

# タンパク質の折畳みの力学\*

北海道大学・理学研究科・数学専攻 松本 健司

(1998年6月30日受理)

## 1 タンパク質の折畳み問題

タンパク質はアミノ酸が一行に結合したものである。タンパク質がその機能を果たすには特定の3次元構造をとる必要がある。生体内ではタンパク質はアミノ酸をひとつづつ片方の端にくっつけていって作るので、タンパク質に特定の3次元構造をとらせることが常に行なわれている。

pHや温度を変えたりして通常的环境と異なる条件に置かれると、3次元構造が変化して、タンパク質は機能を果たさなくなる。これを活性を失う、失活するという。失活したタンパク質をもとの環境にもどすと、自然にまたもとの3次元構造をとって、機能が再生するものがある。これにかかる時間は数十ミリ秒から1秒以下である。

実験的に到達される特定の3次元構造は系のマクロなパラメータをコントロールした下でのマクロな観測のもとで得られたものであるから、熱力学的な自由エネルギーが最小の状態であることにはちがいない。しかし、タンパク質の折畳みは、蒸気機関のように多数の要素が集まって初めて起る現象とは考え難い。その意味で、この現象を熱力学的に扱うことには限界があると思われる。

一説によると [1]、タンパク質のエネルギー極小値があまりに多いため、すべての極小値を巡る時間は、実際の折畳みにかかる時間より、はるかに多いそうである。ここには未知の力学機構が存在していると考えられないだろうか？

これは、この未知の力学機構を分子動力学の長時間シミュレーションにより探ろうとする試みの中間報告である。現在までに計算できた時間は折畳みにかかる時間に5から7桁程足りず、計算機の中で実際に折畳みを起こすには至っていない。しかし、タンパク質のような力学系でどのようなことが起こり得るかのヒントを与える程度の計算ができると考えている。

\* 本稿は、編集部の方から特にお願いして執筆していただいた記事である。

## 2 タンパク質の分子動力学モデル

### 2.1 分子動力学

タンパク質分子内の原子の挙動は量子力学によって記述される。しかし、計算機でこの挙動をシミュレートしようとするときには、はるかに計算量の少ない古典力学による近似が用いられる。分子内の原子を質量をもつ点として扱い、その間のポテンシャルをうまく選んで、量子力学で働く力を近似するようにする [2]。

このためのさまざまなポテンシャルが提案あるいは販売されている。ここでは、その全てのパラメータが以前公開された CHARMM というポテンシャルを使う [3]。これらのポテンシャルは頻繁に更新されるのが常で、ここでの計算に使っているのは、かなり以前に公開されたものであり、最新版でないことをお断わりしておく。ここでの目的は、分子の力学的な性質はどのようなものかを調べることであり、ポテンシャルの改良あるいは優劣を決めることではないし、残念ながらそのような立場にはない。そのために、まずとりあえず1つのポテンシャルから始めようということである。このポテンシャルによる分子の挙動の解析で、方法論などが出そろえば、他のポテンシャルとの比較も可能になる。

### 2.2 CHARMM ポテンシャル

2つの原子が分子内で化学的に結合しているとき、この2つの原子を始点と終点とした空間ベクトルをボンドと呼ぶことにする。CHARMMに限ったことではないが、このようなポテンシャルは分子の形を維持するためのボンドから決まる近距離のポテンシャルと、2つ以上の原子の座標から決まる遠距離のポテンシャルからなる。

CHARMM では短距離ポテンシャルとして、

- ボンドの長さを2つの原子の種類によって決まるある値に維持するポテンシャル、
- 1つの原子から出た2つのボンドの間の角度を3つの原子の種類によって決まるある値に維持するポテンシャル、
- 1つの原子から出た3つのボンドを平面に保つ場合に必要なポテンシャル、
- 3つのボンドで連続的に繋がった4つの原子のうち、はじめの3つの原子から決まる平面とおわりの3つの原子から決まる平面の間の二面角のポテンシャル

を使う。前者3つは極小値が1つの2次のポテンシャルで、強力な力を出すようにパラメータが設定されており、シミュレーションの間にも極小値のまわりで小さく揺らいでいるだけである。最後のポテンシャルは力も比較的弱く、極小値を2つか3つ持っている。タンパク質が非常に多くの極小値を持つのは、このポテンシャルが原因である。

長距離ポテンシャルは

- 3つの原子の位置で決まる水素結合のポテンシャル、
- 2つの原子間の van der Waals 力、
- 荷電原子間の Coulomb 力、

である。前者2つはある距離までは引力で、それ以下だと斥力になる。CHARMM では、ひとつひとつの原子について、その種類と分子中の位置によって、型が定められていて、その型によって電荷が与えられている。これらの荷電原子間には通常の Coulomb 力が働く。長距離力はボンドでつながっている2つの原子の間には働かない、という設定になっている。

CHARMM ではパラメータの値自体は精度が2桁足らずである。また、水素結合しない水素原子は省略することができる。公開されたパラメータが水素原子を省略したものであったので、ここでもそれを用いている。水素原子を省略した分、その水素原子が結合していた原子を拡張原子という型にして、やや大きくしている。しかし、この省略は折畳みの過程においては、重大な結果を生ずるかもしれない。

### 2.3 運動方程式とその積分

分子内の原子  $i$  の質量を  $m_i$ 、速度を  $v_i$ 、座標を  $x_i$ 、前述したポテンシャルの総和を  $U$  とすると、原子  $i$  のニュートン方程式は、

$$m_i \frac{dv_i}{dt} = -\nabla_i U \quad (1)$$

である。

保存系をルンゲクッタ等の特に保存系を意識しない方法で数値積分すると、数値誤差により総エネルギー等の保存量が積分時間に比例して増える、あるいは減るといった現象が見られる。これは短時間の数値積分ではあまり目立たないし、たぶん、他の数値積分をする際に出現する問題に較べて特にこれが大きな問題であるということはないと思われる。しかし、ここでは相当長時間にわたって数値積分するので、エネルギーのグラフなどでは大変目立つし、誤った結論に到達してしまう可能性も大きい。そこで、ここでは保存系のために特別に工夫されて、このような現象を排除したシンプレクティック積分法 [4] を使用した。

## 2.4 温度

温度はそもそも折畳みをコントロールする非常に重要なパラメータである。できれば正確に設定できるといいのであるが、残念ながら2つほど困難な点がある。

正確に温度を測定するには、理想気体と接触させて熱平衡に達したところで理想気体の温度を測定する。しかし、これを数値計算で実行するには、揺らぎが充分小さくなる程度に理想気体の分子の数を大きくしなければならないのと、熱平衡に達する程度に長時間計算しなければならない。これは、ここでの目的であるタンパク質の折畳みを計算するのより、はるかに大規模な計算であり、いまのところ実行不可能である。

そこで、ここでは温度をエネルギー等分配則が原子の運動エネルギーについて成立すると仮定して定義する。総運動エネルギーを原子の総数の3倍で割ったものが温度のボルツマン定数倍の半分である。

また、現実のタンパク質ではまわりに十分な数の水分子があって、タンパク質の温度は結果的に一定に保たれる。これを数値的に実行することも理論的には可能であるが、水分子はタンパク質分子に較べてずっと小さく、タンパク質分子を覆うには非常に多数の水分子が必要となり、その挙動の計算には途方もない時間がかかってしまう。ここでは、タンパク質の挙動を長時間計算することの方を優先し、水分子なしで計算した。その結果、タンパク質がより低いポテンシャルエネルギーの極小値に入ると、温度はかなり上昇してしまうことになる。これをコントロールすることは、ここでの目的であるタンパク質の長時間の挙動を邪魔してしまうことになるので、温度を設定しなおすことは最小限に止めた。

また、温度をあらかじめ定められた値に設定すること自体もかなり難しい。正確に設定することが出来るのは総エネルギーであって、そのうちのどれだけが運動エネルギー(これが温度に比例)になるかは正確には予想はできず、計算結果を長時間平均して初めて温度が何度になったかわかる状態である。

## 2.5 oxytocin

実際に計算するタンパク質の撰択には二つの点しか考慮していない。ひとつは計算時間がアミノ酸の数の2乗に比例するので、できるだけ小さいタンパク質であること。もうひとつは現在唯一の3次元構造の情報源であるX線結晶解析のデータが存在することである。どうせ最終的に折畳むまでの計算は不可能であったので、それ自体だけで折畳むことが実験的に確かめられているかは考えなかった。

その結果、選ばれたのはタンパク質ではなくホルモンの一種のoxytocinである。ここでの計算では、とても機能をうんぬんできる立場ではないので、機能については省略させていただく。

## 2.6 計算機事情

分子が小さいとベクトル計算機はあまり効果がなく、CPU時間はPentium Pro 200MHzと変わらない。また、大規模な計算機はひとつのプログラムで占有できるわけではないが、パソコンならば場合によっては何台も占有することができる。というわけで、ここでの計算はすべてUNIXを載せたパソコンで行なわれている。ここでの計算はPentium Pro 200MHzで二箇月程度かかっている。

このところパソコンのCPUは1年半で性能が2倍になっている。また、複数のCPUを積んだ並列パソコンも一般的になり安価で手に入るようになった。ここで使うようなポテンシャルの計算には並列計算機は非常に有効で、CPUの数だけ速く計算できる。この二つの効果によって現在では2年前の4倍の速度で計算できるようになった。今年中にこれが16倍、さ来年のなかころには50倍になるというのが現在の予定である。

## 3 折畳み初期のタンパク質分子の挙動

### 3.1 初期条件

タンパク質分子はトポロジー的にはバックボーン原子と呼ばれる原子団が一系列に繋がり、そこからアミノ酸1つ当たり1つずつの枝分れが出ている。枝分かれ部分の原子は側鎖原子と呼ばれる。

本当はランダムな初期条件からでも出発すると自然なのだが、これだけの数の極小値があると何が自然なランダムかということ自体、本質的な困難を含む。また、所詮大した数の実験ができないので、ランダムな初期条件といってもあまり意味がない。そこで、最も単純で偏りのないバックボーン原子がほぼ一直線上にならんだ初期条件を採用することにした。この条件では遠距離ポテンシャルが最大になっており、最も非現実的な状態と思われる。速度の初期条件は自然にランダムに取れるので、ある程度のランダム性も保証される。

もう1つの初期条件として、X線結晶解析で得られた原子の座標を用いた。座標データはPDBというデータベースが作られていて誰でも容易に入手できるものもある。X線結晶解析では水素原子の位置はわからないし、精度も原子間距離程度であるし、そもそも結晶状態と水中での分子では相当違いがあると思われるが、これが現在得られる最も自然状態に近いデータである。

速度の初期値は一様乱数でまず与え、重心運動と自転運動を取り除きつつ、エネルギー等分配則から与えられた温度になるようにスケールリングした。

### 3.2 線形初期条件

初期温度を摂氏-20度にして線形初期条件から  $2 \times 10^{-8}$  秒ニュートン方程式に従って時間発展させた oxytocin のポテンシャルエネルギーを図1に示す。これを移動平均したものが図2で、同様に平均した温度は図3である。

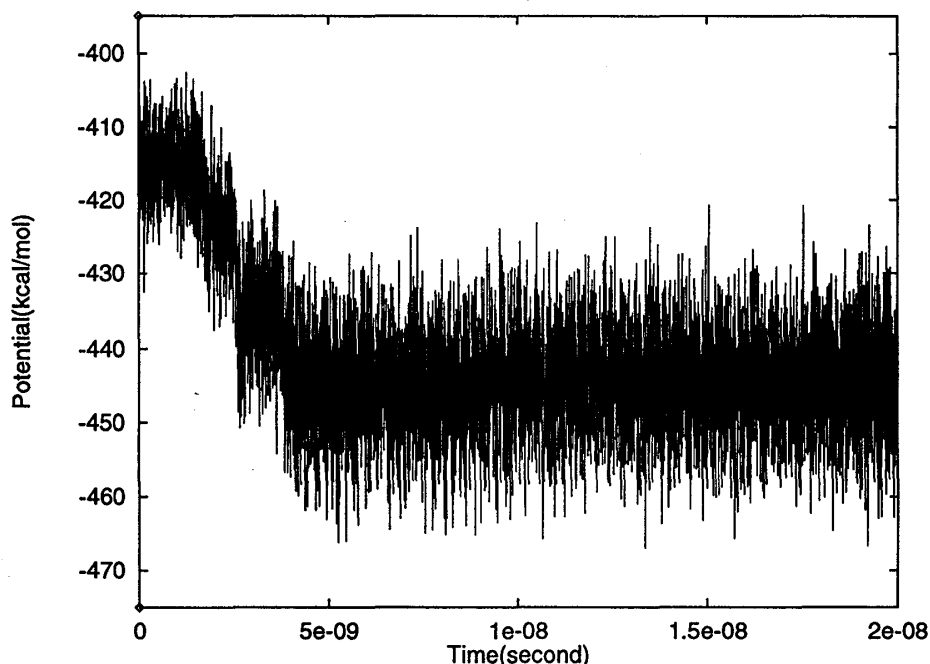


図1: 線形初期条件のときのポテンシャルエネルギーの時間変化

図1、図2に見られるとおり、ポテンシャルエネルギーはニュートン力学に従っているだけで減少していく。総エネルギーは一定なので、温度はどんどん上昇して、最終的にはポテンシャルにとっては非現実的な値の摂氏120度になった。(こうなることを見越して、初期温度をこれも非現実的な-20度にした。)

図からわかるように、これらの変化は徐々に起こるのではなく、ときおり見られるポテンシャルエネルギーの不連続によるものである。これらの不連続は分子がポテンシャルのある極小値の近傍から別の極小値の近傍に移ったことにより引き起こされる。そこで、図のような分子の挙動を理解するには、分子がどのように極小値を移っていったかを記述することが第一歩となる。

### 3.3 極小値へのクエンチ

タンパク質分子内の原子の数を  $n$  とすると、ある時点でのタンパク質のすべての原子の座標は  $3n$  次元の配位空間の一点として表現される。一方、ある時点でのタンパク質の完全

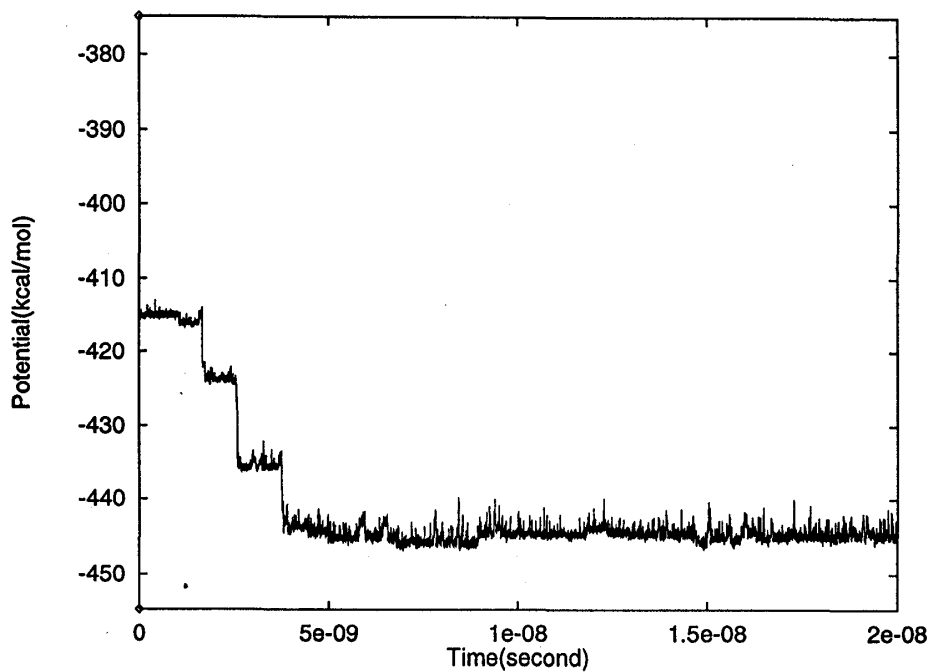


図 2: ポテンシャルエネルギーの移動平均

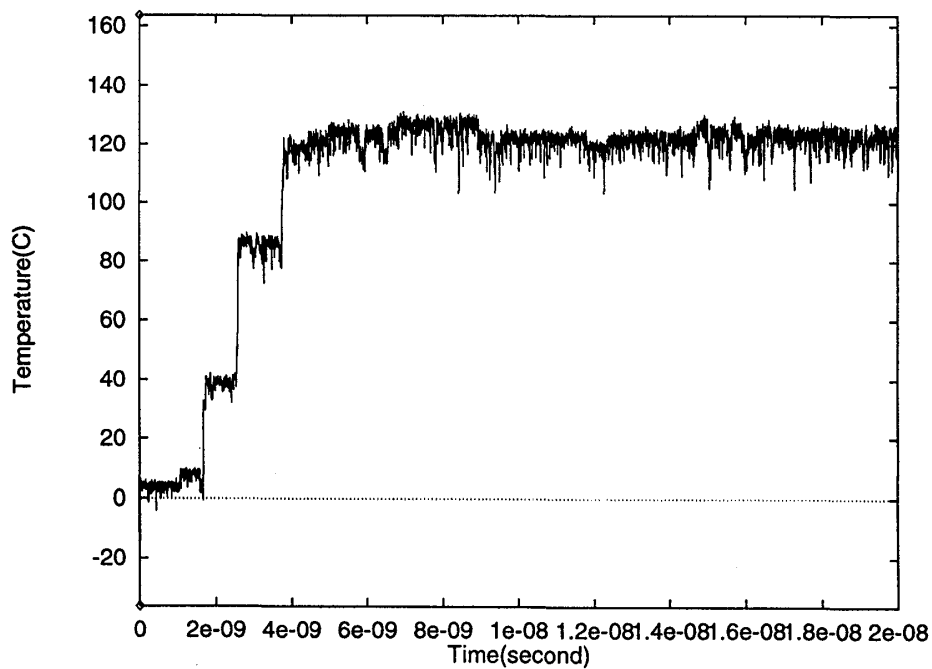


図 3: 線形初期条件のときの移動平均した温度の時間変化

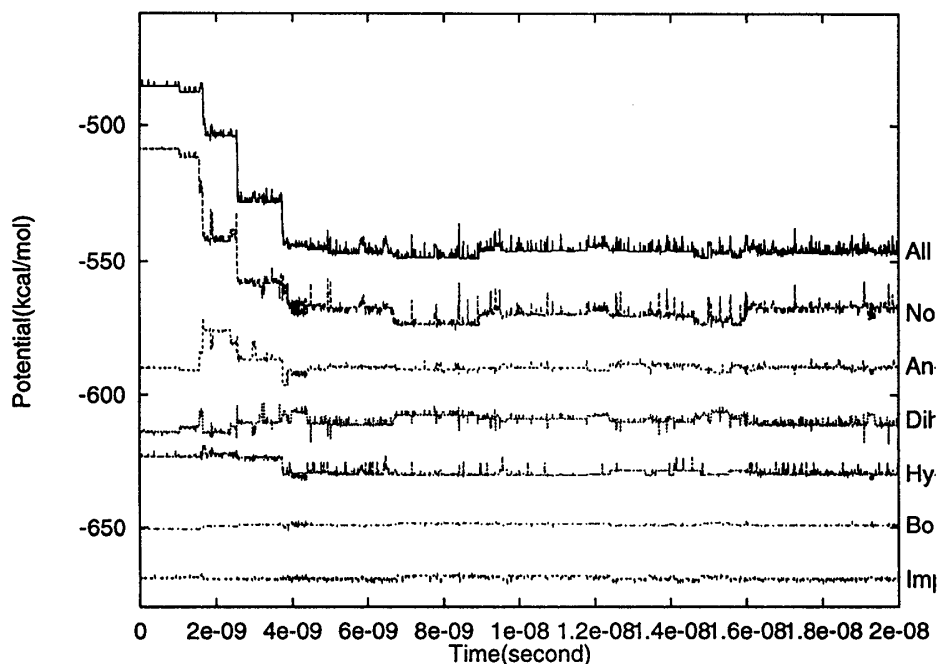


図 4: 線形初期条件のときのクエンチした総ポテンシャルエネルギーと各種ポテンシャルの時間変化(目盛は総ポテンシャル用、その他はスケールのみ正しい) 上から全体、遠距離、角度、二面角、水素結合、ボンド、平面維持ポテンシャル

な状態は、すべての原子の座標と速度をあわせた  $6n$  次元の位相空間の一点で表現される。CHARMM ポテンシャルは配位空間上の関数であり、配位空間内で多数の極小値をもつ。

ある時点で分子が配位空間内のどの極小値の近傍にあるか、あるいは属するか、というには定義が必要である。ここでは大雑把に、次の方程式

$$\frac{dx_i}{dt} = -\nabla_i U \quad (2)$$

で時間無限大で達する極小値に属しているということにする。

実際に極小値を求めるのに上の方程式を積分するとたいそう時間がかかってしまうので、ここでは別の方法を探った。まず、出発点でのポテンシャルのグラディエントを求め、その線上での極小値をニュートン法で求める。さらにその新しい極小値の位置のグラディエントから出発して同様の方法で次々と極小値の位置を求めていく。最終的に新しい極小値の位置がその前のステップの極小値の位置から動かなくなったら、それが求める極小値である。

ここで使用しているようなポテンシャルでは、近距離の部分と遠距離の部分のグラディエントの大きさが極端に違う場所がある。このような場合には、上の方法でも非常に多数のステップが必要となるが、ここでは時間の都合で、エネルギーの値がある程度収束するところまでしか計算していない。



図4は、こうして求めた極小値の時間変化である。一番上が総ポテンシャルエネルギーで、その下に各種のポテンシャル別の時間変化を示す。平均操作はしていない。

ポテンシャルの種類ごとの時系列をみると、ボンドの長さ等には全く変化がなく、ボンドの間の角度と二面角が遠距離ポテンシャルの変化と共に変化しているのがわかる。

### 3.4 内部座標

タンパク質分子の形は配位空間の1点として表わされる。しかし、同じ形の分子でも回転や平行移動させれば配位空間では別の点になるし、1つのボンドを少し回転しただけでも、配位空間では大きな変化になる可能性があり、分子の形を記述するには配位空間はあまり適当ではない。

分子の形を記述するときには、分子全体の平行移動や回転に影響されない内部座標を用いる。内部座標は、すべての回転可能なボンドの角度をある基準角度から測ったものである。

ボンドの長さや2つのボンド間の角度等がすべて個々のポテンシャルで設定された極小値にあれば、内部座標はタンパク質の形態に関して完全な記述になっている。実際の極小値に対応する分子の形では、これは必ずしも成り立っていない。しかし、前にも述べたように、タンパク質分子の無数の局所最小値は二面角ポテンシャルが原因であり、このため内部座標は極小値を見分けるには充分であると思われる。つまり、ここでは内部座標が同じ2つの極小値は存在しないと仮定する。

実際には回転可能なボンドの角度はボンドのまわりの適当な原子を一組選んで、それらの二面角として測定している。

oxytocinの場合で、配位空間は252次元(前述したように拡張原子を使って水素分子をいくつ省略している)のところ、内部座標では次元となり、かなり次元を制限することができる。しかし、それでも非常に高次元であることには変わりなく、グラフにプロットすることさえままならない。

### 3.5 クラスタ解析

タンパク質分子がポテンシャル極小値を移っていくのを記述するために、極小値に出現順に番号を付けていくことにする。しかし、極小値のときの真の内部座標を求めるのは容易なことではない。前述したようにクエンチしても、その方法の限界から完全な極小値にクエンチすることは不可能である。クエンチして得られた内部座標と真の極小値の内部座標との間にはかなりの誤差が出てくる。そこで、ある閾値を設定して、それ以下の距離にあるクエンチした内部座標同士で集合(クラスター)を作り、この集合でひとつの極小値を表わしていると考えことにする。

内部座標  $\Theta$  は回転可能なすべてのボンドの角度の集合  $\{\theta_i\}$  である。2つの角度変数の距離を

$$d_a(\theta, \phi) = \min(|\theta - \phi|, 2\pi - |\theta - \phi|) \quad (3)$$

とすると、内部座標上での距離は

$$d(\Theta, \Phi) = \sqrt{\sum_i d_a(\theta_i, \phi_i)^2} \quad (4)$$

と定義できる。

閾値  $D$  に対してクラスターを定義する。任意の内部座標の集合  $A$  と任意の内部座標  $\Phi$  の距離を  $A$  の  $\Phi$  に最も近い要素と  $\Phi$  の距離と定義する。タンパク質の時系列をクエンチした内部座標の列  $\{\theta_i\}$  を考えよう。この列の任意の要素を取り、これだけからなる集合を  $C_0$  とする。 $C_0$  との距離が  $D$  以下の列の要素の集合を  $C_1$  と定義する。同様に、 $C_k$  が既に定義されているとき、 $C_k$  との距離が  $D$  以下の列の要素の集合を  $C_{k+1}$  とする。 $C_n = C_{n+1}$  となったとき、 $C_n$  が求めるクラスターである。

こうして、タンパク質の時系列をクエンチした内部座標の列  $\{\theta_i\}$  をクラスターに分けることができる。これらのクラスターが、それぞれ1つずつの極小値を表わす。クラスターの出現順に番号を付けると、クエンチした内部座標は、ただ1つのクラスターに属するので、ただ1つの番号が対応する。

図5は、この番号を時間に対してプロットしたものである。

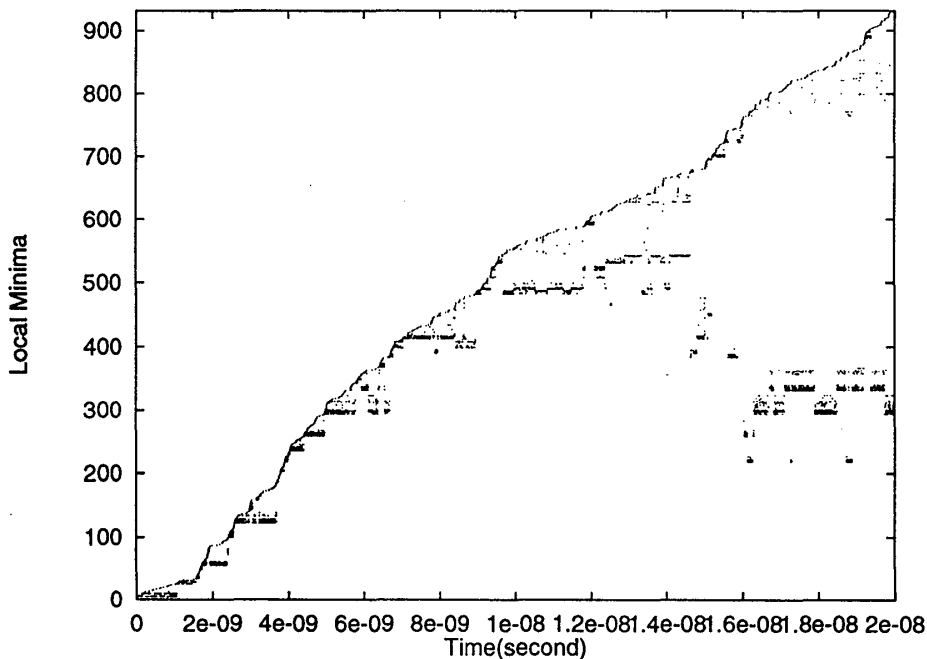


図 5: 線形初期条件のときの極小値を移動するタンパク質分子

図で水平方向に長く連なる点の集合はタンパク質分子がその極小値に長くとどまったことを示している。これらの水平線は、よくみると線ではなくて帯になっていて、複数の極小値の集合であることがわかる。これは、ある種のボンド(特に分子の端近くにあるボンド)が、分子の他の部分とは比較的無関係に自由に回転できるからである。分子全体は同じ形にとどまっていますが、端の方のボンドだけ回転するので、内部座標でみると異なる極小値を頻繁に往き来しているように見える。

図の水平の帯の部分は閾値  $D$  が 0.02 から 0.5 の範囲でほとんど変わらない。範囲の下限の 0.02 というのは今回のクエンチの最小誤差で、閾値がこれ以下になるとクラスターの数が増加する。上限の 0.5 というのは、二面角ポテンシャルから考えて、これ以上だと明らかに 2 つ以上の極小値がつながってしまう。

水平方向の帯の上の方には、しばしば右肩上に連なる点の集合が見られる。これはクエンチのときに十分に極小値に収束しなかった点であると思われる。それらの多くは時系列全体で数度だけしか表われない孤立クラスターであり、温度が高いときや、閾値の小さいときに多く現れるからである。

この図によると、分子はポテンシャルエネルギーのこのシミュレーションの範囲内での最小値に比較的短時間で達している。そして、ややポテンシャルエネルギーの高い、しかしほぼ最小値付近の他の極小値をいくつかまわって、また最小値に戻ることを繰り返しているのがわかる。最小値には 3 回滞在しているが、滞在時間はかなり短い。

このシミュレーションの範囲での最小値のときのタンパク質分子の形は次に述べる X 線結晶解析による座標データとはかなり違っている。つまり、このタンパク質は誤った折畳みをしている可能性が強い。初期条件が非現実的であったために本当に誤った折畳みをしているのか、計算している時間範囲が通常の折畳みにかかる時間より 5 桁以上小さいので、ここまでしか見えていないのか、いずれにしても、折畳みを再現しているかどうかについては、これだけでは確実なことは言えない。

タンパク質分子の力学として、ポテンシャルエネルギーの減少方向の不連続がしばしば見られることが特徴である。

### 3.6 X 線結晶解析による初期条件

これまで見てきたようなポテンシャルエネルギーの不連続は、非現実的な線形初期条件から出発したためかというところ、そうではない。次に X 線結晶解析から得られた原子の座標を初期条件としたシミュレーションの結果を示す。

図 6 はクエンチしたポテンシャルエネルギーで、図 7 はクラスター解析の結果である。

やはり、大きな下向きのポテンシャルエネルギーの不連続が見られる。自然な形に十分近いはずの PDB の座標データからこれほどの変化を起こすというのは、ポテンシャルや

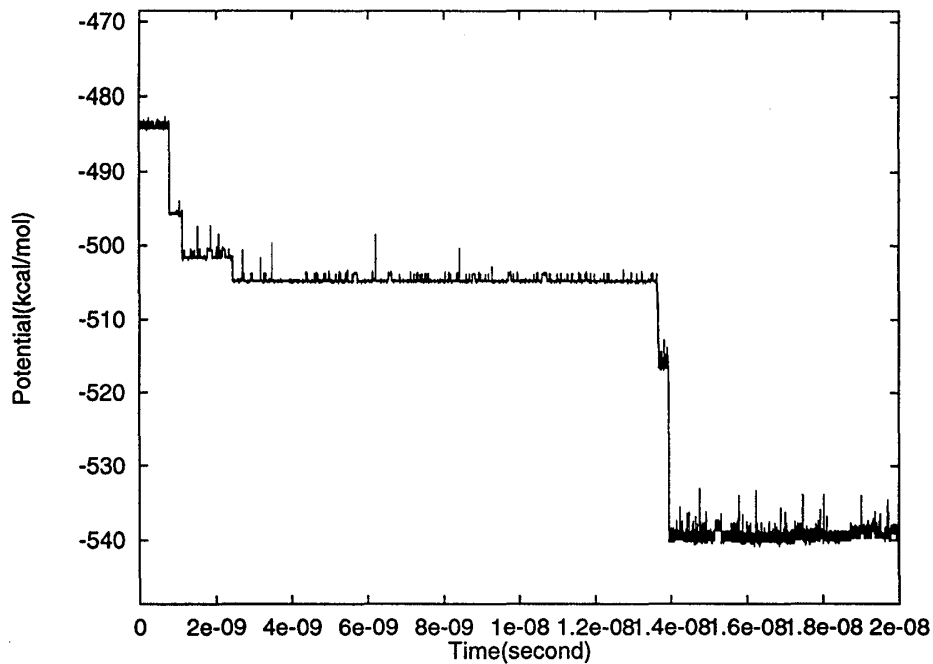


図 6: PDB の座標データを初期条件としたときのクエンチした総ポテンシャルエネルギーの時間変化

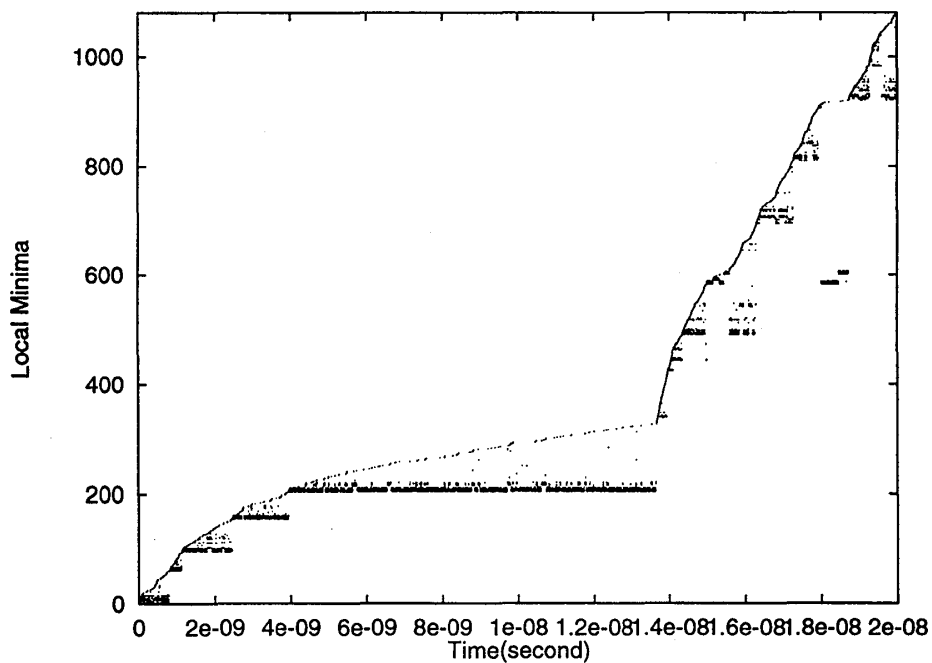


図 7: PDB の座標データを初期条件としたときの極小値を移動するタンパク質分子

プログラムに不安を抱かせるものがある。(確かに長距離ポテンシャルは元の論文ではカットオフの距離以上は計算していないのに比べ、ここではすべて計算しているので絶対値で大き目に出る可能性がある。)しかし、重視したいのは、ニュートン力学がエネルギーの低い状態を探し出してそちらに向かっていることである。

## 4 考察

タンパク質の折り畳みのような問題を古典力学だけで考えることには、二つの困難が存在する。ひとつは、ニュートン方程式のある解の時間を逆にしたものも、やはり解であることである。ある極小値から別の極小値へ移動できる軌道が存在すれば、その逆への軌道も常に存在する。折り畳みに相当する軌道が存在していれば、その逆の失活する軌道も存在する。

もうひとつは Liouville の定理である。位相空間で体積をもった初期条件の集合を考えたとき、ニュートン方程式による時間発展によって、この集合の形は変化するが、体積は常に一定である。最初いろいろな形をしていたタンパク質が折りたたんでただ一つの形になるという場合には、位相空間で広がっていた初期条件が、位相空間の比較的小さな領域に集まることに対応するので、Liouville の定理を破っていることになる。

現実のタンパク質分子は周りを水または脂質の分子によって囲まれている。ここでのシミュレーションのようにタンパク質分子がより小さなエネルギーの極小値に落ち込んで温度が跳ね上がると周りの分子によって冷やされることになる。タンパク質分子の運動エネルギーを取られることになり、こうなると総エネルギーも減少し、もはや最初の等エネルギー面から外れてしまう。そして、この時点でニュートン力学の記述も終わるので、この場合は上の二つの困難は問題にならない。

では、周りの分子がタンパク質分子を冷やすことによって、折り畳みのような一方向的な現象が見られるのかというと、そうともいえない。局所的な温度は常に大きく揺らいでいるので、周りの分子に冷やされることもあれば、暖められることもある。その場合には元の等エネルギー面あるいはそれ以上のエネルギー面に戻ることも可能である。つまり、周りの分子との熱力学的な相互作用を考えた場合でも、この相互作用が原因で一方向的な折り畳みの現象が起こっているのではない。

原因はあくまでもニュートン力学で大きなポテンシャルエネルギーの不連続が下がる方向にしかみられないことによるのである。

## 参考文献

- [1] C. Levinthal, Are there pathways for protein folding? *J. Chem. Phys.*, **65**, 44-45(1968)
- [2] Martin Karplus and Gregory A. Petsko, Molecular dynamics simulations in biology  
*NATURE*, **347**, 631-639(1990)
- [3] B. R. Brooks et. al., Chemistry at HARvard Macromolecular Mechanics *J. Comp.  
Chem.*, **4**, 187(1983)
- [4] 吉田春夫, シンプレクティブ数値解法 数理科学, **384**, 37(JUNE 1983)