Discriminative Analysis of the Conformational Patterns of Protein Amino Acid Residues

Motokazu KAMIMURA
kamimura@molout.tutkie.tut.ac.jp

Yoshimasa TAKAHASHI
taka@molout.tutkie.tut.ac.jp

Laboratory for Molecular Information Systems,
Department of Knowledge-based Information Engineering,
Toyohashi University of Technology, Toyohashi, Aichi 441 JAPAN

Abstract

This paper describes the discrimination of ϕ-ψ conformational pattern classes for protein amino acid residues which were defined in our previous works. Statistical discriminant analysis technique has been employed for the present analysis. Each residue was characterized by its peripheral physicochemical environment. The environment was described in a vector representation of which components involve Van der Waals volume, hydrophobic parameter π, and partial charges of a carbon atom, hydrogen atom of NH and oxygen atom of C'O of ten neighbor residues (five neighbors in each terminal side of the target residue). The discriminant functions obtained with 67 proteins taken from the PDB file correctly discriminated 58.3% of the residues for their conformational pattern classes.

1 Introduction

Protein higher structure prediction is one of very important problems in molecular biology and protein engineering. Much effort has been spent on predicting the higher structures of proteins from their amino acid sequences. However, no fully successful methods for the problem has yet been established. Obviously, this problem requires exhaustive analysis of the relations between the partial conformation of the main chain and the related peptide segments and deep understanding of those relationships. The authors have been investigating the relationships between amino acid sequences and higher structure of proteins, aiming to apply the results to the higher structure prediction of proteins. In our previous paper, the ϕ-ψ conformational clustering of protein amino acid residues showed that there are two to five of typical conformational pattern classes for twenty kinds of amino acid residues[1]. The representative local three-dimensional structural features of the each conformational pattern class were also examined[2]. In the present paper, we will discuss on class discrimination of ϕ-ψ conformational pattern clusters of protein amino acid residues which were obtained in our previous work.

神村基和、高橋由雅：豊橋技術科学大学 知識情報工学系、〒441 豊橋市天伯町字雲雀ヶ丘 1-1

2 Method

The peptide fragments which consists of eleven amino acid residues (i.e. a center residue referred to as a target amino acid residue; TA, and five neighbor residues on each side of the target residue) were prepared from the set of sixty seven proteins taken from the Protein Data Bank. These peptide fragments were grouped into five classes (A, B ,L ,X and Y) which were defined in our previous work according to the ϕ-ψ conformational pattern of the TA for every kind of the TA residue. The development of discriminant models was carried out by ordinary statistical discriminant analysis technique. Environment description vector (EDV) method based on physical and chemical properties of amino acid residues was used for the description of the peripheral environment of each TA . In the present work, the five neighbor residues at each side of the TA were considered in the environment description, and the following five physicochemical parameters (Van der Waals volume, hydrophobic parameter π, and partial charges of a carbon atom, hydrogen atom of NH and oxygen atom of C'O of the main chain) were used as environment descriptors of each residue. Therefore, each EDV is described as a fifty-dimensional vector.

3 Result and Discussion

The discriminant models of the conformational pattern classes of protein amino acid residues were developed by the linear discriminant analysis. The discrimination rate of the conformational pattern classes for 14,723 residues from 67 proteins was 58.3% in the total average. Obviously the obtained models can be used to predict the conformational pattern classes of amino acid residues from the other proteins. The discriminant models were applied to the prediction of the conformational pattern classes of new 26 proteins (3,806 residues). The models correctly classified 53.2% of the total residues. These results suggest that the current approach with environment description vector could be applicable for the prediction of the ϕ-ψ conformational pattern classes.

Acknowledgement

This work was supported by the Ministry of Education, Science and Culture of Japan under the Grant-in-Aid for Scientific Research on Priority Area, 'Genome Informatics'.

References

[1] M.Kamimura and Y.Takahashi, *CABIOS*, 10,163-169(1994).

[2] M.Kamimura and Y.Takahashi, *Proceedings Genome Informatics Workshop IV*, 1993, pp315-324.