

多次元空間解析で迫る生体高分子の 3次元立体構造と Energetics

生物分子工学研究所 肥後 順一*・小野 聡**
大阪大学蛋白質研究所*** 中島 伸介・中村 春木

(受付 2000 年 10 月 2 日; 改訂 2001 年 3 月 21 日)

要 旨

核磁気共鳴実験によって得られる蛋白質・核酸等の生体高分子中の多数の水素原子間距離の情報から、化学的構造制約条件下において高分子の立体構造を再構築する Distance Geometry 計算では、多次元尺度法や多次元空間中のシミュレーティッド・アニーリング (SA) が利用される。一方、生体高分子の立体構造形成のダイナミクスを理解するためには構造空間に射影した自由エネルギー地形が必要であり、最近発展しているマルチカノニカル・アンサンブル等の効率的サンプリング手法が極めて有効である。また、これら計算結果の解析においては、主成分分析が多用される。我々が開発した効率的な構造探索手法を解説し、蛋白質や核酸の構造解析の現場で、これらの手法がどのように利用されているかを紹介する。

キーワード：Distance Geometry 計算，シミュレーティッド・アニーリング，自由エネルギー地形，マルチカノニカル・アンサンブル，主成分分析。

1. 序

蛋白質や核酸などの生体高分子は、遺伝情報に規定された数千～数万個の多数の原子によって構築される。蛋白質は、アミノ酸がゲノム中の遺伝子情報によって、ある一定の並び方で 100～数千個つながった 1 本の高分子鎖であり、主鎖と呼ばれる各アミノ酸で共通な鎖の骨格部分をなす部分と、側鎖と呼ばれるアミノ酸毎に異なる部分とからできている。核酸では、アミノ酸のかわりに、ヌクレオチドと呼ばれる塩基と糖とからなる化合物が、やはり鎖状につながったものである。これらの生体高分子は、生理的環境下では特定の立体構造を形成して機能を発揮することが多いが、一方でダイナミックに変形もする。実際、蛋白質の部分フラグメント構造である、アミノ酸が数残基つながっただけの短いペプチドは、全く同一のアミノ酸配列でも、その周囲の物理化学的環境や蛋白質中の前後のアミノ酸配列に応じて異なる立体構造を持つ場合があり、カメレオン配列として知られている(中村・有坂(1997))。また、プリオン病で代表されるアミロイド形成は、同一のタンパク質が、機能を持つ天然構造とは別の異なる立体構造に折れ畳まれて、繊維状となることによる。

*現 東京薬科大学 生命科学部：〒192-0392 東京都八王子市堀之内 1432-1.

**現 三菱東京製薬(株)：〒227-0033 神奈川県横浜市鴨志田町 1000.

***生体分子解析研究センター：〒565-0871 大阪府吹田市山田丘 3-2.

このように、たいへん複雑な構造・事象ではあるが、X 線結晶学や核磁気共鳴法 (NMR) によって大量の実験値が得られるようになり、本稿で述べる解析法などによって、そこから構造情報を引き出し、精密な構造決定がなされるようになってきた。

第 2 節では、まず、NMR 実験によって得られる、ある環境下において特定の立体構造を形成している生体高分子に対して、その分子中の水素原子間距離の情報を全て満たすような分子構造を高速に探索し、分子構造を決定する手法について概説する。

一方、分子がダイナミックに変形する現象は、位相空間中の広い範囲に渡って、準安定状態 (エネルギー極小構造) がたくさん存在し、しかも各極小点におけるエネルギー値がそれほど変わらないことに起因している。計算機によって実験値を再現し、実験に置き代わるほどの精度の高いリアルな計算を行って、現実には起きていない現象をシミュレートするためには、それぞれの分子のダイナミックな状態の周辺密度、あるいはその対数の符号を変えた量として定義される「自由エネルギー」を知る必要がある。自由エネルギーは熱力学量であり、自由エネルギーの低い状態が実験では観測される。第 3 節では、それら多くの準安定構造と最安定構造との両方を探索し、どのようにそれらの状態が位相空間中に分布しているかを調べる、すなわち「自由エネルギー地形」を俯瞰する我々の試みを紹介し、同一のアミノ酸配列が多様な立体構造を形成する構造多形の機構を提示する。

これらの生体高分子の構造決定および構造多形解析では、多次元空間の表現を用いた手法が威力を発揮する。本稿では、その点についても具体的な応用を紹介する。

2. 多次元尺度法と多次元シミュレーティッド・アニーリング計算で分光学データから生体分子立体構造を組み立てる

2.1 Distance Geometry 計算

核磁気共鳴 (NMR) 実験、特に核オーバーハウザー効果 (NOE) スペクトルから得られる水素原子ペア間の距離の値をもとに、溶液中での蛋白質の立体構造を決定するアルゴリズムとその計算のことを Distance Geometry 計算と呼んでいる。最近の NMR 測定法の進歩は目覚ましく、高磁場の超電導磁石と複数の安定同位体標識を利用することによって、今や、分子量 3 万前後までの生体高分子中の一つ一つの水素原子核に起因する化学シフトを、それぞれ分離して同定することができ、さらに、その同定された水素原子間の距離を NOE スペクトルから算出できる (京極・月原 (1997))。

Distance Geometry 計算では、 i 番目と j 番目の水素原子間の距離 $\{d_{ij}^{obs}\}$ が与えられた場合に、その距離ペアを満たすような立体構造を数学的に構築する。X 線結晶解析や電子顕微鏡による解析が、全原子の絶対座標あるいはそのフーリエ空間座標を観測するのに対し、分光学によって得られるのは原子間距離という相対配置であり、そこから Distance Geometry 計算によって絶対座標を逆算して再構築するのである。これは構造決定のための道具であり、以下の 3 点を検討することが要請される。

- (A) 実験で得られた原子間距離を満たす立体構造はたくさんありうるが、それらをすべて偏りなく検出できるか? (サンプリングの問題),
- (B) 原子間距離を満たす矛盾のない (バイオレーションがない) 立体構造を速く検出できるか? (収束性の問題),
- (C) 計算に要する CPU 時間やメモリー量は適当か? (経済性の問題)

これらを満たす Distance Geometry 計算には、歴史的に従来からいくつかの異なる手法が開発されており、多次元メトリック行列の対角化とその 3 次元空間への投影による手法 (Embedding

法)と、ペナルティー関数をシミュレーティッド・アニーリング (SA) 計算によって最適化する手法の2つに大別される。どちらも多次元空間の利用によって効果的に最終解を得ることができる。

しかし、前者の多次元尺度法の応用によって、確かに蛋白質の立体構造を再構築できるものの、残念ながら、そのアルゴリズムの美しい数学的な完全性に対して、対象となるデータの不確実性のバランスが悪く、精密に行うのはかなりハードな計算が要求される。その理由は、本稿の付録に示すように、再構築したい分子中の全ての原子間距離が既知であることをアルゴリズムが前提としているのに対し、実験的に得られる原子間距離は実際には極めてわずかしかなく、全てのつじつまが合うように残りの原子間距離を推定するための計算(この操作はメトラライゼーションと呼ばれる)が必須とされるからである。このため、後者の SA 計算手法の方が多用されているのが実情である。多次元尺度法の説明とその Distance Geometry 法における問題点に関しては、付録で説明を行うことにし、以下では実用に多く用いられている SA 計算について説明する。

2.2 多次元シミュレーティッド・アニーリング (SA) 計算

構造探索によってペナルティー関数を最適化する場合には、まず、距離制限 (Distance Restraint) に対するペナルティー関数 P_{DR} として一般に次の形のものが用いられている。

$$(2.1) \quad P_{DR} = \sum_{i < j} p_{ij}$$

$$(2.2) \quad p_{ij} = \begin{cases} k^{min} (d_{ij} - d_{ij}^{min})^2 & (d_{ij} \leq d_{ij}^{min}) \\ 0 & (d_{ij}^{min} < d_{ij} \leq d_{ij}^{max}) \\ k^{max} (d_{ij} - d_{ij}^{max})^2 & (d_{ij}^{max} < d_{ij}) \end{cases}$$

d_{ij} の代わりに d_{ij}^2 が用いられることも多い。 k^{min} , k^{max} は重みである。このペナルティー関数が 0 の値を取るような $\{d_{ij}\}$ の値を持つ立体構造が求める正解であるが、実際にはこのペナルティー関数の最小値を探索することによって解である立体構造を再構築する。しかし、多数の変数(各原子の3次元座標の個数から分子全体の並進・回転の6つの自由度を引いた $3N-6$ 個の自由度がある)による関数であるため極小値が無数に存在し、また、共有結合 (Covalent Bond) で各原子が結ばれているという複雑な束縛条件 (P_{CB}) も、 P_{DR} に加えて全体のペナルティー関数には加算されるため、単なる極小値を求める通常的手法では、真の最小値に到達できない。

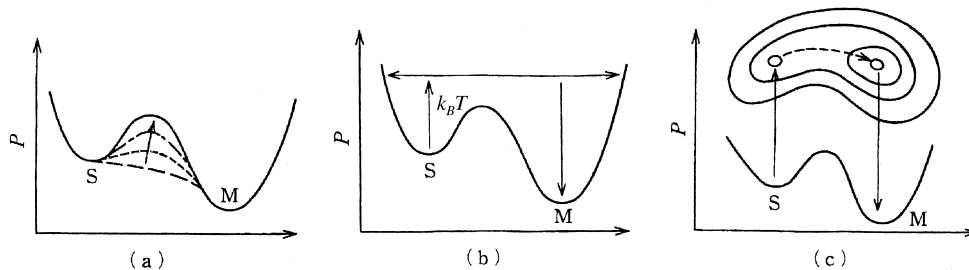


図 1. ペナルティー関数の最小値 M を探索する手法。(a) Variable Target 関数法: 関数の形自体を変形させていく。(b) 距離の制限を課した分子動力学法: 温度 T に対応する運動エネルギーを与えて準安定状態 S から抜け出す。(c) 変数を多次元化することによって、エネルギー障壁を通らずに最小値 M へ到達できるようになる。

この問題は、多自由度がある場合の、最適解の束縛付き探索問題そのものである。この問題の解決のために、いろいろな方法が考えられているが、その内の代表的なものを図 1 に示す。図 1(a) には Variable Target 関数法 (Braun and Go (1985)) を示す。この方法では、ペナルティ関数の形自体を初期にはなめらかにしておき、極小化の計算の進行と同期して次第に正確なペナルティ関数の形にもっていくことで最小値を捜し出す。図 1(b) の制限付分子動力学法 (Nilges et al. (1988)) においては、温度(運動エネルギー)を与えて分子を運動させることによって、ある局所的な極小値(準安定点)にトラップされることなく、最小値に到達させる。図 1(c) に示す多次元化による極小値探索では、少ない次元の空間(図の例では 1 次元)では障壁が高すぎてある局所的な極小値から抜けられない場合でも、次元を拡張(図では 2 次元)するとうまくすり抜けて真の最小値に到達しやすくなる。このように状態空間を拡張して構造探索を効果的に行う手法は、最近になって、蛋白質の格子モデルにおける構造探索にも応用されている (Iba et al. (1998), Chikenji et al. (1999))。以下では、これら 3 つの方法を同時に取り入れた我々の Distance Geometry 計算プログラム (Nakai et al. (1993)) において、特に多次元への拡張による効果について紹介する。このプログラムでは、付録で述べる多次元尺度法も計算可能である。

分子構造を維持するためのペナルティ関数 P_{CB} として、通常分子計算では分子力場によってエネルギーが高くなるように束縛することが多い。しかし、ここでは多次元化への拡張のため、共有結合による分子構造を維持するための束縛条件を、全て N 個の原子における (i, j) 原子ペア間の距離を二次式によって束縛するようなペナルティ関数として表現する。多次元に自然に拡張された距離を用いる。例えば 4 次元では、

$$(2.3) \quad d_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 + (w_i - w_j)^2$$

として x, y, z のデカルト座標以外に新たな w の軸を加えるわけである。さらに、同様の多次元に拡張された距離を用いて、実験値を満たすために、式 (2.1) のペナルティ関数 P_{DR} を用いる。

初期構造としては、3 次元空間中のランダムな N 個の点でも良いし、まず多次元尺度法を行いその結果を投影する際に、主な成分として 3 つでなく、4 つあるいはそれ以上の成分をとった n 次元構造としても良い。そのような N 個の点群を対象にして、高温での n 次元の分子動力学計算を行い徐々に温度を 0K にまで下げる操作(シミュレーティッド・アニーリング—SA—)をする。最終的に得られる構造は 3 次元空間中の分子の立体構造であるため、最後に 3 次元より高次の座標が残らないように、低温に下げる際に、座標の 4 次元 (4D) や 5 次元 (5D) 目に相当する n 次元目の値 $\{r_{nD,i}\}$ がなくなるようにさらに余分のペナルティ関数

$$(2.4) \quad P_{nD} = \sum_i k_{nD} (r_{nD,i})^2$$

を加えておく。 k_{nD} は重みである。

結局、全体のペナルティ関数 P は、

$$(2.5) \quad P = P_{DR} + P_{CB} + P_{nD}$$

となっている。計算効率を良くするため、人為的に大きな質量 (1000 Da) を各原子に与え、50 fs という大きな単位時間ステップ幅での分子動力学計算 (Havel (1991)) として探索効率をあげる工夫も取り入れている。

2.3 実施例と構造探索効率について

冒頭に述べたサンプリング問題の解決を確認するために、アラニン (Ala) という側鎖にメチル基が 1 つしかない単純なアミノ酸が 30 個連なったホモポリマーにおいて、水素原子間に特

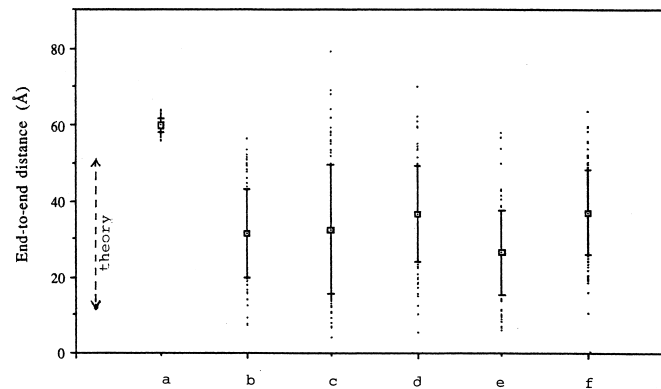


図 2. 距離の制限をつけずに, a-f の 6 つの異なるプロトコルで生成した, アラニン 30 個が連なった鎖状分子(各 100 構造を生成)の末端間距離の分布(Nakai et al. (1993)). 平均値を \bar{x} で, 標準偏差を σ で示す. a. メトライゼーションをしない多次元尺度法 (no-metrization) の直後に構造最適化 (min) を行った場合. b. no-metrization 後に 4 次元 SA(4D-SA) を行った場合. c. メトライゼーションを行った多次元尺度法 (metrization) の直後に min の場合. d. metrization 後に 4D-SA の場合. e. 多次元尺度法を使わず単にランダムな不規則構造 (random-coil) から min の場合. f. random-coil から 4D-SA の場合. 左端に, 3 次元の自由な Gauss 鎖と仮定した時の, 理論的に期待される末端間距離の範囲を点線で示す. 最大の確率となる末端間距離の場合の $1/e$ の確率までの範囲を示す.

別な距離の制限を加えず, どのような立体構造が得られるかを調べてみた. 得られた鎖の構造の両端間の距離 (end-to-end distance) は, 制約を入れていないために, 自由な鎖が取りうる広い範囲に分布するはずである. しかし, 図 2-a に示すように, 多次元尺度法において, 全ての原子間距離がつじつまが合うという保証のない簡略法(付録で述べるメトライゼーションを行わない場合)の後に構造最適化を行うだけでは, 極めて限定した伸びた構造しか再構築できない. 構造最適化は, 鎖の体積補正と分子の幾何学構造を conjugate gradient 法による極小値探索で補正して, 化学的に正しい分子構造を得るために必要な操作である.

一方, メトライゼーションを行えば, 構造最適化後の構造として図 2-c のように確かに様々な構造が得られるが, 上記 2.2 節で述べた多次元の SA(ここでは 4 次元 SA)を利用すると, 出発構造がどのようなものであれ, 図 2-b, d, f に見られるように, 常に同様の自由な鎖の構造が再構築されてくる(Nakai et al. (1993)). 参考として, 自由な Gauss 鎖を仮定した場合の両端間の距離の期待値 (Havel (1990)) の範囲も図 2 中に示しておく.

X 線結晶構造解析で得られている Bovine Pancreatic Trypsin Inhibitor (BPTI) の立体構造 (Wlodawer et al. (1984)) から, NMR 実験で得られると同じような 393 個の距離情報をシミュレートして生成した. まず, メトライゼーションをせずに多次元尺度法を用いてから, その後に, 3 次元, 4 次元, 5 次元の SA 法によって 100 個の構造を作ったところ, 3 次元の SA 法ではペナルティー関数 P が大きな正の値のまま局所的極小値に落ち込み, 与えた原子間距離情報を全ては満たさない構造しか得られなかった. 一方, 4 次元, 5 次元の SA を用いれば, どちらでも, ほとんど全ての距離情報による制約を満たし, 良い構造を再構築できることがわかった. すなわち 3 次元の SA はあまり効果的でなく, 4 次元の SA で本質的に十分効果的であると結論された. これは, 3 次元空間中では, 原子同士のぶつかり合いのために, SA による新たな構造探索が制限されて極小値構造から抜け出せないことが多いのに対し, 4 次元目の次元が利用できることにより, 広い構造探索が可能となっている事情による. 以後の実施例は, 4 次

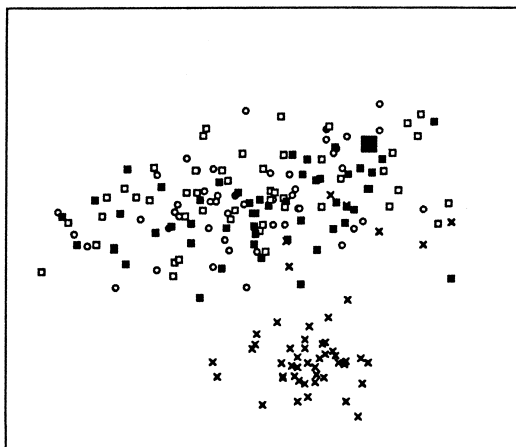


図 3. 主成分解析による 4 つの異なるプロトコルによる Distance Geometry 計算で再構築された蛋白質 BPTI の構造と X 線結晶構造の 2 次元表示 (Nakai et al. (1993)). 4 つのプロトコルの各 50 構造と X 線結晶構造の総計 201 構造の主鎖原子位置のズレの平均二乗偏差の平方根 (r.m.s.d) を総当たりで計算し、各構造間の差異として主成分解析を行った。横軸が主軸であり、縦軸が 2 番目の軸である。大きな \square は X 線結晶構造、 \times が no-metrization + min, \circ が no-metrization + 4D-SA, \square が metrization + 4D-SA, 小さな \square が random-coil + 4D-SA による。各プロトコルは図 2 の説明を参照。

元 SA の操作結果である。

次に、上記の 393 個の距離情報を用いて、いろいろな手法で再構築した BPTI の構造の違いを主軸解析によって行った。4 通りの異なる方法で、それぞれ 50 個の BPTI の構造を再構築し、X 線結晶構造解析で得られている構造を加えた 201 個の BPTI の立体構造をまず用意する。その 201 個の構造から ${}_{201}C_2$ 通りある異なる 2 つの構造を取り出して構造ペアとし、各ペアに対して主鎖原子が最も良く重なるように分子構造を並進・回転した時の、主鎖原子位置のズレの平均二乗偏差の平方根を、構造ペア間の違い (r.m.s.d.) と定義する。この r.m.s.d. から作られるメトリック行列の対角化によって主軸解析を行い、201 個の構造がどのように分布しているかを解析した。この時、主軸と第 2 番目の軸の二次元によって表示したものを図 3 に示す。大きな \square がもとの X 線結晶構造であり、この計算での「正解」であるが、 \times で示すメトライゼーションをせずに構造最適化のみを行ったものは、この「正解」の構造とは少し異なる構造クラスターを形成する。一方、いろいろな初期構造から出発しても、4 次元 SA によって再構築された構造は、「正解」の構造を含む広い構造クラスターを形成している。

具体的な応用例として、核酸と蛋白質からなる複合体分子の構造決定に利用した結果を紹介する。癌原遺伝子 c-myc 産物 (c-Myb)DNA 結合領域は、特異的に AACxG という塩基配列を認識して DNA に結合し転写調節を行う。その AACxG 配列を含む DNA と c-Myb DNA 結合領域との複合体を NMR によって観測し、Distance Geometry 計算によって複合体としての立体構造を得た (Ogata et al. (1994))。通常、DNA と蛋白質の間の分子間距離の情報を与える情報は多くは得られず、この場合もわずか 66 個しか得られていない。しかし、数は少なくとも、DNA 二重らせん構造の溝の底に顔を出している塩基と、直接水素結合の相互作用を行っている蛋白質側の側鎖との間に距離情報が観測されている。この場合、3 次元の SA で立体構造を再構築しようとしても、途中で DNA 側の狭い溝に蛋白質分子が入り込む際、2 つの分子が衝突してしまい、きれいにドッキングした構造が得にくい。しかし、4 次元 SA を用いることに

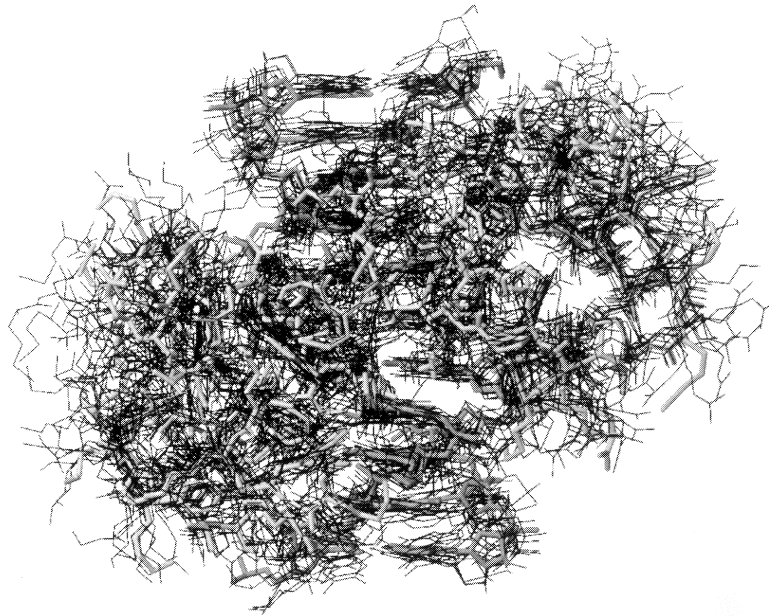


図 4. Distance Geometry 計算によって再構築された、癌原遺伝子 c-Myb 蛋白質 DNA 結合領域(105 残基)と AACxG 配列を含む DNA との複合体構造 (Ogata et al. (1994)). 初期構造としては c-Myb 蛋白質側はランダム鎖とし、DNA 側は典型的な B 型 DNA とした。10 個の異なるランダム鎖から出発して得られた構造を重ね描きしている。中央に上下に走っているのが DNA であり、その周りを c-Myb 蛋白質が取り巻くように結合している。太い白線はこれらの平均構造を示す。PDB データベースのコード 1mse, 1msf に登録されている。

よって、3次元空間ではたとえ衝突していてもすり抜けることが可能なため、ランダム鎖の蛋白質から出発しても、DNA ときれいにドッキングした立体構造が得られた(図 4 (Morikawa et al. (1995))).

3. マルチカノニカル法で多次元自由エネルギー地形を描く

3.1 蛋白質の自由エネルギー地形とマルチカノニカル・アンサンブル

分子構造が大きく複雑な生体高分子の系は、可能と考えられる状態は極めて多いことが特徴である。第 1 節で述べたように、同一のアミノ酸配列が多様な立体構造を形成する構造多形の機構を理解するため、我々は「自由エネルギー地形」を俯瞰しようと試みている。

カノニカル分布においては、温度 T におけるエネルギー E の確率を与えるエネルギー分布確率 $P_c(E, T)$ は、

$$(3.1) \quad P_c(E, T) = n(E) \cdot \exp(-E/k_B T) / Z_c$$

と表される。ここで、 $n(E)$ は状態密度、 Z_c は $P_c(E, T)$ の分配関数であり、

$$(3.2) \quad Z_c = \sum_E n(E) \cdot \exp(-E/k_B T)$$

として和をとったものである(久保(1952))。

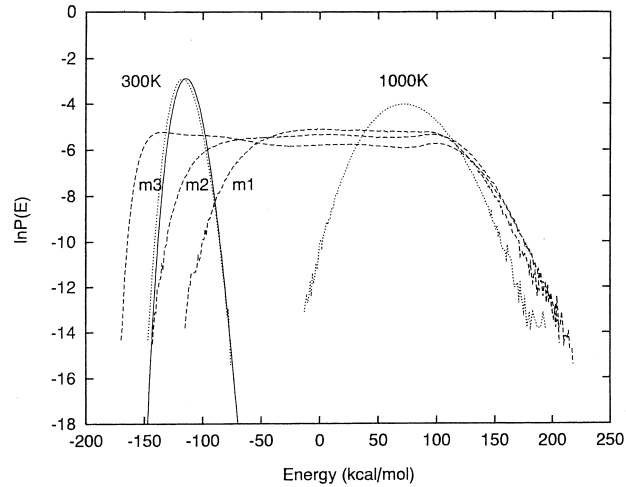


図 5. N 末と C 末とをそれぞれアセチルおよび N-メチルでブロックした 5 残基のペプチド(メト-エンケファリン)の真空中でのマルチカノニカル分子動力学計算の結果得られたエネルギー分布 (Nakajima et al. (1997)). まず 1000K において通常のカノニカル分子動力学計算を行い(1000K の文字の下の点線), 式 (3.3) のポテンシャル・エネルギーに従って分子動力学を行うと, あるエネルギー範囲 (m1 ~ m3) でフラットなエネルギー分布(破線)が実現する. 1000K におけるエネルギー分布 $P_c(E, T_0)$ の, 1000K より低温側の分布関数の見積りの精度を, m1 から m3 にかけて行くに従って, より低温側までフラットな分布が得られるようになる(手法の詳細は, Nakajima et al.(1997)). その後に, 式 (3.6) に従って 300K のカノニカル・アンサンブル(実線)を得る. このマルチカノニカル・アンサンブルから得られた 300K のカノニカル・アンサンブル(実線)は, ある初期構造からカノニカル分子動力学を行った場合の 300K でのエネルギー分布(300K の文字の右の点線)とは若干異なっていることがわかる.

我々の目的は, このカノニカル分布に従うカノニカル・アンサンブルを精度良く求め, それを表現する自由エネルギー地形を最終的に得ることである. このために, Berg らが提案したマルチカノニカル・アンサンブル (Berg and Neuhaus (1991)) が得られるマルチカノニカル分子動力学法を, 我々は開発している (Nakajima et al. (1997), Higo et al. (1997), Nakajima (1998)). この新しい計算法では, エネルギー分布が, 人為的にほぼ一定値を持つフラットな分布 $P_{mc}(E)$ となるように, エネルギー関数 E を以下のように E_{mc} と変形させて分子動力学を行い, 効率の高い構造探索を行う(図 5).

$$(3.3) \quad E_{mc} = E + k_B T_0 \ln P_c(E, T_0)$$

$$(3.4) \quad P_{mc}(E) = n(E) \cdot \exp(-E_{mc}/k_B T) / Z_{mc}$$

$$(3.5) \quad Z_{mc} = \sum_E n(E) \cdot \exp(-E_{mc}/k_B T)$$

この際, 求めたい温度でのアンサンブル間の障壁を越えるほど高い温度に T_0 を設定して, あらかじめカノニカル・アンサンブルのエネルギー分布確率 $P_c(E, T_0)$ を得ておく. こうして, 初期構造に依存せずに多くの可能な構造を探索し, さらにそれら局所構造をどの程度の確率でとるかまで推定する. この際, re-weighting の公式 (Ferrenberg and Swendsen (1988)) によって,

$$(3.6) \quad P_c(E, T) Z_c = P_{mc}(E) \cdot \exp(E_{mc}(E)/k_B T_0 - E/k_B T) Z_{mc}$$

から、 T_0 以下の任意の温度でのカノニカル・アンサンブルを、この人為的なマルチカノニカル・アンサンブルから得られる。 Z_c と Z_{mc} は、式 (3.2), (3.5) のように E に対して和をとっているので定数項として処理される。

以上の操作によって、マルチカノニカル分子動力学の過程で発生した個々の分子構造・分子配置がもつ系全体のエネルギー E から、ある温度における対応する確率 $P_c(E, T)$ が得られ、その系の potential of mean force ($-k_B T_0 \ln P_c(E, T)$) すなわち自由エネルギーが算出されるので、多次元の構造空間中に自由エネルギー地形を描くことができる。ここで紹介する計算は、全て、我々の開発したプログラム PRESTO (Morikami et al. (1992)) を用いた。

ところで、シミュレーション計算に精通されている方は、この手法は、式 (3.3) において、右辺第2項 ($k_B T_0 \ln P_c(E, T_0)$) をアンブレラ・ポテンシャルとした、アンブレラ・サンプリング法 (Torrie and Valleau (1977)) の一つのバリエーションであることがすぐに理解できよう。式 (3.6) も、アンブレラ・ポテンシャルを用いた計算後の re-weighting 操作と全く等価である。ただし、アンブレラ・サンプリング法が、それぞれの対象とする系ごとに最適な反応座標を選択し、アンブレラ・ポテンシャルを作るといふ必ずしも容易でない操作が必要であるのに対し、マルチカノニカル・アンサンブルは、式 (3.3) から自動的に導出できるという大きな利点があることを指摘しておきたい。

3.2 蛋白質構造の多次元自由エネルギー地形

それでは、具体的なペプチド鎖の構造が作る自由エネルギー地形を見てみよう。

まず、N 末と C 末とをそれぞれアセチルおよび N-メチルでブロックした 2 残基のペプチドに対して、水中でのシミュレーション計算を行ってマルチカノニカル・アンサンブルを得た後、300K における自由エネルギー地形を描いた (Nakajima et al. (2000))。図 6 には、側鎖が水素原子だけの最も単純なアミノ酸であるグリシンのダイマーにおいて、両端間の距離 d と、初めのグリシンの主鎖内部回転角 ψ と 2 番目のグリシンの主鎖内部回転角 φ との 3 つの軸で構造空間を表した場合の 3 次元の自由エネルギー地形を示す。番号 1 は最安定状態であり、2~9 の番号がそれぞれ異なる構造に対応する準安定状態である。それらの構造を図 7 に示す。ペプチド鎖内の水素結合によって安定化された種々のターン構造や、水和によって安定化されたジグザグ構造、まっすぐに伸びた構造などが多く出現している。これらの最安定状態、準安定状態の自由エネルギー値にはあまり大きな差がなく、環境によって容易にある構造のみが選択されることがわかる。また、水溶液中で無限に様々な構造を取りえるわけではなく、300K では、ある有限の構造のみを安定状態または準安定状態として偏ってとっており、それらの出現確率は蛋白質中に出現する統計的確率と極めて良い類似があることが見いだされた。

別の例として、7 残基のアミノ酸の N 末と C 末とをそれぞれアセチルおよび N-メチルでブロックしたさらに長いペプチドの水溶液中の構造の 300K における頻度の分布を図 8 に示す (Higo et al. (2001))。この場合には、自由度が多いために、自由エネルギー地形の軸を主成分解析の結果得られた軸とした。すなわち、まず、出現した構造間の違いを、分子中の全ての原子が最も良く重なるように分子構造を並進・回転した時の、原子位置のズレの平均二乗偏差の平方根として定義する。次に、 m 個の構造に対して $m(m-1)/2$ の総当たりで構造間の違いを計算し、主成分解析を行う。その結果得られた、主軸と 2 番目の軸による二次元空間において、頻度が高い構造すなわち自由エネルギーの低い構造を、図 8 では表している。すなわち、本質的に、この図は自由エネルギー地形を表現している。さまざまな立体構造がアンサンブル中に観測され、多くは規則的構造をとっていないが、この不規則構造が主な構造であることは、このペプチドを合成しその水中での実験によって実際に確認されている。ところで、中には、図 8 中に示した代表構造のように、ペプチド内部に水素結合を複数含む α -ヘリックスや β -ヘア

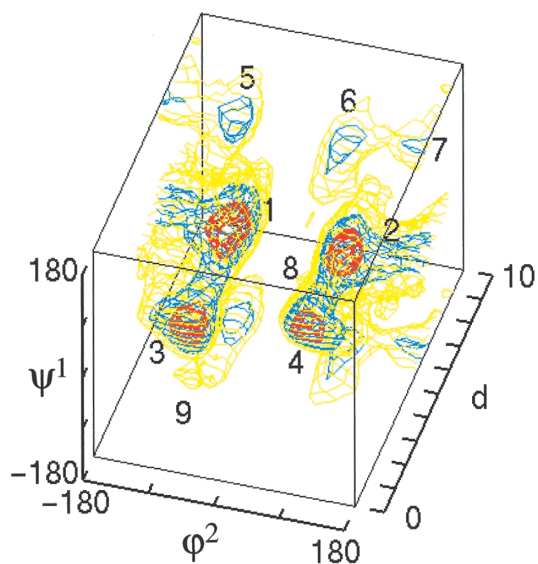


図 6. N 末と C 末とをそれぞれアセチルおよび N-メチルでブロックした水溶液中のグリシン・ダイマー (Ace-Gly-Gly-NMe) による 3 次元の自由エネルギー地形 (Nakajima et al. (2000)). 反応座標として, アセチル基 (Ace) のカルボニル酸素原子と N-メチル基 (NMe) のアミド水素間の距離 d , および最初の残基の主鎖二面角 ψ_{i+1} と次の残基の主鎖二面角と 2 番目のグリシンの主鎖内部回転角 φ_{i+2} の 3 つを用いた. 赤, 青, 黄色の等高線は, それぞれ, RT, 2RT, 3RT のエネルギー値に対応する. 1 の番号の構造が最も自由エネルギー (potential of mean force) が低く, 2 から 9 までの番号は, それぞれが準安定な構造となっているものである.

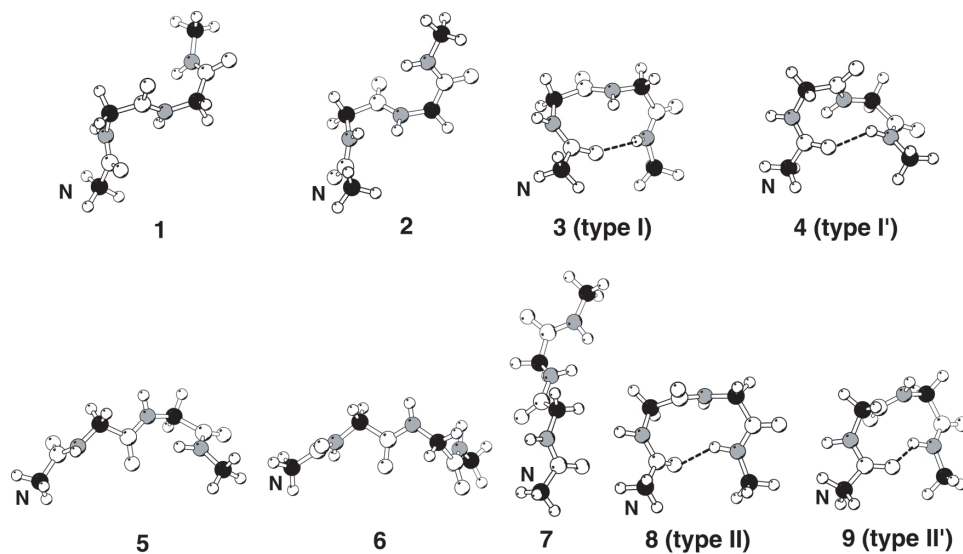


図 7. 水溶液中のペプチド Ace-Gly-Gly-NMe の最安定構造および準安定構造 (Nakajima et al. (2000)). 黒, 灰色の球は, それぞれ主鎖の $C\alpha$, それ以外の炭素原子, 窒素原子とを示し, 小さな白球は水素原子を示す. 点線は水素結合を示す. 各番号は, 図 6 の自由エネルギー地形中の番号に対応しており, 1 が最安定構造である. また, 3, 4, 8, 9 の構造は, それぞれタイプ I 型, I' 型, II 型, II' 型の β -ターンに対応する. Gly は光学活性体でないため, I' 型は I 型の鏡像, II' 型は II 型の鏡像の関係にそれぞれなっている. N という文字で, このペプチドの N 末側を示す.

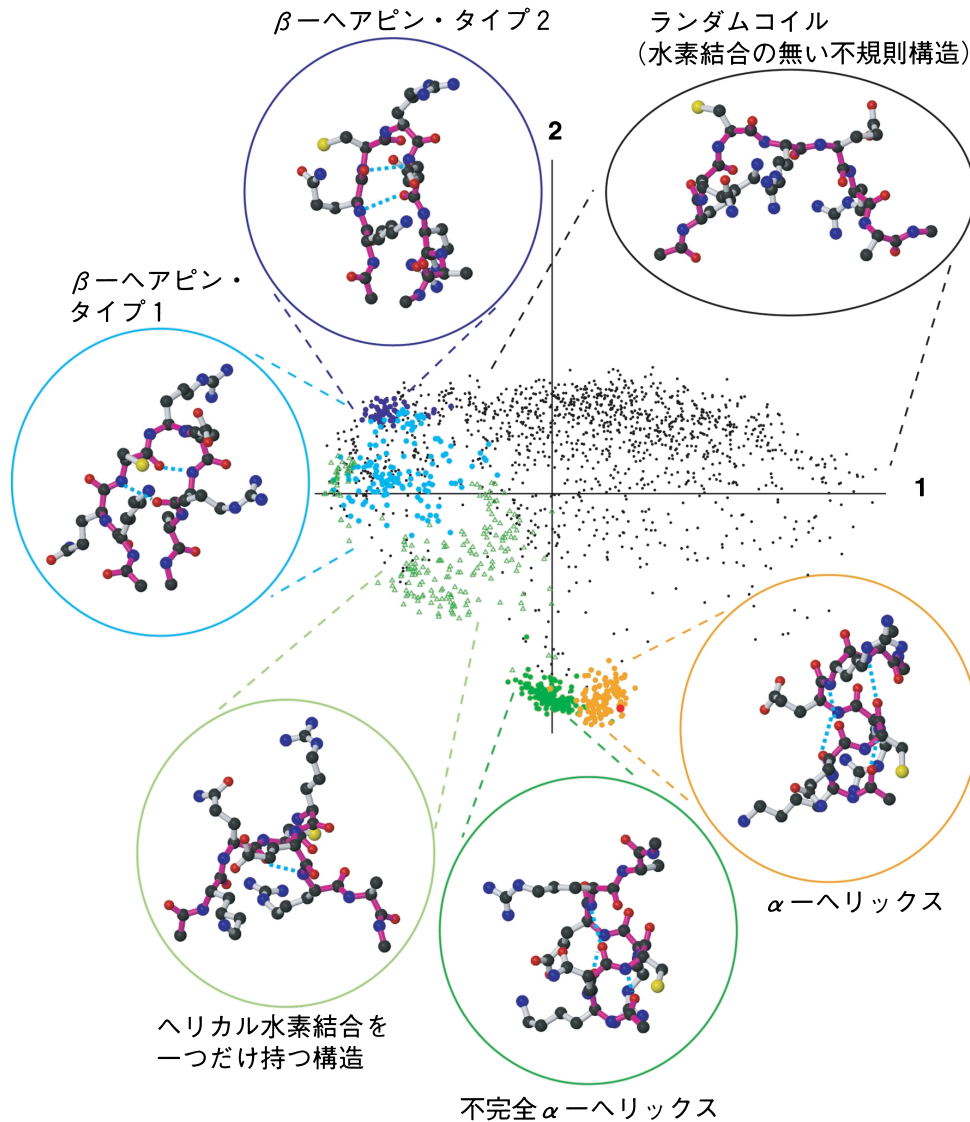


図 8. 水溶液中のペプチド Ace-Lys-Gln-Cys-Arg-Glu-Arg-Ala-NMe の 300K における自由エネルギー地形。出現した構造間の違いを、分子中の全ての原子を最も良く重ねた時の原子位置のズレの平均二乗偏差の平方根として定義し、総当たりで構造間の違いを計算して主成分分析を行った結果。各点の構造は、オレンジ：α-ヘリックス，緑：部分的に壊れたα-ヘリックス，薄緑：中央のヘリカルな水素結合だけができて壊れたα-ヘリックス，青と紫：2種類のβ-ヘアピン構造，黒：不規則構造，赤：同じアミノ酸配列を持つ c-Myb 蛋白質中の部分フラグメント構造(α-ヘリックス)である。各クラスターの代表的な構造を図中に示す。分子の色は、紫色で主鎖を表し，赤，青，黒，黄色が，それぞれ酸素，窒素，炭素，硫黄の原子を表す。水素原子は省略してある。主鎖内の水素結合を水色の点線で示す。

ピン構造等の規則的な構造も出現している．すなわち，実験的に不規則構造と言われている状態は，実は種々の構造の平衡状態であり，環境のわずかな変化によって，容易にある規則構造が他よりも安定となろう．このことが，本章の冒頭で述べた構造多形の現われる理由である．

これらのシミュレーション計算の妥当性は，エネルギー関数として用いられる力場に大きく依存する．実際，図 8 は AMBER param96 という力場で得られたが，別のポピュラーな AMBER param94 という力場を用いると， α -ヘリックスの構造だけが頻出し，実験とは一致しない結果が得られてしまう(Ono et al.(2000) Higo et al.(2001))．

3.3 準安定状態間の遷移状態周辺も含む自由エネルギー地形の観測

それでは，図 6 や図 8 における最安定状態および準安定状態間の高い自由エネルギー障壁はどのようにになっているのだろうか．実際， α -Ala-Pro-ペプチド・ダイマーにおけるプロリン(Pro)は，その直前のペプチド結合においてシス型とトランス型の 2 種類の異性体があり，それらの安定状態と準安定状態間の自由エネルギー差は，3.1 節で説明したマルチカノニカル・アンサンブル解析から良く再現できる(図 9 の点線)．しかし，この 2 つの安定状態間の高いエネルギー障壁部分は，効率化された構造探索法であるマルチカノニカル法においても，十分たくさんの構造が探索されてはいないことが，図 9 には歴然と示されている．理論的には，充分長いトラジェクトリーによってマルチカノニカル・アンサンブルを作れば，このエネルギー障壁部も精度良くカバーされるはずである．しかし，エネルギーの高い状態には極めて多くの歪んだ分子構造があり，このエネルギー障壁に対応する構造は，その中の一部でしかないので，図 9 の横軸に沿った探索は安定状態の探索に比べると，どうしても不十分になってしまうのである．

この障壁は，ペプチド構造の構造的遷移状態に対応しており，より一般的には，蛋白質構造形成におけるある種の遷移状態にも対応している．速度論的実験から得られる構造転移に伴う「活性エネルギー」が，実際にはどのような遷移状態の構造に対応しているかは，未だに良くわかっていない．

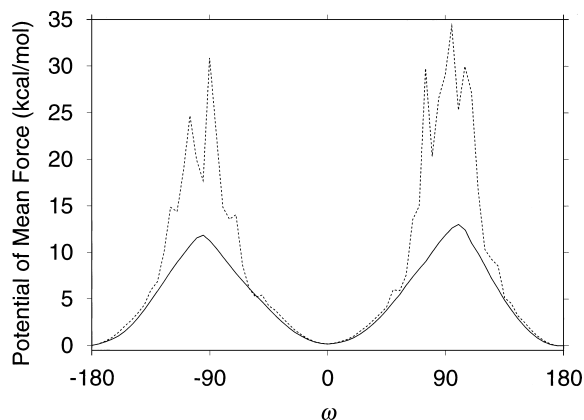


図 9. 真空中の Ace-Ala-Pro-NMe ペプチドにおけるプロリンのシス-トランス異性化反応に伴う自由エネルギー地形(Ono et al. (1999))．Ala と Pro の間のペプチド面の内部回転角 ω を横軸にとつてある．点線が，通常のマルチカノニカル法によるもので，マルチカノニカル WHAM によるものを実線で示す．1 回のシミュレーションで，どれも 10^7 ステップのサンプリングを行っている．

我々は、このように高いポテンシャル障壁を持つ遷移状態周辺の自由エネルギーの値とその地形を高精度に算出し、構造遷移状態の構造を理解するため、従来からよく用いられてきた、反応座標をパラメータとするアンブレラ法と効果的な探索が可能なマルチカノニカル法とを融合した「マルチカノニカル WHAM 法」を最近開発した (Ono et al. (1999)) .

オリジナルの WHAM (Weighted Histogram Analysis Method) 法 (Kumar et al. (1992)) では、正規のポテンシャル E に対して、反応座標 x によるアンブレラ・ポテンシャル U を加えて、

$$(3.7) \quad E_{WHAM} = E + U(x)$$

と変形し、 E だけではエネルギーが高すぎてサンプリングできない構造でも、反応座標 x が満たされるような構造に関してはサンプリングできるようにする。そして、反応座標 x を複数のウィンドーに分割して、各ウィンドー毎に別々にシミュレーションを行う。最後に、 x に沿ってオーバーラップした確率密度を用いて、反応座標に沿った自由エネルギー地形が得られる。WHAM 法の問題点は、反応座標の取り方に任意性があり、最適な反応座標は容易に得られないことである。

マルチカノニカル法を援用することによって、たとえ最適な反応座標でなくとも、その座標に沿って広く周辺の構造も迅速に探索できると期待される。実際、式 (3.7) の E_{WHAM} に、さらに温度 T_0 におけるエネルギー分布 $P_c(E + U, T_0)$ を加え、ポテンシャルを以下のように変形する。

$$(3.8) \quad E_{mcW} = E + U(x) + k_B T_0 \ln P_c(E + U, T_0)$$

と変形する。分割した反応座標 x の各ウィンドー毎に別々に計算した結果から温度 T における各カノニカル分布をまず得て、それらをさらに組み合わせることによって、反応座標に沿った道筋における自由エネルギー地形が得られる。

この手法を分子動力学法に応用し、N 末と C 末とをそれぞれアセチル基および N-メチル基でブロックした-Ala-Pro-ペプチド・ダイマーにおける、プロリン (Pro) のシス-トランス異性化反応に伴う自由エネルギーを計算した (Ono et al. (1999)) . この際、反応座標としては、アラニンの $C\alpha$, O とプロリンの $C\delta$, $C\alpha$ が作る擬二面角 ζ を、アラニンの $C\alpha$, C とプロリンの N, $C\alpha$ が作る通常の二面角 ω の代わりに、より効率的な反応座標として導入した。その結果、図 9 の実線に示されるように、単なるマルチカノニカル法では迅速には探索できなかったエネルギー障壁部分が十分探索され、精度の高い解析が可能となった。図 9 中の potential of mean force の値の精度を見積もるため、 10^7 ステップのサンプリングを独立に 5 回行いその再現性を調べたところ、安定状態においても構造遷移状態においても、標準偏差として約 0.1 kcal/mol を得た。図 9 の点線は 10^7 ステップの通常マルチカノニカル法によって得られた potential of mean force であり、二面角 $\omega = \pm 90^\circ$ 近辺の構造は、 10^7 ステップ中にわずかなイベントしか観測されず、そのためにギザギザの分布となっている。今回のマルチカノニカル WHAM 法では、二面角 $\omega = \pm 90^\circ$ 近辺の構造も十分な数が観測されたため、実線のようにスムーズな分布となり、通常マルチカノニカル法に比べて、はるかに高い精度の計算が行われたことがわかる。

図 10 には、両端間の距離 d とシス-トランス異性化をおこすペプチド面の内部回転角によって表された自由エネルギー地形を示す。シス ($\omega = 0^\circ$) とトランス ($\omega = 180^\circ$) の状態の間の構造遷移状態 ($\omega = 90^\circ$ および $\omega = 270^\circ$) における地形が鞍点 (saddle point) となっており、小さな d の値のまま、つまりターン構造をとったままで遷移状態を越えるパスと、大きな d をもつ伸びた構造で遷移状態を越えるパスの 2 つがあることが良くわかる。

また、得られた活性化自由エネルギーは、速度論的実験から知られている値 (17 ~ 18 kcal/mol) (Grathwohl and Wuthrich (1981)) に近く、遷移状態の構造は、従来提案されていた構造 (Fischer et al. (1994)) に一致していた。

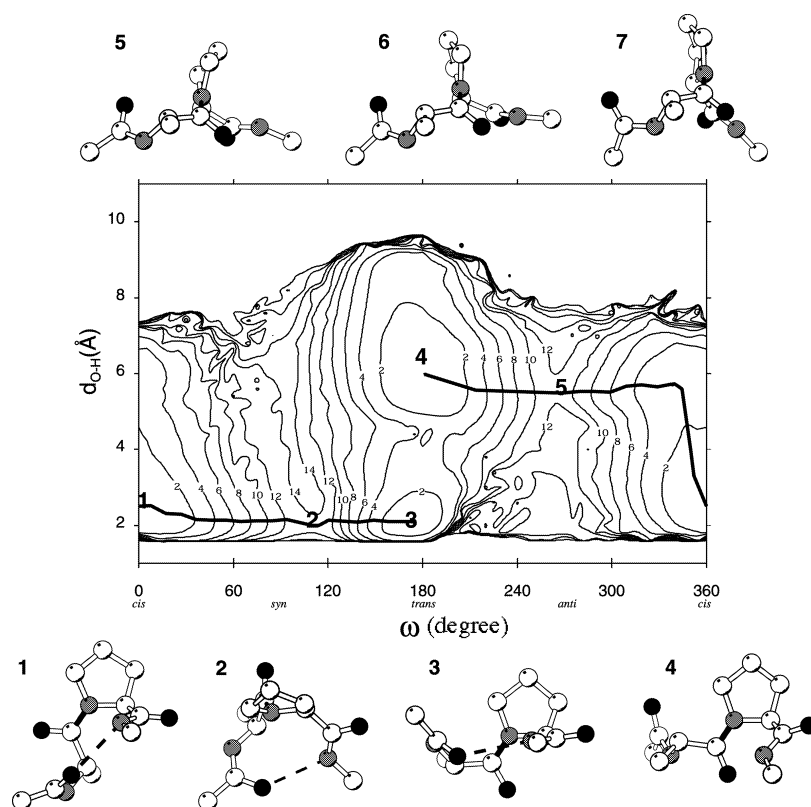


図 10. マルチカノニカル WHAM によって得られた, 真空中の Ace-Ala-Pro-NMe ペプチドにおけるプロリンのシス-トランス異性化反応に伴う自由エネルギー地形 (Ono et al. (1999)). 座標として, Ala と Pro の間のペプチド面の内部回転角 ω と, Ace のカルボニル酸素原子と NMe のアミド水素間の距離 d を用いた. 等高線は 2 kcal/mol 刻みの potential of mean force を表す. 図中の太い数字 1~7 は, 安定構造 (1, 3, 4) と, 構造遷移状態 (2, 5-7) の位置を表し, 対応する分子構造 (水素は省略してある) を欄外に描く. 分子の色は黒が酸素, 灰色が窒素, 白が炭素. 図中に, 遷移状態から最も急な傾斜で安定状態に移るとした場合の経路を, 太線で描く.

4. おわりに

蛋白質や核酸の構造解析の現場では, 数理解析の新しい手法がブレーク・スルーとなることも多いが, 必ずしも数学的に美しい手法が生き残るとはかぎらない. Distance Geometry 計算による生体高分子構造決定の歴史は, その事情を良く示している. 現在は, 構造決定においてもシミュレーションによる解析においても, 効率が高く用途の広い構造探索手法が求められている. この場合, 遺伝的アルゴリズム (genetic algorithm) のようにヒューリスティックで情報科学的な手法よりも, むしろ計算後にカノニカル・アンサンブルを再構築できる統計物理的な手法の方が用途が広い. 実際, 本稿で紹介した自由エネルギー地形の計算は, 生体高分子と薬物とのドッキング・シミュレーションに利用すれば, 単に薬物との複合体モデルを構築するだけでなく, 薬物が結合する強さを表す結合自由エネルギーを算出することも可能であり, 今後は医薬品開発への応用も期待されている.

多次元空間中の自由エネルギー地形を客観的に表現し理解するためには, 得られたアンサン

ブル中のたくさんの構造間の違いを主成分分析によって解析することが好ましい。この場合、得られた軸の構造的・物理化学的意味を調べると、これまで未知であった現象に対応する反応座標に相関があることが理解され、帰納的な発見につながることも多い。実際、図8における主成分分析から、全体としては不規則構造が主な系においても、マイナーではあるが、ある固有の規則構造が存在していることが初めて分かった。また、本稿では省略したが、我々は、同様の手法で蛋白質の部分構造を解析している。例えば、免疫現象に関わる抗体分子が、異物である抗原を認識する部位のループ構造の構造多形を調べると、その部分構造とアミノ酸配列との間に、ある規則性があることが見いだされた(Shirai et al. (1998), Kim et al. (1999))。このように、人間の目ではあらかじめ分かりえないことを発見できるのは、主成分分析の極めて強力な点であり、その利用価値を高く評価している。

謝 辞

本稿を書くにあたり、緒方一博(横浜市大医)、木寺詔紀(京大理・化学)、中井孝尚(鐘淵化学工業)、西村善文(横浜市大総合理)、守川壮一(鐘淵化学工業)の方々(敬称略)の協力を得たことをここに記し、感謝する。

付 録

A.1 多次元尺度法(Embedding法)による蛋白質の立体構造再構築法

多次元尺度法(Crippen (1981))を利用するには、観測された $\{d_{ij}^{obs}\}$ から、中心からの位置ベクトル $\{r_{i0}^{obs}\}$ からなるメトリック $\{g_{ij}\} = \{r_{i0}^{obs} \cdot r_{j0}^{obs}\}$ を要素とする行列を、以下の数学的操作によって $d_{i0} = |r_{i0}^{obs}|$ を求めて作る。

$$(A.1) \quad d_{i0}^2 = \left(\sum_{j=1}^N (d_{ij}^{obs})^2 \right) / N - \left(\sum_{j=2}^N \sum_{k=1}^{j-1} (d_{jk}^{obs})^2 \right) / N^2$$

$$(A.2) \quad g_{ij} = (d_{i0}^2 + d_{j0}^2 - d_{ij}^2) / 2$$

このメトリック行列は、 N 個の水素原子に対して $N \times N$ の正値実対称行列となっていて対角化可能である。次に、この行列を対角化して固有ベクトルを求めるが、この作業は、 N 次元楕円体の主軸を求めることにほかならない。この楕円体のうちで固有値の大きな(主軸の長い)3つまでを選択し、その3つの主軸の張る3次元空間に(N 次元の)ベクトル $\{r_{i0}^{obs}\}$ を投影することによって、3次元座標 $\{R_{i0}\}$ を得る。この操作は、多次元尺度法による主な3つの主成分を選んでいくことと数学的に等価である。

この場合、もし、生体分子を構成するすべての原子に対して、その (i, j) 原子間の距離 $\{d_{ij}^{obs}\}$ が正確に与えられていれば、各原子のデカルト座標を、決定論的に再構築できることが数学的に証明される(Crippen (1981))。しかし、現実には、限られた (i, j) 原子間の距離 $\{d_{ij}^{obs}\}$ のみがNMR実験では観測される。実際、後出するBPTIという蛋白質では、水素も含めた全ての原子数(点の数 N に対応する)は912個であり、全ての原子間の距離情報としては、 ${}_N C_2 = 415,416$ 個が必要である。実際には、 $6 \sim 7 \text{ \AA}$ 以下の近距離にある水素原子間だけに分光シグナルが観測されるため、この場合には多くて500個程度の距離情報しか実験的には得られず、さらに共有結合の情報を加えても高々 N の10倍程度しかない。すなわち、このままでは $\{d_{ij}^{obs}\}$ は、全体でみるとほとんどの要素が不明なままということになっている。この問題の解決法は、以下に述べる。さらに、実験的に観測されるシグナルから原子間距離への翻訳には、ある程度の誤差を見込む必要があり、正確には $\{d_{ij}^{obs}\}$ でなく、それぞれの上限と下限である $\{d_{ij}^{max}\}$ と $\{d_{ij}^{min}\}$

とが実験値として得られる量である。また、後で簡単に補正ができるため大きな問題ではないが、メトリック行列を対角化した後に 3 つの成分だけを選び他を無視することにより、当然、再構築された分子構造の体積は小さくなっており、Radius of Gyration の値で約 20%ほど低く見積もっていることになる。これらが、Distance Geometry 計算において、多次元尺度法をそのままでは利用できない理由である。

具体的な計算のためには、原子間の不明な距離を、距離が既知のペアの情報を用いて、つじつまが合うように推測する。この作業は、バウンダリー・スムージングと呼ばれ、再構築する原子の集合が 3 次元空間中の点群であるための必要十分条件を定義した Blumenthal の定理 (Blumenthal (1970)) に基づいて、 d_{ij} の存在可能範囲を推定する。まず、既知の情報として共有結合から自明なものと実験によって観測されたものから、 $\{d_{ij}^{max}\}$ と $\{d_{ij}^{min}\}$ とを決め、三角不等式を満たすように初期値を作る。以前は、この初期値だけで立体構造の再構築が行われていたが、再構築された立体構造には偏りがあり、他の可能な立体構造を提出できないことが明らかとなった。その後、初期値をもとに再帰的な操作によって、完全に全てのペアが三角不等式を満たすように、 N 個の点間の距離をある幅を有して決めるという操作(メトライゼーションと呼ばれる)が行われるようになった。この操作は、数学的には最短経路問題の解法アルゴリズムである Floyd の方法 (Aho et al. (1983)) が利用される。メトライゼーションを行うことによって偏りが無い構造の再構築が可能とはなった (Havel (1990)) が、この作業には大きな N に対しては膨大な CPU 時間と N^2 のオーダーのメモリー量とを要し、現在では好まれていない。また、値に範囲をもつ $\{d_{ij}^{obs}\}$ から特定の数値を要素とするメトリック行列を作成し数学的な処理をするためには、

$$(A.3) \quad d_{ij}^{min} \leq d_{ij}^{obs} \leq d_{ij}^{max}$$

を満たす $\{d_{ij}^{obs}\}$ を乱数をふって選び出した後、構造構築計算を複数回行い、立体構造の統計的平均をとる作業も不可欠である。

参 考 文 献

- Aho, A., Hopcroft, J. and Ullman, J. (1983). *Data Structures and Algorithms*, Addison-Wesley, Reading, Massachusetts.
- Berg, B. A. and Neuhaus, T. (1991). Multicanonical algorithms for first order phase transitions, *Phys. Lett. B*, **267**, 249–253.
- Blumenthal, L. M. (1970). *Theory and Applications of Distance Geometry*, Chelsea, New York.
- Braun, W. and Go, N. (1985). Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *Journal of Molecular Biology*, **186**, 611–626.
- Chikenji, G., Kikuchi, M. and Iba, Y. (1999). Multi-self-overlap ensemble for protein folding: Ground state search and thermodynamics, *Phys. Rev. Lett.*, **83**, 1886–1889.
- Crippen, G. M. (1981). *Distance Geometry and Conformational Calculations*, Research Studies Press, Chichester; Wiley, New York.
- Ferrenberg, A. M. and Swendsen, R. H. (1988). New Monte Carlo technique for studying phase transitions, *Phys. Rev. Lett.*, **61**, 2635–2638.
- Fischer, S., Dunbrack, R. L., Jr. and Karplus, M. (1994). *Cis-trans* imide isomerization of the proline dipeptide, *Journal of American Chemical Society*, **116**, 11931–11937.
- Grathwohl, C. and Wuthrich, K. (1981). NMR studies of the rates of proline *cis-trans* isomerization in oligopeptides, *Biopolymers*, **20**, 2623–2633.

- Havel, T. F. (1990). The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations, *Biopolymers*, **29**, 1565–1585.
- Havel, T. F. (1991). An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance, *Progress in Biophysics and Molecular Biology*, **56**, 43–78.
- Higo, J., Nakajima, N., Shirai, H., Kidera, A. and Nakamura, H. (1997). Two-component multicanonical Monte Carlo method for effective conformation sampling, *Journal of Computational Chemistry*, **18**, 2086–2092.
- Higo, J., Ito, N., Kuroda, M., Ono, S., Nakajima, N. and Nakamura, H. (2001). Energy landscape of a peptide consisting of α -helix, 3_{10} -helix, β -turn, β -hairpin, and other disordered conformations, *Protein Science* (in press).
- Iba, Y., Chikenji, G. and Kikuchi, M. (1998). Simulation of lattice polymers with multi-self-overlap ensemble, *J. Phys. Soc. Japan*, **67**, 3327–3330.
- Kim, S. T., Shirai, H., Nakajima, N., Higo, J. and Nakamura, H. (1999). Enhanced conformational diversity search of CDR-H3 in antibodies: Role of the first CDR-H3 residue, *Proteins*, **37**, 683–696.
- 久保亮五 (1952). 共立全書『統計力学』, 13–61, 共立出版, 東京.
- Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. and Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *Journal of Computational Chemistry*, **13**, 1011–1021.
- 京極好正・月原富武 (1997). 『構造生物学とその解析法』, 100–155, 共立出版, 東京.
- Morikami, K., Nakai, T., Kidera, A., Saito, M. and Nakamura, H. (1992). PRESTO: A vectorized molecular mechanics program for biopolymers, *Computers and Chemistry*, **16**, 243–248.
- Morikawa, S., Ogata, K., Sekikawa, A., Sarai, A., Ishii, S., Nishimura, Y. and Nakamura, H. (1995). Determination of the NMR solution structure of a specific DNA complex of the Myb DNA-binding domain, *Journal of Biomolecular NMR*, **6**, 294–305.
- Nakai, T., Kidera, A. and Nakamura, H. (1993). Intrinsic nature of the three-dimensional structure of proteins as determined by distance geometry with good sampling properties, *Journal of Biomolecular NMR*, **3**, 19–40.
- Nakajima, N. (1998). A selectively enhanced multicanonical molecular dynamics method for conformational sampling of peptides in realistic water molecules, *Chemical Physics Letters*, **288**, 319–326.
- Nakajima, N., Nakamura, H. and Kidera, A. (1997). Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides, *Journal of Physical Chemistry*, **B101**, 817–824.
- Nakajima, N., Higo, J., Kidera, A. and Nakamura, H. (2000). Free energy landscapes of peptides by enhanced conformational sampling, *Journal of Molecular Biology*, **296**, 197–216.
- 中村春木・有坂文雄 (1997). 『タンパク質のかたちと物性』, 1–20, 共立出版, 東京.
- Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations, *FEBS Letters*, **229**, 317–324.
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices, *Cell*, **79**, 639–648.
- Ono, S., Nakajima, N., Higo, J. and Nakamura, H. (1999). The multicanonical weighted histogram analysis method for the free energy landscape along structural transition paths, *Chemical Physics Letters*, **312**, 247–254.

- Ono, S., Nakajima, N., Higo, J. and Nakamura, H. (2000). Peptide free-energy profile is strongly dependent on the force field: Comparison of C96 and AMBER95, *Journal of Computational Chemistry*, **21**, 748–762.
- Shirai, H., Nakajima, N., Higo, J., Kidera, A. and Nakamura, H. (1998). Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics, *Journal of Molecular Biology*, **278**, 481–496.
- Torrie, G.M. and Valleau, J.P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *Journal of Computational Physics*, **23**, 187–199.
- Wlodawer, A., Walter, J., Huber, R. and Sjolín, L. (1984). Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II, *Journal of Molecular Biology*, **180**, 301–331.

Multidimensional Analyses for Three Dimensional Structures and Energetics of Biopolymers

Junichi Higo and Satoshi Ono

(Department of Bioinformatics, Biomolecular Engineering Research Institute)

Nobuyuki Nakajima and Haruki Nakamura

(Institute for Protein Research, Osaka University)

In order to determine the three dimensional structures of biological macromolecules such as proteins and DNA, distance geometry calculation is applied for the NMR experimental information of distances between hydrogen atom pairs. The distance geometry calculations are based on the algorithms of embedding and multi-dimensional simulated annealing. The principle and applications of the latter method developed by the authors are described. For the analysis of dynamic structure formation of those biopolymers, we need free energy landscapes in the multi-dimensional conformational space, which are obtained by the enhanced conformational sampling methods developed recently. The authors' group has recently developed the multicanonical molecular dynamics method, and applied it to many biological molecular systems, peptides and local fragments of proteins. Principal component analysis is a powerful tool to interpret those complicated free energy landscapes, and very new findings have been noticed.

Key words: Distance geometry, simulated annealing, free energy landscape, multicanonical ensemble, principal component analysis.