Analysis of the conformations of organic compounds in macromolecular structures

Contemporary biophysical and biochemical research is based on the structures of macromolecules, in particular complexes of macromolecules with small molecules (ligands). The study of their interactions allows exploring the causes of civilization diseases such as various types of allergies, Alzheimer's disease or Parkinson's disease. The analysis of complex structures is the basis for modern methods of designing drugs for such diseases as HIV or diabetes. Existing methods of predicting protein structures only based on sequences work better and better when predicting their global topology. Predicting ligand binding (docking) allows for a list of possible interacting chemical compounds, but requires experimental verification. X-ray crystallography is the most popular and trusted method. X-ray diffraction allows accurate assignment of amino acids and ligand atoms to the obtained electron density. The structures obtained in this way are not only used to learn about the arrangement of the peptide chain or conformational changes due to the binding of reaction cofactors by the protein but are the basis for many studies of computational biology. Unfortunately, the crystals used for crystallographic studies are not perfect (long-range ordering) and the resolution of the diffraction data obtained does not allow simple and unequivocal identification of small chemical compounds. An important and often underestimated problem is competition between the components of the crystallization buffer and ligands in binding to active protein sites. This causes, in many cases, the wrong assignment of chemical compounds (ligands) to electron densities. Another mistake, which is also underestimated, is to attribute inadequate substance conformation. Particularly problematic are ring-containing compounds, most especially those with all single bonds. An example of such a ligand is cyclohexane, which may be in a boat or chair conformation. With poor data resolution, it is difficult to unambiguously determine it, which means that some deposits in PDB have an indirect energy conformation assigned to them. This can lead to misinterpretation of intermolecular interactions, which affects subsequent stages of drug design.

The algorithm I created to analyze the correctness of the structures of small ligands containing rings will identify the most common problems. The use of both information about the geometry and chemical knowledge about interactions will make the assessment of conformation correctness more reliable, which means that errors in ligand structures will appear much less frequently.