PDBeChem: Searching for small molecules and small molecule fragments

Gary Battle [1]

- Proteins
- Chemical biology
- Structures
- Beginner
- 1 hour

PDBeChem allows you to search for chemical components (ligands, small molecules and monomers) that appear in PDB entries, and discover which protein structures bind a particular ligand. This course will show you how to use PDBeChem and what you can do with it.

An undergraduate degree in a life science subject would be an advantage. You may wish to look at the PDBe quick tour [2] and Biomacromolecular structures [3] course as an introduction to this tutorial.

Learning objectives:

- Evaluate what PDBeChem is and what is can be used for
- Be able to launch PDBeChem and use different ways to search for small molecules
- Use the SMILES molecular editor to draw parts of your chemical structure to explore related small molecules on PDB
- Evaluate your search results from the chemical component summary page

What is PDBeChem?

PDBe [4] provides a range of services for exploring the binding of small molecules [5] such as drugs, co-factors, co-enzymes, prosthetic groups and biomacromolecules [6].
One such service is PDBeChem [7], which stores information regarding all chemical species (small molecules, ligands, metal ions, amino acids and nucleotides) found in the Protein Data Bank (PDB [8]) archive.

Small molecules are dealt with in the PDB archive as formally equivalent to macromolecular residues. Each one is given a unique three-letter code but these are often arbitrary and so you can’t necessarily guess what the compound might be from the code.

This mini-tutorial uses PDBeChem to explore compounds in the PDB archive that are chemically similar to the benzothiophene [9] derivatives bound to Christmas Factor [10] in PDB entries 3lc3 and 3lc5. You can find out more about Christmas Factor in the accompanying Quips article [11].

The search methods it shows should be of general use to you as you explore small molecules in PDB structures.

Starting up PDBeChem from the PDBe homepage

To start searching small molecules using PDBeChem [7], click on the ‘PDB compounds link’ under the services menu on the PDBe homepage [12] (red box in Figure 1).

![Figure 1 The PDB compounds link on the PDBe homepage.](image-url)
Searching for a chemical component

The chemical components search page allows you to search using five different parameters (Figure 2).

The first option on the search page is useful if you actually know the three letter code of the small molecule you’re interested in, and the second if you know the molecule name. In both cases, changing the pull-down menu to the left of the input box from ‘=’ to ‘like’ means the search results will find hits containing the term you’ve input, rather than matching it exactly. So entering ‘MSE’ for example takes you straight to the chemical component page for selenomethionine.

![Figure 2 The chemical components search page.]

The third option allows you to input the molecular formula into the search box, so searching for ‘C6 H12 O6’ will find all the hexose sugars. This search box defaults to formula range but you can also explicitly define a range. Inputting ‘C6-7 H12 O6’, for example, will find all those molecules containing six or seven carbons, twelve hydrogens and six oxygens.

Searching using the SMILES molecular editor

Starting up the molecular editor SMILES

For our purpose in this tutorial, it is the next search option ‘non-stereo SMILES’ which is the most useful. In this search box you can input the SMILES [13] string of the molecule or molecule fragment of interest. SMILES stands for Simplified Molecular-Input Line-Entry Specification [14].
Figure 3 The molecular editor for chemical substructure search.

Try it for yourself....

More than likely, you won’t know the SMILES [13] string of your small molecule, so instead click the ‘edit’ button (arrowed in the top of Figure 3) next to the ‘non-stereo SMILES’ input box.

This will open up another window with the JME molecular editor tool [15] (bottom of Figure 3). You’ll find a link with tips on how to use the editor just below the sketch box.

Using the SMILES molecular data editor (!)

As an example we are going to look at the PDB [8] entry 3lc3, which contains the inhibitor 1-[5-(3,4-dimethoxyphenyl)-1-benzothiophen-2-yl] methanediamine (shown graphically in Figure 4 below).
Figure 4 The chemical structure of the inhibitor 1-[5-(3,4-dimethoxyphenyl)-1-benzothiophen-2-yl] methanediamine.

Let's sketch just the double-ringed benzothiophene [9] moiety in the editor.

Using the SMILES molecular data editor (II)

Try it for yourself....

1. Firstly, select the benzene ring icon (number 1 in Figure 5 below) and drag it into the editor.
2. Select the five membered ring (2) and while hovering over a bond in the benzene ring, click the mouse to fuse this second ring to the benzene ring.
3. Now select ‘S’ from the left-hand side menu (3) and click the mouse over the right place in the 5-membered ring to swap the carbon for a sulphur.
4. Then select the the double bond icon (4) and add a double bond into the five membered ring to form benzothiophene [9].
5. Check the ‘discard bond order’ box (highlighted in red) to relax the search criteria.
6. Now hit ‘submit’ (arrowed) to take you back to the search page. The SMILES [13] string has automatically been filled in for the structure you have drawn.
There are other useful options in the editor tool. For example you can upload a molecule to alter in the editor, which might make it more easy if you want to search for a complex fragment. You can also decide whether you want to search for fragments as part of a ring or optionally to disregard bond order.

Using the SMILES molecular data editor (III)


You will see a results page similar to that below (Figure 6). For our search, you can see that there are almost 100 molecules containing benzothiophene, listed with the three letter code, the molecule’s name, and an image of the structure. Hovering over the image will give you a larger view.
Figure 6 The PDBBeChem search results. Our chemical component of interest 1-benzothiophene-2-sulfonamide (highlighted in red) is at the top of the list.

Try it for yourself....

Click on the three letter code 03T (arrowed) to bring up more information about 1-benzothiophene-2-sulfonamide (Figure 7).

**Chemical Component Summary page**

The Chemical Component Summary page contains the names and synonyms of the molecule as well as several different chemical descriptors, formula and molecular weight (Figure 7). The links to the left allow you to explore the atom names and bonds in the molecule, and to see a list of PDB entries which contain this molecule.
Chemical Component Summary page for 03T.

The 'Download Links' section is where you can download the coordinates of the molecule and (its dictionary) should you want to use this molecule in your own structure determination.

**Summary**

- The PDBeChem database stores information regarding all chemical species found in the PDB archive.
- There are several ways to search for your chemical interest, including by molecule name, formula or using the SMILES molecular editor.
- The chemical component summary page provides you with useful information about each molecule from your search results. This includes: names and synonyms, chemical descriptors and molecular weight.

**Your feedback**

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

**Learn more**

**Find out more**

You can read the accompanying Quips article to this tutorial on the [PDBe website](https://www.ebi.ac.uk/training/online).
Recommended courses

Train online courses available:

- **Biomacromolecular structures** [3]: An introduction to EBI resources;
- **PDBePISA** [16] tutorial - explore macromolecular interfaces and predict the quaternary structure of your protein;
- **PDBeFold** [17] tutorial - search, analyse and examine the characteristics or protein-ligand interactions, binding environments and structural motifs.

Get help and support on PDBeChem

EMBL-EBI's **PDBe team** [12] develops and maintains the EBI's **Protein Data Bank in Europe** [18].

Support

- A list of frequently asked questions about PDBe can be found on the **PDBe website** [19].
- For general enquiries about the **PDB** [8], contact the pdbhelp [at] ebi.ac.uk (PDB help desk)

Collaborators

PDBe collaborates with the **x-ray crystallography** [20], nuclear magnetic resonance (NMR) spectroscopy and cryo-**electron microscopy** [21] (EM) communities. To keep abreast of new developments in the NMR community, PDBe has participated in EU projects and continues to contribute to the **Collaborative Computational Project for the NMR community** [22] (CCPN). PDBe also operates **EMDB** [23], the international repository for density maps, which are created using high-resolution biological transmission electron microscopy in collaboration with **RCSB** [24] and **Baylor College of Medicine** [25]. EMDB contains both macromolecular images and structures reconstructed using the single-particle method and images of sub-cellular regions from **electron tomography** [26].

Funding

Core support for PDBe is provided by the **European Molecular Biology Laboratory** [27] and **The Wellcome Trust** [28]. Specific projects are funded by the **European Union** [29], the **UK Biotechnology and Biosciences Research Council** [30] (BBSRC) and the **National Institutes of Health** [31] (NIH).

Contributors
Gary Battle [1]

EMBL-EBI
Project Leader Outreach - Kleywegt team: Protein Data Bank in Europe (EMBL-EBI alumni)

Gary was the outreach coordinator for the Protein Data Bank in Europe (PDBe). He was responsible for helping users with a broad range of backgrounds and interests to make the most of macromolecular structural data. Gary has a PhD in synthetic organic chemistry from the University of Warwick. Before joining the EMBL-EBI, he worked for over 10 years at the Cambridge Crystallographic Data Centre where he gained a wealth of experience in supporting software tools and data resources for pharmaceutical discovery, life science research and materials design.

Abhik Mukhopadhyay [32]

EMBL-EBI
Scientific Database Curator - Kleywegt team: Protein Data Bank in Europe

Source URL: https://www.ebi.ac.uk/training/online/course/pdbechem-searching-small-molecules-and-small-molec
[13] https://www.ebi.ac.uk/training/online/glossary/smiles
[15] https://www.ebi.ac.uk/training/online/glossary/jme-molecular-editor-tool
[16] https://www.ebi.ac.uk/training/online/course/pdbe-identifying-and-interpreting-likely-biological
[17] https://www.ebi.ac.uk/training/online/course/pdbe-searching-structural-homologues-protein-using
[18] https://www.ebi.ac.uk/training/online/glossary/term/533
[19] http://www.ebi.ac.uk/pdbe/#m=4&amp;h=0&amp;e=0&amp;r=0&amp;l=2&amp;a=0&amp;w=0
[20] https://www.ebi.ac.uk/training/online/glossary/x-ray-crystallography
[21] https://www.ebi.ac.uk/training/online/glossary/electron-microscopy
[22] http://www.ccpn.ac.uk/
[23] https://www.ebi.ac.uk/training/online/EMDB
[25] https://www.bcm.edu/
[26] https://www.ebi.ac.uk/training/online/glossary/electron-tomography
[27] http://www.embl.org/
[28] http://www.wellcome.ac.uk/
[32] https://www.ebi.ac.uk/training/online/trainers/abhik