

非格子モデル/GAによるタンパク質立体構造の予測

2T-2

金井 邦夫 元木 達也†

新潟大学大学院自然科学研究科情報・計算機工学専攻‡

あらまし タンパク質立体構造を予測する問題は重要だが、非常に難解な問題として認識されている。Unger 達 [1,2,3] は各アミノ酸残基の存在位置を立体格子に制限した簡単なタンパク質モデルを使用し、GA により安定した折りたたみ構造を探索した。この論文では、Unger 達の「立体格子」という制限を外したモデルを考え、GA により安定した折りたたみ構造を探索する。

1 タンパク質立体構造の予測

タンパク質の性質の立体構造はその機能に大きく関わっているため、その立体構造を予測する問題は非常に重要な問題として認識され様々なアプローチで研究されている。しかし、この問題は熱力学的に最も安定する様に、すなわち自由エネルギーが最も低くなるようにアミノ酸残基を配置することと同等であり、非常に難問としても知られている。実際、結合の角度を3通りと見積もったとしても、長さが100のタンパク質だと、 $3^{98} \approx 5 \times 10^{46}$ 通りもの配置が可能であり、自由エネルギーを最小にする構造を探索するには何らかの工夫が必要になる。

2 非格子モデル/GAによる立体構造予測

Unger 達 [1,2,3] は、

- (1) 各アミノ酸残基を3次元格子に配置し、
- (2) 自由エネルギーとしては疎水性残基同士の近接による安定性、2つ以上の残基が同一の格子に存在することによる不安定性、の2つだけを考慮にいたした、

非常に単純な格子モデルを考え、このモデルの下でGAを用いて安定した折りたたみ構造を探索する実験を行った。彼らの研究によって、探索空間を制限すれば自由エネルギー最小/極小の構造も探索可能となることが明らかになった訳であるが、得られた安定構造は自然なものとは言い難い。

*Using a lattice-free model and GA to search for the native protein conformation

†Kunio Kanai and Tatsuya Motoki

‡Information and Computer Engineering Course, Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan

そこでこの論文では、上記(1)の「格子上」の制限を外して、Unger 達と同様のGAによる実験を行う。

4 個体表現、遺伝的操作

個体表現 各折りたたみ構造(図1)は、図2のように、先頭のアミノ酸から2番目のアミノ酸までの相対的な位置情報、2番目のアミノ酸から3番目のアミノ酸までの相対的な位置情報、.....の配列によって表す。アミノ酸の相対位置を示す4つの遺伝子はそれぞれ、z軸を中心としたxy平面での回転方向、回転角度、z軸方向に変位する向き、z軸方向への変位距離を示している。

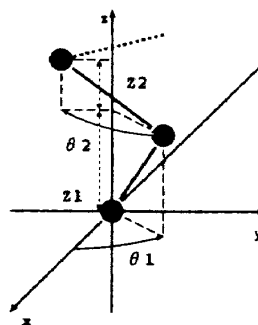


図1：表現型

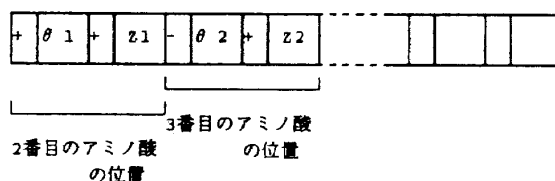


図2：遺伝子型

適合度 タンパク質の立体構造が決定されるための要素は色々あるが、ここでは疎水性相互作用とアミノ酸同士の衝突のみに着目する。まず、疎水性アミノ酸残基同士が近くにある(hitと呼ぶ)方が疎水性相互作用の性質により、構造エネルギーを低くなる。これを適合度を決定する要素として取り入れる。しかし、アミノ酸残基同士が近付きすぎる(衝突と呼ぶ)と斥力が働くため、構造エネルギーが急上昇する。これも適合度を決定する要素として取り入れる。これら2つの要素を元に次の様に適合度を定める。

$$\text{適合度} = C1 * (\text{hitの個数}) - C2 * (\text{衝突の個数})$$

交叉の方法 次の方法を基本的なものとする。

方法1: 一点交叉を行う。

しかし、通常の交叉を行なうと、個体内のアミノ酸同士で衝突が多発することが考えられ、個体がうまく進化しないおそれがある。そのため、この実験では次の様な2つの交叉方法も考えてみる。

方法2: 一点交叉。但し、交叉点に選ばれたアミノ酸の符号部分の遺伝子を4通りに変化させる。それぞれに対して一点交叉し、生まれた4つの子の中で最も適合度が高い個体を子とする。

方法3: 一点交叉。但し、適合度を評価する時に遺伝子の書き換えを行なうことができるだけ衝突を避ける。

5 実験

集団の大きさを200、親の選択方法をサイズ5のトーナメント方式、 $C1=2$ 、 $C2=2$ 、各アミノ酸間の距離を2.0、衝突と見なす距離を0.8以下、hitと見なす距離を1.8以上2.2以下に固定し、交叉確率と突然変異確率を調節して方法1~3の実験をそれぞれ250回行った。また、比較のため、同じ条件で格子モデルによる実験も行った。

この実験では Unger 達 [2,3] に倣って wwbwbwww wwbwbwww (w は親水性アミノ酸残基、b は疎水性アミノ酸残基を示す) で表されるタンパク質の立体構造予測を行った。図3には各世代の最良適合度を250回の実行について平均したものを、表1には250回の実験の中で生まれた最良個体の適合度を示す。これらの実験結果は、全て交叉確率90%、突然変異確率5%のものである。これらを見ると、方法2により進化させた集団が一番優れている事がわかる。このようになった原因は、方法2では大きさ200から4種類の子が生成され、新しい大きさ800の集団が形成され、その後、大きさ800の集団の中から良い個体を200個を選んでいるような動作をしていると考えられるからであろう。方法1で行った実験結果は予想通り方法2、方法3の両方より劣る結果となった。

格子モデルによる実験では、1次元配列上で偶数個離れた疎水性アミノ酸残基同士は、どうしてもhitになり得ないので、通常の交叉を行った結果より劣った結果になったようだ。

6 おわりに

GAに局所探索を取り入れた集団は、通常の交叉

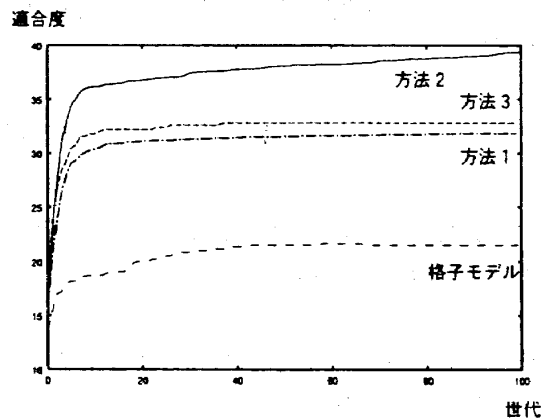


図3: 各世代における最良適合度の平均

表1: 最良適合度

実験方法	最良個体の適合度
格子モデル	22
方法1	42
方法2	52
方法3	48

のみを行い進化させた集団より、適合度の高い個体が多く現れた。また、アミノ酸の位置の制限を外すことにより、局所探索を取り入れた集団は高い適合度をもった個体を含む集団に進化した。

しかし、出来上がったタンパク質の立体構造をグラフィカルに表示するためのインターフェースがないため、タンパク質の評価が適合度によるものだけになってしまった。また、タンパク質の疎水性コアがどのように構成されているかわからないままである。

参考文献

- [1] R.Unger & J.Moult: A Genetic Algorithm for Protein Folding Simulations, *Journal of Molecular Biology*, vol.231 pp.75-81,1993.
- [2] R.Unger & J.Moult: A Genetic Algorithm for 3D Protein Folding Simulations, *Proc.of the Fifth Annual International Conference on Genetic Algorithms*, pp.371-408,1993.
- [3] A.L.Patton, W.F.Punch III & E.D.Goodman: A Standard Approach to Native Protein Conformation Prediction, *Proc.of the Sixth Annual International Conference on Genetic Algorithms*, pp.574-581,1995.
- [4] 中村春木 & 中井謙太: バイオテクノロジーのためのコンピュータ入門, コロナ社, 1995
- [5] C.Branden & J.Toozee: タンパク質の構造入門, 教育社, 1992