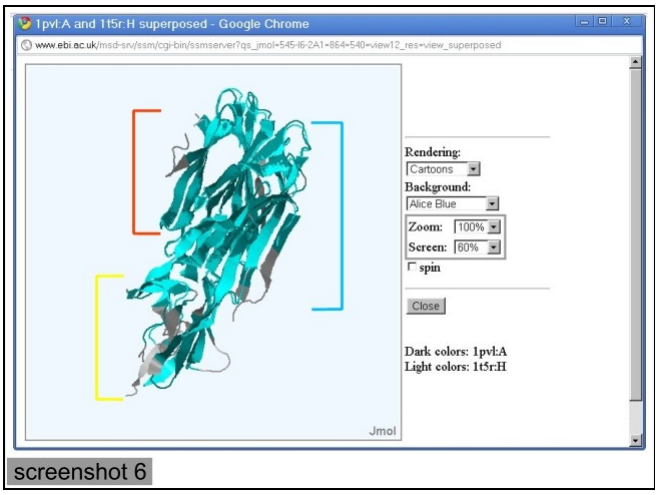
At the bottom of the results page is a detailed matching giving residue ranges. Although complicated, this can be very helpful if, for example, you want to write out a residue selection to make an image with equivalent colouring between matched parts of two structures.

Finally below the residue ranges summary is a detailed table of matches between the query and the hit on a residue-by-residue basis. This can be helpful in preparing a structure-based alignment of the protein sequences. Or it may highlight significant differences where a residue in the equivalent three-dimensional position in a domain has changed its character markedly - perhaps to adapt the fold to a new function.

Comparing the matches with your query

The '**view superposed**' button (arrowed in screenshot 5) launches a molecular viewer window showing the superimposed domains similar to that in screenshot 6. In your own search session this is an interactive viewer so you can rotate the superimposition to get a good view of the toxin functional subdomains, similar to that shown here.

Jmol graphic display of superimposed **1pvl** and **1t5r**. Bracketting shows regions of β-sandwich (blue), stem (red), and rim (yellow) subdomains.



screenshot 6

In the view here we have also used coloured brackets alongside to show the positions of the **stem** (red), **rim** (yellow), and **β-sandwich** (blue) subdomains as explained in the accompanying **PDBeQuips** article. Rotating the superimposition in your own session you should see that the elements of the query **1pvl** are in darker colours and those of the matched **1tr5** structure are in lighter colours. Elements that could not be matched in the superimposition are shown in dark and light grey (for the query and the hit, respectively). Loops connecting matched **SSE**s are also coloured provided that their backbone atoms superimpose well.

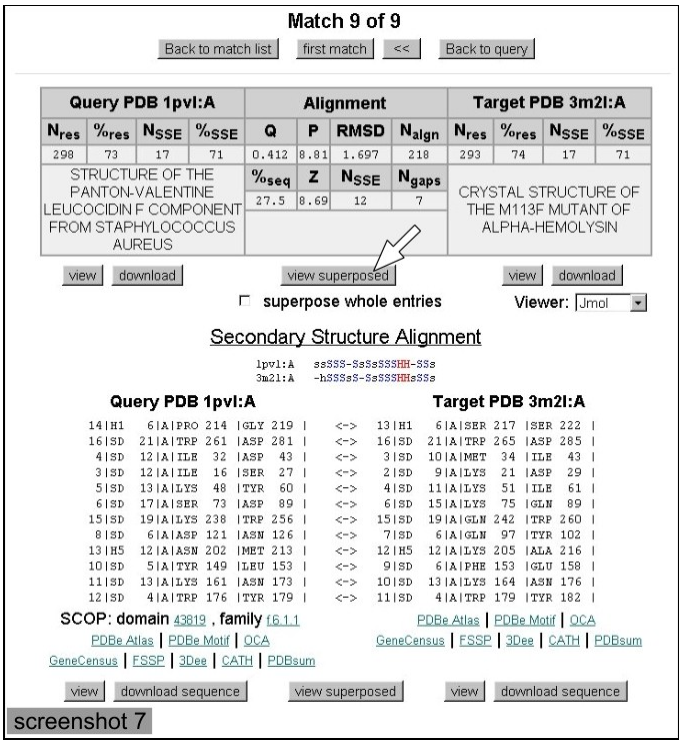
Very often you can inspect the superimposition and recognize β-strands or α-helices that are most likely equivalent but which **PDBeFold** has not matched as they have altered position too much between the two structures. The program has left them out to improve its **Q-score** for the match but sometimes looking by eye you will realise that the local structural differences are biologically significant - as the next example

shows.

Looking at an α-hemolysin superimposition

α-hemolysin similar to the entry **7ahl** mentioned in the **PDBeQuips** article appears here only as entry **3m2l** which is a slightly mutated form. However, changing the search parameters can retrieve more examples. However for now you can have a look at this example. Choose the number for the **3m2l** match from your **PDBeFold** results table (it was returned as match 9 here).

Details of **3m2l** matched to **1pvl**.
This is an α-hemolysin subunit.
Click on the **'view superposed'** button (arrowed).

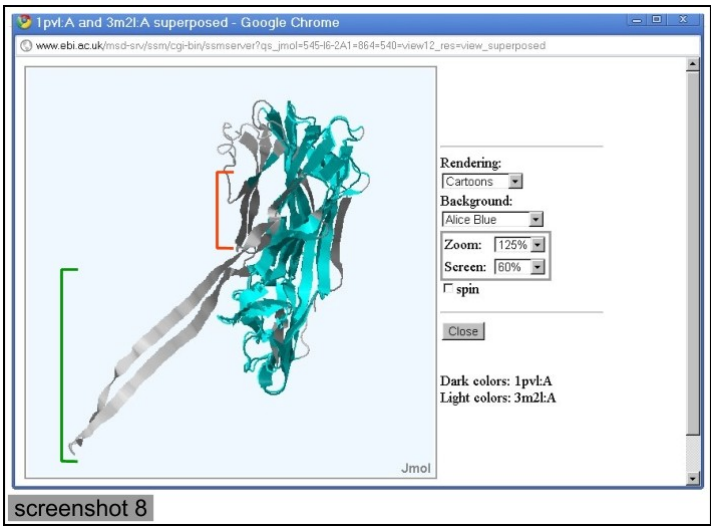


screenshot 7

The summary page (screenshot 7) for this match shows a 1.7 Å RMSD for 12 matched **SSE**s structure elements. Obviously this includes only the residues that have been selected as matching between the two structures but this does in fact represent most of the two structures.

As before you can click on '**view superposed**' button (arrowed in screenshot 7) to launch a viewer window.

Details of the **3m21** matched to **1pvl**.
This is an α-hemolysin subunit.
Click on the **'view superposed'** button (arrowed).



screenshot 8

If you have a live session running then use your mouse to rotate the structures to get a view of the toxin's functional subdomains. This time only the **rim** and the **β-sandwich** subdomains of the two proteins have been matched.

The screenshot here (number 8) indicates the position of the **stem** domain of **1pvl** with a red bracket and the extended **stem** domain of **3m2l** with a green bracket.

The large difference in conformation of the **stem** domain is due to fact that the **Panton Valentine Leukotoxin** (**1pvl**) crystallised in its soluble form whereas the structure of **α-hemolysin** (**3m2l** here) is in its membrane-bound form. As you can read in the **PDBeQuips** article this extended stem domain in the **α-hemolysin** structures participates in the β-barrel pore.