47. FILTREST3D: program for discrimination of protein structure models that match the restraints from experimental data

Michal J. Gajda, Marta Kaczor, Anastasia Bakulina, Janusz M. Bujnicki

Laboratory of Bioinformatics and Protein Engineering,
International Institute of Molecular and Cell Biology in Warsaw, Poland

We developed a program for efficient discrimination of a large number of models of protein structure (decoys) against a set of restraints derived from experimental analyses.

Contemporary methods for protein structure prediction from protein sequence usually generate not just one model, but from several (in fold-recognition) to tens of thousands (in de novo modelling) of alternative models. As demonstrated in CASP, among these sets of alternatives there usually exist models that somewhat resemble the native structure, but they are often extremely difficult to distinguish from completely wrong models using solely on computational analyses (e.g. by assessing the alignment or based on energy or knowledge-based potentials). It is known that inclusion of sparse experimental data as spatial restraints can greatly improve the selection of near-native protein structure models among the available alternatives. Relationships between sequence-structure and function of proteins are commonly studied by low-resolution methods: functionally important residues that cluster together in space (such as the active sites) can be identified by mutagenesis, protein surfaces or ligand-binding sites can be discovered by chemical modification, the crude topography of protein structure can be obtained by intra- or inter-molecular crosslinking and identification of crosslinked peptide fragments by mass spectroscopy, finally the shape of the molecule can be studied by electron microscopy or SAXS and SANS techniques. These experiments produce data that are more ambiguous, fuzzy and of much lower resolution than NMR or crystallography, and cannot be used to solve the protein structure on their own, but when combined with bioinformatics methods can pinpoint models with a correct global fold and architecture of functionally important regions. FILTREST3D is a new method, implemented as a web server (freely available at http://genesilico.pl/~mkaczor/score/), which allows to evaluate to which extent a user-defined model of protein structure agrees with the provided "fuzzy" distance restraints that can be derived from mutagenesis, chemical modification, and crosslinking. If a set of models is submitted then the method generates a ranking according to the degree of violation of the restraints. The currently implemented types of restraints include: permitted range of distances between the residues, amino acid burial/exposition to the solvent, and secondary structure. We are currently extending the method to use shape restraints and discriminate between alternative multimolecular - protein-protein and protein-ligand - docking models. FILTREST3D has been tested on a set of decoys generated for several restriction enzymes using ROSETTA and a set of experimental data generated by site-directed mutagenesis and CD spectroscopy.