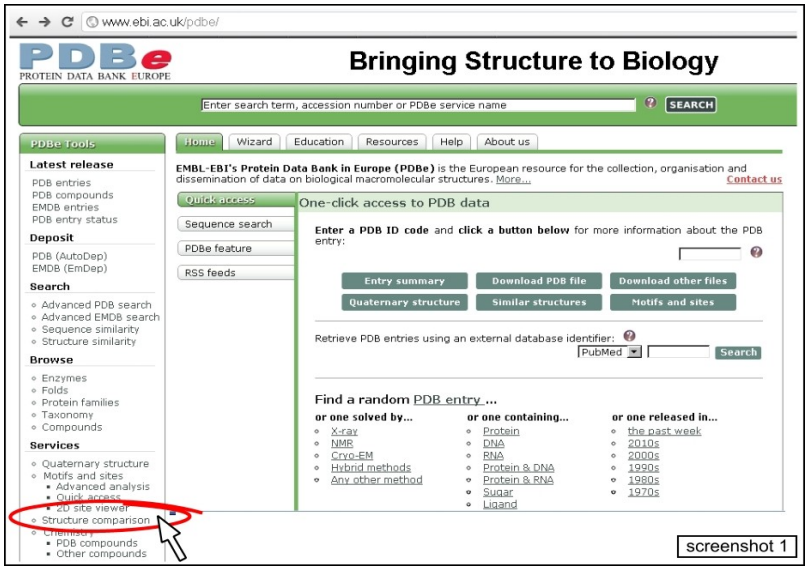
Using PDBeFold for aligning multiple members of a protein structural family

Very often in your analysis of structures you will be interested to see a family of structures superimposed on each other. These superimpositions are possible owing to the conserved regions of evolutionarily-related proteins. **PDBeFold** uses an algorithm based on matching corresponding secondary structure elements in related proteins. For an introduction to that algorithm applied to database searches you can refer back to the previous mini-tutorial on **PDBeFold**, which accompanied the first **Quips** article on leukotoxin structures. This new mini-tutorial shows you how to produce a multiple structure superimposition of dehydrogenases similar to that shown in the latest **Quips** article on Molecular Replacement with **Phaser** (appearing as the animated **View-2**).

To start on multiple superimposition you can launch **PDBeFold** from a link in the **Sevices** menu on the lefthand side of the **PDBe** home page (screenshot 1, below).

PDBeFold link in the Services menu. Click on **Services> Structure comparison** link (circled in red).

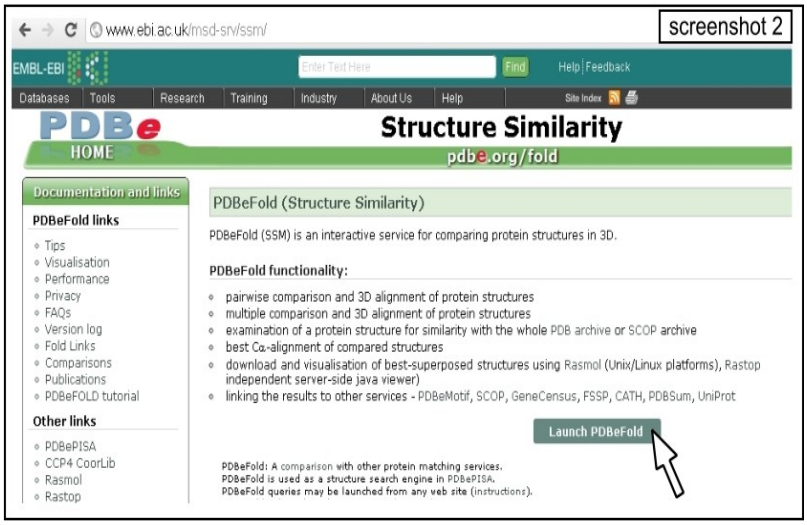


Launching PDBeFold for multiple superimpositions

The Structure Similarity page (screenshot 2) explains the functions available from **PDBeFold** and indicates that it uses the **SSM** secondary structure matching algorithm. You can see multiple comparison listed as the second distinct function of the **PDBeFold** service.But for this special function you will have to select a special set of pages and forms **after** you start the service. So launch **PDBeFold** straightaway from the button on this page (arrowed in screenshot 2).

Launch PDBeFold from the Structure similarity page.

Multiple superimposition is a special function of PDBeFold that you will select **after** launching the service (arrowed button).



Selecting the multiple superimposition function

Multiple superimposition is a special function of **PDBeFold**. To access the submission form for this you should select the appropriate 'radio' button at the top of the submission page (circled in red and numbered (1) in screenshot 3 below).

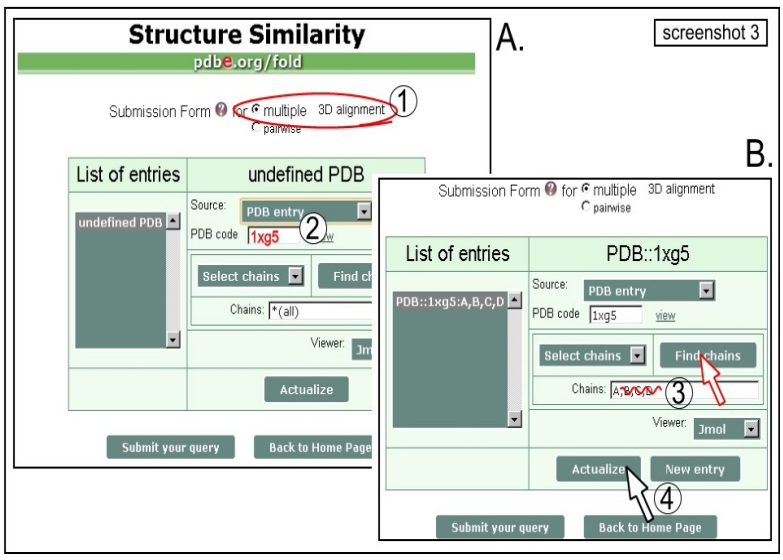
The page is then refreshed and allows you to define the structures that you want to superimpose.

To see how the selection works, follow the steps shown below. Enter the four letter identifier of the human dehydrogenase **1xg5** into the **'PDB code'** field (step (2) in screenshot 3). Then click on the **'Find chains'** function button (red arrow (3)). This populates the chain selection field just below with a list of protein chains present in your chosen entry. In this case delete all those except A (red wavy line(3)).

Finally add you selection **'1xg5 chain A'** to the submission list by clicking on **'Actualize'** (numbered (4) in screenshot 3).

Making a selection for submission.

Clicking on 'Multiple superimposition' button (circled in red step (1)) redraws the submission form. Enter the pdb id 1xg5 (red at step (2)) panel A. To restrict to one chain click on 'Find chains' (step (3), panel B) and delete back to 'A' (red correction line). Hit the step (4) arrowed button to add the selection to the list.



Adding more structures for superimposition

To complete the list of multiple entries for superimposition you repeat this selection process for each entry that you want to add.

The addition of the second entry is shown below: first click on **'New entry'** (number (1), screenshot 4). This is **very** important as otherwise your new selection will overwrite the previous one. Selecting **'New entry'** gives you a new line labelled 'undefined PDB'. Now you can go through the selection steps again. Enter the PDB code **1edo** (shown red at step (2) below) and click on 'Find chains' (labelled (3)). You will then see that **1edo** has only a single chain, so you can click on 'Actualize' straightaway (arrowed (4)).

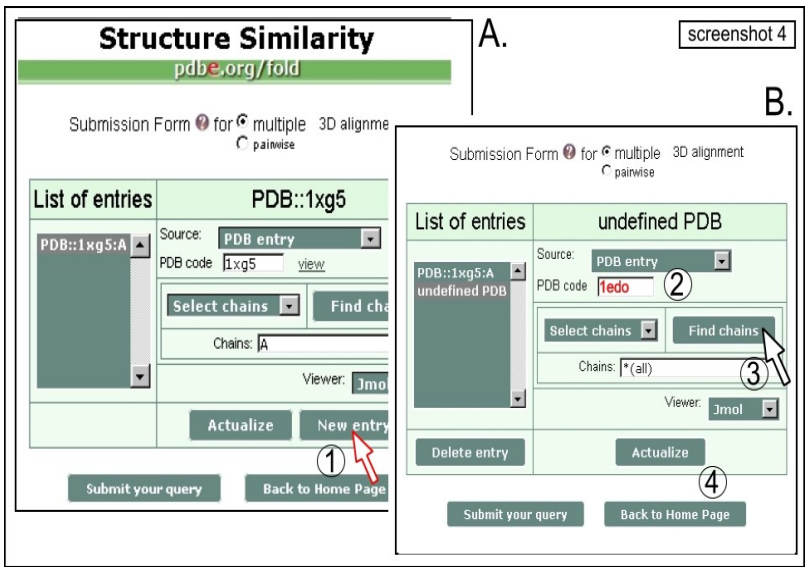
If you are unsure during this process on which chain you need, then below the entry field you'll find a link to launch a viewer to inspect your chosen entry.

If you add a selection to the growing list that you want to change then you can use the **'Delete entry'** button at the bottom of the list to remove it.

Remember that the SSM algorithm used in **PDBeFold** is only applicable to protein and peptide chains. If you select an entry with associated nucleic acid chains then you should remove them from your selection before running a superimposition. Also, unlike the database search function of **PDBeFold**, the multiple superimposition will not separate out multiple protein chains in a selection. So if you want to separately superimpose additional copies of a protein in a single entry then you should explicitly enter each chain as a separate line in the submission list: so for example **'1xg5:A'**, then a separate line for **'1xg5:B'** for two chains A and B.

Adding further protein selections.

Continue populating the list by selecting **'New entry'** (red arrow (1)). Type in **1edo** (red text, step (2)) and click **'Find chains'** (arrowed (3)), then **'Actualize'** (step (4)).



Completing the submission list

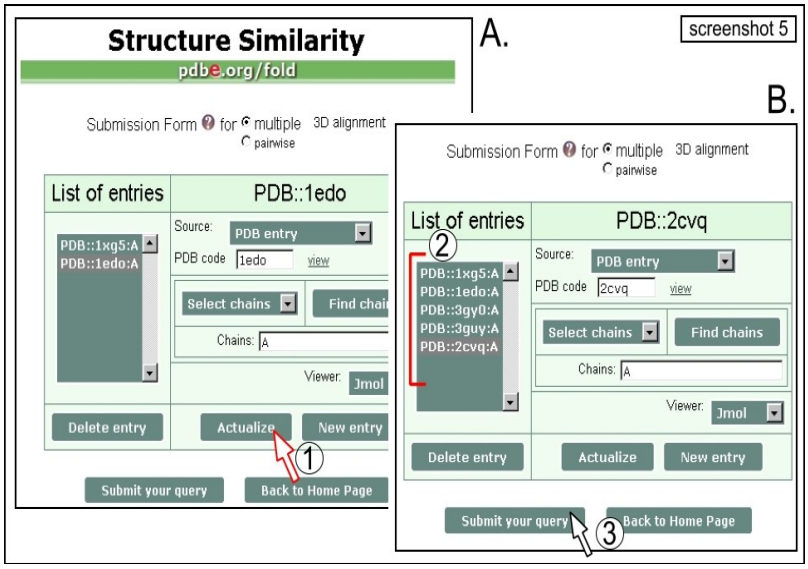
To complete the dehydrogenase superimposition you should continue adding the entries mentioned in the associated **Quips** article. Repeat the selection process for **1edo** for the following entries (picking chain A in each case): **3gy0**, **3guy**, and **2cvq**.

Screenshot 5 below shows **1edo** being added (red arrow (1)) and then on the

righthand side gives an example of how the completed list should look (highlighted in red in panel B (2)). Once your list is complete you can start the superimposition job using the **'Submit your query'** button (step number (3)).

Building the submission list.

Following screenshot 4 steps adds **1edo** as shown here (numbered step (1)). Repeating the selection steps for the other dehydrogenases gives the final list (numbered (2), bracketed in red). Click **'Submit...'** to start the job (arrowed, step (3))



Inspecting the results from the superimposition

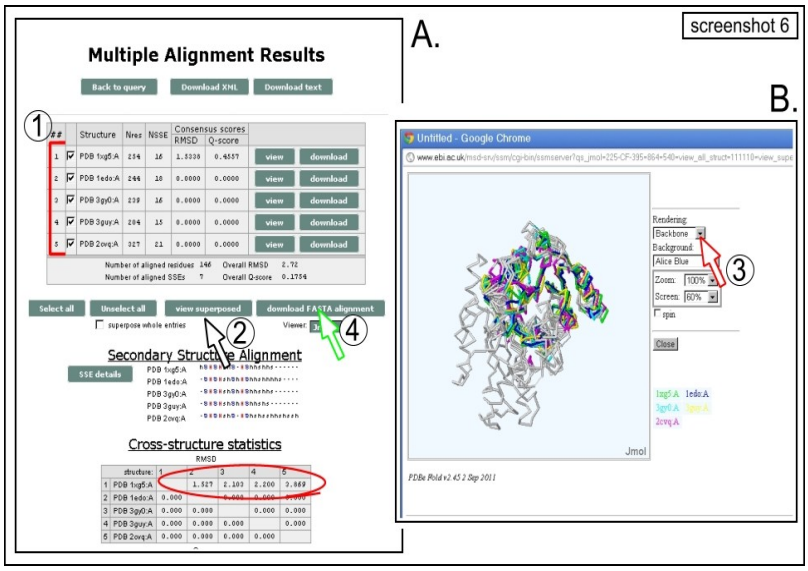
After the multiple superimposition has run, the results are presented as a page for you to browse. Screenshot 6 below shows the top of the results page. This lists the input files as a selectable list of checkboxes (bracketed in red, number (1)). If you want to select a subset for analysis you can uncheck entries here.

The first result you will be interested to see is, of course, the actual superimposed structures. You can view these by clicking on **'view superimposed'** (step number (2) below). A viewer window launched in this way is shown to the right of screenshot 6 (panel B). Here the default viewer **Jmol** has been used but the default representation changed (pulldown arrowed as step (3)). Because of the superimposition's complexity a simple backbone is a better way to look at it. This is an interactive viewer so you can rotate and see where the entries overlay well and where they differ.

The complete multiple superimposition shown cannot be downloaded directly as a single file. This is owing to the likely clash in residue numbers and chains in the superimposed structures. You can, however, download each superimposed structure separately using the **'Download'** link in the selection list. You can then read them separately into a molecular graphics viewer to recreate the superimposition.

Multiple alignment results.

After the superimposition **PDBeFold** presents the results as a complete page of alignments. You can select a subset for analysis using checkboxes (bracketted in red (1)). Click to launch a viewer (arrowed (2)). In the viewer it helps to change to backbone (arrowed (3)). The RMSD results are shown (circled in red). The superimposition yields a sequence alignment downloadable in FASTA format (arrowed in green (4).



Digging deeper into the results

This set of entries was specifically chosen to show the effects of increasing sequence divergence on structural similarity within this dehydrogenase family of proteins. The results here show the increasing RMSD from the human **1xg5** in the set of dehydrogenases (circled in red, screenshot 6). Further down the result page you will find the percentage identity shared with the human dehydrogenase **1xg5** for comparison with the RMSD values.

It is important to remember that all the **PDBeFold** statistics relate to the matched structural regions. Larger RMSD values would result if the complete structures including divergent loops were included. For the same reason the percentage identity values are lower than for the alignment of the complete sequences of the dehydrogenases (so the values differ from the ones you will find in the **Quips** article).

You can browse the alignment of matched residues corresponding to the superimposition in a table at the bottom of the results page (not visible in the screenshot). In this residue alignment table, the unmatched residues, which are in regions of divergent structure, are shown in **black**. You can download the residue-to-residue analyis as a multiple alignment in FASTA format using the button 'download FASTA alignment' (green arrow, labelled (4), screenshot 6).

The FASTA file aligns residues that were matched and superimpose well in space. These are shown in **uppercase**. This can seem odd when you see a residue conserved in sequence appearing in lower case. The likeliest explanation is that a structural change has moved it so far in space that **PDBeFold** has not matched it.

Finally, it is also possible to submit pre-prepared lists of PDB entries, or to add a structure not in the PDB, or to submit a set of such novel structures. You can refer to the **PDBeFold** help section to get detailed in instructions on uploading these.