PDBePISA: Identifying and interpreting the likely biological assemblies of a protein structure

Gary Battle [1]

- Proteins
- Structures
- Beginner
- 1 hour

PDBePISA allows you to explore macromolecular (protein, DNA/RNA and ligand) interfaces and predict the quaternary structure of your protein. This course will show you how to use PDBePISA and what you can do with it.

An undergraduate degree in a life science subject and some background knowledge in protein structure 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 PDBePISA is and what it can do
- Be able to launch PDBePISA and search for interfaces related to your structure of interest
- Analyse and interpret probable assemblies of your structure of interest

What is PDBePISA?

PDBePISA [4] is a sophisticated service which analyses the interfaces between macromolecules in their crystal environment allowing you to predict the quaternary structure of your protein.

PDBePISA does a calculation on the likely strength of all contacts between macromolecules and will cut down these assemblies to the most likely biologically relevant assembly. If crystal symmetry operations are needed to build up biological assemblies then PDBePISA will do that for you.

When to use PDBePISA

PDBePISA can be used to answer the following questions:
What quaternary structure can my crystal structure have?
What are the crystal contacts and interfaces in my structure?
What is the size of the crystal interfaces?
What are the energetics that keep my quaternary structure together?
Are there any other structures in the PDB [5] that have similar interfaces?

This is not a molecular docking prediction; PDBePISA uses only interactions and macromolecular contacts that are already present in the crystal structure. Unfortunately this means it is not applicable to structures determined by non-crystallographic methods.

**Where does the data come from?**

Each X-ray structure is solved in a particular crystal symmetry and the PDB [5] requires authors to deposit only the smallest or asymmetric unit from any crystal form. As a result, the PDB entry in the archive often does not contain the assembly that is most likely to exist under physiological conditions.

In many cases, crystal symmetry operations need to be applied in order to obtain the biologically active form of a macromolecule or complex. Conversely, some macromolecules crystallise with several copies of the biologically relevant assembly in the asymmetric unit. This means that the deposited entry contains extra copies of the biological assembly. Contacts in the crystal hold these copies alongside each other but are too weak to survive in solution.

**PDBePISA for analysing the NGF structure 1bet**

The nerve growth factor (NGF) structure 1bet is a candidate for PDBePISA analysis since (as described in the accompanying Quips article [6]) this crystal structure has only a single chain deposited in the PDB [5] entry but is active as a dimer [7].

This mini-tutorial shows you how to identify and access this assembly in the PDBePISA analysis of this entry.

**Starting the PDBePISA service**

Starting the PDBePISA service (i)

From the PDBe [8] homepage, you can enter an entry identifier [9] such as 1bet directly in the quick access box (circled in red in Figure 1) and then press the 'Quaternary structure' button (arrowed in red). This takes you straight to a ready completed PDBePISA analysis form for the entry shown as Figure 3.
PDBePISA: Identifying and interpreting the likely biological assemblies of a protein structure
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 1 Accessing PDBePISA from the PDBe homepage.

Try it for yourself....

Alternatively, you can get a blank PDBePISA submission form by clicking the 'Quaternary structure' link in the Services section on the left-hand side of the PDBe [8] homepage (red box marked with an arrow in Figure 1).

Starting the PDBePISA service (ii)

Try it for yourself....

Next, click on the 'Launch PDBePisa' button (arrowed in Figure 2) and fill in the PDB [5] ID code for the structure you are interested in - 1bet in this case.

The submission form is shown in Figure 3 on the next carousel page.
PDBePISA: Identifying and interpreting the likely biological assemblies of a protein structure
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Starting the PDBePISA service (iii)

As this is an existing entry, you will get immediate access to pre-calculated results - summarised as '1 amino acid chain in ASU' and 'most probable assembly: 2mer’ (highlighted with a blue bracket in Figure 3).

If you want a PDBePISA analysis for your own crystal structure too, you can choose the option 'Coordinate file' on this submission page and PDBePISA will process this file.

Figure 2 Interfaces and Assemblies page.

Figure 3 PDBePISA submission form. '1bet' entered in the entry field (circled in red) produces an immediate summary of the PDBePISA analysis (blue bracket).
Try it for yourself....

Once you have typed '1bet' into the entry field, click the interfaces button (mouse arrow) to display the Interfaces results (Figure 4).

**Interfaces formed by symmetry in PDB entry 1bet**

As you can see, PDBePISA has listed five crystal interfaces for 1bet in order of calculated stability (Figure 4). The table of results gives details of each contact surface from a particular Structure 1 (red box) on one side of each interface to a Structure 2 (green box) on the other side.

Each Structure 2 has been produced by applying a specific crystal symmetry operation to Structure 1. The symmetry is indicated in the Structure 2 section - for example the top (most stable) assembly is generated by mapping coordinates (x,y,z) to (-y+1,-x+1,-z+1/6) and in addition a translation \([1 1]^{T}\) to position the Structure 2 in an adjacent unit cell.

All PDBePISA pages give extensive help for the results so you can click on table headings to get a summary explanation.

![PDBePISA Interface List](image)

**Figure 4** PDBePISA listing of 1bet interfaces. Structure 1 section is boxed in red and Structure 2 is boxed in green. The symmetry code producing Structure 2 is circled in green. Details of the most stable interface are underlined in yellow.

You will now appreciate that it is good to have PDBePISA taking care of these details - who could have anticipated both rotation and a unit cell shift would produce the NGF biological assembly?!

This assembly is present multiple times in the crystal unit cell - every single copy of NGF exists with a dimeric partner - PDBePISA just presents this for the one that the authors have chosen to deposit. The unit cell translation is hidden in the Sym.ID \([12]\) four number code 10,665 (circled in green in Figure 4). Here the first number relates to the symmetry operations possible for the crystal while the second set of numbers relate to translations.

Try it for yourself....
Click on ‘Assemblies’ (mouse arrow in Figure 4) to see the dimer [7] produced by this interface.

1_555, a special code for doing nothing to a structure

One of the most common symmetry operations is not to do anything and take the macromolecule ‘as is’: no rotation and no translation [11] - this operation has the special code 1_555 so watch out for that when you check PDBePISA output. Code 1_555 next to a structure means that this component is already present in deposited entry. This is not the case for any significant contacts in 1bet - as the entry only has a monomer [13] there can be no contacts with unrotated/untranslated molecules.

The interface analysis (underlined in yellow, Figure 4) for 10_665 Structure 2 interface with Structure 1 shows that it has a very sizable area (1369.9 Å² - anything approaching 1000 Å² represents a good interface between macromolecules) and that it has a favourable calculated energy of -21.7 kcal/mol.

Finally, PDBePISA prints [14] out the contribution of this interface to any predicted biological assemblies - the CSS [15] value, which is calculated as a fractional contribution to the assemblies in the crystal. A CSS value of 1.0 means that this interface represents 100% of the interactions for the predicted assembly. This is the NGF dimer [7] we are interested in.

You might like to take the time looking through the other entries in the PDBePISA table of found interfaces (Figure 4). The calculated interface areas and energies should convince you that these are much weaker interactions. These contacts hold the crystal together but are not likely to persist in solution.

Downloading and analysing the NGF dimer

Downloading and analysing the NGF dimer (i)

Clicking on the arrowed link at the top of the screen (Figure 4 [16] - the PDBePISA ‘assemblies’ button) confirms that the 1bet multimeric [17] state is a dimer [7] A2 (circled in orange, Figure 5). You can now download PDBePISA’s version of the dimeric assembly by clicking on the ‘download’ link (Figure 5). Helpfully, PDBePISA will have relabelled the second NGF molecule in the dimer as chain B so you can use the downloaded file for making pictures with the halves of the dimer coloured distinctly in Rasmol [18], Pymol [19], Jmol [20], or CCP4mg [21] for example.
Figure 5 PDBePISA probable assemblies for 1bet. A dimer is the only likely assembly (circled in orange).

Try it for yourself....

- Clicking the 'view' button (Figure 5) launches a Jmol [20] interactive graphics window showing the NGF dimer [7] with the contact surface coloured up in red and green atoms.
- Click on the 'details' link (arrowed in Figure 5) to examine the NGF dimer we are interested in. You can also download these details by clicking on the download link.

**Downloading and analysing the NGF dimer (ii)**

The first assembly summary page (Figure 6) repeats details of the interfaces and components derived from the PDBePISA analysis. A more detailed analysis of the interface in the dimeric assembly is available by clicking the 'details' button (arrowed in Figure 6).
Figure 6 PDBePISA summary of the dimeric assembly.

Try it for yourself....

Click on details of selected interface for a residue-by-residue and bond-by-bond listing (arrowed in Figure 6).

**Downloading and analysing the NGF dimer (iii)**

Unlike the interfaces analysis shown earlier, which included all crystal interfaces (Figure 4), the detailed analysis here (Figure 7) relates only to contacts and bonds in the predicted biological assembly. Scroll down to examine the residue-by-residue analysis shown in the Figure 7 snippet. All residues involved in non-covalent bonds across the interface are listed.

Unlike the NGF monomer [13], there are no disulfide bonds involved in stabilising the NGF dimer [7] (circled in orange in Figure 7). So, what gives the NGF dimer its very high stability in solution? One important contribution is provided by the hydrogen bonds listed here. But, in addition, by checking through the residue-by-residue output, you can start to see that hydrophobic sidechains such as those of tryptophan (e.g Trp 21, circled in blue in Figure 7) and phenylalanine (Phe) make a large contribution to the interface of the NGF dimer.
Figure 7 Extracts from the details of 1bet probable assembly interface. List of interactions shows H-bonds across the interface and reveals no disulfide bonds (circled in orange). The tryptophan side chain (circled in blue) is a key contributor to the NGF dimer's

In the NGF dimer [7], the same component is on each side of the interface, and so each Structure 1 residue in the interface also contributes to the Structure 2 side.

Other important information to notice from the table at the bottom of Figure 7:

- All residues making contacts in the interface are shown with a yellow background.
- Residues making hydrogen, or optionally disulfide or salt bridges [22], are coloured brown.
- Each residue's contribution to the buried surface area, from which solvent is excluded, is calculated and shown graphically with bars.

Close-fitting hydrophobic sidechains such as Trp and Phe contribute to macromolecular interfaces by excluding solvent. It is the entropic gain of the excluded solvent moving into free solution that produces a 'hydrophobic interaction' stabilising the interface.

As you browse assemblies in the PDB [5] archive you will frequently uncover heterodimeric [23] interfaces between different molecules. PDBePISA will help you view the fine detail in any crystal structure that you need to examine for interfaces and significant assemblies.
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.

Summary

- PDBePISA allows you to explore macromolecular interfaces and predict the quaternary structure [24] of your protein.
- PDBePISA does a calculation on the likely strength of all contacts between macromolecules and will show you the most likely biologically relevant assembly.
- You can use PDBePISA to view the fine detail in any crystal structure that you need to examine for interfaces and significant assemblies.
- For interface analysis, PDBePISA gives you a complex significance score (CSS) which indicates how significant for assembly formation the interface is.

Learn more

Find out more

You can read the accompanying Quips [25] article to this tutorial on the PDBe website [6].

Recommended courses

Train online courses available:

- Biomacromolecular structures [3]: An introduction to EBI resources;
- PDBeFold [26] tutorial - identify structural homologues in the PDB [5];
- PDBeChem [27] tutorial - search for chemical components that appear in PDB entries, and discover which protein structures bind a particular ligand [28].

Get help and support on PDBePISA

The EBI's PDBe team [29] develops and maintains the EBI's Protein Data Bank in Europe [30].

Support

- A list of frequently asked questions can be found on the PDBePISA website [31].
- For general enquiries about the PDB [5], contact the pdbhelp [at] ebi.ac.uk (PDB helpdesk)
Collaborators

PDBe collaborates with the X-ray crystallography [32], Nuclear Magnetic Resonance [33] (NMR) spectroscopy and cryo-Electron Microscopy [34] (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 [35] (CCPN). PDBe also operates EMDB [36], the international repository for density maps, which are created using high-resolution biological transmission electron microscopy in collaboration with RCSB [37] and Baylor College of Medicine [38]. EMDB contains both macromolecular images and structures reconstructed using the single-particle method and images of sub-cellular regions from electron tomography [39].

Funding

Core support for PDBe is provided by the European Molecular Biology Laboratory [40] and The Wellcome Trust [41]. Specific projects are funded by the European Union [42], the UK Biotechnology and Biosciences Research Council [43] (BBSRC) and the National Institutes of Health [44] (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 is 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 [45]