Poster I-99

FILTREST3D: a simple method for discrimination of multiple structural models against spatial restraints

Authors:
Michal J. Gajda (*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell*)
Marta Kaczor (*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell*)
Anastasia Bakulina (*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell*)
Andrzej Kasperski (*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell*)
Alan M. Friedman (*Department of Biological Sciences, Purdue University, West Lafayette, IN, USA*)
Chris Bailey-Kellogg (*Department of Computer Science, Dartmouth College, Hanover, NH, USA*)
Janusz M. Bujnicki (*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell*)

Short Abstract: FILTREST3D http://filtrest3d.genesilico.pl/ is a new method for discrimination of alternative models against mixed sets of restraints (e.g. spatial distances) derived from experimental analyses as well as from computational predictions (e.g. solvent accessibility) or arbitrarily defined by the user. It can be used to analyze models of individual proteins and complexes.

Long Abstract:

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 modeling of alternative models. As demonstrated in CASP, among these sets of alternatives there usu-ally exist models that somewhat resemble the native struc-ture, but they are often extremely difficult to distinguish from completely wrong models using solely 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 func-tion 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; the shape of the molecule can be studied by electron microscopy or SAXS techniques. These experiments produce data that are more ambiguous, fuzzy and of much lower resolution than X-ray

crystallography or NMR spectroscopy, and cannot be used to solve the protein structure on their own. However when combined with bioinformatics methods, they can pinpoint models with a correct global fold and architecture of functionally important regions.

The aforementioned experimental methods have been previously explored for the discrimination of a set of protein models. In the case of crosslinking, for example, we have previously investigated model discrimination by detailed examination of model geometry (Potluri, et al., 2004) and probabilistic evaluation of experimental data coupled to efficient planning of optimal crosslinking experiments (Ye, et al., 2004).

FILTREST3D takes the alternative approach of evaluat-ing restraints by the simple counting of violations. This allows the rapid evaluation of the extent to which a user-defined model of protein structure agrees (lacks violations) with the provided "fuzzy" structural restraints derived from experiments or to rank a set of models 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, the overall content of secondary structure, specific local secondary structure, and the shape of the molecule. The method is applicable to individual protein domains as well as multi-subunit models (protein-ligand, protein-protein, protein-nucleic acid etc.) obtained by docking.

For instance, we used FILTREST3D to identify a docking model of a complex between two independently modeled domains of a tRNA methyltransferase TrMet(m2G10) and the crystal structure of its substrate tRNA, which agrees with a set of experimental restraints obtained by RNA mutagenesis and studies of protein-RNA interactions, by limited proteolysis of the enzyme and chemical modification to identify surface-exposed residues. In the analysis described in detail by Gabant, et al. (2006), we generated a set of 1000000 combinations between the docking models obtained independently for each protein domain and the tRNA molecule by low-resolution docking and identified only two very similar models that fulfilled nearly all restraints.

Our preliminary tests suggest that FILTREST3D is a very useful tool for preliminary protein fold assignment, which could be useful in structural genomics, e.g. to improve the discrimination between candidates for ‘old’ and ‘novel’ folds, as well as for the characterization of quaternary structures in macromolecular complexes, including those that are difficult to crystallize. Obviously, the utility of our method, e.g. its ability to discriminate between native-like and native-unlike models, critically depends on the quantity and quality of experimental data and the interpretation required to translate the results of experiments into spatial restraints. In the future studies we will focus on the determination of the robustness of FILTREST3D with respect to the coverage and correctness or (un)certainty of restraints derived from different types of data.

References:

Gabant, G., Auxilien, S., Tuszynska, I., Locard, M., Gajda, M.J., Chaussinand, G., Fernandez, B., Dedieu, A., Grosjean, H., Golinelli-Pimpaneau, B., Bujnicki, J.M. and Armengaud, J. (2006) THUMP from archaeal tRNA:m22G10 methyltransferase, a genuine autonomously folding domain., Nucleic Acids Res, in press.

Potluri, S., Khan, A.A., Kuzminykh, A., Bujnicki, J.M., Friedman, A.M. and Bailey-Kellogg, C.

(2004) Geometric analysis of cross-linkability for protein fold discrimination, Pac Symp Biocomput, 447-458.

Ye, X., O'Neil, P.K., Foster, A.N., Gajda, M.J., Kosinski, J., Ku-rowski, M.A., Bujnicki, J.M., Friedman, A.M. and Bailey-Kellogg, C. (2004) Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure, Protein Sci, 13, 3298-3313.