

## Probing details, limitations and uncertainties in structural information of protein databases

Lessandra Eller<sup>1</sup>; Luiz F. O. Rocha<sup>1\*</sup>

<sup>1</sup>Faculty of Pharmaceutical Sciences at Ribeirão Preto, University of São Paulo.

\* luiz@fcfrp.usp.br

**Introduction:** Nowadays, we have known more and more about atomic details, molecular compound characteristics, and cellular infrastructures and communications in living systems. In this context, it has been fundamental the cumulative and thorough knowledge on polymeric molecules, whose abundant structural information are publically available in the database of the Protein Data Bank, PDB, with primary data and in other related databases that categorize the data distinctively. **Objective:** However, as shown here, it can be a challenge to get and use these data, since they contain many subtle features, intricate details, crucial limitations and intrinsic uncertainties. **Methods:** In order to suitably reveal such cases and situations, we make use of close and detailed inspections in files and outputs from the PDB library and in derived databases. **Results:** In this work, we are presenting the main dares and bottlenecks for the suitable exploration and utilization of the biomacromolecules data and information, in addition to their presumable causes. These include, missing parts of molecules, structures with only alpha-carbon coordinates, archives having different formats, and many PDB entries for a same protein. **Conclusion:** The precise understanding and explanation of such circumstances are essentials in specific explorations and subsequent employments by theoretical, experimental and computational approaches, and in practical applications.

Keywords: protein data bank, structural analyses, atomic coordinates, biological assemblies, structure quality, biomolecules.