

Many PDB files (or mmCIF) files from the RCSB have multiple copies of the protein of interest. It's likely this is not the oligomer state in solution. For HullRad to give you an accurate analysis the PDB file should represent exactly what you have in solution. You should always look at a PDB file using a molecular graphics program before using the file for anything else (e.g., PyMOL, Chimera, VMD, Jmol, etc.). You can edit the PDB file to contain only the copy (chains) as it exists in solution.

Many PDB files also have missing residues. HullRad may not notice this and the calculated hydrodynamic properties will not reflect the solution structure. Sometimes the crystallographers will include only the backbone atoms but not the side chain. HullRad needs at least the CB atom and will not work if the entire side chain is missing. A solution to the above for proteins is to use the AlphaFold structure. For a folded protein it will be an accurate model. Note that if your protein has flexible loops, or an IDR, the single structure will likely not represent what exists in solution. But the AlphaFold structure will be a good start for modeling an ensemble of conformations.

Beware that the AlphaFold structure has all the amino acids in the open reading frame of the DNA. If your protein is processed by the cell, the actual protein you are studying in solution will be different from the AlphaFold one.

To find both the AlphaFold structure and the sequence of the mature, processed protein go to UniProt (<https://www.uniprot.org/>). Find your protein in this database, click on the Entry page, and you will find links to the AlphaFold structure and the sequence of the mature form.