Conformational Search and Analysis of β-hairpin Formation by High-Speed Exhaustive Tree Search

MAKOTO ANDO†, YUTAKA AKIYAMA†† and HIDEO MATSUDA†††

We investigate the stability of a conformational motif called tyrosine corner, which is a conformational feature found in most Greek key β-barrel proteins, by using an exhaustive conformational search. Our conformational search model is based on an exhaustive tree search that can be performed on a massively parallel computer, and the strong point is that it can rapidly generate feasible conformations of the target peptide depending solely on steric clash detection and distance/angle constraints on the atoms of the target peptide. The proposed method has a great potential to help us understanding the mechanism of folding of a protein, which we could not realize by other methods such as a molecular dynamics or a Monte Carlo simulated annealing.

1. Introduction

The folding mechanism of proteins from linear chains of amino acids to specific tertiary structures has been of great interest to biological scientists for a long time. The amino acid sequence of a protein can be divided into individual segments in the folded structure, that form small secondary structural units such as α-helices and β-sheets. Connecting these segments are regions that ultimately form loops, turns, hairpins, or areas of nonordered structure in the folded protein. The diversity of loops makes a large contribution to the multiformity of protein functions, but also causes difficulties in protein structure studies.

In this research, we investigate the stability of a hairpin structure called "tyrosine corner" by using ESCAPE[1] (Exhaustive Search system for the Conformational Analysis of PEptides), a tool for predicting the conformations of peptides such as loop regions of proteins. ESCAPE was designed for the structural analysis of peptides of a few amino acids in atomic resolution, based on tree search operations in parallel, such as branch operations and pruning operations[1].

Conformation prediction based on exhaustive tree search has many advantages over other approaches. Firstly, because the search operations do not have to depend on any energy calculation, the systems can analyze conformations even if the physicochemical environment around the target peptide is not precisely and accurately known. In such cases, the systems based on the exhaustive tree search model can analyze the conformations more precisely and accurately than systems based on energy minimization approaches, such as molecular dynamics or a Monte Carlo simulated annealing[2]. Secondly, exhaustive tree search operations do not have to depend on the knowledge obtained from PDB[3),4], limited information by the size of the database. The systems using exhaustive tree search can analyze conformations of peptides even if any similarity in sequence or structure is not found in the known structure database.

†NKK Corporation
††Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)
†††Graduate School of Information Science and Technology, Osaka University

2. Methods

Many of the single-covalent bonds in a peptide allow free rotation of the atoms they join, giving the peptide great flexibility. We modeled such flexibility by using an exhaustive tree search. In our model, we

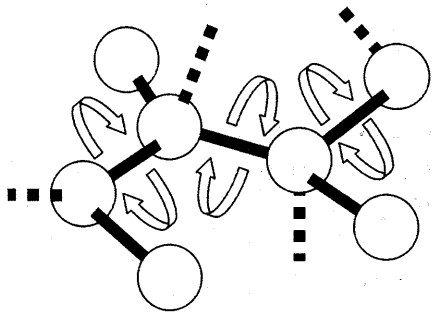

Fig. 1 Conformational search model

assume that the atoms are hard spheres with certain radii, and the single-covalent bonds are rotatable axes, each with a fixed length and fixed bond angles (Figure 1). Conformational isomers are generated by systematically varying the torsion angles of all the rotatable axes according to the search tree. When the torsion angles of one or more rotatable axes are set to certain values, relative positions of some atoms around these axes are fixed. The fixed atoms are then checked for feasibility: steric clashes, distance constraints, and angle constraints are examined. When any of the requirements is not satisfied, the corresponding node in the search tree is pruned. Two kinds of constraints can be defined; distance constraints on any pairs of atoms, and constraints on the angles defined by any three atoms in the peptide.

In the implementation of ESCAPE, we added a function to calculate the potential energy of each generated conformation of the target peptides in order to assign the information regarding their stabilities. The calculation basically follows the commonly used force field for molecular dynamics calculations, such as AMBER[5], but some calculation terms are neglected in order to realize a high-speed calculation. We incorporated the following two terms into our

$$U = \sum_{i>j} \frac{q_i q_j}{r_{ij}} + \frac{1}{SC_{14ELC}} \sum_{1\text{-}4pairs} \frac{q_i q_j}{r_{ij}} \qquad (1)$$

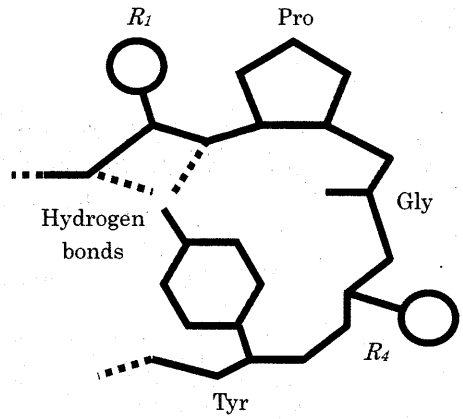

Fig. 2 Tyrosine corner

program: (a) Coulomb, and (b) 1-4 Coulomb (Eq. 1).

3. Results

3.1 Tyrosine corners

To demonstrate the effect of the proposed conformational search model, we focused on the conformational search of a peptide segment of five amino acid residues called a tyrosine (Tyr) corner[6]. The tyrosine corner is a conformational motif in which a Tyr (residue R_5) near the beginning or end of an antiparallel ß strand makes a hydrogen bond from its side-chain OH group to the backbone NH and/or CO of residue whose position is R_1 (Figure 2). The "Δ4 tyrosine corner", the most typical tyrosine corner, usually consists of the five amino acids as follows:

- R_1 = a hydrophilic amino acid
- R_2 = always a proline
- R_3 = usually a glycine whose left-handed ß or very extended ß conformation helps the backbone curve around the tyrosine ring
- R_5 = always a tyrosine

Thus, the consensus sequence is xPGxY. According to the earlier research[6], it seems likely that the tyrosine corners contribute to the stability of a Greek key, and also that they may aid in the process of folding up Greek key structures.

Table 1 Sequences used in the examination.

Amino acid sequences	PDB ID	Chains	Positions
QPGAY (Gln-Pro-Gly-Ala-Tyr)	1PAZ	-	70-74
VPGAY (Val-Pro-Gly-Ala-Tyr)	1DI0	A	53-57
QGGAY (Gln-Gly-Gly-Ala-Tyr)	1BCP	C	16-20
QAGAY (Gln-Ala-Gly-Ala-Tyr)	2DOR	A	36-40
QLGAY (Gln-Leu-Gly-Ala-Tyr)	1CP2	A	108-112
QDGAY (Gln-Asp-Gly-Ala-Tyr)	1QKK	A	96-100
QEGAY (Gln-Glu-Gly-Ala-Tyr)	1F3I	A	53-57

Table 2 Search results

Sequences	# Conformations	# Backbone
QPGAY	18,110	5
VPGAY	80	5
QGGAY	404,070	105
QAGAY	229,760	40
QLGAY	1,358,840	40
QDGAY	2,133,960	65
QEGAY	8,909,230	65

3.2 Potential energy distributions of tyrosine corners

Table 1 shows the sequences used in the examination. A distance constraint was set for the conformational search of these sequences :

- O of R_1 - O_η of Tyr: 0.8-4.8Å

This distance constraint corresponds to a hydrogen bond that is found in natural proteins. Using the distance constraint, feasible conformations of the target peptides (shown in Table 1) are searched by using ESCAPE. Compaq AlphaServer ES40 (Alpha 21264 500MHz x 4) was used for the experiments. The rotation intervals for all the rotatable axes were set to 60 degrees.

Table 2 shows the search results: the numbers of feasible conformations and the numbers of the conformations of the backbones are shown. Figure 3 is the superposition of the histograms of potential energy distributions for the feasible conformations of the test sequences generated by ESCAPE.

4. Discussion

In order to discuss the stability of the tyrosine corner, we focus on the following two points. Firstly, according to the Table 2, the "# conformations" column (i.e., the total number of feasible conformations) showed small numbers when the sequences have a proline in the second position (R_2) of the five amino acids sequences due to its rigid structure. The backbone of a proline has less flexibility than that of other amino acid residues. This fact implies that the proline in R_2 restricts the variation of feasible conformations. This restriction made the shapes of the plots for the QPGAY and VPGAY sharper than other plots (Figure 3).

Secondly, we focus on the energy distribution. For the sequence QPGAY, the potential energy of all the feasible conformations fell into relatively low values. On the other hand, the potential energy of all the feasible conformations of VPGAY fell into relatively high values. The difference between the two sequence is R_1. This implies that R_1 must be a hydrophilic amino acid to form a tyrosine corner, and that the peptide do not form a tyrosine corner when R_1 is a hydrophobic amino acid residue.

The observations described above agree with the physicochemical facts: this implies that our computational model (both the peptide modeling and the potential energy function) and the experimental conditions (rotation intervals, distance constraints, etc.) are correct, and our method is proper for searching and analyzing the conformations of peptides, such as loop regions of proteins.

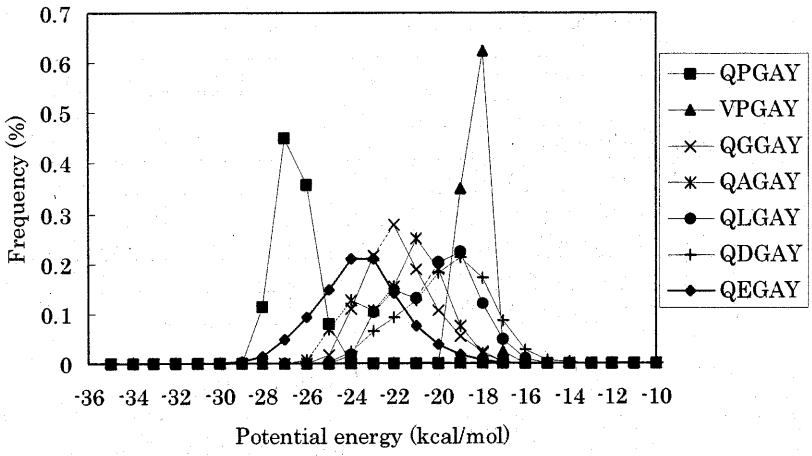

Fig. 3 Histograms of potential energy distribution

In summary, the combination of a thorough (exhaustive) conformational generations and a potential energy calculation is powerful for analyzing short peptides, with only a few knowledges from its environment such as distance/angle constraints obtained from structural features such as hydrogen bond information or receptor structure information. This new method has a great potential to help us understanding the mechanism of folding of a protein, which we could not realize by other methods such as a molecular dynamics or a Monte Carlo simulated annealing.

References

1) Ando, M., Akiyama, Y. and Matsuda, H.: Hierarchical Approach to Parallel Tree Search for Protein Conformational Analysis, Proc. 2000 International Conference on Parallel and Distributed Processing Techniques and Applications, pp.643-649 (2000).

2) Kinoshita, M., Okamoto, Y. and Hirata, F.: First-Principle Determination of Peptide Conformation in Solvents: Combination of Monte Carlo Simulated Annealing and RISM Theory, J. Am. Chem. Soc., Vol.120, pp. 1855-1863 (1997).

3) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E.: The Protein Data Bank, Nucleic Acids Research, Vol.28, pp.235-242 (2000).

4) van Vlijmen, H. and Karplus, M.: PDB-based Protein Loop Prediction: Parameters for Selection and Methods for Optimization, J. Mol. Biol., Vol.267, pp. 975-1001 (1997).

5) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A.: A second generation force field for the simulation of proteins and nucleic acids, J. Am. Chem. Soc., Vol.117, pp.5179-5197 (1995).

6) Hemmingsen, J. M., Gernert, K. M., Richardson, J. S. and Richardson, D. C.: The tyrosine corner: A feature of most Greek key beta-barrel proteins, Protein Sci., Vol.3, pp.1927-1937 (1994).