

PDBe-KB: a community-driven resource for structural and functional annotations

M.Varadi, PDBe-KB consortium

European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK

mvaradi@ebi.ac.uk

The Protein Data Bank in Europe - Knowledge Base (PDBe-KB, <https://pdbe-kb.org>) is a community-driven, collaborative data resource that provides literature-derived, curated and predicted structural and functional annotations of molecular structure data [1]. Consortium members provide annotations including catalytic sites, ligand binding sites, protein flexibility, post-translational modification sites, and the effect of genetic variability or mutations (Fig 1.).

PDBe-KB aims to increase the visibility and reduce the fragmentation of these annotations and place macromolecular structure data in their biological context, thus facilitating their use by the broader scientific community in fundamental and applied research. PDBe-KB currently collaborates with 31 resources from 11 countries, and we integrate their annotations with core PDB structural data in a novel and distributable PDBe graph database. Researchers can access all the annotations either by using the graph database or programmatically via API endpoints [2]. We have also created web pages called “Aggregated Views” that provide an overview of all the structure data related to a full-length protein (i.e. UniProtKB accession). These views are better at displaying the biological context of proteins instead of the conventional PDBe entry-page focus on a single PDB entry.

We have been continuously improving and expanding PDBe-KB since its inception in 2018. Last year we rolled out a significant update to support the global effort to tackle the COVID-19 pandemic by creating dedicated pages gathering all the structural information for the new coronavirus SARS-CoV-2 (e.g. <https://www.ebi.ac.uk/pdbe/pdbe-kb/protein/P0DTC2>). We have also added new functionalities, such as viewing the superposition of every PDB chain that corresponds with a protein of interest. These views are instrumental in identifying ligand-binding hotspots and conformationally flexible regions. Most recently, we are linking to and display data from predicted model providers such as AlphaFold DB and SWISS-MODEL.

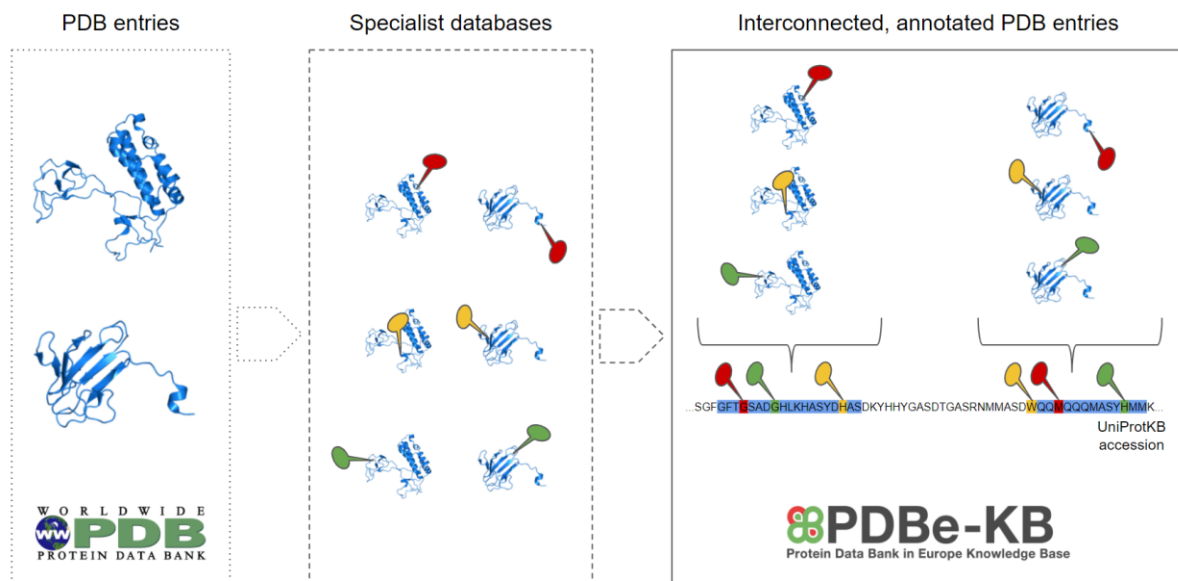


Figure 1 – Schematic overview of data integration in PDBe-KB

[1] Varadi, M., et al. (2019). Nucl. Acids Res. 48, D344–D353.

[2] Nair, S. & Varadi, M. et al (2021). Bioinformatics.

Keywords: protein structures, functional annotations, knowledge base