CONFLEX: Conformational Behaviors of Polypeptides as Predicted by a Conformational Space Search

H. Goto[*], T. Takahashi[*], Y. Takata[*], K. Ohta[**] and U. Nagashima[***]

[*] Toyohashi University of Technology, Japan, gotoh@cochem2.tutkie.tut.ac.jp
[**] Conflex Corporation, Yokohama, Japan, k-ohta@conflex.co.jp
[***] National Institute of Advanced Industrial Science and Technology, Japan, u.nagashima@aist.go.jp

ABSTRACT

A new method for searching folding pathways of small polypeptides from over thousands conformers, which were generated by our original conformational space search program, CONFLEX, has been developed. In order to approximate an interconversion between similar conformers, conformational distance in an angular space based on all peptide's back-bone torsion angles (ϕ and ψ) were used as a conformational similarity index. A folding pathway that consists of a series of conformers between predefined starting and target conformers was determined by using a single-linkage clustering technique based on the similarity indices of conformational distance matrix. This method has been applied to finding a folding pathway from seven thousands conformers of alanine octamer. Interesting conformational behaviors in this folding process have been discussed.

Keywords: Conformational Space Search, Peptide Conformation, Polyalanine, Conformational Analysis, Parallelization Technique, Folding Process

1 INTRODUCTION

Elucidation of the stable conformations and the folding process of proteins is one of the most fundamental and challenging goals in life science. While some of the most common secondary structures (e.g., a few types of helix, beta-strand, coil, and etc.) have been well-known, precise analyses for thousands order of the chemically important conformers and picoseconds order of their conformational interconversions via the transition states on the potential energy surface will be required for the study of the microseconds order of folding process toward the tertiary structure formations.

CONFLEX has been known as one of the most efficient reliable conformational space search programs [1]. Recently, we have applied high-performance computing techniques into this program, and suggested that the performance of the parallelized CONFLEX enables to explore the lower-energy region of the conformational space of small polypeptides within an available elapsed time using PC cluster [2]. In this paper, we propose a new method for predicting a folding pathway of alanine octamer based on the seven thousands conformers found by "Parallel CONFLEX".

2 METHODS

Although our new method provides attractive views of a folding process and conformational behaviors, the reliability is strongly dependent on finding all conformers of low-energy regions. Therefore, algorithms and techniques used in our conformational space search CONFLEX and parallelized CONFLEX are also briefly explained below.

2.1 Algorithm of Conformational Space Search

The basic strategy of CONFLEX is the exhaustive search of only the low-energy regions. In order to actualize the strategy, the original CONFLEX is organized to contain the following four major steps: 1) selection of an initial structure from among the already found and unique conformers sorted in a conformational database (only the beginning of a search execution, input structure is used for the first initial structure), 2) generation of trial structures by using local perturbations to the selected initial structure, 3) geometry optimization for those trial structures, 4) comparison of those successively optimized (trial) structures with the other conformers stored in a conformation database, and preservation of newly found unique conformers into the database (Figure 1).

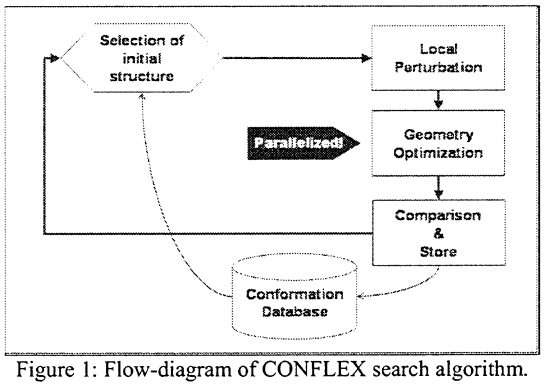

Figure 1: Flow-diagram of CONFLEX search algorithm.

In these repeating procedures, two unique strategies are incorporated. The first strategy is the local perturbations: corner flapping and edge flipping, the both for ring atoms, and stepwise rotation for side-chain or backbone chain, that are highly efficient in producing several good trial structures. These perturbations can be considered to mimic a barrier-crossing step in the elementary process of the thermal conformational interconversion [1]. Actually, perturbations are so weak and localized. Hence, the search by perturbations of an initial structure is corresponding to precise performance around the space of the initial structure.

"Lowest-Conformer-First", the selection rule of the initial structure, is another strategy for directing the conformation search expanded to low-energy regions. The definition is that the initial structure is selected the lowest energy conformers stored in conformation database, but never perturbed as the initial structure. This rule proved effective in going down the search space towards the lower and lower energy region, like a water stream filling an empty reservoir quickly finds the lowest point while overflowing local dips on the way. Therefore, *Reservoir Filling Algorithm* is our named term for these tactical procedures of CONFLEX search [1].

In order to stay in the low-energy region and perform exhaustive search, the search limit (SEL), that determines the maximum energy of the initial structures, is pre-defined. Gradually increasing SEL allows to search only the low-energy conformers and to avoid straying into unnecessarily high-energy regions. According to our experience, SEL of 7-10 kcal/mol from the global energy minimum is generally sufficient for the exhaustive search covered in chemically meaningful regions [1].

2.2 Parallelization of CONFLEX

So far we had applied CONFLEX to the molecules of less than 100 atoms system [3]. In order to apply to over 100 atoms, CONFLEX was improved by using the high performance and parallel computing techniques. In general, when parallel computing techniques will apply to existing software, the most time-consuming part of its procedures must be known. In CONFLEX search algorithm, heaviest task was easy to find, because geometry-optimization procedures always spend 90% of the elapsed time of search execution. Therefore, we have parallelized this optimization part by using Master/Slave parallelization technique (see also Figure 1). Modified search procedures are as follows: After trial structures are generated (step 2), they are stored in task pool on the master machine for a while. Then, each slave machine is dynamically supplied with one trial structure from the master. After finishing its optimization on each *slave*, the *master* immediately supplied again with another trial structure. When all the trial structures related to a given initial structure are optimized, only *master* procedure goes to comparison part.

Examination of the parallelization speedup of CONFLEX was performed limited conformation searches for alanine octamer by using one to sixteen AMD Athron 1900+ processors with MPICH parallelization environment library on Linux. So far, performance of parallelized CONFLEX with 16 processors has reached 11 times in comparison with a serial (non-parallel) CONFLEX [2].

In order to complete requirements for searching a folding pathway, exhaustive conformation search of low-energy regions for alanine octamer was also executed. Both terminals of alanin octamer were protected by acetyl function for N-terminal and methyl esterized to C-terminal, Ac-$(Ala)_8$-OMe, and its β-strand conformation was used for the input structure as the first initial structure. Search limit finally reached to 10 kcal/mol from the instant global energy minimum, and stepwise rotation was applied to all rotatable torsions (ϕ and ψ) of backbone bond. Extended MM2 force field [4] was used as a geometry-optimizer. Finally, we found 7,213 conformers of alanin octamer within 7 days with 16 processors of PC cluster as mentioned above.

2.3 Algorithm for Searching a Folding Pathway

Now we obtained over seven thousands conformers of alanin octamer on potential energy surface of the reliable force field. If we know the energy-potential surface, at least, all the important transition states between the minima, we can understand their dynamical conformational interconversion processes based on the transition state theory [5]. However, transition states of such large molecule are usually unknown. Therefore, we propose a rough approximation of the barrier of conformational interconversion process.

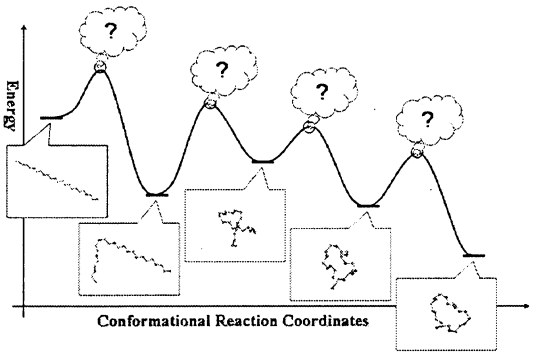

Figure 2: Illustration of Conformational Space

Assuming that an interconversion between similar conformers is easy to occur, some conformational similarity indices can approximately evaluate the barrier of the interconversion. As one of the candidate equations, conformational distance in an angular space based on all peptide's back-bone torsion angles (ϕ and ψ) are useful for such evaluation [6].

Based on this assumption, we developed a simple program for finding a series of conformers between widely

separated conformers in a conformational space by using a modified single-linkage clustering technique based on the similarity indices of conformational distance matrix. The summary is as the follows:

1) The first, define the starting and the target conformers.
2) Minimized the maximum distance on all pathways between the starting and the target conformers.
3) If some pathways have the same value of the maximum distance, and then minimize the second maximum distance.
4) The same procedure applied to third maximum, forth, and so on.

In this algorithm, we additionally considered the interconversion process with upward energy changing. Bond-crossing process is also watched out for avoiding any impossible conformational interconversion in fact.

3 RESUTS AND DISCUSSION

Our new method for finding a folding pathway was programmed based on the algorithm mentioned above, and applied to 7,213 conformers of Ac-$(Ala)_8$-OMe found by parallelized CONFLEX, where β-strand and α-helix conformations are selected as the starting and target structures, respectively, as shown in Figures 3 and 4, steric energies in vapor phase calculated by a modified MM2 force field were used for evaluation of the relative stability among conformers, and threshold of upward energy change was set to 5 kcal/mol.

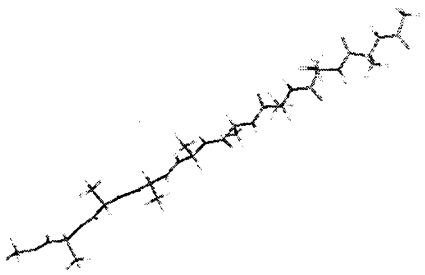

Figure 3: β-strand conformation of Ac-$(Ala)_8$-OMe (ΔE=58.30 kcal/mol)

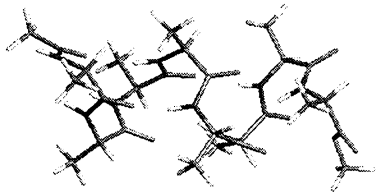

Figure 4: α-helix conformation of Ac-$(Ala)_8$-OMe (ΔE=4.72 kcal/mol, 1749 degrees).

Figure 5 shows energy changes of conformers on the course of a folding process from β-strand to α-helix conformation of Ac-$(Ala)_8$-OMe. It is interesting to note that the sum of conformational distances between the neighboring conformers along the reaction coordinates corresponds to the reaction time of this folding process, if the conformational distance between two minima is correlated with the corresponding transition state. Systematic studies on the possible relationship, in general conformational interconversions, between barrier heights and the conformational distances will be interesting [6].

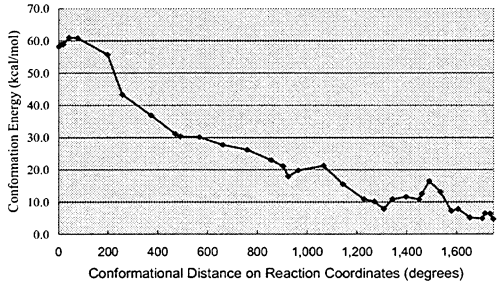

Figure 5: Energy changes of a folding process from β-strand to α-helix conformation of Ac-$(Ala)_8$-OMe.

Figures 6-10 show structures of various conformers appeared at unique stationary points on the folding pathway. Relative energy to the global energy minimum and conformational distance from the starting β-strand conformation are indicated in each caption.

It is noted that the whole folding process can separate into two periods: the first period is a dynamical formation of a hair-pin like conformation via a bending form, and the second period is, as a matter of course, an expedition toward α-helix conformation just as our intention. Additionally, it is by no means easy to say that appearance of a complicated globule conformation as the halftime event connected with two periods is insignificant. It is also interesting to note that a half-completed helix conformation, which has a helical part only on the C-terminal side of back-bone, appears at the beginning of the second period reached to α-helix.

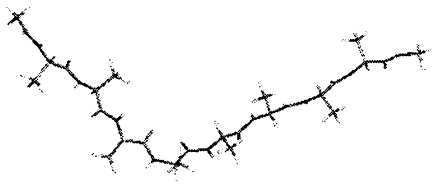

Figure 6: Bended conformation (ΔE=60.94 kcal/mol, 41 degrees)

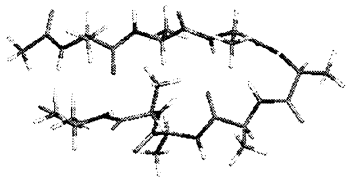

Figure 7: Hair-pin like conformation
(ΔE=30.15 kcal/mol, 567.degrees)

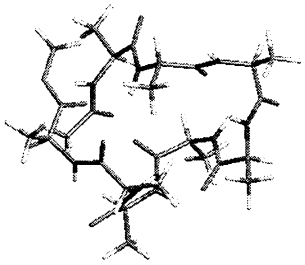

Figure 8: Complicated globule conformation
(ΔE=17.88 kcal/mol, 925.degrees)

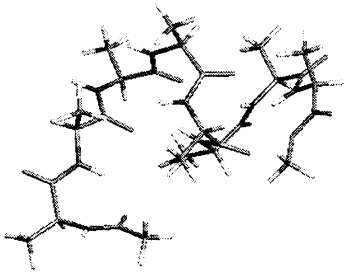

Figure 9: Half-helix conformation
(ΔE=21.26 kcal/mol, 1,066.degrees)

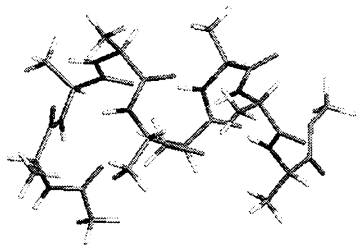

Figure 10: α-helix like conformation
(ΔE=7.85 kcal/mol, 1309.degrees)

4 CONCLUSION

A new method for searching folding pathways of small polypeptides has been developed and reported here. Application of this method to over seven thousands conformers of alanine octamer generated by parallelized CONFLEX has found out a folding pathway between its β-strand to α-helix conformation, and provided interesting information about conformational behaviors in this folding process. We believe that the combination of parallelized CONFLEX and this folding pathway search method will be useful for approach to the most fundamental and challenging goals in life science.

ACKNOWLEDGEMENTS

Financial support by Japan Ministry of Education, Culture, Sports, Science and Technology (Grant-in-Aid for the Development of Innovative Technology: No.12413) is highly acknowledged. This work is also partly supported by ACT-JST (Research and Development for Applying Advanced Computational Science and Technology of Japan Science and Technology Corporation.

REFERENCES

[1] H. Goto and E. Osawa, "Corner Flapping: A Simple and Fast Algorithm for Exhaustive Generation of Ring Conformations", *J. Am. Chem. Soc.*, **111**, 8950-8951, 1989; H. Goto and E. Osawa, "An Efficient Algorithm for Searching Low-energy Conformers of Cyclic and Acyclic Molecules", *J. Chem. Soc., Perkin Trans. 2*, 187-198, 1993; H. Goto and E. Osawa, "Approach to Global Minimum Problem", *J. Mol. Struct. (THEOCHEM)*, *285*, 157-168, 1993.

[2] H. Goto, Y. Takata, K. Takagi, T. Takahashi, Y. Kamakura, K. Ohta, N. Nakayama, U. Nagashima, "CONFLEX: Practical Conformational Space Search - Approach to Proteins", *6th World Congress of World Association of Theoretically Oriented Chemists*, Lugano, Switzerland, Palazzo dei Congressi, 2002.

[3] For example, H. Goto, Y. Kawashima, M. Kashimura, S. Morimoto and E. Osawa, "Origin of Regioselectivity in O-Methylation of Erythromycin with the Aid of Computational Conformational Space Search", *J. Chem. Soc., Perkin Trans. 2*, 1647-1654, 1993.

[4] H. Goto, unpublished; For MM2 force field, U. Burkert and N. L. Allinger, "Molecular Mechanics", American Chemical Society, Washington, D. C., 1982.

[5] H. Goto, "A frontier mode-following method for mapping saddle-points of conformational interconversion in flexible molecules starting from the energy minimum", *Chem. Phys. Lett.*, *292*, 254-258, 1998.

[6] H. Goto, "A Revised Nomenclature for the Ring Conformation and a Note on Conformational Distance in Cyclododecane", *Tetrahedron*, *48*, 7131-7144, 1992.