

3次元off-latticeタンパク質モデルの構造最適化手法の開発
(北大院理)○竹内 浩

【序】タンパク質のアミノ酸配列から立体構造をいかに計算するか、その方法の解明が本研究の目的である。この問題を解決するには、莫大な数の安定構造の中から最安定構造を効率的に探す必要がある。本研究では、タンパク質を単純化した3次元off-latticeモデルを用いて、タンパク質の構造最適化手法の開発を行った。

我々は、昨年度の本討論会で、Lennard-Jones (LJ) クラスターの構造最適化アルゴリズムについて報告した。その後、クラスターサイズを500から561に増やし、6つのクラスターについて新規の最適化構造¹⁾を求めるのに成功した。これにより、我々が開発した最適化アルゴリズムの効率が非常に高いことが示された。本研究では、このアルゴリズムを3次元off-latticeタンパク質モデルに拡張した。得られた結果から最適化アルゴリズムについて検討した。

【off-latticeモデル】アミノ酸を疎水基と親水基の2種類のunit (AとB) に単純化し、それらの間の結合距離は1であるとする(図1)。N個のunitから構成されるタンパク質のエネルギーは、以下の式で表される。²⁾

$$E = \sum_{i=1}^{N-2} \vec{r}_{i,i+1} \cdot \vec{r}_{i+1,i+2} - \frac{1}{2} \sum_{i=1}^{N-3} \vec{r}_{i,i+1} \cdot \vec{r}_{i+2,i+3} + 4 \sum_{i=1}^{N-2} \sum_{j=i+2}^N C_3 (r_{ij}^{-12} - r_{ij}^{-6}) \quad (1)$$

ここで、 $\vec{r}_{i,i+1}$ は、 i から $i+1$ へのベクトル、 r_{ij} は i と j の距離である。また、AA 対では、 $C_3 = 1.0$ であり、それ以外の対では、 $C_3 = 0.5$ である。

【最適化】タンパク質の表面で unit を移動させるオペレーター (S_m , 図2) とタンパク質の内部に移動させるオペレーター (I_m , 図3) を用いて、構造に変化を与え、その構造を準ニュートン法により局所最適化する。移動する m 個の unit としては、それらの unit が式

1 内で寄与するエネルギーを計算し、その値

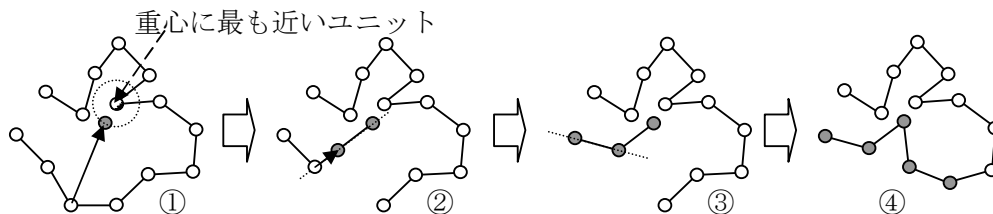


図3 $I_{m=1}$ オペレーター

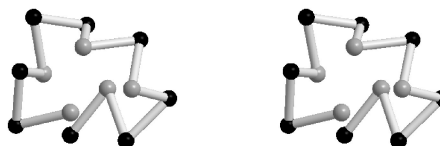


図1 off-lattice タンパク質モデル
ABBABBABABBAB : 疎水基 A を
灰色, 親水基 B を黒色で示す。

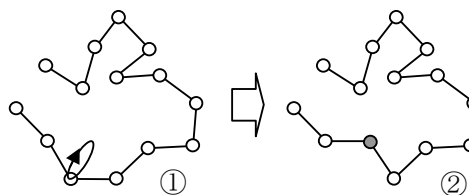


図2 $S_{m=1}$ オペレーター

が最大となる unit の組み合わせを選択する。計算では、 S_m で $m \leq 4$ とし、 I_m で m は 5 以下の乱数とする。また、 $S_{m=1}$ の場合には、エネルギーの最も高い unit に加え、2 と 3 番目に高い unit も選択する (対応するオペレーターをそれぞれ $S_{m=1}^1$, $S_{m=1}^2$, $S_{m=1}^3$ と表記する)。 I_m の最初の段階 (図3の①) では、移動させる unit の位置は、タンパク質の重心に最も近い unit の周

囲にランダムに決定する。この手順で移動する unit 数が m 個であり、これらの移動に付随してさらに複数の unit も移動する (図 3 では、④で見られるように 4 個である)。

計算の主な手順を次に示す。(1) タンパク質の初期構造をランダムに生成し、それを局所最適化する。選択する unit を疎水基に限定する。(2) 得られたクラスターに対して I_m オペレーターを作用させ、その構造を局所最適化する。これでエネルギーが下がらないときは、次のステップに進む。エネルギーが低くなれば、新しい構造を採用し、このステップを繰り返す。(3) 上のステップで得られた構造に対して、 $S_{m=1}^1$ オペレーターを作用させる。局所最適化後のエネルギーが低くなれば、新しい構造を採用し、このステップを繰り返す。そうでない場合には、 $S_{m=1}^2$, $S_{m=1}^3$, $S_{m=2}$, $S_{m=3}$, $S_{m=4}$ オペレーターを順次作用させ、構造を局所最適化する。この過程でエネルギーが低くなった時には、新しい構造を採用し、このステップを初めから実行する。(4) $S_{m=4}$ オペレーターによりエネルギーに改善が見られない時には、選択する unit を親水基に限定してステップ 3 を行い、計算を終了する。

A と B の配列は Fibonacci 配列³⁾に従い、現在までに unit 数が 13, 21, 34 について計算を行った。上記の手順を 1 サイクルとして、このサイクルを各タンパク質モデルに対して 100, 2000, 2000 回繰り返した。13-unit のモデル (S13) の最安定構造を図 1 に示す。

【結果と考察】表 1 に得られた最安定構造のエネルギー値を示す。文献 3 の結果と比較すると、S13 では Liang の最適値を再現し、S21 と S34 については、文献値より安定な構造を求めることができた。したがって、クラスター用構造最適化アルゴリズムは、3次元 off-lattice タンパク質モデルに対しても有効であり、効率的に配座検索を実行することがわかった。

表 2 に最適化構造の検出率をタンパク質モデルと LJ クラスターで比べた。LJ クラスターよりタンパク質モデルの構造の自由度が $N +$

表 1 off-lattice タンパク質モデルの最安定構造のエネルギー

	Liang	Backmann ら	Kim ら	本研究
S13	-26.5066	-26.498	-26.4714	-26.5066
S21	-51.7575	-52.917	-52.7865	-52.9339
S34	-94.0431	-97.261	-97.7321	-98.0757

1 少ないにもかかわらず、検出率はタンパク質の方がクラスターよりも低い。これは、エネルギー曲面における谷の形状の違いを反映していると思われる。

最適化構造が得られたサイクルについて、各オペレーターのエネルギー低下効率を調べた。その結果、 I_m オペレーターのエネルギー低下効率が高く、クラスターと同様に、このオペレーターがアルゴリズムの効率を高く

していると考えられる。今後は、S55 について計算を行い、アルゴリズムの効率をより詳細に検討する予定である。

表 2 最安定構造の検出率 P (最安定構造を得た回数/計算した構造の総数)

Protein	P	Cluster	P
S13	3.6×10^{-3}	LJ13	2.5×10^{-1}
S21	1.1×10^{-4}	LJ21	5.3×10^{-2}
S34	6.1×10^{-6}	LJ34	6.1×10^{-3}

【参考文献】 1) H. Takeuchi, *J. Chem. Inf. Model.*, in press. 2) A. Irbäck et al., *J. Chem. Phys.*, 107, 273 (1997). 3) F. Liang, *J. Chem. Phys.*, 120, 6756 (2004). M. Bachmann et al., *Phys. Rev. E*, 71, 031906 (2005). S. Kim et al., *Phys. Rev. E*, 72, 011916 (2005).