PDBe TUTORIAL

PDBeFold (SSM: Secondary Structure Matching)
http://pdbe.org/fold/

This PDBe tutorial introduces PDBeFold, an interactive service for comparing protein structures in 3D. This service provides:

- Pairwise and multiple comparison and 3D alignment of protein structures
- Examination of a protein structure for similarity with the whole Protein Data Bank (PDB) archive or SCOP.
- Best Cα-alignment of compared structures
- Download and visualisation of best-superposed structures using various graphical packages

PDBeFold structure alignment is based on identification of residues occupying "equivalent" geometrical positions. In other words, unlike sequence alignment, residue type is neglected. The PDBeFold service is a very powerful structure alignment tool which can perform both **pairwise** and **multiple** three dimensional alignment. In addition to this there are various options by which the results of the structural alignment query can be sorted. The results of the **S**econdary **S**tructure **M**atching can be sorted based on the Q score (Cα-alignment), P score (taking into account RMSD, number of aligned residues, number of gaps, number of matched Secondary Structure Elements and the SSE match score), Z score (based on Gaussian Statistics), RMSD and % Sequence Identity.

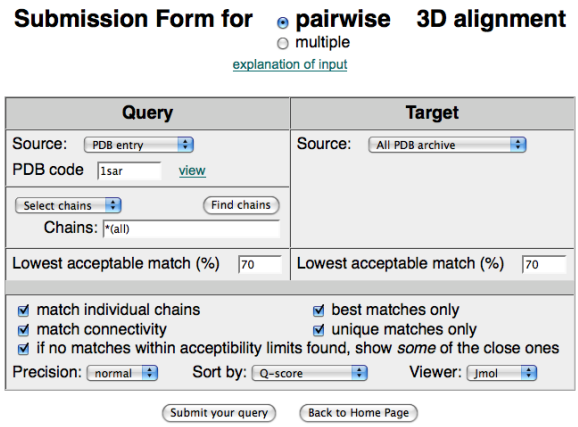
It is hoped that at the end of this tutorial users will be able to use PDbeFold for the analysis of their own uploaded structures or entries already in the PDB archive.

Tutorial

PDBeFold can may be accessed from multiple locations on the PDBe website. From the PDBe home page (http://pdbe.org/), there are two access points for the program as shown below.

The link highlighted in blue may be used to automatically get all structures related to a given PDB code. In order to start the service, click on the link shown in red above. This will open up an introductory PDBeFold page containing additional information, tips and help regarding the service. On this page, now click on the "Start PDBeFold" button.

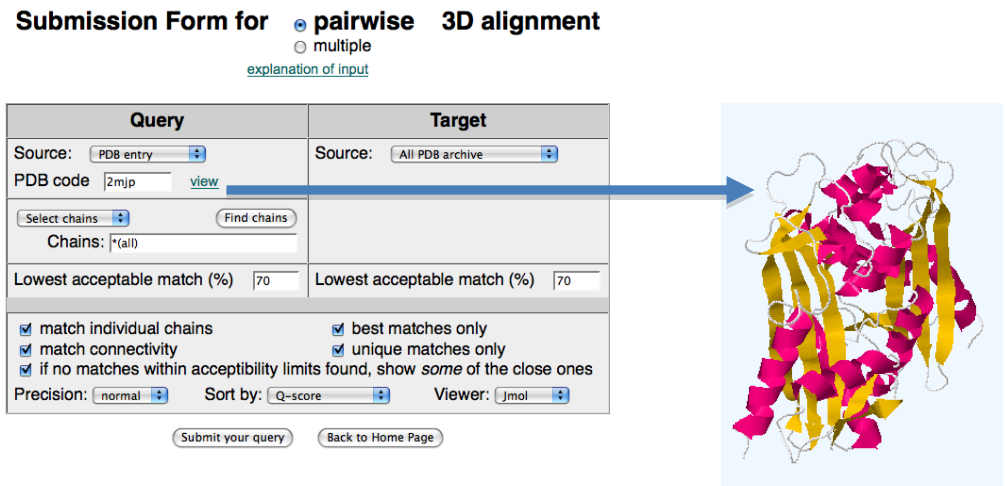
This will start up the PDBeFold service, which will appear in the form of a submission form as shown below.



The form contains parameters that may be adjusted or tweaked in order to customize the results as required. The "Query" contains information regarding

the structure/structures that need to be compared. This could be an existing PDB code, an uploaded coordinate file, SCOP entry or pair of existing PDB entries. In addition to this, if an entry contains more than one chain, the chain that has to be compared can be chosen. Similarly the target for the search could be another PDB entry, uploaded coordinate file or SCOP set or a set of files. The "lowest acceptable match" boxes tell the program what cut-off to use in the matching process. The default is 70% both for the query and target, meaning that in order to list a comparison as a match, at least 70% of the secondary structure of the query must match 70% of the target structure. These settings can be adjusted. Setting this to a lower value will result in a large number of hits and the opposite may result in only identical structures found.

We will use the PDB entry 2MJP (http://pdbe.org/2mjp) (STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226) as an example for demonstrating PDBeFold.



Click on the "Submit your query" button. This will submit an alignment job on the computer farm which will take up to 2-3 minutes depending on server load. Once the results are calculated, a new page will be displayed on the browser, similar to that shown below.

##	Q	P	Z	Rmsd	N_{align}	N_g	$\%_{seq}$	$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	26.8	15.5	0.00	184	0	100	100	2mjp:A	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
2	1.00	25.4	15.0	0.20	184	0	100	100	1b78:A	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
3	0.95	22.3	14.1	0.49	182	1	99	93	1b78:B	93	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
4	0.95	20.0	13.3	0.55	182	1	99	86	2mjp:B	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
5	0.85	15.3	11.6	1.14	182	2	50	93	2dvn:B	100	186	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF IMP FROM PYROCOCCUS HORIKOSHII
6	0.68	9.1	8.9	1.92	181	3	49	93	1v7r:A	100	186	☐	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3
7	0.67	8.5	8.6	1.96	181	3	49	93	2dvo:A	100	185	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF ITP FROM PYROCOCCUS HORIKOSHII
8	0.67	9.0	8.9	1.95	181	3	49	93	2dvn:A	100	186	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF IMP FROM PYROCOCCUS HORIKOSHII
9	0.66	6.7	7.7	1.93	177	5	50	79	2dvp:A	92	184	☐	STRUCTURE OF NTPASE FROM PYROCCOUS HORIKOSHII
10	0.65	7.9	8.3	2.08	181	3	49	93	2e5x:A	100	185	☐	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3
11	0.65	8.7	8.8	1.42	168	7	35	71	2car:B	83	194	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
12	0.64	8.6	8.8	1.48	170	7	35	71	2car:A	83	196	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
13	0.64	8.8	8.8	1.41	168	7	35	71	2i5d:A	83	195	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATE PYROPHOSPHATASE
14	0.63	5.4	7.0	1.81	173	6	35	71	1vp2:A	91	189	☐	CRYSTAL STRUCTURE OF PUTATIVE XANTHOSINE TRIPHOSPHATE PYROPHOSPHATASE/HAM1 PROTEIN HOMOLOG (TM0159) FROM THERMOTOGA MARITIMA AT 1.78 A RESOLUTION

The results here are sorted based on Q-score (Quality of alignment, with 1 being the highest score) for this entry. The pairwise alignment result between **2mjp** and **2e5x** is highlighted in the above figure. There is a **49% amino acid sequence identity** between the two proteins, whereas they have **93% secondary structure identity**. Clicking on the number link on the left hand side of the page will return a residue-by-residue description about the structural alignment between the two proteins. *Note: as the PDB archive continuously adds new entries every week, it is possible that this result may not appear in the same position as shown.*

Match 11 of 29

Back to match list | first match | << | >> | last match | Back to query

Query PDB 2mjp:A				Alignment				Target PDB 2e5x:A			
N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$	Q	P	RMSD	N_{align}	N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$
184	98	14	93	0.650	7.93	2.081	181	185	98	13	100
STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226				$\%_{seq}$	Z	N_{SSE}	N_{gaps}	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3			
				48.6	8.35	13	3				

view | download | view superposed | view | download

☐ superpose whole entries Viewer: Jmol

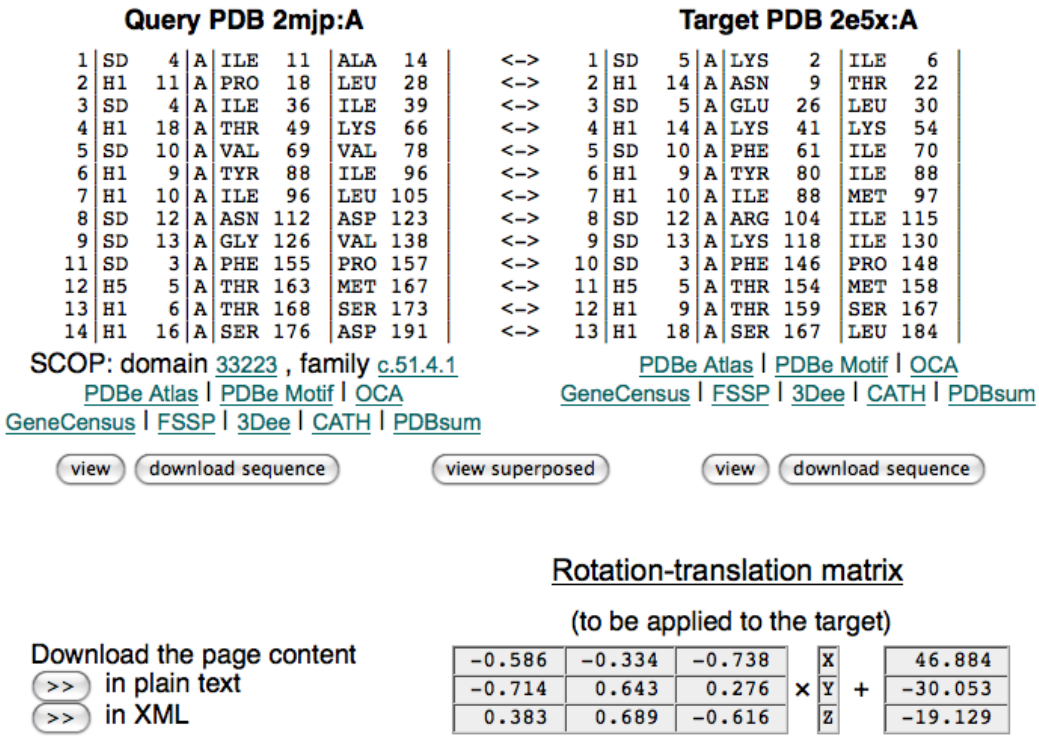
Secondary Structure Alignment

```
2mjp:A    SHSHSHHSShSHHH
2e5x:A    SHSHSHHSS-SHHH
```

Click on the "View Superposed" button to view the secondary structure alignment.

The two structures 2MJP and 2E5X appear share a high level of secondary structure identity despite only having <50% sequence identity. This indicates that the two proteins belong to the same structural family.

A little further down the page is the summary of the alignment showing the secondary structure elements that match between the two structures, with residue ranges and the matrix required to move and align the target to the query structure.



Query PDB 2mjp:A

1	SD	4	A	ILE	11	ALA	14	<->	1	SD	5	A	LYS	2	ILE	6
2	H1	11	A	PRO	18	LEU	28	<->	2	H1	14	A	ASN	9	THR	22
3	SD	4	A	ILE	36	ILE	39	<->	3	SD	5	A	GLU	26	LEU	30
4	H1	18	A	THR	49	LYS	66	<->	4	H1	14	A	LYS	41	LYS	54
5	SD	10	A	VAL	69	VAL	78	<->	5	SD	10	A	PHE	61	ILE	70
6	H1	9	A	TYR	88	ILE	96	<->	6	H1	9	A	TYR	80	ILE	88
7	H1	10	A	ILE	96	LEU	105	<->	7	H1	10	A	ILE	88	MET	97
8	SD	12	A	ASN	112	ASP	123	<->	8	SD	12	A	ARG	104	ILE	115
9	SD	13	A	GLY	126	VAL	138	<->	9	SD	13	A	LYS	118	ILE	130
11	SD	3	A	PHE	155	PRO	157	<->	10	SD	3	A	PHE	146	PRO	148
12	H5	5	A	THR	163	MET	167	<->	11	H5	5	A	THR	154	MET	158
13	H1	6	A	THR	168	SER	173	<->	12	H1	9	A	THR	159	SER	167
14	H1	16	A	SER	176	ASP	191	<->	13	H1	18	A	SER	167	LEU	184

SCOP: domain 33223 , family c.51.4.1

PDBe Atlas | PDBe Motif | OCA

GeneCensus | FSSP | 3Dee | CATH | PDBsum

view | download sequence | view superposed | view | download sequence

Target PDB 2e5x:A

PDBe Atlas | PDBe Motif | OCA

GeneCensus | FSSP | 3Dee | CATH | PDBsum

Download the page content

>> in plain text

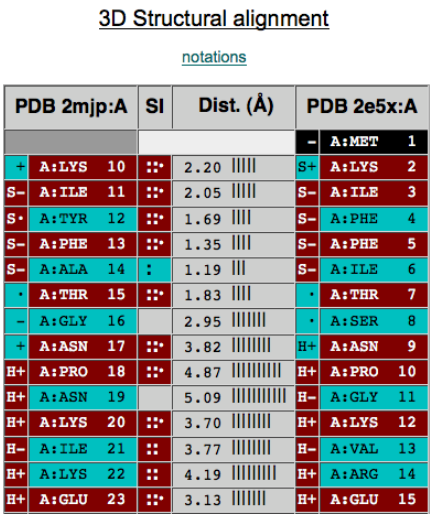
>> in XML

Rotation-translation matrix

(to be applied to the target)

-0.586	-0.334	-0.738	X	+	46.884
-0.714	0.643	0.276	× Y	+	-30.053
0.383	0.689	-0.616	Z		-19.129

If you scroll down the page, there is a 3D structural alignment between the residues from the corresponding PDB entries. This alignment shows the extent of superposition between similar secondary structure folds. A **red** background indicates that matched residues have same residue name; a **cyan** background is used for other matched residues. Unmatched residues are indicated by a **black** background. The letters H and S represent helix and β-strands respectively.



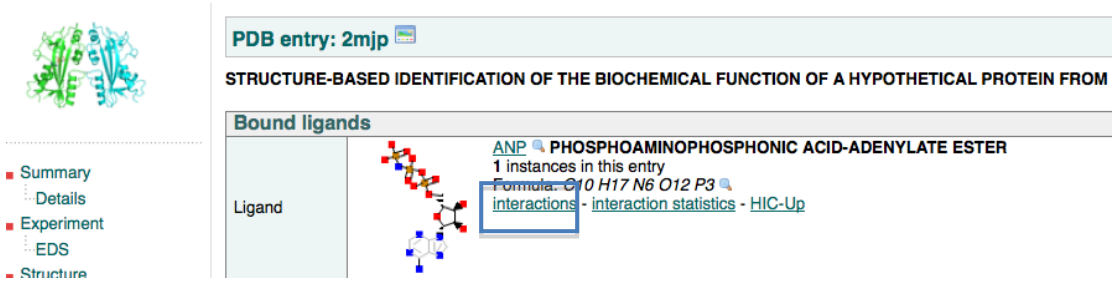
3D Structural alignment

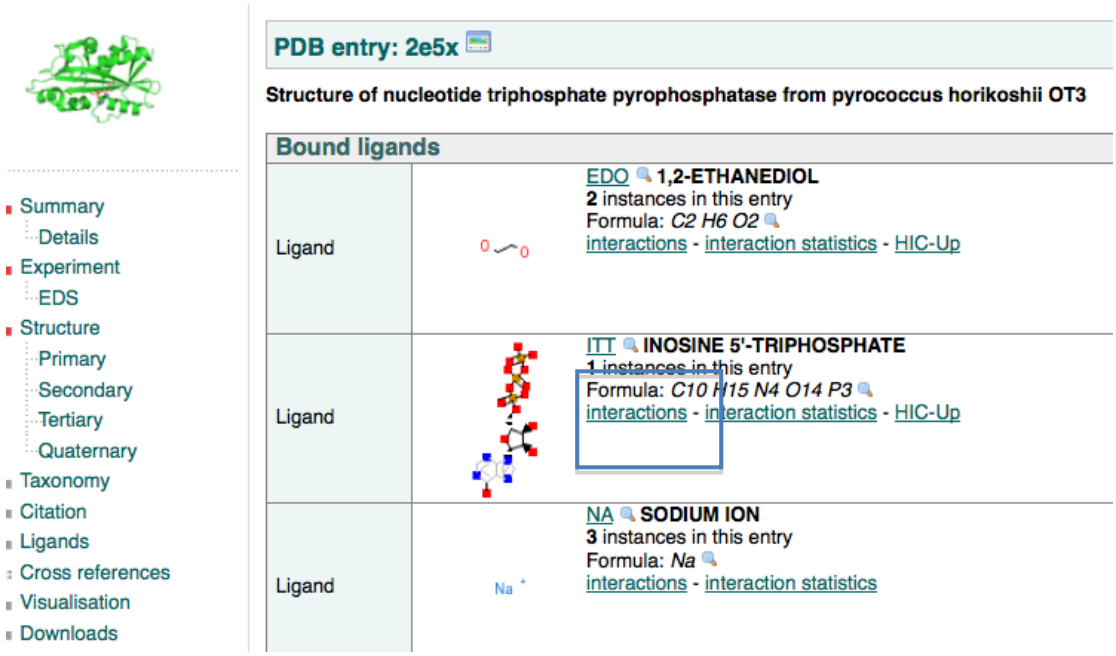
notations

PDB 2mjp:A	SI	Dist. (Å)	PDB 2e5x:A
			- A:MET 1
+ A:LYS 10	::	2.20	S+ A:LYS 2
S- A:ILE 11	::	2.05	S- A:ILE 3
S• A:TYR 12	::	1.69	S- A:PHE 4
S- A:PHE 13	::	1.35	S- A:PHE 5
S- A:ALA 14	:	1.19	S- A:ILE 6
• A:THR 15	::	1.83	• A:THR 7
- A:GLY 16		2.95	• A:SER 8
+ A:ASN 17	::	3.82	H+ A:ASN 9
H+ A:PRO 18	::	4.87	H+ A:PRO 10
H+ A:ASN 19		5.09	H- A:GLY 11
H+ A:LYS 20	::	3.70	H+ A:LYS 12
H- A:ILE 21	::	3.77	H- A:VAL 13
H+ A:LYS 22	::	4.19	H+ A:ARG 14
H+ A:GLU 23	::	3.13	H+ A:GLU 15

The residue-by-residue mapping provides a very useful tool for the analysis of the structure function relationship between the two entries.

Let us now concentrate on the proteins themselves to get more information about them. You can go to the summary pages for these entries in new browser windows or tabs. Go to http://www.pdbe.org/2mjp in one window and http://www.pdbe.org/2e5x in another. On each of these pages, click on the "Ligands" link in the left-hand sidebar.



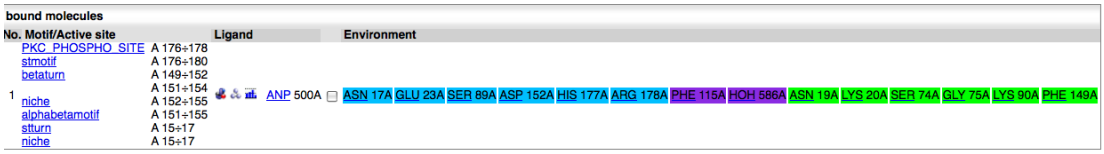


As you can see from the above screenshots, both entries contain nucleotide-type ligands. The ligand bound to 2MJP is ANP (PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER). For 2E5X, the bound nucleotide is ITT (INOSINE 5'-TRIPHOSPHATE). Given that both entries share a high degree of structural similarity and also bind nucleotide-like ligands, do they also share the same or similar binding sites?

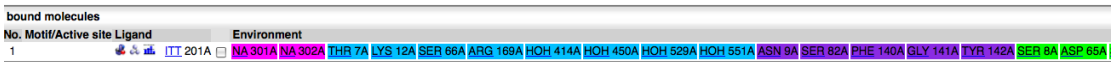
It should be possible to answer that question by looking at the ligand interaction results for these two ligands from the corresponding entries. Click on the "interactions" link for each of the two ligands as shown above. This takes you to

the details of the residue interactions in both entries using another PDBe service (PDBeMotif: http://pdbe.org/motif/).

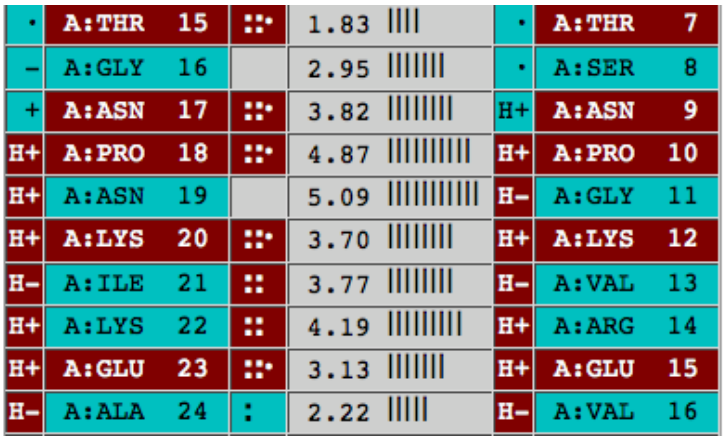
2MJP



2E5X



Going back to the residue-by-residue 3D-mapping results provided by PDBefold, we can make some very interesting observations. For example, ASN 17 and LYS 20 interacts with the ligand ANP in 2MJP, and also aligns with ASN 9 and LYS 12 from PDB entry 2E5X. Both ASN 9 and LYS 12 are also involved in interactions with ITT in a similar manner.



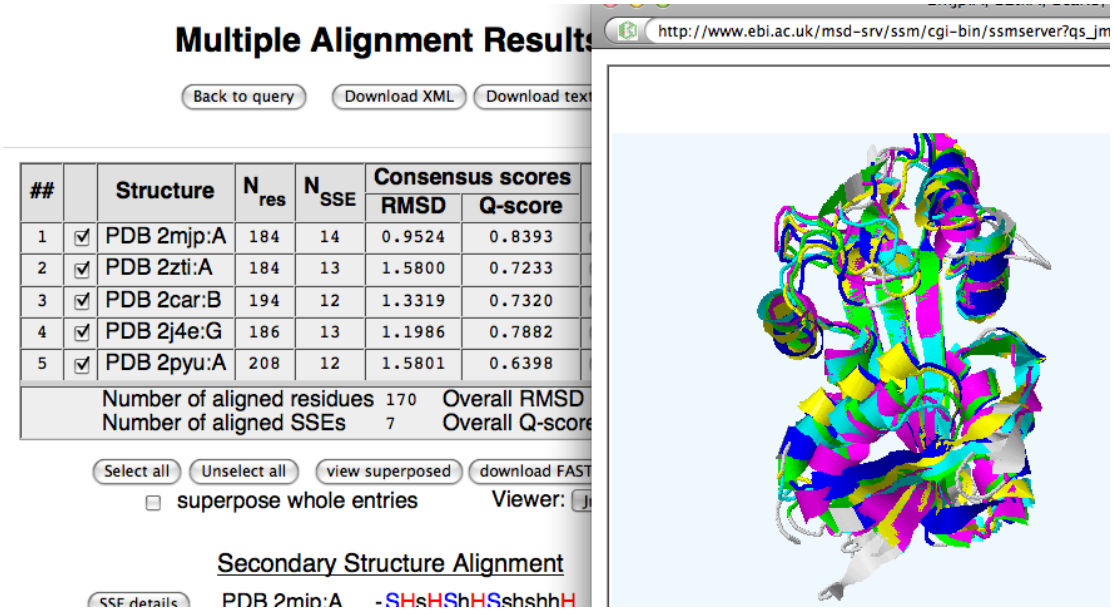
In this way if we compare the rest of the residues present in the binding environment for ANP and ITT, it will be very clear that both the proteins try to adopt a similar binding environment in order to interact with a nucleotide ligand. This observation also indicates that the hypothetical protein represented by the PDB entry 2MJP is a potential pyrophosphatase similar to the protein under PDB entry 2E5X. **Therefore in this case, the function of the protein**

molecule dictates its overall structural fold rather than its sequence identity.

Go back to all the results of the structural comparison of 2MJP with the PDB.

9	0.66	8.2	8.5	2.04	181	3	49	93	2zti:A	100	184	☑	STRUCTURES OF DIMERIC NONSTANDARD NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3: FUNCTIONAL SIGNIFICANCE OF INTERPROTOMER CONFORMATIONAL CHANGES
10	0.66	6.7	7.7	1.93	177	5	50	79	2dvp:A	92	184	☐	STRUCTURE OF NTPASE FROM PYROCCOUS HORIKOSHII
11	0.65	7.9	8.3	2.08	181	3	49	93	2e5x:A	100	185	☐	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3
12	0.65	8.7	8.8	1.42	168	7	35	71	2car:B	83	194	☑	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
13	0.64	8.6	8.8	1.48	170	7	35	71	2car:A	83	196	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
14	0.64	8.8	8.8	1.41	168	7	35	71	2i5d:A	83	195	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATE PYROPHOSPHATASE
15	0.63	5.4	7.0	1.81	173	6	35	71	1vp2:A	91	189	☐	CRYSTAL STRUCTURE OF PUTATIVE XANTHOSINE TRIPHOSPHATE PYROPHOSPHATASE/HAM1 PROTEIN HOMOLOG (TM0159) FROM THERMOTOGA MARITIMA AT 1.78 A RESOLUTION
16	0.63	5.3	7.0	1.87	174	5	34	71	1vp2:B	91	189	☐	CRYSTAL STRUCTURE OF PUTATIVE XANTHOSINE TRIPHOSPHATE PYROPHOSPHATASE/HAM1 PROTEIN HOMOLOG (TM0159) FROM THERMOTOGA MARITIMA AT 1.78 A RESOLUTION
17	0.62	7.9	8.5	1.78	169	8	35	71	2j4e:G	77	186	☑	THE ITP COMPLEX OF HUMAN INOSINE TRIPHOSPHATASE
18	0.61	8.2	8.6	1.76	169	7	34	71	2j4e:E	83	190	☐	THE ITP COMPLEX OF HUMAN INOSINE TRIPHOSPHATASE
19	0.60	3.8	6.8	1.81	177	7	33	71	1k7k:A	91	207	☐	CRYSTAL STRUCTURE OF RDGB- INOSINE TRIPHOSPHATE PYROPHOSPHATASE FROM E. COLI
20	0.60	3.4	6.7	1.79	176	6	32	71	2pyu:A	83	208	☑	STRUCTURE OF THE E. COLI INOSINE TRIPHOSPHATE PYROPHOSPHATASE RGDB IN COMPLEX WITH IMP

The check box shown on the results page can be used to select entries to submit for multiple alignment from the pairwise alignment pages. Check a few of these results and choose “Submit for Multiple Alignment” from the bottom of the results page. This will do a multiple alignment between the chosen entries and the results will be shown similar to that seen for the pairwise alignment. You may choose the “View Superposed” button to see all your chosen entries aligned.

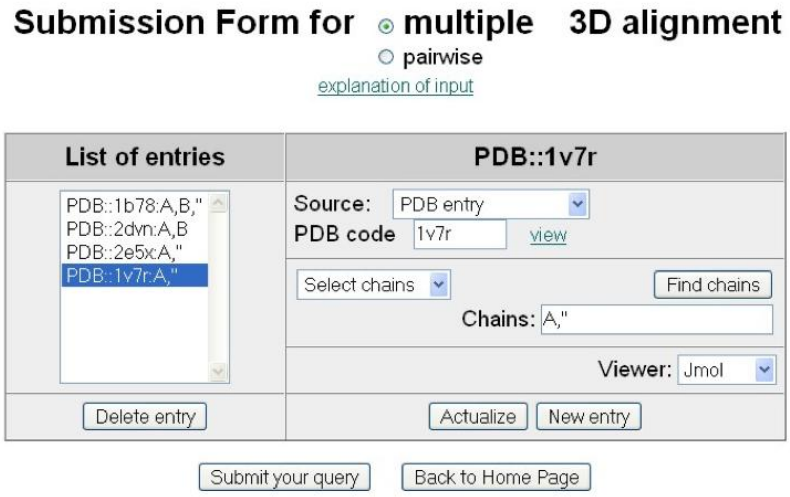


As can be seen all our chosen entries appear to share the same overall structure.

Multiple Alignments using PDBeFold

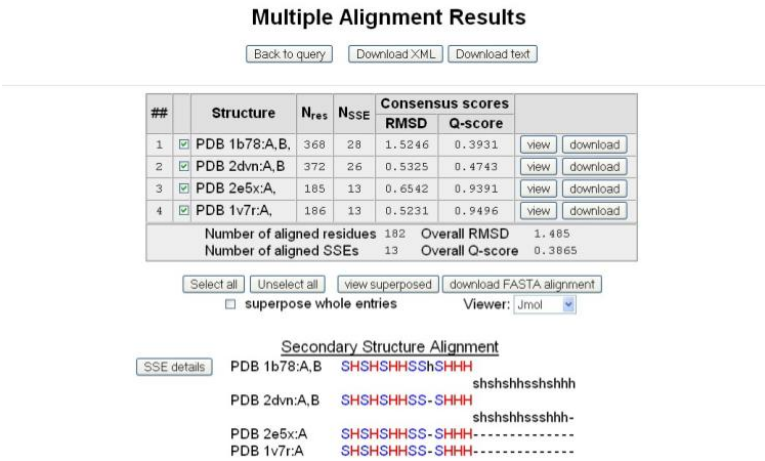
From the PDBeFold submission form, choose the radio button for "Multiple Alignment" to bring up the submission form for multiple alignments.

The multiple alignment option provided by the PDBeFold service is very useful when you have more than two proteins for which you need to compare structural folds.

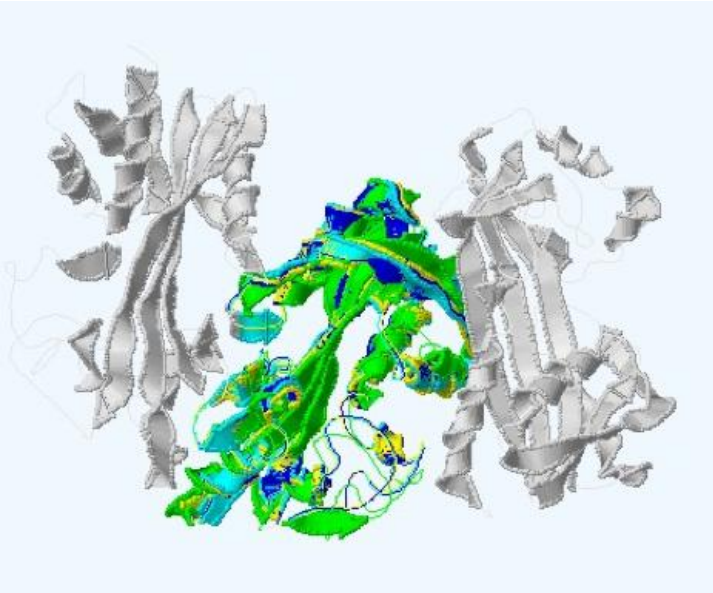


In the query page for "multiple alignment", enter the PDB ID of your choice and click on the "Find chains" button. This will upload the ID code in the "List of entries" column as shown above. Once an entry is loaded, click on the "new entry" button to upload the next entry in a similar way. The "delete entry" button enables the user to remove the highlighted PDB ID code from the list of entries.

The **Actualize** button updates the data for the highlighted entry from the given list using details input on the right section of the Submission Form. Once all the PDB entries are uploaded, click on the "submit your query" button.



The multiple alignment results page shows structural alignment among the four entries that were uploaded (1b78, 2dvn, 2e5x and 1v7r). You can also view the superposed entries in a graphics viewer by pressing the “Show Superposed” button.



The central domain region in all proteins appears to be highly similar. Analysis of the residue-by-residue mapping data (in a similar manner as previously done for pairwise alignment) also indicates a high degree of similarity observed in the active binding site (nucleotide binding) of all these proteins.

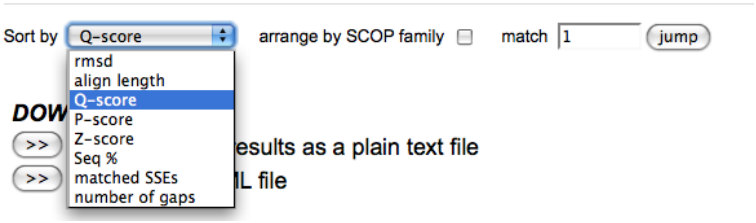
These results obtained from PDBeFold provide encouraging prospects of understanding possible roles of a hypothetical protein or structural genomics proteins whose function is yet to be determined.

Other examples

Secondary structure alignments can often show relationships that are not immediately obvious from sequence identity alone. Here are a few examples which you may find interesting.

a) Alpha-lactalbumin (PDB entry 1A4V http://pdbe.org/1a4v).

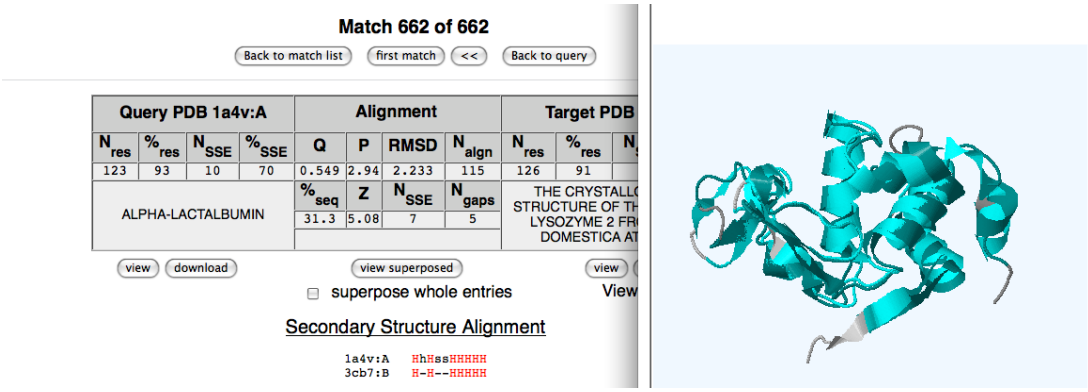
Start a PDBeFold comparison for all entries in the PDB archive against 1A4V. Once the results are shown, scroll to the bottom of the page and sort by %seq instead of Q-score.



Now scroll to the last page and choose one of the results from the last page.

660	0.76	6.5	7.5	1.45	122	5	34	80	21z2:A	100	129	☐	THE THREE DIMENSIONAL STRUCTURE OF TURKEY EGG WHITE LYSOZYME AT 2.2 ANGSTROMS RESOLUTION
661	0.57	2.9	5.1	2.16	114	6	33	70	2h5z:A	100	122	☐	CRYSTALLOGRAPHIC STRUCTURE OF DIGESTIVE LYSOZYME 1 FROM MUSCA DOMESTICA BOUND TO CHITOTETRAOSE AT 1.92 A RESOLUTION
662	0.55	2.9	5.1	2.23	115	5	31	70	3cb7:B	100	126	☐	THE CRYSTALLOGRAPHIC STRUCTURE OF THE DIGESTIVE LYSOZYME 2 FROM MUSCA DOMESTICA AT 1.9 ANG.

Entry 3cb7 has 31% sequence identity but 70% structure similarity between the two proteins. Look at the details of the match. The structures are highly similar.

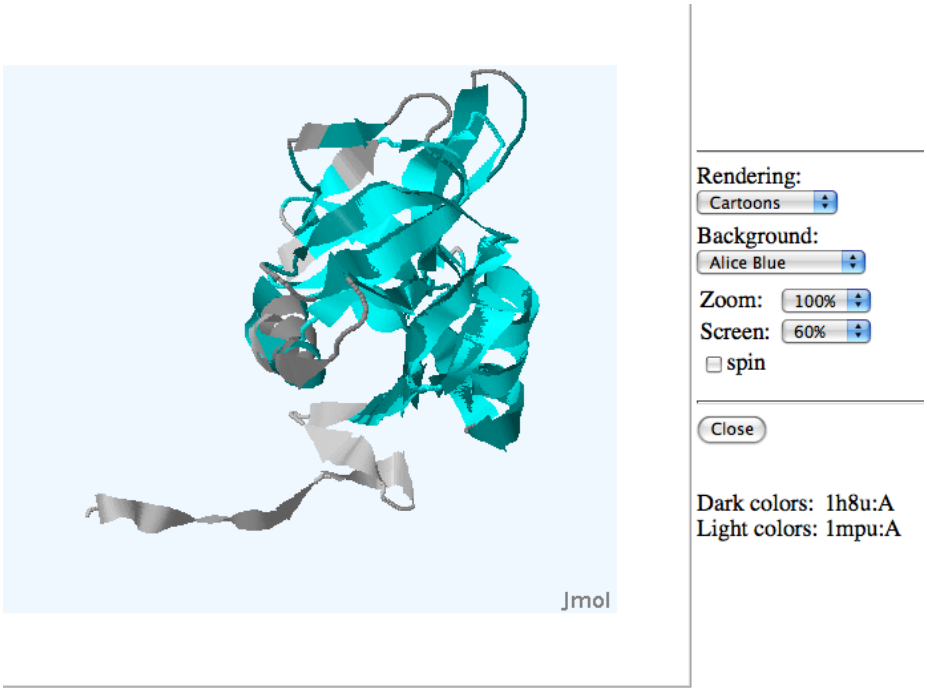


b) Eosinophil Major Basic Protein (PDB entry 1H8U http://pdbe.org/1h8u).

Do a search for structural similarity for PDB entry 1H8U against the whole PDB archive. Once the results are in, resort the results by %seq identity as in the previous example. Scroll to the last page.

191	0.39	5.7	7.4	1.65	85	7	14	88	1kcg:B	70	123	☐	NKG2D IN COMPLEX WITH ULBP3
192	0.38	0.6	3.1	2.01	93	8	14	75	1kg0:C	75	136	☐	STRUCTURE OF THE EPSTEIN-BARR VIRUS GP42 PROTEIN BOUND TO THE MHC CLASS II RECEPTOR HLA-DR1
193	0.43	5.9	7.5	1.89	94	8	14	88	1mpu:A	78	128	☐	CRYSTAL STRUCTURE OF THE FREE HUMAN NKG2D IMMUNORECEPTOR

The last hit in the list is PDB entry 1MPU. This entry has 14% sequence identity with our query structure while sharing 88% structural identity. Look at the details of this hit and view the superposed entries as previously.



Scroll down the details page and look at the residue-by-residue listing. All the CYS residues between the two structures are conserved and at the same equivalent positions. These residues form the disulphide bonds in the two structures that keep the scaffold of the protein intact. Both these proteins belong to a large family of sugar-binding proteins called c-type lectins. All c-type lectins share the same overall structure constrained by the disulphide bonds, and not all proteins in this family actually bind sugars.

c) PDB entry 1TIM and 2O55. Do a pairwise alignment between 1TIM and 2O55 using the pairwise alignment form. The results will show that the two proteins share 4% sequence identity and 70% structural similarity.

Structure Alignment Results

explanation of output

Query: pdb entry **1tim:A** , 247 residues
STRUCTURE OF TRIOSE PHOSPHATE ISOMERASE FROM CHICKEN MUSCLE

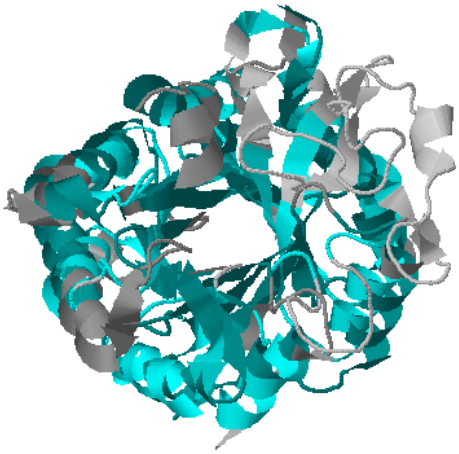
Examined 1 entry (1 chain),
Match 1 of 1.

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	0.19	-0.0	3.1	3.62	171	14	4	70	2o55:A	82	254	☐	CRYSTAL STRUCTURE OF A PUTATIVE GLYCEROPHOSPHODIESTER PHOSPHODIESTERASE FROM GALDIERIA SULPHURARIA

Examined 1 entry (1 chain),
Match 1 of 1.

Back to query



Rendering: Cartoons
Background: White
Zoom: 100%
Screen: 60%
spin

Close

Dark colors: 1tim:A
Light colors: 2o55:A

Both proteins have minimal sequence identity and yet belong to the same fold class (the TIM barrel).

This ends our tutorial on PDBeFold. We hope you found this useful and will be able to use this tool in your future research and analysis. It should be clear from the examples given in this tutorial that fold space is more restricted than sequence space and most proteins tend to fall broadly into one of the many fold classes already found in the PDB. If you need to get in touch with the PDBe regarding any aspect of the program, please email pdbehelp@ebi.ac.uk and we will try to assist you in any way possible.