[MS04-P09] ValiFrag: validation of fragments during automated protein model building.
Joana Pereira, Tim Wiegels, Victor Lamzin

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, Hamburg 22607, Germany
E-mail: joana.pereira@embl-hamburg.de

X-ray diffraction data from flexible macromolecules and their multi-component assemblies is rarely measured to a resolution better than 3 Å. The number of observations at such resolution is limited, and thus an atomic model cannot be straightforwardly constructed. The lack of observations requires the use of additional restraints and causes smoothing of density maps and a loss of detectable atomic features. Therefore, the determination of lowresolution structures is beyond the current operational range of crystallographic software and necessitates a large amount of manual intervention. Automated protein building at low resolution generates at best incomplete and fragmented models. ARP/wARP [1] version 7.4 generates structures that are up to 85% complete at 3.0 Å, but the completeness drops sharply as the resolution gets worse.
Reduction of the model completeness is complemented with an increase in the number of fragments built, which therefore become shorter. Such fragments are applicable for further model building when they are correct. However, when they are wrong, fully or in part, they may cause the formation of incorrectly built regions in the final model. Therefore, there is a need to improve the fragment quality before automated model completion is applied. Given the vast amount of structural information deposited in the Protein Data Bank (PDB), it should be possible to make use of it for structural validation of built fragments. Specifically, we evaluate the conformation of each fragment. If the conformation is present in several different protein models in the PDB, it is likely to be modelled correctly in the model being built and is accepted. On the contrary, if it cannot be found in any PDB model, it is likely incorrect. Here we present the software implementation of this validation, called ValiFrag, which checks the validity of automatically built protein chain fragments by evaluating their occurrence in the PDB.
Proteins on the PDB were broken into fragments from five to 25 residues in length and conformational parameters for each of these fragments were subsequently stored in a database. For each automatically built fragment, ValiFrag computes the number of its occurrences in the database. It, therefore, assesses which fragments are likely to be structurally incorrect and should possibly be modified, or even removed, to improve the final model.

[1] Langer, G., Cohen, S. X., Lamzin, V. S., &. Perrakis, A. (2008), Nature Protocols, 3(7), 1171-1179

Keywords: automated model building, fragment validation, software implementation