**Introduction**

The Protein Data Bank (PDB) is a free accessible global data bank that houses macromolecular structures that are experimentally determined (Velankar *et al.,* 2021) or from computer structure models (CSM). PDB was established 52 years ago and has grown to such international relevance that it is managed by “four partners - RCSB Protein Data Bank, the Protein Data Bank Japan, the Protein Data Bank in Europe, and BioMagResBank” (Armstrong *et al.,* 2020; Burley *et al.,* 2019; Kinjo *et al.,* 2018, Ulrich *et al.,* 2008).

The PDB archive is a resource that allows researchers (“depositors”) to deposit their research even without being published in peer-reviewed journals. PDB has found its relevance in different areas such as structural biology, medical and biomedical research, physics, enzymology, mathematics, art, and education (Velankar *et al.,* 2021).

A screenshot of a website

Description automatically generated

A

B

A screenshot of a computer

Description automatically generated

**Figure 1:** *PDB home page. A. The Research Collaboratory for Structural Bioinformatics (RCSB PDB) home page is the United States operated version of PDB. B. WWPDB home page operated by all the consortiums. The wwPDB consortium in conjunction with structural biology and other non-structural biology users develops policies and processes for data deposition, biocuration, validation, and data distribution (Velankar et al., 2021).*

The PDB contains 3D structures of biomolecules that are experimentally determined using Nuclear Magnetic Resonance (NMR), X-ray diffraction, and cryo-EM. As of the time of writing this article, 210,836 experimentally determined structures and 1,068,577 Computer Structure Models (CMS) were deposited on the RCSBPDB. The RCSB PDB has a tutorial – PDB-101 that introduces new users to the basics of this publicly available data bank.

The PDB has metamorphosed from the 1969 deposited structure of myoglobin with a resolution of 2 Armstrong (Figure 2A) to a rsKiiro pump with a resolution of 1.5 A. The advancement in PDB structure is due majorly to advancements in technological instruments used for determining structure and the growth in the use and management of the data sector has also improved the user experience on PDB.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Objective**

1. Develop an R-based pipeline for batch extraction of resolution values from PDB files.
2. Conduct exploratory data analysis to identify trends and patterns in resolution values.
3. Assess the impact of resolution on structural quality and its relevance to various research areas.

**Methodology**

a. Data Extraction: Utilize the biocLite and rpdb packages to access the PDB database, download PDB files, and extract resolution values.

b. Data Preprocessing: Clean and standardize data to ensure uniformity and quality.

c. Data Analysis: Employ various R packages for statistical analysis, visualization, and hypothesis testing.

**References**

Armstrong DR, Berrisford JM, Conroy MJ et al (2020) PDBe: improved findability of macromolecular structure data in the PDB. Nucleic Acids Res 48:D335–D343

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000). The Protein Data Bank, Nucleic Acids Research, Volume 28, Issue 1, Pages 235-242.

Burley SK, Berman HM, Bhikadiya C et al (2019) RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47:D464–D474

Kinjo AR, Bekker GJ, Wako H et al (2018) New tools and functions in data-out activities at protein data Bank Japan (PDBj). Protein Sci 27:95–102

Ulrich EL, Akutsu H, Doreleijers JF et al (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408

Velankar, S., Burley, S.K., Kurisu, G., Hoch, J.C., Markley, J.L. (2021). The Protein Data Bank Archive. In: Owens, R.J. (eds) Structural Proteomics. Methods in Molecular Biology, vol 2305. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1406-8\_1