数理モデル化と問題解決 30－6
（2000．6．26）

並列木探索によるタンパク質立体配座解析の階層的アプローチによる拡張

安藤 誠[†] 秋山 泰[††,☆] 松田秀雄[†††]

本研究では，階層的に拡張された並列木探索によるタンパク質立体配座解析のモデルを提案する．本モデルでは，すでに枝刈りが行われた幅の狭い部分探索木を用いて原問題の探索木の再構築を行うため，部分問題において枝刈りされた不要な探索空間は，原問題の探索空間からあらかじめ取り除かれる．よって，本モデルの階層的探索は，存在可能な立体配座を見逃しなく探索できるという強みを犠牲にすることなく，無駄な探索操作を大幅に減らすことができる．本稿では，実際のタンパク質に見られるtyrosine cornerと呼ばれる構造の解析を行い，本モデルの有効性，および並列木探索の効果を実証する．

Hierarchical Approach to Parallel Tree Search for Protein Conformational Analysis

MAKOTO ANDO,[†] YUTAKA AKIYAMA[††,☆] and HIDEO MATSUDA[†††]

We report on an enhanced parallel tree search model for protein conformational analysis. By using this model, the original problem of a conformational analysis can be divided into several subproblems, and the search tree of the original problem can be reconstructed using the results of the subproblem analysis. Because the search spaces obtained from the results of the subproblem analysis do not contain any branches for invalid conformations, we can reduce the search space of the original problem. We demonstrate the effect of the new model by analyzing a tyrosine corner, which is found in natural proteins, using the new version of our system based on our enhanced model. We also report on the parallel version of our system that showed excellent speedups over the sequential version when analyzing the conformations of a tyrosine corner.

1. Introduction

Precise conformational analysis of short proteins (peptides) at atomic resolution has been playing an important role in molecular biology and pharmacology. Our research group has been putting a great deal of emphasis on conformational analysis by using a *parallel tree search model*[1),2)].

In this paper, we report on our enhanced parallel tree search model for protein conformational analysis. We also discuss the experimental evaluation of the new version of our system called *ESCAPE/Hi* (Exhaustive Search system for the Conformational Analysis of PEptides with Hierarchical Search).

In a study of exhaustive search of conformational space, a similar approach was used by Bruccoleri and Karplus[3)]. Because their approach was intended to solve the problems of homology modeling of proteins based on structural similarities, a special restriction called the chain-closure procedure was used in the process of fixing the backbone structure. Although this restriction greatly reduced the search space, it also reduced the range of problems that can be solved by this approach.

Because our model does not depend on any energy calculation, it can analyze conformations even if information on the physicochemical environment around the target peptide, such as polarity and hydropathy, is not precisely and accurately known. In such cases, a system using our model can analyze the conformations more precisely than systems based on other models, such as molecular dynamics, a simulated annealing approach[4)], and side-chain

† 日本鋼管（株）
NKK Corporation, Japan
†† 技術研究組合 新情報処理開発機構
Real World Computing Partnership, Japan
††† 大阪大学大学院 基礎工学研究科
Graduate School of Engineering Science, Osaka University, Japan
☆ 現在，工業技術院 電子技術総合研究所
Presently with Electrotechnical Laboratory, AIST, Japan

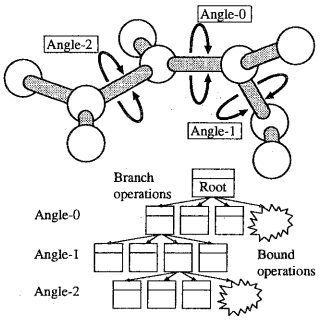

Fig. 1 Search tree for conformational analysis.

Table 1 van der Waals radii.

atom	r(Å)	atom	r(Å)
H (H_N on backbone)	0.5	N	1.6
H (others)	1.2	C	1.7
O	1.45	S	1.8

packing by Dead End Elimination[5].

Our model does not depend on any statistical rotamer information obtained from existing protein structure databases, such as Dunbrack's Rotamer Libraries[6]. A system using our model can analyze peptide conformations, which contain exceptional residues (other than the 20 common amino acids).

2. Overview of Model

2.1 Basic Design Principles of Previous Model

In our model[1),2)], we assume that the atoms are *hard spheres* with certain radii, and the single-covalent bonds are *turnable axes*, each with a fixed length and bond angles. *The conformational isomers* are generated by varying the torsion angles of all the turnable axes according to the search tree (Fig. 1). To detect steric clashes, we use the *van der Waals radii* (Table 1) to see if the distance between any two atoms is too short.

2.2 Hierarchical Search

We added a new feature called *hierarchical search* to reduce the search space and to expand the range of problems that our model can analyze.

The original problem of a conformational analysis can be divided into several subproblems. Our model does not put any restriction on how a problem is divided: it can be divided into amino acid residues, a backbone and side-chains, *etc.*

The subproblems are analyzed individually, and the results are stored in text files (*conformation files*) on file systems for reuse (left of Fig. 2).

The original search tree is reconstructed using

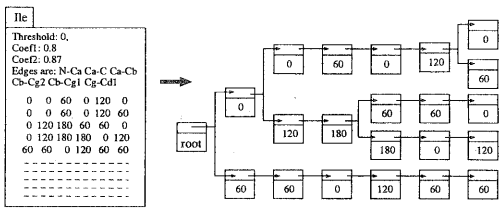

Fig. 2 Reconstruction of search space from subproblem analysis results

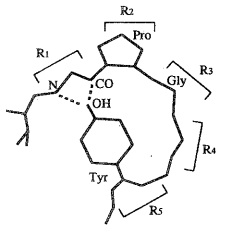

Fig. 3 The Δ 4 tyrosine corner. Dotted lines represent hydrogen bonds.

the torsion angles stored in the files. Because the reconstructed tree does not contain any branches for invalid conformations, we can reduce the search space. For example, using the file in Fig. 2, the system does not have to create nodes for 60, 180, 240, and 300° for the second torsion angle (for a rotation interval set to 60°) when the first torsion angle is 0.

3. Results and Discussion

3.1 Division of Problems

To study the effects of the hierarchical search, we compared two different approaches; dividing a problem into (1) amino acid residues, and (2) a backbone and side-chains.

In section 3.3, we compared the effects of dividing a problem into (1) amino acid residues, and (2) a backbone and side-chains by analyzing a peptide segment of five amino acid residues. The conformation files generated in section 3.2 were used in the analysis in section 3.3.

3.2 Conformation Files for Amino Acid Residues

We generated the conformation files for all the amino acid residues[7]. The conformation files for Asp, Glu, Asn, Gln, Arg, Pro, Gly, Ala, and Tyr[7] were used to analyze the conformations of tyrosine corners in section 3.3.

3.3 Analysis of Tyrosine Corner

To study the effects of dividing a problem into (1) amino acid residues and (2) a backbone and

Table 2 Distance constraints.

atoms	range (Å)
N of R_1 ↔ O of Tyr	0.3-6.5
O of R_1 ↔ O of Tyr	0.3-6.5

side-chains, we analyzed a peptide segment of five amino acid residues called a *tyrosine corner*[7]. A tyrosine corner is a conformation in which a tyrosine (residue R_5) near the beginning or end of an antiparallel β 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 (Fig. 3). The "Δ4 tyrosine corner" is the most common - R_2 is Pro and R_3 is Gly. R_1 is a hydrophilic amino acid residue. We put the five different residues to R_1: (1) DPGAY (Asp-Pro-Gly-Ala-Tyr), (2) EPGAY (Glu-Pro-Gly-Ala-Tyr), (3) NPGAY (Asn-Pro-Gly-Ala-Tyr), (4) QPGAY (Gln-Pro-Gly-Ala-Tyr), and (5) RPGAY (Arg-Pro-Gly-Ala-Tyr).

The demonstration of the sequential version of ESCAPE/Hi was performed on a Sun EnterPrise 4000 (250 MHz UltraSPARC II, Solaris 2.6). The common backbone structure of the peptide segments has 11 turnable axes. Its two distance constraints[7] are shown in Table 2. The conformation file generated for the backbone (*i.e.*, collection of valid conformations of the backbone) has 1,716 entries. It took 526 seconds to analyze these conformations.

Table 3 shows the number of nodes in the search tree created during the analyses and the elapsed times in seconds. The label "No conformation files" indicates the cases where no conformation files were used for analyses. For the "AA (Amino Acid) conformation files" label, we divided the problems into (1) amino acid residues, and used the corresponding conformation files for each peptide segment: Asp, Glu, Asn, Gln or Arg for R_1, Pro for R_2, Gly for R_3, Ala for R_4, and Tyr for R_5. The "Backbone conf. file" label indicates the analyses done with the common backbone conformation file; the problems are divided into (2) a backbone and side-chains. The number of turnable axes in the segments is also shown. Two distance constraints (Table 2) were used to eliminate the conformations that did not form tyrosine corners (*i.e.*, no hydrogen bonds). Although the typical distance between two atoms that make a hydrogen bond is 2.9±0.1 Å, we set the upper limit of the constraint to 6.5 Å to compensate for the quantization error.

For all the cases, both the number of nodes and

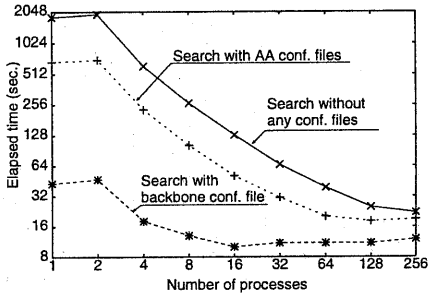

Fig. 4 Elapsed time.

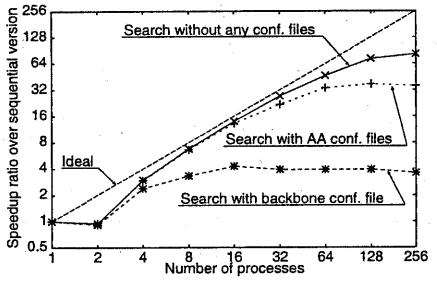

Fig. 5 Speedup of parallel version over sequential one.

the elapsed times were reduced by using the conformation files. These effects were especially notable when using the backbone conformation file. The subproblems which contained distance constraints were effective for reducing the elapsed times.

3.4 Speedup by Using Parallel Version

In our model, we use the *master-worker approach*[7] for dynamic load balancing among multiple processors. The master process distributes the nodes in the search tree to the worker processes by message passing (MPI). To prevent degradation of performance due to the frequent communications, each worker expands nodes several times locally (*e.g.* 1024 times), then reinserts the chunk of created nodes into the master process.

We analyzed the DPGAY tyrosine corner (Asp-Pro-Gly-Ala-Tyr) using the parallel version of ESCAPE/Hi. The analysis was performed on a Hitachi SR2201 (150 MHz PA-RISC 1.1 + PVP-SW, 256 processing units, HI-UX/MPP). We set coef1[7] to 0.85, coef2[7] to 0.90, and the rotation interval of all turnable bonds to 60°. The two distance constraints of Table 2 were used. Each worker process expanded a node locally 1024 times or until all the nodes were pruned, whichever happened first.

Fig. 4 shows the elapsed times for the analyses of the DPGAY tyrosine corner. Fig. 5 shows the observed speedups of the parallel version over the sequential version for up to 256 processes. Because the number of processes includes the master pro-

Table 3 Effects of reducing the number of nodes created and elapsed times.

peptide	No conformation files		AA conformation files		Backbone conf. file	
	# of nodes created	elapsed time (sec.)	# of nodes created	elapsed time (sec.)	# of nodes created	elapsed time (sec.)
DPGAY (15)	566,765	693	156,434	292	12,046	16
EPGAY (16)	205,293	244	58,237	109	8,838	12
NPGAY (15)	213,758	254	56,853	108	2,970	4
QPGAY (16)	194,773	229	55,465	105	7,998	11
RPGAY (19)	330,632	431	152,493	283	158,022	234

Numbers after peptide names are the total tunable axes.
coef1[7] = 0.85, coef2[7] = 0.90, rotation interval = 60°

cess, the parallel version using two processes took almost the same amount of time as the sequential one did because only one worker process performed the tree search operations.

Fig. 4 and 5 show that the problem of a tyrosine corner has enough parallelism for 64 or 128 processes to be used when the problems are analyzed with amino acid conformation files. These speedup figures also show that the search space of the problem became narrow (had low parallelism) when using the backbone conformation file, and that ESCAPE/Hi can solve larger problems by using the backbone conformation files.

4. Conclusion

We demonstrated that conformational analysis by using an exhaustive parallel tree search has a strong advantage over other methods in the analysis of short peptides precisely and accurately. Moreover, the hierarchical search feature of our model effectively reduces the search space (number of nodes created for the search operations) by dividing a problem into several subproblems and reconstructing the search tree of the original problem by using the results of the subproblem analysis.

The effective use of conformation files improved the performance of ESCAPE/Hi. With ESCAPE/Hi, we analyzed a tyrosine corner, an important peptide segment for understanding the protein folding. For the analysis done with amino acid conformation files and backbone conformation files, ESCAPE/Hi effectively reduced the search space and generated valid conformations within several seconds. Subproblems which had distance constraints were especially effective for reducing the search space and elapsed times for our analyses. The parallel version of ESCAPE/Hi showed excellent speedups over the sequential one.

We are planning to generate and evaluate the conformation files for "segments of two amino acids" (400 files, 20 x 20), which we believe as effective at reducing the search space and elapsed times for analyses.

References

1) Ando, M., Akiyama, Y., Onizuka, K. and Noguchi, T.: ESCAPE: Parallel Tree Search System for Conformational Analysis of Peptides, *Proc. the 1999 International Conference on Parallel and Distributed Processing Techniques and Applications*, pp. 1537–1543 (1999).

2) Ando, M., Akiyama, Y. and Matsuda, H.: ESCAPE 2.0: Parallel Exhaustive Conformational Search System of Peptides with Conformational Libraries as Building Blocks, *Proc. the 4th International Conference on High-Performance Computing in Asia-Pacific Region*, pp. 1097–1102 (2000).

3) Bruccoleri, R. and Karplus, M.: Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling, *Biopolymers*, Vol. 26, pp. 137–168 (1987).

4) 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).

5) Desmet, J., Maeyer, M. D., Hazes, B. and Lasters, I.: The dead-end elimination theorem and its use in protein side-chain positioning, *Nature*, Vol. 356, pp. 539–542 (1992).

6) Bower, M., Cohen, F. and Dunbrack, R. Jr.: Prediction of Protein Side-chain Rotamers from a Backbone-dependent Rotamer Library: A New Homology Modeling Tool, *J. Mol. Biol.*, Vol. 267, pp. 1268–1282 (1997).

7) Ando, M., Akiyama, Y. and Matsuda, H.: Hierarchical Approach to Parallel Tree Search for Protein Conformational Analysis, *Proc. the 2000 International Conference on Parallel and Distributed Processing Techniques and Applications* (2000) (to appear).