Amino acid propensities for secondary structures

Costantini S (1,2), Colonna G (2), Facchiano AM (1,2)

(1) Laboratorio di Bioinformatica e Biologia Computazionale, Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italy
(2) Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università di Napoli, Italy

Motivation

Propensity for secondary structures represents an intrinsic property of amino acid, and it is used for generating new algorithms and predictive methods. Our work has been aimed to investigate what is the best protein dataset to evaluate the amino acid propensities, either larger but not homogeneous or smaller but homogeneous sets consisting of only all-alpha, only all-beta or only alpha-beta proteins.

Methods

All analyses were performed using a set of experimentally determined, non-redundant protein structures in the PDB with mutual sequence homology <25%. The secondary structure for every PDB entry was assigned by the DSSP algorithm and was used to assign the secondary structural class, according to two definitions of structural classifications. The residue propensity values in different secondary structural types (Pij) were determined with an original software from the ratio of the residue's frequency of occurrence in helices, beta-strand and coil versus its frequency of occurrence in the PDB. An original prediction strategy was applied, based on the residue propensity for the secondary structures. For each protein amino acid, the average value of helix, beta strand and coil propensity has been determined by considering their surrounding amino acids in a window of length n. For each secondary structure, it has been evaluated the better window as well as the better coefficient to be applied to the average propensity of the segment. Finally, by comparing the average helix, beta strand and coil propensities for the segment under examination, the higher value was the criterion to assign the secondary structure of the amino acid in the middle of the segment. The quality of these predictions was examined by resubstitution and jackknife tests.

Results

We evaluated amino acid propensities for helix, beta-strand and coil in more than 2000 proteins from the PDBselect dataset. With these propensities, secondary structure predictions performed with a method very similar to the Chou and Fasman gave us results better than the original one, based on propensities derived from the few tens of X-ray protein structures available in the 1970s. Moreover, we subdivided the PDBselect dataset of proteins in three secondary structural classes, i.e. all-alpha, all-beta and alpha-beta proteins. For each class, the amino acid propensities for helix, beta-strand and coil, have been calculated and used to predict secondary structure elements for proteins belonging to the same class by using resubstitution and jackknife tests. The results were improved in comparison to the predictions for the whole PDB dataset (1). Therefore, amino acid propensities for secondary structures result more reliable depending on the degree of homogeneity of the protein data set used to evaluate them. Indeed, our results indicate also that all algorithms using propensities for secondary structure can be still improved to obtain better predictive results. We are developing a web-service to predict the secondary structure of proteins from their amino acid sequence using the amino acid propensities for secondary structures calculated for this study.

The service will be available at the lab web server: http://bioinformatica.isa.cnr.it/

Availability: http://bioinformatica.isa.cnr.it/

Contact email: susan.costantini@unina2.it , angelo.facchiano@isa.cnr.it

References

1. Susan Costantini, Giovanni Colonna and Angelo M. Facchiano: "Amino acid propensities for secondary structures are influenced by the protein structural class.", Biochemical and Biophysical Research Communications (2006), 342, 441-451.