

## 並列木探索を用いたタンパク質立体配座解析システムの構築

安藤 誠† 秋山 泰†  
鬼塚 健太郎† 野口 保†

我々は、以前に行った 0-1 ナップサック問題を解く並列プログラムの実装で得られた知見をもとに、より実用的で大規模な木探索問題であるタンパク質立体配座解析問題を解くシステムを構築した。本システムでは、回転可能なすべての共有結合を回転させ、それによって得られるすべての立体配座異性体について実際に存在可能かどうかを調べあげる網羅的手法を採用した。本手法により、アミノ酸 1 残基レベルについて、高精度の解析結果を得ることができた。また、探索の並列化によって速度を向上させることにも成功した。タスク管理の処理と木探索処理を別のプロセスに行わせることによって、処理のオーバーラップによる高速化も実現された。今後は、本成果を階層的なアプローチによって拡張し、創薬分野において重要なアミノ酸 3~7 残基程度の立体配座解析を目指す。

### Parallel Tree Search-based Protein Conformation Analysis System

MAKOTO ANDO,<sup>†</sup> YUTAKA AKIYAMA,<sup>†</sup> KENTARO ONIZUKA,<sup>†</sup>  
and TAMOTSU NOGUCHI<sup>†</sup>

We have developed the Parallel Protein Conformation Analysis System, based on our implementation techniques used in parallel tree search programs. We employed an exhaustive search method for conformation search, rotating all the rotatable bonds by specified grain. We have implemented the system, and succeeded in analyzing the conformations of 20 amino acids individually. The parallel version of our program showed good speedups over the sequential version. We also succeeded in overlapping task management operations onto tree search operations by assigning them to multiple processes. We are aiming at analyzing the conformations of several amino acids that is useful in drug design.

#### 1. はじめに

昨年度我々は<sup>1)</sup>、木探索問題の典型例である 0-1 ナップサック問題<sup>2)</sup>を解くプログラムを並列計算機上に実装し、並列に木探索問題を解くことの有用性を実証した。さらに、木探索問題の並列化においては、探索空間の広がりにより偏りが生じるため、タスクプール方式等の動的な負荷分散が重要であるとともに、分散メモリ型並列計算機では、部分問題の展開や枝刈りなどの木探索処理を、スレーブの内部でまとめて行うなどの通信回数低減の工夫も必要であるという知見を得た。

本研究では、上記の研究で得られた知見を利用して、より実用的で大規模な木探索問題に取り組むことにした。今回は、タンパク質立体配座解析問題をとりあげる。この問題は、対象となるタンパク質について、それを構成する原子(球)の集まりとそれらを結合する共有結合(リンク)の集まりから、そのタンパク質の

とり得る立体配座(補遺 1)を求める。我々はタンパク質立体配座解析を行う並列システムを実装し、いくつかのデータに対して実験を行った。

本稿では、第 2 節でタンパク質立体配座解析の概要を他の手法と比較しながら解説する。第 3 節で開発したシステムの設計について述べ、第 4 節で実装方法を述べる。第 5 節では、実装されたシステムを用いて、実際に並列計算機上で 1 残基レベルのアミノ酸について立体配座解析を行った結果を報告する。第 6 節では、立体配座解析の結果や並列システムの性能について考察を行い、最後にまとめを述べる。

#### 2. タンパク質立体配座解析

##### 2.1 タンパク質の構造

タンパク質は、20 種類のアミノ酸残基が一行につながった鎖状の分子である。それぞれのアミノ酸は、主鎖と呼ばれる共通の構造を持っており、この部分が同じ向きに連結することによってタンパク質を構成する。20 種類のアミノ酸は、側鎖と呼ばれる部分の構

† 技術研究組合 新情報処理開発機構  
Real World Computing Partnership

表 1 タンパク質立体構造予測または解析手法の比較

手法	考慮する制約 衝突	力場	環境情報 必要度	探索空間	ローカル ミニマム回避	適用範囲
分子動力学法	○	○	高	局所的探索	×	数〜数百残基 大規模タンパク質
モンテカルロ法	○	○	高	部分的探索	△	数〜数十残基 小規模タンパク質
総当たり法 (本手法)	○	×	不要	全解探索	◎	数残基 ペプチド

成の違いによって区別される。

図 1 に、アミノ酸残基の例としてフェニルアラニン (PHE) の構造を示す。主鎖の結合は灰色で、側鎖の結合は黒で表現されている。

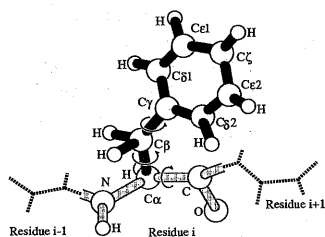


図 1 フェニルアラニン (PHE) の構造

本研究におけるタンパク質立体配座解析では、タンパク質の構造を以下のようにモデリングする。

## 2.2 探索空間

タンパク質中の共有結合のうち、一重結合のまわりでは、その両端の原子は自由に回転することができ、さまざまな配座をとることが可能である (補遺 1)。

与えられたタンパク質の回転可能なすべての結合について、その二面角 (補遺 2) を指定した角度刻みで回転させ、その立体配座が存在可能か調べる場合、回転可能な結合の数を  $N$ 、それぞれの結合における二面角のサンプリング数 (回転の範囲/角度刻み) を  $p_i$  ( $0 \leq i < N$ ) とすれば、探索に必要な組み合わせ数 (枝刈り前) は  $M = \prod_{i=0}^{N-1} p_i$  となる。

なお、図 1 の PHE に注目すると、 $C_\beta - C_\gamma$  を軸にベンゼン環を回転させると 180 度の対称性があることが分かる。同様に、タンパク質内には 120 度の対称性を持つ部分も多く存在する。これらの場合、回転範囲を 0 度から 120 度あるいは 180 度に限定し、探索する組み合わせの数を減らすことができる。

## 2.3 制約条件

2 つの原子が衝突していることを判定するために、van der Waals 半径 (補遺 3) を用いる。ただし、以下の理由により、van der Waals 半径の合計値に 1 より小さい係数 (0.8~0.9) を乗じたものを使用する。

- 各原子の van der Waals 半径は、相手原子としてあらゆる種類を考えた時の衝突距離の平均値であり、厳密な値ではない。
- 本研究では「結合長は固定」を前提としているが、厳密には多少伸縮する。よって、van der Waals 半径を小さくし伸縮誤差を吸収したほうが、現実と適合し見逃しが無い。

## 2.4 他の手法との比較

表 1 に、本手法と他の手法との比較を示した。

分子動力学法やモンテカルロ法といった物理学的手法では、クーロン力や水素結合等を考慮した厳密な力場計算で、エネルギー最小の構造を求める<sup>3)4)</sup>。よって、周囲の環境情報 (溶媒など) が未知であると、正確なエネルギー計算がなされず、精密な結果を得ることはできない。しかし、本手法では、力場計算は一切行わないので、周囲の環境情報が未知の場合でも、精密な存在可能配座を列挙することができる。また、物理学的手法で問題となるローカルミニマムへのトラップも、本手法では起きない。

ただし、本手法では力場を考慮しないため、多数の候補が出力される。本手法における組み合わせ爆発の問題は、枝刈りと並列化でかなり対応できる。しかし、本質的には避けられないので、1~2 残基程度の部分構造ごとに本手法を用いて、それ以上のものには部分解を組み合わせる等のアプローチ (「考察」参照) をとる必要がある。

## 3. 設 計

本問題を解くアルゴリズムには、0-1 ナップサック問題を解くアルゴリズム<sup>1)</sup>と多くの共通点がある。以下に、本研究で開発したシステムの設計について述べる。

### 3.1 探索木

0-1 ナップサック問題では、与えられた荷物について、一つずつ順番に「入れる」または「入れない」の状態に固定することで、原問題をより小さな部分問題へ展開していく。それぞれの荷物には、効率 (価値/重さ) を基準に優先度が付けられており、優先度の高いものから順に状態が固定される。本問題でも同様に、回転可能な結合について、順にその結合のまわりの原子の位置を固定していく。それぞれの結合には、0-1 ナップサック問題と同様に優先度を導入する。一般に、原子が密に存在する部位においては立体障害が発生しやすく、容易に枝刈りすることができる。このような部位の結合に高い優先度をつけることで、探索木の根本付近から大きく枝刈りすることができる。

一方、0-1 ナップサック問題では、各荷物の状態は 2 つしかないので、探索空間は二分木で表現できた。しかし、タンパク質中の一重結合の二面角は複数の値をとりうるので、探索木は多分木となり、0-1 ナップサック問題の探索木に比べると横幅が広がる。

**3.2 マスタ-スレーブモデルによる負荷の動的分配**  
本システムでは、0-1 ナップサック問題の分散メモリ型並列計算機上でのプログラミングと同様に、マスタ-スレーブモデルを採用した。マスタはシステム内に常に1つ存在し、部分問題をプールのリストの管理を行う。スレーブは、マスタから得た部分問題についてさらなる部分問題の展開と枝刈りを、指定された回数だけローカルに行い、新たに生成された部分問題群をマスタ上のプールに戻す。この操作を、部分問題の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、かなり均一な負荷分散が可能になる<sup>1)</sup>。

## 4. 実装

前節に掲載した設計に基づいて、逐次版と並列版の立体配座解析プログラムを実装した。

### 4.1 システムの構築

本システムの構築にあたっては、当研究室で開発した PAPIA ライブラリを用いた<sup>5)6)</sup>。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++言語のクラス階層としてライブラリ化したものである。本システムにおいては、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 ライブラリの機能で実装されている。なお、今回開発した立体配座解析の関数群も、PAPIA ライブラリの新しい機能として追加することを計画中である。

### 4.2 初期データ

立体配座解析に必要な初期データとして、下記の2種類の情報が格納されたファイルを起動時に読み込む。

#### 4.2.1 原子情報

対象となるタンパク質を構成する原子に関する情報で、各エントリは下記の属性を持つ。

- 原子の識別子 (例: 1VAL-CA)
- 空間内での初期座標 ( $x, y, z$ )
- 原子の van der Waals 半径

各原子の初期座標については、商用のタンパク質構造解析ソフトウェア Insight II<sup>7)</sup>によって生成する。

#### 4.2.2 結合情報

原子間の共有結合に関する情報で、各エントリは下記の属性を持つ。

- 結合の識別子 (例: 1VAL-C-2TYR-N)
- 結合の両側の原子の識別子
- 回転可能かどうか (y or n)
- もし回転可能なら
  - － 二面角の基準となる両端の原子の識別子
  - － 回転の始点、終点、刻み(度)
  - － 木探索における優先度

### 4.3 結合解析結果を用いた回転操作の高速化

本システムでは、初期データの読み込みの際にすべ

ての共有結合のつながり方を調べ、回転可能なすべての結合について、その両側にある原子団のリストをあらかじめ用意するようにした。結合を回転させる際には、より数の少ない原子団を回転させたほうが計算が少なくて済むが、このリストを利用すればどちらの原子団を回転させるべきかを瞬時に判断できる。

### 4.4 キャッシュによる衝突検査の高速化

プログラム内部に、衝突検査の結果を格納するソフトウェアによるキャッシュを持たせた。検査対象の原子のあらゆる組について、衝突しているかないかをキャッシュに記憶させる。そして、それらの位置関係が変わらない限り、キャッシュの内容を再利用する仕組みである。衝突検査のたびに2原子間の距離を計算し直すのに比べ、かなりの高速化が期待できる。

## 5. 実験結果

### 5.1 実験に用いた入力データ

本システムは、創薬分野において重要といわれるアミノ酸3~7残基程度(ペプチドレベル)の立体配座解析を目標としている。例えば、タンパク質が生体内の受容体と呼ばれる構造に結合する問題を考えるとき、実際にはタンパク質内の3~7残基程度の部分が主に結合に関与していることが知られている。これまでの手法では、この程度の長さの立体配座を網羅的に求めるのは困難であったが、本システムでは許容される立体配座の列挙が可能になる。その結果に実験的知見などを融合して、真の構造に絞っていくことができる。

我々は、まずアミノ酸3~7残基の立体配座解析への第一段階として、1残基レベルの解析を行った。使用したデータは水素原子を考慮に入れたもので、考慮に入れない場合に比べると計算量は大幅に増えるが、より厳密な解析が期待できる。

また、解析をより現実に近いにするために、それぞれの残基の主鎖の窒素側の末端に1つの炭素を、炭素側の末端に1つの窒素と2つの水素を付加した(補遺4)。

### 5.2 逐次版による解析

本実験では、20種類のアミノ酸に対して、1残基レベルでの立体配座解析を行った。逐次版は、Sun Enterprise 4000 (250MHz UltraSPARC II) で実行した。各入力データにおける原子数、回転軸数、全組み合わせ数  $M$ (枝刈り前)、および存在可能な配座数と計算時間を表2に示した。

### 5.3 並列版による速度向上

表2に示した入力データのうち、全組み合わせ数の大きい MET, GLN, GLU, LYS の4つについて、Hitachi SR2201(150MHz PA-RISC 1.1 + PVP-SW, 256 CPUs, HI-UX/MPP MPI) 上で並列解析を試みた。解析に用いた条件は、表2に示したものと同一とした。図2に、逐次版に対する並列版の速度向上比を示す。横軸は使用 CPU 数、縦軸は、逐次版の実行

表 2 1 残基レベルでの立体配座解析 (水素を考慮)

記号	名称	原子数	回転軸数	全組み合わせ数 $M$	存在可能配座数	比率 (%)	実行時間 (秒)
GLY	グリシン	11	1	24	19	79.2	2.65e-4
PRO	プロリン	18	1	24	7	29.1	3.71e-4
ALA	アラニン	14	2	192	63	32.8	1.87e-3
PHE	フェニルアラニン	24	3	6,912	160	2.31	1.88e-2
CYS	システイン	15	3	13,824	728	5.27	1.57e-2
TRP	トリプトファン	28	3	13,824	308	2.22	3.39e-2
HIS	ヒスチジン	21	3	13,824	209	1.51	1.18e-2
SER	セリン	15	3	13,824	1,344	9.72	3.59e-2
VAL	バリン	20	4	36,864	160	0.434	2.13e-2
THR	トレオニン	18	4	110,592	485	0.439	2.43e-2
TYR	チロシン	25	4	331,776	2,120	0.639	1.01
ASN	アスパラギン	18	4	331,776	9,593	2.89	2.05
ASP	アスパラギン酸	17	4	331,776	24,171	7.29	4.07
LEU	ロイシン	23	5	884,736	2,716	0.307	1.08
ILE	イソロイシン	23	5	884,736	2,473	0.280	1.02
MET	メチオニン	21	5	2,654,208	27,614	1.04	8.08
GLN	グルタミン	21	5	7,962,624	180,529	2.27	41.1
GLU	グルタミン酸	20	5	7,962,624	160,804	2.02	39.0
LYS	リジン	25	6	191,102,976	85,072	0.0445	43.0
ARG	アルギニン	27	8	110,075,314,176	66,198,629	0.0601	33458

注: 逐次版の実行結果 (Sun Enterprise 4000). 全組み合わせ数の少ない順に掲載. 水素原子を考慮した厳密な計算. 回転はすべて 15 度刻み, van der Waals 半径への係数は 0.88. 原子数は前後に付加した C と  $NH_2$  を含む

時間を基準にした速度向上比である. 使用 CPU のうちの 1 台がマスタプロセスに割り当てられ, 残りがスレーブプロセスに割り当てられる.

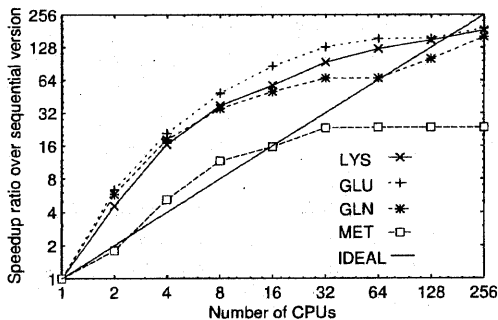


図 2 並列版の逐次版に対する速度向上比

なお並列版では, スレーブ内では木探索の操作を指定した回数だけまとめて行う (3.2 参照) が, 上記の 4 つのアミノ酸については, いずれも 1024 回を指定したとき最高の速度向上を得られた.

## 6. 考 察

### 6.1 PDB との比較

本システムの解析結果の正当性を示すために, タンパク質立体構造データベース PDB (Protein Data Bank)<sup>8)</sup> から得られる統計的情報との比較を行った.

図 3 に, 一例として TRP の  $C_\alpha - C_\beta$ , および  $C_\beta - C_\gamma$  のとり得る二面角の関係を示した. 左が本シ

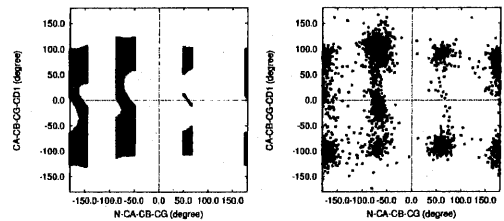


図 3 TRP の存在可能配座 (左) と PDB 上の統計的分布 (右)

ステムによる配座解析で得られた存在可能配座, 右側が PDB に登録されているあらゆる TRP から抽出した二面角である. それぞれの横軸は  $C_\alpha - C_\beta$  のとり得る二面角を, 縦軸は  $C_\beta - C_\gamma$  のとり得る二面角を示している. 解析に用いた条件のうち, 角度刻みは 2 度とし, それ以外は表 2 に示したものと同一とした. これをみると, PDB データから得られた二面角の分布に現れている多くの特徴が, 我々の実験結果にも現れていることが分かる.

なお, 実験結果では許容領域とされていないが, PDB データではプロットされている二面角が存在する. これについては, 以下の理由が考えられる.

- PDB 内には, 誤ったデータや, 結晶回析の分解能の低いデータが存在し, 二面角に誤差が生じた.
- van der Waals 半径に乗じた係数では吸収しきれない共有結合長の伸縮があった.
- 今回考慮していない原子間引力によって, 原子が van der Waals 半径の合計値よりも近付いた.

なお, アミノ酸 1 残基の例では, 本システムを用いても, PDB から統計的情報を抽出しても似たような

表 3 タスクオブジェクト再利用によるメモリ新規割当の低減効果

残基種	プロセスの種類	新規割当回数	再利用回数	合計	新規割当発生比率 (%)
MET	逐次	109	70152	70261	0.155
	並列:マスタ	109	28293	28402	0.383
	並列:スレーブ	4	71042	71046	0.005
GLN	逐次	707	419899	420606	0.168
	並列:マスタ	707	181497	182204	0.388
	並列:スレーブ	5	422273	422278	0.00118
GLU	逐次	630	415569	416199	0.151
	並列:マスタ	630	162089	162719	0.387
	並列:スレーブ	4	418107	418111	0.00095
LYS	逐次	334	287302	287636	0.116
	並列:マスタ	334	87037	87371	0.382
	並列:スレーブ	3	289929	289932	0.00103

結果が得られるが、アミノ酸 2 残基以上に対しては、PDB からだけでは十分な情報が得られないため、本システムによる配座解析計算が重要になる。

## 6.2 オブジェクト再利用による速度向上

図 2 によると、4つの残基種のすべてについて、並列版の 2CPU での処理速度が、逐次版を上回った。2CPU を使用した場合、木探索の処理は 1 プロセス内で行われており、展開する部分問題の個数やそれらを展開する順序は逐次プログラムと全く同一である。また、部分問題を展開する順序が同じであるため、衝突検査のキャッシュへのヒット率も全く同じである。むしろ、マスター-スレーブ間で通信するため、その時間が余計にかかるはずであるが、実際には時間は短縮している。

そこで、我々は使用されているメモリに着目した。本システムでは、個々の部分問題を表現するために「タスク」というオブジェクトを使用している。

タスクオブジェクトは、部分問題の生成や枝刈りといった操作によって、作成と削除が頻繁に繰り返される。よって、タスクオブジェクトの効果的な再利用を図るため、C++言語のデフォルトの new/delete 演算子を使わずに、フリーリストを使った独自の new/delete 演算子を用意した (今回は、256 個のタスクオブジェクトを一度に作成するようにした)。

表 3 に、各実験におけるタスクオブジェクトの再利用の実測値を載せた。スレーブでは、内部でローカルに行う木探索処理の回数が決められているので、必要なタスクオブジェクトの数には上限がある。したがって、数回の新規割当後は、フリーリストからの再利用だけで木探索処理を進めることができる。これに対し、タスクをプールするマスタでは、必要なタスクオブジェクトの数に上限はなく、処理の進行にしたがって次々とメモリの割当を行わなければならない。

本システムの並列版は、マスタにおけるメモリ割当の処理とスレーブにおける木探索の処理をオーバーラップさせることができ、それらをオーバーラップできない逐次版に比べ処理が高速になったと考察できる。

## 6.3 より大きな問題に向けての対応

本研究は、創薬分野において有用な 3~7 残基程度

の立体配座解析が可能なシステムの構築を目標としている。しかし、本手法では組み合わせ爆発を本質的には避けられないので、1~2 残基程度の部分構造の解析に本手法を用い、それ以上のものには部分分解を組み合わせる等のアプローチをとらなければならない。現在、1 残基レベルでの存在可能な配座のうち有望な代表点を選び、それらの組み合わせを用いてさらに解析するような階層的なアプローチを検討している。

また、存在可能な配座の領域は連続していることが多いので、連続した領域の内部の計算を省略して計算量を減らす工夫についても検討中である。

## 7. まとめ

並列木探索を用いて、タンパク質立体配座解析問題を解くシステムを構築した。本研究では、回転可能なすべての共有結合を回転させ、それによって得られるすべての立体配座異性体について実際に存在可能かを調べあげる網羅的手法を採用した。本手法を実装し、アミノ酸 1 残基レベルについて立体配座解析を行った結果、高精度の解析結果を得ることができた。さらに、並列プログラムによって速度を向上させることにも成功した。また、並列プログラムでは、タスク管理の処理と木探索処理のオーバーラップが可能になり、プログラムの実行がより高速になった。

本手法は、力場計算を必要としないので、環境情報の同定が必要な既存の解析手法に比べ、1 残基レベルで確実に良い結果を得られることが確認された。今後は、階層的なアプローチによって、3~7 残基程度のタンパク質の解析が可能になるようにシステムを拡張する予定である。

## 参考文献

- 1) 安藤, 田中, 久保田, 松田, 秋山, 佐藤, "Knapsack 問題における共有メモリ型/分散メモリ型並列計算機の性能比較", 情報研報 97-HPC-67-9, pp. 49-54 (1997).
- 2) S. Martello, P. Toth, "KNAPSACK PROBLEMS - Algorithms and Computer Implemen-

tations”, Wiley-Interscience Series in Discrete Mathematics and Optimization (1989).

- 3) 中村, “生体高分子の解析における構造探索の新しいアプローチ”, 日本化学会情報化学部会 *CICSJ Bulletin*, **15**, No. 2 (1997).
- 4) M. Kinoshita, Y. Okamoto, F. Hirata, “First-Principle Determination of Peptide Conformation in Solvents: Combination of Monte Carlo Simulated Annealing and RISM Theory”, *Journal of American Chemical Society* **120**, pp. 1855-1863 (1997).
- 5) 鬼塚, 野口, 斎藤, 秋山, “タンパク質立体構造研究支援のための並列統合解析システムの構築”, 情処研報 97-HPC-68-8, pp. 45-50 (1997).
- 6) 秋山, 鬼塚, 野口, 安藤, 斎藤, “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”, 情処研報 98-HPC-70-6, pp. 31-36 (1998), <http://www.rwcp.or.jp/papia/>
- 7) <http://www.msi.com/>
- 8) F. Bernstein, T. Koetzle, G. Williams, E. Meyer, Jr., M. Brice, J. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, “The Protein Data Bank: a Computer-based Archival File for Macromolecular Structures”, *Journal of Molecular Biology* **112**, pp. 535-542 (1977).
- 9) M. O’Leary, 中島訳, “有機化学”, 東京化学同人, pp. 301-330 (1981).
- 10) IUPAC-IUB Commission on Biochemical Nomenclature 1969, “Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains”, *Biochemistry* **9**, pp. 3471-3479 (1970).
- 11) G. Schulz, R. Schirmer, 大井監訳, “タンパク質—構造・機能・進化—”, 化学同人, pp. 36-37 (1980).

### 補遺 1 — タンパク質の立体配座

立体配座 (conformation) とは, 原子同士の結合の回転 (ねじれ) によって, その結合の両端の原子団を構成する原子の相互位置が変化して生じる種々の形態のことである。

タンパク質中の共有結合のうち, 主鎖の O—C—N 付近に見られる共鳴部分では, その結合の回転は制限されている。これに対し, 一重結合, 例えば主鎖の N—C $\alpha$  や C $\alpha$ —C, 側鎖中の C $\alpha$ —C $\beta$  などは, 自由に回転することができる。このため, 同一のタンパク質について, 複数の配座異性体の存在が可能である<sup>9)</sup>。

### 補遺 2 — 二面角

タンパク質が空間的にどのように折りたたまれているかを記述するとき, 各原子の (x, y, z) 座標で表現すると, 1 原子につき 3 つの変数が必要となる。これに対し, 原子間の共有結合長を固定として考えると, 二面角と呼ばれる表示法を使用することができ, より少ない変数で座標を表現することができる。図 4 において, 点 A, B, C, D をタンパク質分子中の連続する 4 点とすると, 直線 BC を回転軸として点

ABC が作る平面に対して, 点 BCD が作る平面が時計回りに回転した角度  $\phi$  を二面角という<sup>10)</sup>。

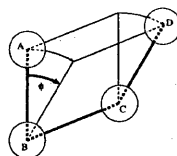


図 4 二面角

### 補遺 3 — van der Waals 半径

2 つの原子間には, 図 5 に示す形状をしたポテンシャルエネルギーがかかることが知られている。

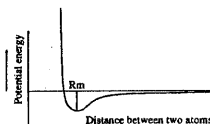


図 5 van der Waals ポテンシャル

2 つの原子は平衡距離  $R_m$  よりも近付くと非常に大きな反発力を受けてそれ以上近付かない (立体障害と呼ぶ)。この性質を利用して, 各原子の種類毎に van der Waals 半径という概念を導入することができる。表 4 に今回の実験で使った値を示す<sup>11)</sup>。

原子の種類	半径 (Å)	原子の種類	半径 (Å)
H(芳香環内)	1.0	N	1.6
H(上記以外)	1.2	C	1.7
O	1.45	S	1.8

### 補遺 4 — 隣接アミノ酸残基の原子の考慮

タンパク質は, アミノ酸が複数連結することによって構成される。よって, 1 残基レベルの解析でも, 前後に隣接するアミノ酸の原子との立体障害について考慮したほうが, 解析をより現実的に近付けることができる。そのため本実験では, 対象となるアミノ酸の主鎖の窒素側の末端にさらに 1 つの炭素 (C) を, 炭素側の末端にさらに 1 つの窒素と 2 つの水素 (NH<sub>2</sub>) を付加した (図 6 の斜線で示した原子)。これにより, 前後にアミノ酸が連結されている状態に近くなり, 実際のタンパク質でみられる立体障害が検知され, より正確な結果が期待できる。

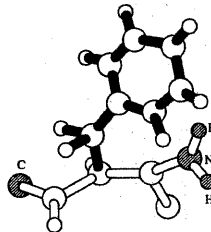


図 6 隣接アミノ酸の原子を考慮した構造