



Contribution ID: 6

Type: Poster

## A new approach for assigning names and persistent, unique identifiers to macromolecular complexes in the Protein Data Bank

Macromolecular complexes are key functional components in nearly all biological processes, and achieving an atomic-level understanding of them is crucial for unravelling their molecular mechanisms. The Protein Data Bank (PDB) is the primary worldwide repository of experimentally determined macromolecular structures, offering valuable insights into the dynamics, conformations, and functional states of biological assemblies. However, the current annotation practices within the PDB lack consistent naming conventions for these assemblies and do not assign persistent, unique identifiers to them, making it challenging in downstream data analysis to identify all the PDB entries that represent the same assembly.

We propose a novel approach that uses external data resources, including the Complex Portal, UniProt, and Gene Ontology, to precisely describe macromolecular assemblies. Using our approach, we successfully assigned standardised names to more than 90% of unique assemblies in the PDB and provided each assembly with persistent, unique identifiers. To further enrich our understanding of these assemblies, we also computed symmetry data for each using the AnAnaS program. This annotation is valuable as symmetry plays a key role in the structural and functional characterisation of macromolecular complexes, offering insights into their biological roles and interactions.

In addition to making the assigned identifiers and assembly names accessible via our FTP area, we also integrated this information to enhance the user experience of the PDB search system, making it easy to find all the PDB entries for a specific assembly composition.

The naming and standardisation of assembly data enhances the PDB, makes biological assemblies easier to discover, and facilitates a deeper understanding of macromolecular mechanisms.

### Submitting to:

8th CAPRI assessment meeting

**Primary authors:** APPASAMY, Sri (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr BERRISFORD, John (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))

**Co-authors:** GABOROVA, Romana (CEITEC – Central European Institute of Technology); Mr NAIR, Sreenath (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Mr ANYANGO, Stephen (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr GRUDININ, Sergei (Univ. Grenoble Alpes, CNRS); Mr DESHPANDE, Mandar (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr ARMSTRONG, David (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); PIDRUCHNA, Ivanna (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Mr ELLAWAY, Joseph (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr LEINES, Grisell (Protein Data Bank in Europe, European Molecular Biology Laboratory, European

Bioinformatics Institute (EMBL-EBI)); Dr GUPTA, Deepti (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr HARRUS, Deborah (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr VARADI, Mihaly (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr VELANKAR, Sameer (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))

**Presenter:** APPASAMY, Sri (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))