

Application of Covariation Analysis in 3D Structure Prediction of G-Protein Coupled Receptors

Wataru Nemoto¹

rb007973@se.ritsumei.ac.jp

Takashi Imai¹

t-imai@se.ritsumei.ac.jp

Takeshi Kikuchi²

tkikuchi@chem.kusa.ac.jp

Norihisa Fujita¹

norihisa@se.ritsumei.ac.jp

¹ Department of Science and Technology, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

² Department of Chemical Technology Kurashiki University of Science and The Arts, 2640 Tsurejima-cho, Hishiura, Kurashiki, Okayama 712-8505, Japan

Keywords: covariation analysis, G-protein coupled receptor, contact prediction, evolutionary trace

1 Introduction

Genes encoding the large family of G binding protein coupled receptors (GPCRs) occupy perhaps 3 % of human genome. GPCRs represent one of the most important target class of proteins for drug. Over 30% of clinically marketed drugs are active at this family. Homology or comparative modeling based on the crystal structure of bovine rhodopsin [3] is currently the most frequently used for the GPCR-targeted drug design. However, the absence of significant sequence identity to the rhodopsin makes it more uncertain to predict other GPCRs' structure. GPCRs are characterized by seven transmembrane α -helices. We have reported that seven transmembrane α -helical regions were predictable with high accuracy by adequate combination of some methods, HMM, SOSUI and Fourier Transform analysis. However, their topology is hardly predicted. To improve the accuracy of the predicted structure, the topologies of secondary structure are essential. Covariation analysis was shown to predict residue-residue contacts [2]. Therefore, we attempted covariation analysis to improve the topology prediction.

2 Method and Results

2.1 Detection of Covariation

An alignment of 527 Opsin family sequences was used in this study. The alignment was constructed by searching of the sequence databases using BLAST. To account for over-representation of certain closely related families of sequences within the alignment, a weighting algorithm was used when the residue occurrence frequencies were calculated for each position in the alignment [1]. In this way, the calculated residue frequencies were more representative of the family as a whole. Covariation between positions was calculated by comparing the expected probability of two residues occurring together in any one sequence to the frequency with which they actually do appear together. Phi coefficients were also calculated and used to evaluate the strength of covariation as has been described [2].

2.2 Detected Residue Pairs and Clusters

Covariation residue pairs were adopted with a p value less than 0.001. Among them, we searched the residues possessing higher phi values more than 0.50 and obtained 17 residue pairs of which distance between the center of mass of their side-chains was within 8Å (Table 1). Interestingly, these pairs were localized in the following regions; extracellular and intracellular region, transmembrane region (Figure 1).

Table 1: Detected residue pairs and their localization.

Residues	Cluster	phi	Residues	Cluster	phi	Residues	Cluster	phi
Q28-Q184	I	0.627522	P23-Q184	I	0.508058	E249-N310	IV	0.552795
P23-Y102	I	0.618758	E122-H211	II	0.776137	M253-M309	IV	0.544543
F9-P23	I	0.559843	C167-M207	II	0.754542	A234-R252	IV	0.517067
V20-Y30	I	0.53516	I48-T92	III	0.531888	L79-N302	IV	0.51269
F9-I179	I	0.508676	M257-N302	IV	0.677791	L76-N302	IV	0.500932
Q28-Y102	I	0.503638	I75-S127	IV	0.634841			

3 Discussion

As shown in Figure 1, the highest density of covarying pairs is localized to the transmembrane region close to cytoplasm (cluster IV). This region is considered to be common to whole GPCR family and plays crucial role for transferring the intramolecular signal to the corresponding G protein. But it is not concerned with a ligand binding. The second highest density of covarying pairs locates in extracellular region (cluster I). The information is almost absent about the extracellular structure. Depending on a ligand, the structure of this region is divided into two. One is bulky enough to form a binding site for a large ligand such as a peptide, while the other is compact enough to have a small ligand such as an amine passed into the helical bundle. Rhodopsin belongs to the latter family. The third high density of covarying residues locates near the extracellular site of transmembrane helices (cluster II) surrounding the π -ionone ring of retinal. E122 appears specifically in rhodopsin family and plays an important role. Evolutionarily it was substituted from Q to E when the cone β -pigment family diverged to rhodopsin family. This result suggests that covariation analysis is useful to detect subfamily-specific residue pair and provides the information of structural difference between subfamilies. The last is between Helix I and II (cluster III). This is the hydrophobic residue pair near the hydrogen bond. Several hydrophobic residue pairs showing relatively high phi value locate in the neighborhood. This result was well corresponding to the previous report that the covariation analysis presented hydrophobic interaction in SH3 domain [2].

In conclusion covariation analysis can be applied to identify functional sites and interacting residues in the representative structure. The covariation analysis currently produces a high fraction of false positives. However, as recently reported about Evolutionary Trace, covariations ranked at the top of phi values also assembled to clusters and they had important biochemical functions [4]. Thus the clustering should have significant meanings.

References

- [1] Henikoff, S. and Henikoff, J.G., Position-based sequence weights, *J. Mol. Biol.*, 243:574–578, 1994.
- [2] Larson, S.M. *et al.*, Analysis of covariation in an SH3 domain sequence alignment: Applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions, *J. Mol. Biol.*, 303:433–446, 2000.
- [3] Palczewski, K. *et al.*, Crystal structure of rhodopsin: A G protein-coupled receptor, *Science*, 289:739–745, 2000.
- [4] Madabushi, S. *et al.*, Structural clusters of evolutionary trace residues are statistically significant and common in proteins, *J. Mol. Biol.*, 316:139–154, 2002.

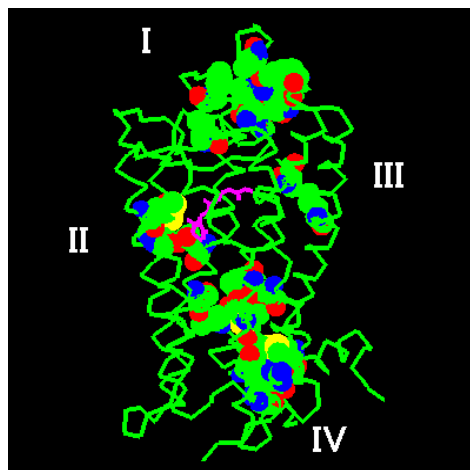


Figure 1: Detected four clusters. Each residue pair has high score of phi association coefficient.