

**MS01-2-1 PDBe-KB: collaboratively defining the biological context of structural data**  
**#MS01-2-1**

**D. Armstrong**<sup>1</sup>

<sup>1</sup>PDBe (EMBL-EBI) - Cambridge (United Kingdom)

**Abstract**

The Protein Data Bank in Europe – Knowledge Base (PDBe-KB, <https://pdbe-kb.org>) is a collaborative resource between world-leading specialist data resources striving towards a two-fold goal: (i) to increase the visibility and reduce the fragmentation of annotations contributed by specialist data resources, and to make these data more findable, accessible, interoperable and reusable (FAIR) and (ii) to place macromolecular structure data in their biological context, thus facilitating their use by the broader scientific community in fundamental and applied research. Since we described PDBe-KB in 2019, there have been significant improvements in the variety of available annotation data sets and user functionality. Here we highlight all these additional annotations and new features such as a bulk download data service and a novel superposition service that generates clusters of superposed protein chains weekly for the whole PDB archive.

**Scientific Justification**

To date, the PDB contains over 190,000 structures of 57,000 distinct proteins. However, structures by themselves are not necessarily of intrinsic biological value, unless they can be related to additional functional information. Many data resources can derive additional annotations related to structures. While most of these annotations are openly accessible, they may not be easily findable, and the lack of standard data formats often hinders interoperability and reusability. To make these annotations FAIR (i.e. findable, accessible, interoperable, reusable), we established PDBe-KB in 2018 through a global collaboration between PDBe and leading specialist data resources. Currently, PDBe-KB integrates data from 30 partner resources providing over 1.2 billion residue-level annotations. These annotations are made available on PDBe-KB web-pages, programmatically via an API, and distributed as a Neo4J graph database.

The PDBe-KB web pages- Aggregated Views of Proteins recently reached over a million unique users since its first release. These aggregated views offer a “one-stop-shop” to users providing a comprehensive view of the functional context of the protein structure. We are actively updating these pages, and a few recent features include (i) a superposition service to visualise protein chains clustered by structural similarity; (ii) annotations for small molecules and macromolecular interaction partners. (iii) a section highlighting all the mature and processed proteins for a polypeptide; and (iv) a bulk download service that provides easy access to all the coordinate files, validation reports, sequences for a given protein. ECCB is a great platform to showcase all these new features to the scientific community. It provides a good opportunity to enhance further collaborations and collate valuable feedback which will help us improve our PDBe-KB service offerings.

**References**

PDBe-KB consortium, PDBe-KB: collaboratively defining the biological context of structural data *Nucleic Acids Research, Database Issue* (2022)

PDBe-KB consortium, PDBe-KB: a community-driven resource for structural and functional annotations. *Nucleic Acids Research, Database Issue* (2020)

Overview of PDBe-KB

**Data enrichment and integration in PDBe-KB**

