This PDBe tutorial introduces the PDBMotif service, which is an integrated search system for protein and nucleic acid structures. PDBMotif allows users to retrieve crucial information regarding different small molecule binding sites and secondary structure elements present in their protein of interest. The nature of interactions between the ligands and macromolecules (proteins, nucleic acids) is highly significant in terms of their structure-function relationships and possible roles in cellular mechanism. In addition, PDBMotif returns a list of other structures that share similar secondary structure fold/patterns with the query protein and also have analogous binding environment for ligands. All the above information can be very useful to a user trying to navigate through a macromolecular structure of unknown function.

Currently PDBMotif is the only tool that combines protein sequence, chemical structure and 3D data in a single speedy search. The aim of this tutorial is to introduce the users to this very powerful research and analysis tool by using a few examples. 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.

**How to use this tutorial**

Please keep this tutorial guide open in a browser or PDF viewer; then start the tutorial itself by opening a new browser window and following the steps outlined in the tutorial guide.

**Tutorial**

PDBMotif can be accessed from multiple locations on the PDBe home page ([http://www.pdbe.org/](http://www.pdbe.org/)) as shown in the screenshot below. You may also go directly to the PDBMotif system by pointing your browser to [http://pdbe.org/motif](http://pdbe.org/motif).
The PDBeMotif start page will look as below:

The PDBeMotif service provides a number of powerful search engines that may be used separately or in combination. You can construct your query based on the following options:

- Define search by PDB ID code – PDB Header Search Option
- Define search by author last name - PDB Header Search Option
- Define search by type of experiment (X-ray, NMR etc) - PDB Header Search Option
- Define search by three letter code for the ligand (e.g. HEM, NAG, BMA) – Molecule Binding Option
- Define search by metal site geometry (octahedral, tetrahedral etc.)- Molecule Binding Option
- Define search by ligand environment (same/similar) and Prosite patterns – 3D environment Option
- Search on Pair Binding Statistics (e.g. how does ATP interacts with LYS?) – Pair Bonds Option

Let us begin by using the ‘PDB Header Search’ option from the main page (as shown above). To begin exploring PDBeMotif, type in ‘1a0t’ in the PDB Codes box. The PDB entry 1a0t represents a sucrose specific porin, which contains sucrose and Ca2++ ion as bound ligand.

Click on the first icon in the results list as highlighted above to view this structure graphically.

This porin is an outer membrane protein, a homotrimer where each monomer is made up of 18- stranded antiparallel beta barrel. The beta sheets are connected by short turns on the periplasmic side and longer loops facing towards the exterior of the cell membrane with the exception of long loop L3, which protrudes into the beta barrel of each monomer about half way through the membrane and forms a constriction zone with in the channel. The loop L3 (residues 185-204) is directly involved in substrate recognition and binding (Journal of Bacteriology, Vol 181, pg 1920-23, 1999).
You can click through the various checkboxes on the side to view specific motifs present in the structure. The terms themselves provide more information regarding the motif or pattern. The entire sequence of Sucrose specific porin is mapped against PROSITE motifs, small 3D motifs, MEROPS sequence and as well as well known catalytic sites.

To view the results in more detail, go back and click on the link as shown by the orange circle in the results screenshot. The Results Page for the PDBeMotif is divided into 4 different sections. They are as follows:

1. Sequences
2. 3D Motifs
3. Ligand Environment
4. Ligand Bonds
The **Sequences Section**

The *Sequences* section provides information about the secondary structure elements present across the entire amino acid sequence for each chain. It also highlights the individual amino acid residues that are involved in interactions with the small molecules/ligands.

The Sequence results for PDB entry 1a0t indicates that there are 3 chains (P, Q, R) present in the asymmetric unit of the crystal. The residues are colour-coded based on their chemical nature *(hydrophobic/hydrophilic/intermediate)*.

The secondary structure description for the sequence uses the standard DSSP notation:

- **H**: alpha helix
- **B**: residue in isolated beta-bridge
- **E**: extended strand, participates in beta ladder
- **G**: 3-helix (3/10 helix)
- **I**: 5-helix (pi helix)
- **T:** hydrogen bonded turn
- **S:** bend
- **.:** no assigned structure

The amino acid residues that participate in the binding of small molecules/ligands are colour-coded based on their nature of interactions (covalent, ionic, H-bond, van der Waals, planar). These are represented as coloured ‘*’s above the sequence and align with the rows of ligands (shown below).

The Motif Search criteria at the top of this page help the user to explore the different motifs/patterns present in the PDB entry to a greater detail. It is possible to specify a range of residues and visualize the fragment or to search the entire PDB archive based on the following:

- Sequence of φ/ψ angles
- Sequence (residue codes)
- Sequence formed of secondary structure elements

The mapping of the protein sequence in this structure against PROSITE motifs, small 3D motifs, MEROPS sequence and well-known catalytic sites is provided on the left-hand side of the page. These may be visualised from here if rasmol is setup on your computer. The small graph icon next to each feature does a search for all ligands that have been found to bind to that feature. Click on the “TYR_PHOSPHO_SITE” graph to conduct a search.
This will give the results in graphical form as shown below:

Each row in this graph shows the ligands which bind to this motif, as well as the nature of the interaction colour-coded based on the type of interaction. Clicking on any of the bars will do a further search and show all entries that contain the interaction of the given ligand with the TYR_PHOSPHO_SITE motif. The results may be further chosen for local alignment and display.

**The 3D Motifs Section**

Go back to the Sequence view for 1a0t. The 3D motifs section displays each chain and its associated ligands present in the PDB entry 1a0t individually. This page provides detailed information and the opportunity to view the different 3D motifs and patterns present throughout the sequence and the amino acid
composition of those motifs. There is also a Search option that allows the user to search across the PDB archive for motifs with similar 3D configuration or with similar amino acid sequence.

In this example, we perform a 3D motif configuration based search for the following residues: Betaturm 217:WDNG as shown below.

Click on search to get a set of entries that have a similar 3D configuration. The results page will look something like the one shown below. Choose as many of these as you wish and view the aligned motifs graphically.
The Ligand Environment Section

The nature of interactions between the ligands and macromolecules (proteins, nucleic acids) is highly significant in terms of their structure-function relationships and possible roles in cellular mechanism. The Ligand Environment section of the search provides detailed three-dimensional knowledge about the active site of the biological macromolecules.

Our PDB entry 1a0t has Ca2+ ions and sucrose molecules bound to the protein. The interacting residues are coloured based on the nature of the interaction (described in the legend highlighted below).

For each of the ligands displayed in the Ligand Environment page, there is a link to Jmol/Rasmol which enables the user to visualise only the interactions between the ligand and its environment. Choose to view the environment for SUC 1R.

There are two sucrose molecules bound to each monomer. The Glucose and Fructose ring of the sucrose disaccharide are aligned almost parallel to the aromatic sidechains and are involved in van der Waals and planar interactions with these aromatic amino acids.
This type of hydrophobic interaction helps the sugar molecule to glide through the channel of the porin and reach the periplasmic side of the membrane. Now look at the environment around Ca 10R.

The Calcium ion displays tetrahedral coordination geometry. This calcium ion is situated in a loop at the exterior side of the membrane where many negatively charged residues are present. This calcium is therefore thought to partially compensate the negative charge. Click on the graph icon next to this Calcium on the Ligand environment page to see all ligands that bind to the residues present around this calcium (ANLD).

These residues predominantly appear to be present in an environment that binds ligands such at NAD, GDP etc. Further down on this page you can find CA (calcium). Click on this bar to see other entries that bind calcium and contain the same residues in the binding site.

**The Ligand Bonds Section**

This section of the search result page contains detailed description of the bonds between ligands and other residues that make up its binding site. As always all bond information is coloured based on the nature of interaction. The details of the atoms that make the interactions between the ligand and neighbouring residues are shown in tabular form.
To find other entries in the PDB archive that have the same relative positions of the residues that make up the Calcium binding site, check all the magenta residues for CA 10R, and search by Cα position. This search may take a few minutes to complete.

Once the results are available, choose a few entries from the list and view them graphically using the menu at the bottom of the page.

The results could look like those shown below if only the hits are aligned, where
the same residues appear to be in the same juxtaposition in all the entries. However, if you choose to view the alignments for the whole chain, a different picture emerges! These entries have totally different fold and secondary structures.

This example merely indicates that similar 3D configurations could occur between totally unrelated proteins and mere secondary structure analysis may not necessarily be enough to analyse the function of a protein fully.

**Constructing Custom Queries in PDBeMotif**

PDBeMotif allows users to construct their own queries using different search elements (e.g. small molecules, secondary structure pattern, phi-psi angles, interactions etc). Go back to [http://pdbe.org/motif](http://pdbe.org/motif) and click on the “Search” option from the top of the page.

In the following example, we will construct a query based on small molecule and protein sequence pattern. Let us find PDB entries that represent nucleotide-binding proteins that take part in signalling process.
Click on the link for Protein sequence pattern icon on the right side of the page and paste the following sequence GXXXXGKT, which is the well-known signature pattern for Walker A motif. The Walker A motif is implicated in nucleotide-binding in signal transduction proteins. Click on the button ‘Add it to the search area’ which is highlighted in red.

In a similar manner click on the Search element named small molecule and add ATP to the search criteria. If you “load into the editor” and then add to search criteria, PDBeMotif will ignore chirality and search for all variants of ATP instead.

The query page will look like that shown below.
Click on the “Search” button to carry out a search for all proteins that contain the Walker Motif and bind ATP.

Check a few entries as shown and from the bottom of the page choose the 3D alignment and viewing option. The best way to see these is by choosing “the hit” and viewing in either Jmol or AstexViewer.
As can be seen from the results above, all the entries that we chose have a similar fold and bind ATP in the same position. The entries themselves come from different UniProt sequences and have different 3D structures. Mouse over the list of PDB codes in the result list to see small thumbnail pictures of the proteins.

**Finding Specific Interactions in the PDB**

Often scientists are interested in finding specific interactions between atoms and residues. PDBBeMotif provides a powerful search functionality whereby the whole PDB archive can be analysed for a specific kind of interaction and the results downloaded or displayed graphically.

In order to do this kind of analysis, please start PDBBeMotif from [http://pdbe.org/motif](http://pdbe.org/motif). From the main page, choose the “Pair Bonds” link from the top of the page. This will open up a search interface into which the relevant information may be input.
We are going to search all interactions between HEM ligand with MET amino acid across the whole of the PDB. Fill in the boxes as shown above and click on “Build Statistics Chart”. The results will be available after the search has completed, usually within minutes.

Interactions between HEM and met are shown in the form of a bar chart with the different bars highlighting the kind of interaction between these two residues. The red bar on this chart corresponds to all covalent interactions between the two residues. Click on this red bar to be presented with a result list of entries where this interaction has been seen.

Choose a few of the hits and click on the 3D alignment option at the bottom of the page to see your chosen list of entries in a graphical browser.
The results clearly show that methionine only makes covalent contact with the Heme porphyrin ring from one direction and all the hits align with each other quite well.

You could also specify the specific atoms that interact with each other rather than give just the residue 3-Letter code. For example, to search for all interactions between CYS SG atom and DTT S1 atom, input your search with the following syntax: CYS.SG amino acid and DTT.S1 ligand and click on “Build Statistics Chart”. Further help on the naming syntax is available by clicking on the “…” icons just below the search input box.

**Conclusion**

As you can see from this short presentation PDBeMotif is a very powerful application that can yield valuable information present in the PDB archive. There is extensive online help available regarding the use of this tool. If you need any help regarding the use of this tool or further information, please get in touch with us via email to pdbehelp@ebi.ac.uk.