

改良ローカルマッチング法によるタンパク質分子の標準構造の生成

安部晴男、水戸三千秋、日比野真也
(西日本工業大学)

1. まえがき

分子モデリング・システムでは、最初に、コンピュータの内部に、取り扱うタンパク質分子の立体構造を生成する必要がある。従来、初期立体構造を生成するための基となるデータとしては、Protein Data Bank(PDB)¹⁾に蓄積されている、X線結晶構造解析によるタンパク質分子の構成原子の三次元位置座標のデータ(これを参照構造と呼ぶ)が用いられている。しかし、X線結晶解析における分解能不足のため、水素原子が欠落していて、分子内の原子結合距離・結合角が、誤差を含んでいる。そこで、水素原子を含んでいて、標準的な原子結合距離・結合角をもつ原子団からなる剛体と見なせるユニット(これを標準ユニットと呼ぶ)を、順々に連結し、参照構造と類似の構造を生成する(これを標準構造と呼ぶ)ことを考える。この問題は、全ての原子の位置を変数とする非線形多変数最小2乗問題となる。

我われは、この近似解法として、既に、ローカルマッチングによる標準構造生成法を提案し、数残基の生成結果²⁾、及び Glucagon に適用した結果³⁾を報告した。局所的には参照構造と類似の標準構造を生成することができたが、N末端からC末端へ順々に生成するにつれて、元の参照構造と新しくできつつある生成構造がグローバルな座標系では、少しずつ隔たってくる傾向にある。そこで、今回は、標準ユニットを用いて生成した後、更に、標準ユニットのいくつかの集合からなるユニットを新たなユニットとみなして、参照構造とグローバルに一致するように補正を行った。この改良ローカルマッチング法を実際のタンパク質分子に適用した結果について報告する。

2. 標準ユニットとその集合からなるユニット

少数の原子(水素原子を含む)から構成され、標準の原子結合距離・結合角をもつ、剛体とみなせる最小の構造単位である、手・足・ボディからなる標準ユニットのボディ間は軸回転しうる化学結合(ボンドと呼ぶ)で結ばれていて、同一の標準ユニットに属する原子はどのボンドの回転に対しても

相互の空間的位置関係は変化しない。標準ユニットの局所座標系は、足を原点、脚をZ軸正方向にとるように定めておく。具体的な原子座標値は、ECEPP⁴⁾のデータを用いて、20種類のアミノ酸残基毎に作成した。図1に、標準ユニットと標準ユニットの集合からなるユニットの例を示す。

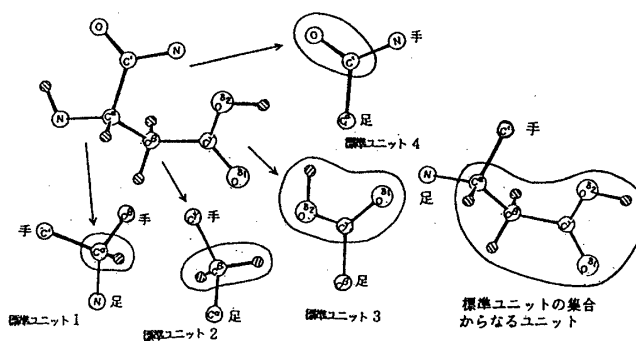


図1 標準ユニットとその集合からなるユニットの例

このように標準ユニットを定義すると、タンパク質分子は、標準ユニット U_{i-1} の腕であり、次の標準ユニット U_i の脚でもある軸回転し得るボンドを共有し、それらが次々に連結されて構成されているものであるとみなすことができる。

3. 改良ローカルマッチングによる標準構造の生成

3.1 標準ユニットによる生成

標準構造の生成は、最初のユニットから順に連結した2つのユニットを参照構造に最適マッチングさせるように行う。いま、(i-1)番目のユニット U_{i-1} まで生成を終わっているとする。この生成ユニット U_{i-1} の腕

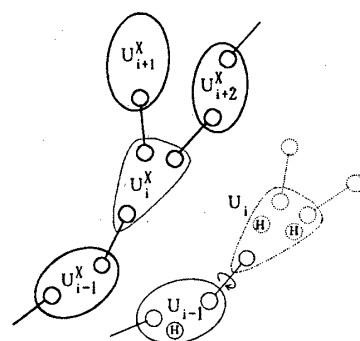


図2 参照ユニットと生成ユニット

に、参照構造 U_{i-1}^X に対応する標準ユニット U_{i-1}^S の脚を重ねて連結することによって U_i を生成する手順の要点は次の通りである³⁾。

- (1) 参照ユニット U_{i-1}^X の脚がZ軸、足が原点になるように U_{i-1}^X と U_{i-1}^S の2つのユニットを座標変換する。

- (2) U_{i-1} を座標変換して、次の生成ユニット U_i の脚 (= U_{i-1} の腕) が Z 軸に、足が原点になるようにする。
- (3) 前参照ユニット U_{i-1}^x と前生成ユニット U_{i-1} が最適マッチングするように参照構造を Z 軸回転する(付録 A)。
- (4) U_i^x に対応した標準ユニット U_i^s をもってきて、足と脚軸を一致させる。
- (5) 標準ユニット U_i^s と参照ユニット U_i^x との最適マッチングを行う(付録 A)。
- (6) この結果得られた U_i^s が生成ユニット U_i となる。

3.2 標準ユニットの集合からなるユニットによる補正

次に、生成された標準構造において、標準ユニットのいくつかの集合からなるユニットを新たなユニットとみなし、これを用いて、参照構造とグローバルに一致するように3.1の手順に従って補正を行う。

4. 参照構造と標準構造の重ね合わせ

生成された標準構造と元の参照構造とを比較するために、2つの構造の座標重心を一致させた後、対応する原子間の距離の2乗和を最小にするような正規直交回転行列を求めて2つの構造を重ね合わせる(付録 B)。

5. あとがき

本方法に基づいて、マイコンPC-9801RA上で Borland C++ によりプログラムを作成し、X線結晶構造データ¹⁾(参照構造データ)を用いて実際のタンパク質に対して標準構造の生成を行った。マイコン上で実用的な速度で実行できることが確かめられた。得られた標準構造と元の参照構造の重ね合わせの結果、及び、この方法の有効性については講演で述べる。

6. 参考文献

- 1) Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M., "A computer-based archival file for macromolecular structures", J. Mol. Biol., 112, 535-542(1977).
- 2) 安部晴男、水戸三千秋、本田宗生、廣田悦司: "タンパク質分子の標準立体構造の高速生成法"、情報処理学会第44回全国大会、3U-3(1992).
- 3) 安部晴男、水戸三千秋、日比野真也、原田勝: "タンパク質の標準構造の生成"、平成4年度電気関係学会九支連大、No. 245(1992).
- 4) Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A., "Energy parameters in polypeptide. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for naturally occurring amino acids." J. Phys. Chem., 79, 2361-2381(1975).

【付録A: 2つのユニットをZ軸回りの回転により重ね合わせる正規直交回転行列の求め方】

2つのユニットの対応する原子間の距離の2乗和を最小にするような、Z軸回りの正規直交回転行列 $T(\theta)$ を求める。いま、2つのユニットの原子位置行列を P 、 Q とすると

$$p_i = \begin{bmatrix} p_{ix} \\ p_{iy} \\ p_{iz} \end{bmatrix}, \quad q_i = \begin{bmatrix} q_{ix} \\ q_{iy} \\ q_{iz} \end{bmatrix} \quad \text{として}$$

$$P = [p_1 \ p_2 \ \dots \ p_n]$$

$$Q = [q_1 \ q_2 \ \dots \ q_n]$$

ここで p_i 、 q_i は、対応する原子の位置ベクトルである。

Z軸回りの正規直交回転行列 $T(\theta)$ は次式で表される。

$$T(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

この成分は次のようにして求めることができる。

$\Delta = TP - Q$ において、 $\Delta^t \Delta$ を最小化する。

$$\frac{\partial \text{tr}(\Delta^t \Delta)}{\partial \theta} = 0, \quad \frac{\partial^2 \text{tr}(\Delta^t \Delta)}{\partial \theta^2} > 0$$

これより、 $T(\theta)$ の成分が次のように決まる。

$$\cos \theta = \frac{\alpha}{(\alpha^2 + \beta^2)^{1/2}}, \quad \sin \theta = \frac{\beta}{(\alpha^2 + \beta^2)^{1/2}}$$

$$\alpha = \sum_{i=1}^n (p_{ix}q_{ix} + p_{iy}q_{iy}), \quad \beta = \sum_{i=1}^n (p_{ix}q_{iy} - p_{iy}q_{ix})$$

【付録B: 参照構造と標準構造の重ね合わせ】

対応する n 個の原子からなる標準構造と参照構造とを空間回転行列 R によって重ね合わせようとするとき、対応する原子間の距離の2乗和を最小にするような正規直交回転行列 R を次のように求める。

$$\text{標準構造: } P = [p_1 \ p_2 \ \dots \ p_i \ \dots \ p_n]$$

$$\text{参照構造: } Q = [q_1 \ q_2 \ \dots \ q_i \ \dots \ q_n]$$

ここで p_i 、 q_i は、対応する原子の位置ベクトルである。

P を正規直交回転行列 R で回転したとき、対応する Q の原子座標の誤差の2乗和 δ^2 は、

$$\delta^2 = \|RP - Q\|^2 = \text{tr} \{ (RP - Q)(RP - Q)^t \}$$

従って、 δ^2 を最小にする意味での最適マッチングを与える R は、次の条件のもとでの δ^2 の最小化問題となる。

$$RR^t = I \quad (t: \text{行列の転置}, I: \text{単位行列})$$

これは、ラグランジュの未定乗数法により解くことができる。

この結果は、 QP^tPQ^t (対称行列) を固有値分解し、固有値を成分とする対角行列を D 、固有ベクトルを列にもつ正規直交行列を W とすると、 R は次式で表される。

$$R = PQ^tWD^{-1/2}W^t$$