PDBe
Protein Data Bank in Europe
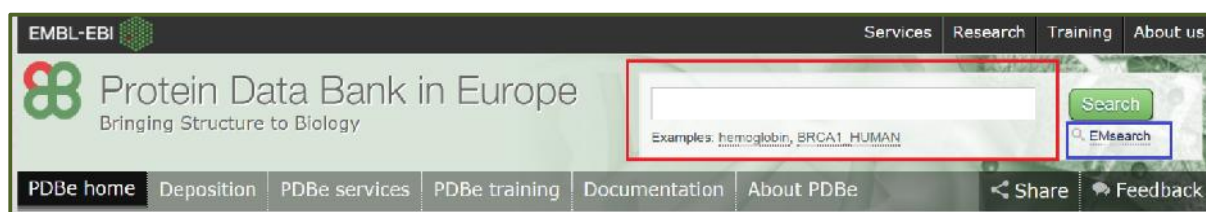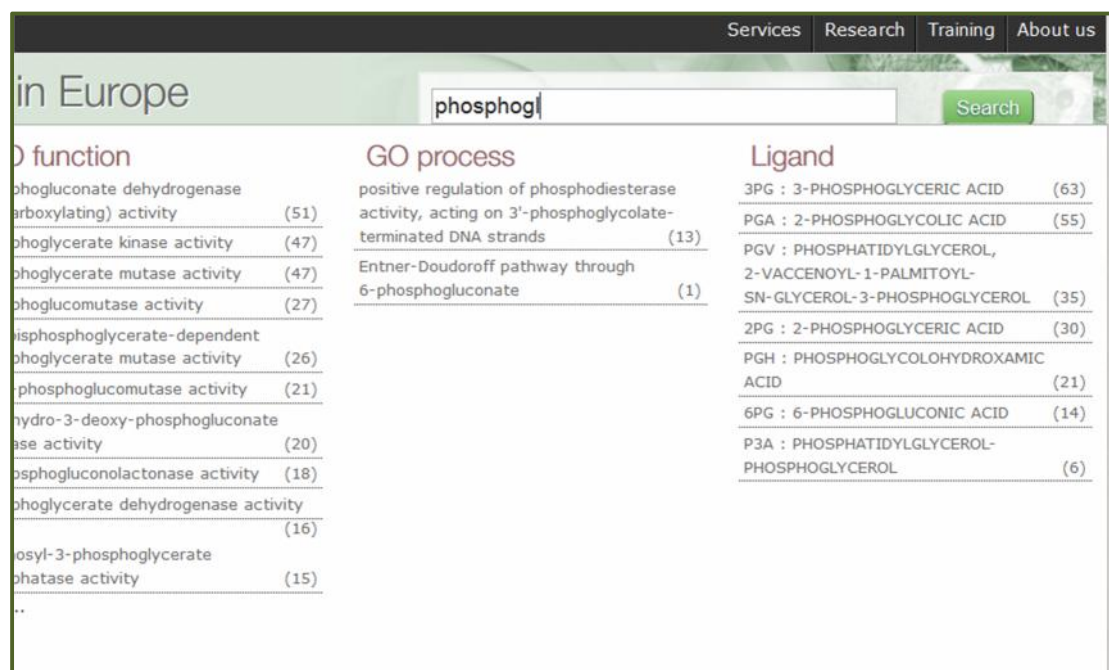
email us: pdbhelp@ebi.ac.uk

How to search the PDB

- The searchbox on the PDBe homepage allows you to switch between searching and downloading data from the PDB and the EMDB archive

- The search box on the PDBe website provides an autocomplete feature, which displays suggested search terms. These are categorised by context, e.g. author, molecule name, species etc.

- Search results are organised in a way that highlights different features of the search. They are sorted by unique macromolecules, type of macromolecules, interacting macromolecules, experimental methods, assembly composition, species names etc.

- The best representative structure from a particular protein family is always the top hit (PDB entry 3NML in this example). This is followed by the remaining entries from the same family in order of model and data quality.

Useful Tips for searching

1. The search /query can be saved for future use using the 'Save search' button.

2. You can choose to add multiple entries into your basket, which can be downloaded later without having to create any user login account.

3. It is useful to look at the results from different viewpoints, e.g.:

    a. I need to know all the entries returned by the query – 'Entries' tab;

    b. What unique macromolecules (i.e. protein/DNA/RNA) are found in these entries? – 'Macromolecules' tab;

    c. What sort of compounds are associated with the entries? – 'Compounds' tab;

    d. What protein families do these entries belong? – 'Protein families' tab.

Search Exercises

Last week Nobel prize in Chemistry 2015 was awarded jointly to Tomas Lindahl, Paul Modrich and Aziz Sancar "for mechanistic studies of DNA repair".
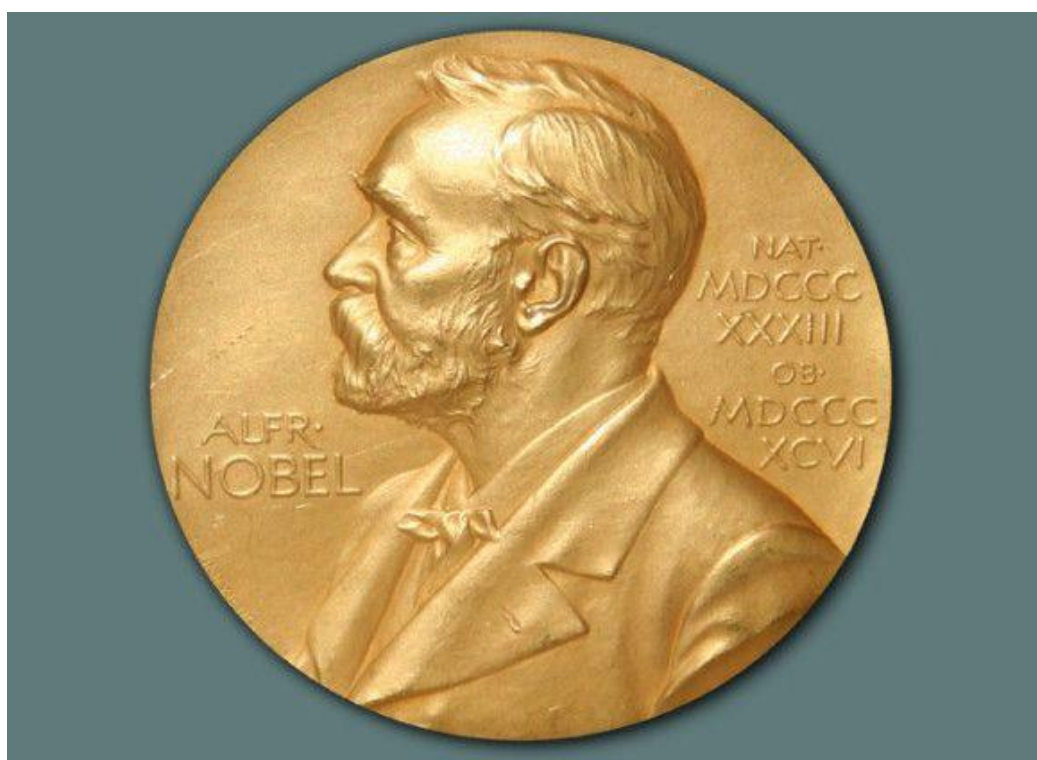
Their research has been focused on the three main areas described below:

a. Base excision repair - Base excision repairs DNA when a base of a nucleotide is damaged, for example cytosine

b. Nucleotide excision repairs DNA-injuries caused by UV radiation or carcinogenic substances like those found in cigarette smoke.

c. Mismatch repair - When DNA is copied during cell division, mismatching nucleotides are sometimes incorporated into the new strand. Out of a thousand such mistakes, mismatch repair fixes all but one.

Exercise 1:

Using the searchbox at the top of the PDBe homepage (www.pdbe.org) search for the following terms –

- Base excision repair (categorised under GO process)

  o What are the proteins that are involved in the base excision repair process?

  o How many of the proteins come from Human?

  o How many structures are there which represent a complex between DNA and protein?

  o Refine your query further by choosing for 'DNA/protein complex' from the left-hand panel of your results (check under Assembly composition).Click on the "protein families" view and find out how many structures in the PDB belong to "Protein family: AlkA_N". Which is the best representative structure in this protein family?

- Nucleotide excision repair, DNA incision (categorised under GO process)

  o What protein is responsible for the DNA incision? Is it an enzyme? (check if there are EC numbers)

  o What are the types of interacting molecules? (e.g. Protein/DNA/RNA)

- o Which is the best representative structure for this protein in the PDB archive? (click the Macromolecules tab).

- mutS gene

    - o What is the name of the DNA mismatch repair protein encoded by the mutS gene?

    - o What are the source organism of these protein structures that are present in the PDB?

    - o Name the Pfam families and the CATH and SCOP class for these proteins (check the left hand pane of the search result for the answer)

Exploring a PDB entry
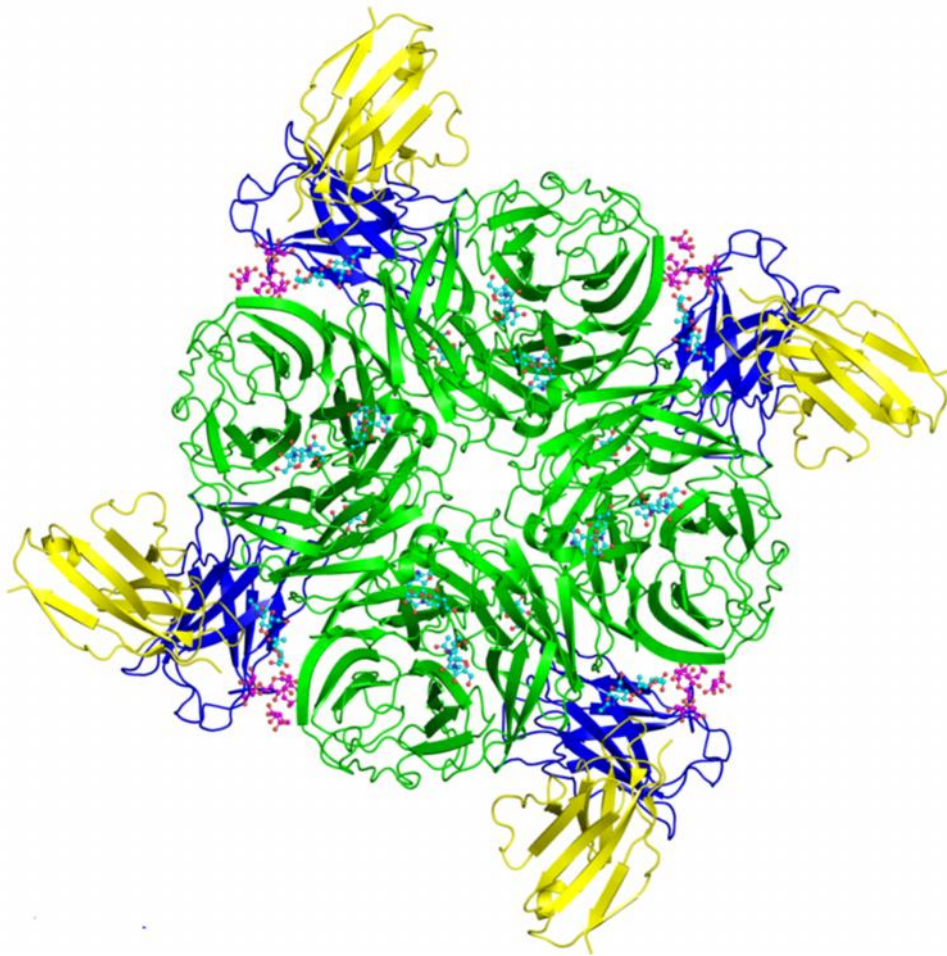
Things to remember

- A typical PDB entry contains 3D coordinates of the macromolecular structures that are experimentally derived using X-ray/NMR/electron microscopy.

- It also contains information about secondary structure, biological assemblies, sequence mapping information (UniProt for protein and GenBank for RNA) for the proteins and nucleic acids that are present in the entry.

- The PDB entry can be further explored for binding environment of ligands and other small molecules.

- The information from a PDB entry is structured into five major sections on the PDBe website to make the data easier to browse.

- Every PDB entry page comes with a portfolio of images - displaying the contents of the entry from various structural, functional and chemical perspectives.

- The citation page not only provides information about primary and associated publications related to the entry, it also displays figures and figure legends for open access publications.

- The 'Structure analysis' section enables you to interactively explore proteins in 1D/2D/3D layout.

- The 'Ligands and Environments' section lists all the chemically distinct ligands that are present in a particular PDB entry and shows their binding sites.

- There are several different download options and file formats present from a PDB entry page. In addtion to the PDB archive files, every entry page provides download options for biological assembly files (which can be instantly viewed using any visualisation tool like Chimera, pymol etc.), sequence files and SIFTS information.

Exercise 2:

Quick facts about influenza: The influenza virus particle is made up of the viral RNA genome wrapped in a lipid membrane. The membrane, or envelope, contains three different kinds of viral proteins. The hemagglutinin molecule (HA) attaches to cell receptors and initiates the process of virus entry into cells. Sialic acid is present on many cell surface proteins as well as on the viral glycoproteins; it is the cell receptor to which influenza virus attaches via the HA protein. The neuraminidase protein (NA) removes sialic acid from glycoproteins. An important function of the NA protein is to remove sialic acid from glycoproteins. Sialic acid is present on many cell surface proteins as well as on the viral glycoproteins; it is the cell receptor to which influenza virus attaches via the HA protein. The sialic acids on the HA and NA are removed as the proteins move to the cell surface through the secretory pathway. Newly released virus particles can still potentially aggregate by binding of an HA to sialic acid present on the cell surface. The essential nature of the NA for virus production has been exploited to develop new drugs designed to inhibit viral release.

- Search and explore the PDB entries 4b7j, 2c4a, 3ti3 and 3b7e containing the structure of the neuraminidase enzyme. (pdbe.org/4b7j, pdbe.org/2c4a, pdbe.org/3ti3, pdbe.org/3b7e)

- From the function and biology section of PDB entry 2c4a find out the biochemical reaction catalysed by neuraminidase. What is the EC number for the enzyme?

- What is the CATH domain classification for all the 4 entries?

- Explore the "Molecule details" section for PDB entry 4b7j. Find the location and identity of the amino residue that covalently binds to NAG A 521 (N-Acetyl D-Glucosamine) in the 1D/2D/3D alignment of the protein.(checkout the regions with the coloured dots on the 1D representation of chain A – look out for ASN residue lurking around). Does the residue NAG always bind to the same amino acid type in all the 4 entries?

- What is the UniProt coverage for PDB entry 4b7e? (go to the "molecular details" section)

- Find out other protein structures in PDB which share similar 3D fold with the entry 4b7e. (go to "molecular details" and then click for "similar 3D structures").

- What is the strain of the influenza virus used in PDB entry 3ti3?

- Which of the entries have open source citation information? (check pdbe.org/3ti3, pdbe.org/4b7j)

- What is the preferred assembly type for the neuraminidase enzyme in all the structures? (check the "Assembly composition" under structural details)

- Analyse the binding environment for ligands G39 (Oseltamivir), LNV (Laninamivir), SIA (Sialic acid), ZMR (Zanamivir). Which one is the natural substrate for neuraminidase enzyme? Compare the chemical structure of Sialic acid with the other three Ligands. How do they differ? Do you think the neuraminidase enzyme will be able to carry out its sialidase activity in cases where it is bound to Oseltamivir, Laninamivir, Zanamivir?

Validation Tutorial–

http://xray.bmc.uu.se/gerard/embo2001/modval/index.html

More Help

- PDBe: Searching the Protein Data Bank
  (http://www.ebi.ac.uk/training/online/course/pdbe-searching-protein-data-bank)

- PDBe: Exploring a Protein Data Bank (PDB) entry
  (http://www.ebi.ac.uk/training/online/course/pdbe-exploring-protein-data-bank-pdb-entry)

- PDBeFold: Search for a structurally homologous protein
  (http://www.ebi.ac.uk/training/online/course/pdbefold-searching-structural-homologues-protein)

- PDBePISA: Identifying and interpreting the likely biological assemblies of a protein structure (http://www.ebi.ac.uk/training/online/course/pdbepisa-identifying-and-interpreting-likely-biolo)

- PDBeChem: Searching for small molecules and small molecule fragments
  (http://www.ebi.ac.uk/training/online/course/pdbechem-searching-small-molecules-and-small-molec)