

階層的な大規模並列木探索による タンパク質立体配座解析システム ESCAPE/Hi

安 藤 誠^{†,††,†††} 秋 山 泰^{††} 松 田 秀 雄^{†††}

本論文では、実用的なペプチドの立体配座を大規模並列計算機上で網羅的かつ高速に解析するための、階層的な並列木探索による新しいタンパク質立体配座解析手法を提案する。提案する手法では、マスターワーカーモデルに基づく並列処理の導入によって、探索木に広がりがある場合でも高速な解析を可能にした。また、原問題をいくつかの部分問題に分解し、部分問題の解析結果から原問題の探索木を構築することで、立体配座を網羅的に探索できるという強みを犠牲にすることなく、無駄な探索操作の大幅な減少を実現した。また、本論文では、大規模並列計算機上に実装したシステム ESCAPE/Hi を用いて、実際のタンパク質に多く見られるチロシン・コーナーと呼ばれる構造の解析を行い、作成ノード数削減や実行時間短縮における新しい手法の有効性を実証する。さらに、探索木の枝の広がり方が不均一であるにもかかわらず、動的な負荷分散の採用により、128 プロセッサ使用時に最高で約 72 倍の高速化が達成されたことを示す。

ESCAPE/Hi: A Massively Parallel Protein Conformational Analysis System Using Hierarchical Tree Search

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 infeasible 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 report on the parallel version of our system based on a master-worker approach that showed 72-fold speedup over the sequential version when using 128 processing units.

1. はじめに

タンパク質は、20 種類のアミノ酸が一次的につながった鎖状の高分子で、生体のさまざまな生理現象に関わっている¹⁾。多様な長さを持つタンパク質の中でも、短いタンパク質、あるいはタンパク質中の短い断片（以下では、まとめてペプチドと呼ぶ）は、酵素活

性のしくみの理解や薬物の設計に深い関わりがあり²⁾、我々はこれまでもペプチドの構造予測に取り組んできた^{3),4)}。

ペプチドは、一重の共有結合が持つねじれの自由度により、理論的には無数の形状（立体配座）をとれるが、それ自身および周囲の溶媒・他分子との非共有結合力（クーロン力、水素結合、疎水相互作用など）により、それぞれがある決まった形状をとる。したがって、ペプチドとその環境を構成する原子の間で力場計算を行い、最小のエネルギーを持つ構造を求める研究がこれまでに多くなされてきた^{5),6)}。しかし、細胞の受容体や酵素といった巨大な分子を含む系での力場計算は複雑であり、また力場計算による構造最適化を行うための前段階としての初期構造の探索が必要である。そのため、ペプチドの一重結合を系統的に回転させ、と

[†] 日本鋼管株式会社情報システム部
Information Systems Department, NKK Corporation

^{††} 産業技術総合研究所生命情報科学研究センター
Computational Biology Research Center (CBRC),
AIST

^{†††} 大阪大学大学院基礎工学研究科情報数理系専攻
Department of Informatics and Mathematical Science,
Graduate School of Engineering Science, Osaka
University

りうる構造を網羅的に探索するアプローチが注目されている^{7)~9)}。しかし、従来のいずれの手法も、共有結合のねじれ角のサンプリングについては、計算量の低減のためにタンパク質立体構造データベース PDB¹⁰⁾からの統計情報に基づいた代表値を使用しており、経験に基づかない真の意味での網羅的探索とはいえない。

そこで本論文では、上記の問題点の克服のために、階層的な並列木探索によるタンパク質立体配座解析の手法を提案する。本論文で提案する手法では (I) 大規模並列計算機上での木探索処理の並列化と (II) 階層的探索による探索空間の縮小化により、従来の手法よりもさらに網羅的な探索を目指す (I) 探索処理の並列化では、マスター-ワーカモデルに基づく並列木探索によって、大規模計算機上での効果的な並列木探索処理を実現する (II) 階層的探索では、与えられた問題をあらかじめいくつかの小さな部分問題に分解し、部分問題の探索によって得られた立体配座に関する情報を利用して、無駄な探索空間が削除された新しい探索木を構築する。

本論文では、2章で階層的な並列木探索によるタンパク質立体配座解析の手法について述べ、3章で新しい手法に基づいたシステム ESCAPE/Hi の実装について述べる。4章では ESCAPE/Hi によって得られた実験結果を示し、新しい手法の有効性を議論する。5章で結論を述べる。

2. 階層的な並列木探索による立体配座解析

2.1 木探索による立体配座解析の問題点

木探索による立体配座解析の手法^{3),4)}では、回転可能な結合 (回転軸) にあらかじめ付けられた優先度 (図1の rank) に従って、すべての回転軸を系統的に回転させる。この操作は、図1の下図に示された探索木によって表現できるが、探索の途中で生成されるノード数は、探索木の深さ、つまり回転軸の総数の増加に対して指数的に増加する。このため、生成されるノード数が増加した場合、実際のプログラムでは以下に列挙する問題が発生する。

- (1) プログラム実行中のノードの生成・削除は、メモリ空間における領域の割当て・解放に該当するため、ノードの頻繁な生成・削除はプログラムの実行速度を低下させる。
- (2) 個々のノードは、解析対象のタンパク質におけるすべての二面角 (結合のねじれ角¹¹⁾) の状態を格納しており、生成されるノード数の増加は、必要とするメモリ空間の増大を意味する。

以下では、これらの問題を解決するための方法につ

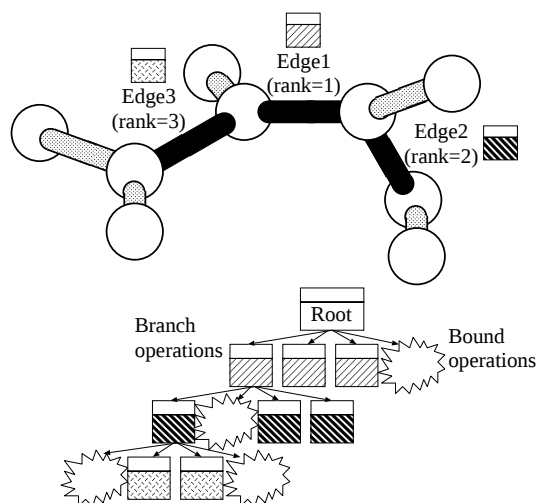


図1 モデル化されたタンパク質と探索木

Fig.1 The protein structure model and the search tree.

いて述べ、大規模並列計算機上に実装されたシステムを用いて、それらの効果を評価する。

2.2 工夫1: 木探索の並列処理

処理すべきノード数の増加や必要メモリ空間の増大に対しては、大規模並列計算機上において木探索処理を複数のプロセッサに分散させることが効果的であると考えられる。ただし、立体配座解析の探索木においては、van der Waals 半径を用いた立体障害の検出、および距離制約・角度制約の検査によって探索空間が大きく枝刈りされるため^{3),4)}、探索木の広がり方に偏りが生じることが多い。このため、探索木を静的に空間分割する並列処理方式ではタスクの公平な分配が困難であり、タスクプール方式などの動的な負荷分散機能を持った並列処理方式の導入が必須である。

そこで本研究では、過去の並列木探索の研究¹²⁾において実績のあるマスター-ワーカモデルに基づいた並列木探索処理機構を採用した。マスタプロセスの内部に、処理されるべきノードがプールされるようにした。ワーカプロセスは、マスタプロセスから受け取った1つのノードに対し分枝限定操作を行い、その結果生成されたノードをマスタのプールに返却する。この操作をノードのプールが空になるまで繰り返し行うことで、すべてのワーカに公平にノードを分配する。

2.3 工夫2: 立体配座ファイルを用いた階層的な探索

本論文で提案する新しい手法では、あらかじめ分解しておいた部分問題の解析結果を利用して、原問題の探索木を再構築する。分解されたそれぞれの部分問題の解析結果は立体配座ファイルに格納される。図2の

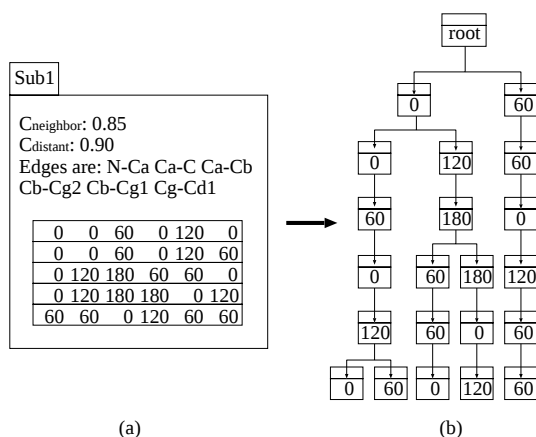


図2 (a) 立体配座ファイルと (b) 立体配座ファイルから生成された部分探索木

Fig. 2 (a) Conformation file and (b) the subtree obtained from the conformation file.

(a) に示すように、立体配座ファイルでは、1つの行が、部分問題における1つの存在可能な立体配座（ねじれ角で表現）を表している（図の例では5つの立体配座が含まれている）。立体配座ファイルから生成される新しい探索木（図2の(b)）には、部分問題の探索の際に生成され枝刈りされた無駄なノードは含まれていないので、この探索木を用いて原問題の探索木を再構成すれば、探索空間がより小さい探索木を構成することができる。

なお、一度生成された立体配座ファイルは、各原子の半径や、回転軸の角度刻みなどの解析条件が同じである限り、幅広く再利用することが可能である。よって、あらかじめ典型的な問題分割の手法を決めておき、それに対応する立体配座ファイルを作成しておけば、それらをライブラリとして繰り返し利用できることになり、効率の良い高速解析が実現できる。

2.4 問題の分割方法

タンパク質は、20種類のアミノ酸が任意個、任意の順序で結合した分子であるため、それぞれのアミノ酸を、タンパク質を構築するブロックとしてとらえることができる。したがって、アミノ酸1残基、あるいは、実際のタンパク質に高い頻度で見られる2~3残基の組合せパターンを立体配座ファイルとしてライブラリ化しておくことは、ごく自然な考え方である。なかでも、アミノ酸の側鎖¹³⁾と呼ばれる部分に環状構造を含むもの（図3のチロシンなど）や、側鎖の根本付近に枝分かれがあるイソロイシン（図3）などは、複数組み合わせることでより大きな立体障害が起こり、立体配座ファイルとして効果的であると予想される。

一方で、タンパク質には水素結合が随所に見られ、

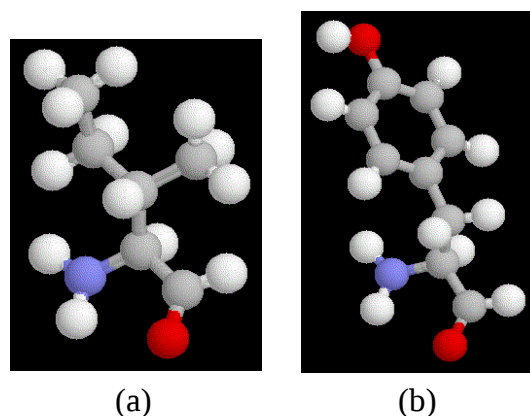


図3 (a) イソロイシンと (b) チロシンの構造
Fig. 3 Structures of (a) an Ile and (b) a Tyr.

タンパク質構造、特に主鎖¹³⁾の構造を決める一因であると考えられている。我々の手法では、水素結合によって位置関係が固定された原子対の距離を制約としてモデル化することができる。立体配座ファイルを生成するための部分問題には、できるだけ多くの枝刈り処理が含まれているほうが好ましいが、水素結合を含む主鎖の構造を部分問題とした立体配座ファイルは、距離制約による枝刈りが多く含まれることになる。よって、主鎖と側鎖を分割し、主鎖部分を立体配座ファイルとしてライブラリ化することは効果的であると予想される。また、主鎖と側鎖の分割は、ある決められた主鎖構造に対して種々の側鎖を配置する問題¹⁴⁾にも適用可能である。

2.5 関連研究

ペプチドの立体配座を網羅的に探索するアプローチとして、Brucoleriらの手法¹⁵⁾が有名である。この手法では、“chain-closure procedure”¹⁶⁾と呼ばれる手続きを導入している。この手続きは、与えられた2つの点（座標）を結ぶペプチドの立体配座を高速に探索するものであるが、探索空間のサイズの低減に大きく寄与している一方、彼らの手法の適用範囲をホモロジーモデリングの問題への適用に限定してしまうという欠点を生み出している。

Dunbrackらの backbone dependent rotamer library^{17),18)}は、存在可能な側鎖の立体配座のライブラリであり、これは我々の立体配座ファイルをアミノ酸1残基単位で作成した場合と類似している。しかし、彼らのライブラリは、タンパク質立体構造データベースPDBからの統計情報から作成されており、網羅的探索によって得られた我々のライブラリはより広い探索空間をカバーしていると考えられる。

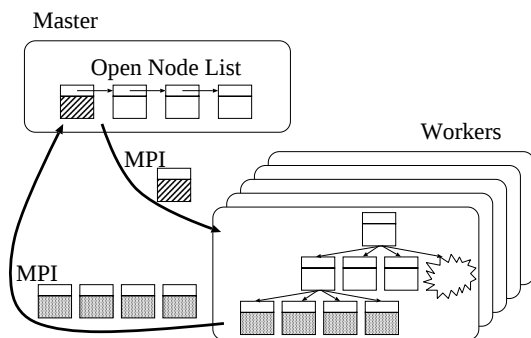


図4 マスターワーカーモデル
Fig. 4 Master-Worker model.

3. ESCAPE/Hiの実装と並列化

新しい手法に基づいたシステム *ESCAPE/Hi* (Exhaustive Search system for the Conformational Analysis of PEptides with Hierarchical Tree Search)の実装には、*PAPIA* ライブラリ^{19),20)}を利用した。*PAPIA* ライブラリは、タンパク質の情報解析に用いられる共通のプログラムモジュールを、C++のクラス階層としてライブラリ化したものである。3次元空間での原子の位置変換（並行移動、回転）、類似構造検索、配列アライメント、多変量解析などのよく利用される計算は、対応するクラスのメンバ関数として実装されている。*ESCAPE/Hi*においては、原子をある回転軸を中心に指定した角度だけ回転させる操作を頻繁に行うが、これらはすべて *PAPIA* ライブラリのメンバ関数を組み合わせることで容易に記述できるため、プログラム開発の効率化が実現された。

*ESCAPE/Hi*の並列処理機構については、2.2節で述べたとおり、マスターワーカーモデルを採用した(図4)。マスタはシステム内につねに1つ存在し、探索木の管理を行う。マスタ内のオープンノードリストは、探索木の葉（探索の各時点での最末端部分）を一次元のリスト構造で表現したものである。ワーカプロセスは、通信頻度の低減のために、マスタから得たノードに対して分枝操作と限定操作を指定された回数（たとえば1,024回）だけローカルに行い、新たに生成されたノード群をマスタ上のオープンノードリストに戻す。この操作を、オープンノードリストが空になるまで続ける。この方法により、探索木を静的に空間分割する方法に比べ、均一な負荷分散が可能になる¹²⁾。

4. 実験結果と考察

原問題の分割方法については、2.4節で(1)アミノ酸残基単位の分割と(2)主鎖と側鎖の分割という2

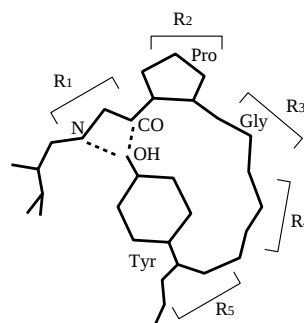


図5 Δ4チロシン・コーナー
Fig. 5 The Δ4 tyrosine corner.

通りの方法を提案した。本章では、それらの効果の比較を行うために、チロシン・コーナー²¹⁾と呼ばれるアミノ酸5残基のペプチドの解析を行った。

4.1 チロシン・コーナー

チロシン・コーナーは、反平行のβシート¹³⁾の始点あるいは終点付近にあるチロシン(Tyr, 図5のR₅)の側鎖の水酸基(OH)が、R₁の位置にある残基の主鎖の窒素(N)や酸素(O)と水素結合して作られる立体配座である。このようにタンパク質の主鎖部分が大きく折れ曲がった部分はターン構造¹³⁾と呼ばれ、タンパク質のフォールディングに大きく寄与する重要な部分である。Δ4チロシン・コーナーが最も一般的で、R₁には親水性残基、R₂にはPro、R₃にはGlyが配置されることが知られている。本論文では、上記の5残基だけを解析したときに、期待される構造（つまり、ターン構造）が解析結果の中に含まれているかを検証する。今回の実験では、R₁を5種類のアミノ酸残基で置換する：(1)DPGAY (Asp-Pro-Gly-Ala-Tyr) (2)EP-GAY (Glu-Pro-Gly-Ala-Tyr) (3)NPGAY (Asn-Pro-Gly-Ala-Tyr) (4)QPGAY (Gln-Pro-Gly-Ala-Tyr)、および(5)RPGAY (Arg-Pro-Gly-Ala-Tyr)である。

4.2 立体配座ファイルの生成

表1は、20種類のアミノ酸立体配座ファイルと1種類的主鎖立体配座ファイルの生成の概要を示している。厳密には、主鎖立体配座ファイルには、主鎖部分の構造だけでなく、R₅のTyrの側鎖も含んでいる。これは、主鎖のR₁とTyrの側鎖の間に距離制約(表2)を設定するためである。生成にあたっては、精密な解析結果を得るために、水素原子を考慮に入れた解析を行った。*ESCAPE/Hi*では、タンパク質を構成する原子を、ある決められた半径を持つ硬い球とモデル化するが、各原子の半径として表3のvan der Waals半径を使用した。表1の総組合せ数C_tは、回転軸数をN、

表 1 生成された立体配座ファイルの概要
Table 1 Outline of the generated conformation files.

残基種	回転軸数	総組合せ数 C_t	生成配座数 C_g	$C_g/C_t(\%)$	実行時間 (秒)
Pro	1	6	4	66.7	1.00e-2
Gly	2	36	36	100	1.00e-2
Ala	3	108	102	94.4	2.00e-2
Phe	4	864	156	18.1	1.40e-1
Cys	4	1,296	480	37.0	1.10e-1
Trp	4	1,296	252	19.4	2.80e-1
His	4	1,296	228	17.6	1.70e-1
Ser	4	1,296	678	52.3	1.30e-1
Val	5	1,944	264	13.6	1.80e-1
Thr	5	3,888	444	11.4	1.90e-1
Tyr	5	7,776	528	6.79	3.40e-1
Asn	5	7,776	2,742	35.3	6.60e-1
Asp	5	7,776	2,832	36.4	5.90e-1
Leu	6	11,664	960	8.23	4.90e-1
Ile	6	11,664	582	4.99	6.10e-1
Met	6	23,328	2,826	12.1	1.39
Gln	6	46,656	11,982	25.7	3.01
Glu	6	46,656	12,402	26.6	2.98
Lys	7	279,936	3,216	1.15	2.61
Arg	9	10,077,696	317,394	3.15	125
主鎖	11	362,797,056	1,716	4.73e-4	526

角度刻み = 60 度

表 2 設定した距離制約
Table 2 Distance constraints.

原子	制約 (Å)
R_1 の N $\leftrightarrow$ Tyr の O	0.3-6.5
R_1 の O $\leftrightarrow$ Tyr の O	0.3-6.5

表 3 使用した van der Waals 半径
Table 3 van der Waals radii.

原子種	r (Å)	原子種	r (Å)
H (主鎖の N)	0.5	N	1.6
H (上記以外)	1.2	C	1.7
O	1.45	S	1.8

各回転軸のねじれのサンプル数を s_i ($1 \leq i \leq N$) とした場合の, $\prod_{i=1}^N s_i$ の値である. 表 1 に掲載されたファイルのうち, Asp, Glu, Asn, Gln, Arg, Pro, Gly, Ala, Tyr, および主鎖立体配座ファイルについては, 4.3 節の解析に利用した.

表 1 の解析を含め, ESCAPE/Hi の逐次版の実行は, Sun Enterprise 4000(250 MHz UltraSPARC II, Solaris 2.6) で行った.

4.3 階層的木探索による探索空間の広がり抑制

表 4 は, 探索の最中に作成されたノードの数と, 実行時間を掲載したものである. 表中のファイルなしラベルは, 立体配座ファイルをいっさい利用しなかった場合である. アミノ酸残基ラベルは, 原問題を (1) アミノ酸残基単位で分割した場合で, 4.2 節で生成したアミノ酸立体配座ファイル(表 1)を利用している. ま

た, 主鎖ラベルは (2) 主鎖と側鎖の分割に関する実験で, 4.2 節で生成した主鎖立体配座ファイル(表 1)を利用している. 作成ノード数および実行時間の低減率は, 立体配座ファイルをいっさい利用しなかった場合の値に対する減少効果を示したものである. 表 2 に掲載された距離制約を用いて, チロシン・コーナーを形成しない立体配座を枝刈りしている.

表 4 に示したとおり, すべての実験において, 立体配座ファイルを利用することで, 作成されたノード数および実行時間を低減させることに成功した. この効果は, 主鎖立体配座ファイルを利用した場合に顕著である. 主鎖立体配座ファイルを利用しない場合, 距離制約による枝刈りは, 制約がかけられている 2 原子の相対的位置関係が決まって初めて行われるので, 探索開始から 2 原子の相対的位置関係が決まるまでの間, 制約を満たさないかなりの数の無駄なノードが作成されることになる. そして, 位置関係が決まった時点で, 大規模な枝刈りが起こる. 一方, 主鎖立体配座ファイルを利用した場合には, これら大量の無駄なノードはあらかじめ削除されているので, 探索空間の広がりが抑制できる. このことより, 距離制約を含む立体配座ファイルが, 作成ノード数および実行時間の低減により効果的であると考察される. ただし, 解析に必要な計算量は, パラメータ(距離制約など)のとり方の組合せで大きく変わるため, 本論文の例題(5 残基程度)より大きな問題を解析するためには, いかにうまく制

表 4 作成ノード数および実行時間減少の効果

Table 4 Effects of reducing the number of nodes created and elapsed times.

ペプチド	回転軸数	ファイルなし		アミノ酸残基			
		作成ノード数	時間 (秒)	作成ノード数	低減率	時間 (秒)	低減率
DPGAY	15	566,765	693	156,434	27.6%	292	42.1%
EPGAY	16	205,293	244	58,237	28.4%	109	44.7%
NPGAY	15	213,758	254	56,853	26.6%	108	42.5%
QPGAY	16	194,773	229	55,465	28.5%	105	45.9%
RPGAY	19	330,632	431	152,493	46.1%	283	65.7%

ペプチド	主鎖				配座数
	作成ノード数		時間（秒）		
		低減率		低減率	
DPGAY	12,046	2.13%	16	2.31%	7,824
EPGAY	8,838	4.31%	12	4.92%	5,724
NPGAY	2,970	1.52%	4	1.57%	1,272
QPGAY	7,998	4.11%	11	4.80%	5,400
RPGAY	158,022	47.8%	234	54.3%	108,408

角度刻み=60度.

作成ノード数および実行時間の低減率は,
「ファイルなし」の値を基準としたもの.

約条件を見つけて、枝刈りを行うかも重要である.

なお、表中の「配座数」は、解として得られた立体配座の数である。いずれの実験においても、アミノ酸残基立体配座ファイルや主鎖立体配座ファイルを利用した場合と、立体配座ファイルを利用しない場合とで、得られた立体配座数は一致した。

4.4 並列処理による解析速度の向上

DPGAY チロシン・コーナー (Asp-Pro-Gly-Ala-Tyr) の解析における、ESCAPE/Hi の並列版による速度向上について考察する。実験は、Hitachi SR2201 (150 MHz PA-RISC 1.1 + PVP-SW, 256 プロセッサユニット, HI-UX/MPP) で行った。角度刻みなどの設定はこれまでの実験と同じ条件で、表 2 に掲載した 2 種類の距離制約を設定した。ワーカプロセスは、マスタから得たノードについて分枝操作と枝刈りを指定された回数だけローカルに行うが (3 章参照), その回数を 1,024 に設定した場合に最も良い性能が得られたので、今回は 1,024 に設定して以下の評価を行った。

図 6 は、DPGAY チロシン・コーナーの解析に要した実行時間を示したものである。また、図 7 は、逐次版の実行時間を基準にしたときの、並列版使用による速度向上の比を示したものである。なお、図 6 および図 7 における横軸 (プロセスの個数) にはマスタプロセスも含めている。よって、逐次版で実行した場合と、並列版で 2 プロセスを使用した場合とは、双方とも 1 プロセスのみが木探索処理を行うため、両者の実行時間は、ほぼ同じとなる。

これらの図から、同じ問題を解く場合でも、立体配

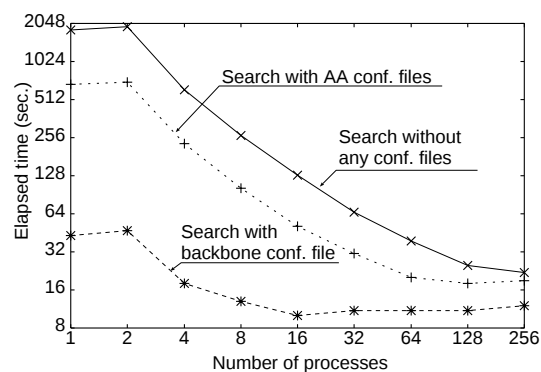


図 6 実行時間。「AA conf. files」は、アミノ酸残基立体配座ファイルを利用した解析を表す

Fig. 6 Elapsed time. "AA conf. files" stands for "Amino Acid conformation files".

座ファイルを利用するか否かの違い、および問題の分割方法の違いによって、探索空間の広がり方に著しい違いが生じることが分かる。立体配座ファイルを利用しない場合では、128 プロセス使用時に約 72 倍の高速化を達成しており、大規模並列計算機の持つ性能を有効に利用できるだけの並列性が探索木に存在していると考えられる。よって、枝刈りや立体配座ファイルといった理論的な手法での探索空間の縮小がうまくいかない場合でも、並列処理によって十分な高速化が可能になることが示せた。

一方、主鎖立体配座ファイルを利用した場合には、探索空間の幅方向の広がり方が抑えられ、探索空間に並列性がほとんどなくなってしまうと考えられる。こ

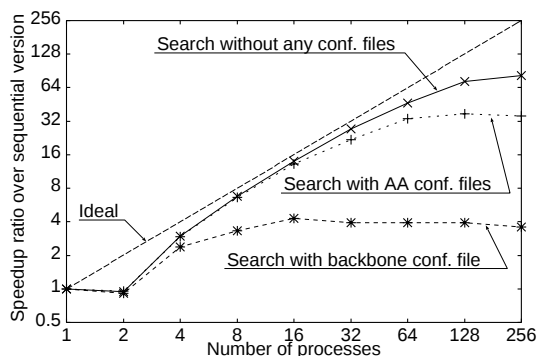


図7 逐次版に対する並列版の速度向上比
Fig.7 Speedup of parallel version over sequential one.

のことから逆に、立体配座ファイルの効果的な利用は、探索空間の広がりをも十分に抑制しており、回転軸をより多く含む問題を解析可能であることを示していると考察される。

4.5 チロシンコーナーの既知構造との比較

ここでは、解析に使用したペプチドのうち、X線結晶解析法によってその構造が明らかにされているQP-GAY (Gln-Pro-Gly-Ala-Tyr) について、ESCAPEによって生成された立体配座とX線解析法で得られている構造(既知構造)との重ね合わせを行ってみた。

表4に示したとおり、角度刻みを60度にした場合、QPGAYは5,400個の立体配座が有効であることが分かった。これらの構造には、分子力場などによる構造最適化の処理は行っていない。また、5,400個の立体配座を互いに重ね合わせたと、5,400個のバリエーションはすべて側鎖のばらつきによるもので、主鎖の立体配座は、角度刻みが60度の場合には1通りしか見つからなかった。

この主鎖の立体配座を、PDBの1PAZの70-74残基目と重ね合わせを行ったところ、主鎖の C_{α} による重ね合わせの場合にはRMSD (Root Mean Square Deviation) = $0.93\text{\AA}$ であり、ほぼ同じ立体配座をとっているものと考えられる。また、PDB 1PAZにおけるTyrの χ_1 は g^- 配置(-60 度付近)をとっているが、ESCAPEによって得られた5,400個の立体配座についても、すべてが g^- 配置をとっていることが分かった。ESCAPEによる5残基だけの立体配座解析によって、チロシンコーナーの特徴ある構造を導くことができた。図8は、既知構造との全原子(水素以外)による重ね合わせにおいてRMSDが最小である立体配座(RMSD = $1.69\text{\AA}$)を、既知構造と重ね合わせた様子を示している。

ESCAPEによって得られた結果に対して、分子力

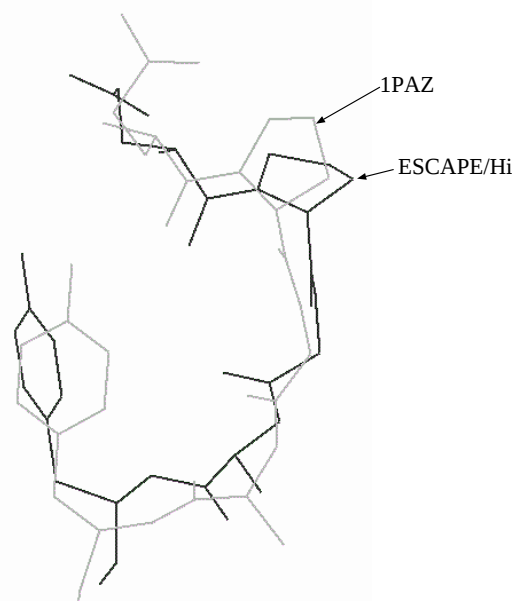


図8 1PAZとESCAPE/Hiによる解析結果の重ね合わせ
(insightIIによる描画)
Fig.8 Superposition of 1PAZ and a result obtained from ESCAPE/Hi (illustrated by insightII).

場などによる構造最適化の処理を行うことで、より詳細な構造解析を行うことができる。つまり、ESCAPEは、構造最適化の前段階としての初期構造を高速に生成することに利用できると考えられる。

5. おわりに

本論文では、階層的な並列木探索による立体配座解析の手法を提案した。新しい手法は、次のような工夫によって、タンパク質の立体配座を網羅的かつ高速に解析することを可能にした。

- 工夫1: マスター-ワーカモデルに基づく、動的な負荷分散機能を持った並列木探索処理の導入
- 工夫2: 立体配座ファイルを用いた階層的な木探索による、作成ノード数の低減

新しい手法の有効性を示すために、チロシン・コーナーと呼ばれるペプチドの解析を行い、以下に列挙する効果が得られた。

- すべての実験において、立体配座ファイルを利用することで、作成されたノード数および実行時間を低減させることに成功した。
- 特に、部分問題に距離制約を含む場合など、部分問題の探索木で大きな枝刈りが起こる場合には、作成ノード数の減少および実行速度の向上の面で立体配座ファイル利用の効果は著しく、本研究の

所期の目標を達成できた。

- 木探索処理の高速化については、立体障害の検出や、距離制約の検査、立体配座ファイルの利用といった理論的な手法での探索空間の縮小に加え、探索木に広がりがある例でも、マスターワークモデルによる並列処理によって、処理速度の向上を達成することができた。

参考文献

- 1) Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.: *MOLECULAR BIOLOGY OF THE CELL*, 3rd edition, Garland Publishing, Inc. (1994).
- 2) 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).
- 3) 安藤 誠, 秋山 泰, 鬼塚健太郎, 野口 保: 木探索アプローチによるタンパク質立体配座解析と大規模並列計算機上での高速解析システムの構築, 情報処理学会論文誌: 数理モデル化と応用, Vol.40, No.SIG2 (TOM1), pp.91–104 (1999).
- 4) Ando, M., Akiyama, Y., Onizuka, K. and Noguchi, T.: ESCAPE: Parallel Tree Search System for Conformational Analysis of Peptides, *Proc. 1999 International Conference on Parallel and Distributed Processing Techniques and Applications*, pp.1537–1543 (1999).
- 5) 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).
- 6) Okamoto, Y.: First-Principles Protein Folding Simulations, *Molecular Simulation*, Vol.24, pp.351–368 (2000).
- 7) Moult, J. and James, M.: An Algorithm for Determining the Conformation of Polypeptide Segments in Protein by Systematic Search, *Proteins*, Vol.1, pp.146–163 (1986).
- 8) Sudarsanam, S., DuBose, R., March, C. and Srinivasan, S.: Modeling protein loops using a ϕ_{i+1} , ψ_i dimer database, *Protein Sci.*, Vol.4, pp.1412–1420 (1995).
- 9) Deane, C. and Blundell, T.: A Novel Exhaustive Search Algorithm for Predicting the Conformation of Polypeptide Segments in Proteins, *Proteins*, Vol.40, pp.135–144 (2000).
- 10) Bernstein, F., Koetzle, T., Williams, G., Meyer, E. Jr., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T. and Tasumi, M.: The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures, *J. Mol. Biol.*, Vol.112, pp.535–542 (1977).
- 11) IUPAC IUB Commission on Biochemical Nomenclature 1969: Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains, *Biochemistry*, Vol.9, pp.3471–3479 (1970).
- 12) Ando, M., Tanaka, Y., Kubota, K., Matsuda, M., Akiyama, Y. and Sato, M.: Performance Characterization of Shared- and Distributed-Memory Multiprocessors on a Tree Search Problem, *Proc. 3rd High Performance Computing Asia*, pp.620–629 (1998).
- 13) Schulz, G., Schirmer, R., 大井龍夫 (監訳): タンパク質—構造・機能・進化, 化学同人 (1980).
- 14) 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).
- 15) Brucoleri, R. and Karplus, M.: Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling, *Biopolymers*, Vol.26, pp.137–168 (1987).
- 16) Gō, N. and Scheraga, H.A.: Ring closure and local conformational deformations of chain molecules, *Macromolecules*, Vol.3, pp.178–187 (1970).
- 17) Dunbrack, R. Jr. and Karplus, M.: Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction, *J. Mol. Biol.*, Vol.230, pp.543–547 (1993).
- 18) 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).
- 19) 秋山 泰, 鬼塚健太郎, 野口 保, 安藤 誠, 斎藤 稔: 並列タンパク質情報解析 (PAPIA) システムの PC クラスタ上での実現, 情報処理学会研究報告 97-HPC, Vol.70, No.6, pp.31–36 (1997).
- 20) Akiyama, Y., Onizuka, K., Noguchi, T. and Ando, M.: Parallel Protein Information Analysis (PAPIA) system running on a 64-node PC Cluster, *Proc. 9th Genome Informatics Workshop*, pp.131–140 (1998).
- 21) 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).

(平成 13 年 2 月 14 日受付)

(平成 13 年 4 月 24 日採録)



安藤 誠（正会員）

昭和 42 年生．平成 4 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本鋼管（株）入社，現在に至る．この間，平成 6 年より 2 年間米国コンベックスコンピュータ社（テキサス州，現ヒューレットパッカード社）に派遣．平成 8 年より平成 10 年まで技術研究組合新情報処理開発機構に出向，並列分散システムパフォーマンスつくば研究室研究員．平成 10 年大阪大学大学院基礎工学研究科情報数理系専攻博士後期課程入学，同課程に在学中．現在は，産業技術総合研究所生命情報科学研究センターに勤務し，並列計算機上でのタンパク質立体構造解析等の研究に従事．日本バイオインフォマティクス学会会員．



秋山 泰（正会員）

昭和 36 年生．平成 2 年慶應義塾大学大学院理工学研究科電気工学専攻博士課程修了．工学博士．同年電子技術総合研究所研究官．平成 4 年京都大学化学研究所助教授．平成 8 年新情報処理開発機構並列応用研究室長．平成 12 年電子技術総合研究所生命情報科学ラボ主任研究官．平成 13 年産業技術総合研究所生命情報科学研究センター長，現在に至る．並列計算機を用いたタンパク質立体構造および遺伝子配列情報解析等の研究に従事．電子情報通信学会，日本バイオインフォマティクス学会，日本生物物理学会，分子生物学会，神経回路学会，IEEE 各会員．



松田 秀雄（正会員）

昭和 34 年生．昭和 57 年神戸大学理学部物理学科卒業．昭和 59 年同大学院工学研究科システム工学専攻（修士課程）修了．昭和 62 年同大学院自然科学研究科（博士課程）修了．同年同大学工学部助手となり，同大学講師，助教授を経て，平成 6 年 10 月より大阪大学基礎工学部情報工学科助教授，平成 9 年同大学大学院基礎工学研究科情報数理系専攻助教授（組織変更），現在に至る．この間，平成 3 年 4 月より 10 カ月間米国アルゴンヌ国立研究所客員研究員．学術博士．ゲノムデータのコンピュータ解析等の研究に従事．日本バイオインフォマティクス学会，日本分子生物学会，電子情報通信学会，IEEE CS，ACM 各会員．