

# 主成分分析を使って眺めた蛋白質の エネルギー地形

京都大学\* 北尾 彰 朗

( 受付 2000 年 10 月 2 日 ; 改訂 2000 年 12 月 22 日 )

## 要 旨

蛋白質はある少数の方向に変形しやすい．そのため主成分分析は蛋白質のゆらぎやそれを規定しているエネルギー面のかたち，いわゆるエネルギー地形を眺めるための方法として非常に有用である．蛋白質は多自由度・多成分系であるので，そのエネルギー地形は複雑である．蛋白質は天然の状態においても多数の準安定な状態をもち，それぞれの状態に対応するエネルギー地形のくぼみ (= エネルギー極小) がある．これらのくぼみの形や分布を調べ蛋白質のゆらぎとエネルギー地形を研究するための方法として，筆者らは主成分分析の発展形である JAM モデルを提案している．このモデルと分子シミュレーション・実験データを駆使して明らかになってきた蛋白質のエネルギー地形について報告する．

キーワード：蛋白質，エネルギー地形，JAM モデル，調和性，非調和性．

## 1. 蛋白質の何が問題となっているか？

現在，ヒトをはじめとする様々な生物の遺伝子(ゲノム)の解読が進んでいる．既に数十種の生物の遺伝子が完全に解読されている．これまではその複雑さゆえに，システムとして生命の全体像が明らかになることはなかった．しかし，遺伝子の全貌が解明されることで，生命の設計図は情報として全て得られることとなった．ヒトの遺伝子の総数は  $10^4 \sim 10^5$  個程度であり，それぞれが蛋白質のアミノ酸配列(1次構造)を決定している．現段階では，配列が決定されているが機能が予想できない蛋白質が大量にある．蛋白質の3次元立体構造は現在精力的に調べられているが，構造が決定されているものは全体の中でまだ少数である．その中でも機能のメカニズムが詳細にわかっているものは非常に少ない．

大量の設計図が得られたあとは，いったい何が問題になるかという点(いわゆるポストゲノム)を考えてみよう．ポストゲノムの段階では，大量情報処理だけではわからなかったより原理的なレベルからの理解が求められることになる．蛋白質が動作して機能を発揮するメカニズムは蛋白質のゆらぎと密接に関係している．化学エネルギーを用いて機能を発揮する筋肉などの蛋白質もあるし，熱ゆらぎの範囲での機能を果たす蛋白質もある．どちらの場合にも蛋白質立体構造のゆらぎは機能を発揮するのに重要だと考えられている．例えば酵素が基質を認識する場合，以前は酵素と基質は鍵と鍵穴の関係になっていて，形状が一致する場合に基質を認識するという考え方があった．しかし，最近の研究では蛋白質分子が変形することによって基質

\*大学院 理学研究科化学専攻：〒606-8502 京都市左京区北白川追分町．

と結合し反応がおこりうる状態になること、このような変形は熱ゆらぎと密接に関係していることがわかってきた。本稿では蛋白質が「分子機械」として動作するメカニズム解明のために主成分分析を用いる。熱ゆらぎの特徴を探ったり、機能を発揮するためにおこる蛋白質の変形を調べたりするために主成分分析はとても有用である。ゆらぎの特徴を決めているのは蛋白質のエネルギー面がどのような形状をしているかという「エネルギー地形」である。逆にゆらぎを詳細に調べることで蛋白質のエネルギー地形がわかる。本稿では主成分分析やそのバリエーションがなぜ蛋白質のゆらぎとエネルギー地形を解析するのに有用なのか、どのように使われているのか、その結果どういことが明らかになってきたのかを解説していくことにする。

## 2. 蛋白質という物質の特徴

### 2.1 ゆらぎの異方性

蛋白質はアミノ酸残基がつながった鎖(ポリペプチド)からなるが、天然の状態においてはある特定の構造の近傍でゆらいでいる。この構造を天然構造と呼ぶ。天然構造では蛋白質の鎖は折れ畳まされており、原子が高密度で充填(パッキング)された状態にある。その密度は最密充填と同程度である。このため蛋白質が大きくゆらく際には原子のパッキングの仕方が大きく変わる(エネルギー障壁を越える必要がある)、パッキングがあまり変わらないように変形する必要がある。2.3節で述べるようにパッキングの状態が蛋白質のゆらぎの異方性を規定しているといつてよい。

蛋白質の天然状態でのゆらぎは大まかに見ればある状態のまわりでゆらいでいるとみなしてよく、そのゆらぎは異方性が高い。この特徴があるので、蛋白質のゆらぎを解析するには蛋白質の各原子のデカルト座標(原子座標)ではなくて、蛋白質の異方的なゆらぎを特徴的に表わすため原子座標の線形結合で表わされる集団座標を用いるのがよい。そういう理由で、主成分分析は蛋白質のゆらぎを探るために重要な方法となっているのである。

### 2.2 調和性而非調和性

蛋白質は複雑な相互作用をする多自由度系であるので、その振る舞いは非調和的だと思われるやすい。しかし、実際には観測の仕方(とそれに応じた空間解像度や時間解像度)や温度などに依りて、調和的な側面が現れてきたり、非調和的な現象が現れてきたりする。調和的な側面は、当然のことながら低温であらわれやすい。結晶状態で蛋白質のゆらぎの大きさを温度の関数としてプロットすると多くの蛋白質で220 K程度までゆらぎが温度に対して線形に増加する。従って、この範囲ではほぼ調和的に振舞っているように見える。一方、それより高い温度ではゆらぎの大きさは温度の関数として急激に増加し、非線形性が顕著になる。この現象はいわゆるガラス転移だと考えられている。蛋白質のような多自由度・多成分系では天然状態であっても唯一の深いエネルギー極小を持つのではなく、エネルギー的にはほぼ等価な多数のエネルギー極小を持つ(多重極小問題)。つまり、エネルギー地形には多数の同じような深さのくぼみがあることになる。十分低温では、各蛋白質分子の状態点はあるひとつのエネルギー極小点近傍に拘束されてゆらいでいる。しかし、ある程度温度が上がってくると他の極小に移動することが可能になる。ガラス転移の原因については、蛋白質自身の性質によるものか、水和している水の影響なのか、現在でも議論が分かれており決着していない。いずれにせよ、調和的な振る舞いと非調和的な振る舞いを含んだものが蛋白質のゆらぎである。それらをどのように解析しているのかを以下で見てみよう。

### 2.3 調和性を見る方法：基準振動解析

基準振動解析は蛋白質のゆらぎを調和振動の重ねあわせとして表わす方法である．基準振動を求めるには，まずエネルギー地形のなかであるひとつのエネルギー極小点を求める必要がある．蛋白質のエネルギー状態を表わす関数は粗視化のレベルに応じて様々なものがあるが，通常は蛋白質を構成する原子を最小単位とし，原子間の相互作用によってエネルギーを定義するいわゆる分子力場が用いられることが多い．例えば分子力場のひとつ AMBER (Cornell et al. (1995)) は以下のような関数形を持っている．

$$(2.1) \quad E = \sum_{bond\ i} k_i(r_i - \bar{r}_i)^2 + \sum_{angle\ i} k_i(\theta_i - \bar{\theta}_i)^2 + \sum_{torsion\ i} V_i(1 + \cos(n\varphi_i + \delta_i)) \\ + \sum_{nonbond\ i,j} \left[ \frac{q_i q_j}{\epsilon r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

ここで， $r_i$  は化学結合長 ( $\bar{r}_i$  は平衡値)， $\theta_i$  は結合角 ( $\bar{\theta}_i$  は平衡値)， $\varphi_i$  は二面角である． $q_i$  は原子の部分電荷， $\epsilon$  は誘電率であり， $k_i, V_i, n, \delta_i, A_{ij}, B_{ij}$  はそれぞれの部位に応じて決められたパラメータである．はじめの3項は共有結合に関わるエネルギー項で，それぞれ化学結合長・結合角・二面角の変化に伴うエネルギー変化をあらわし，第4項は非共有結合に関わる静電相互作用と Lennard-Jones 相互作用のエネルギーを表わす．共有結合に関わる3項はこの場合高次項や交差項を加えてより振動エネルギーを正確に計算する分子力場もあるが，蛋白質の場合その必然性は低いので通常はこのような関数で十分である．基準振動解析をおこなうためにはエネルギー関数のあるひとつの極小点を求めた後，そこでエネルギーの2次微分行列を計算すればよい．ここで2次微分行列  $F = \{f_{ij}\}$  を

$$(2.2) \quad f_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j}$$

と書くこととする．ただし座標  $x_i$  にはそれぞれの原子質量の平方根をかけた原子座標，いわゆる質量重み付きの座標を用いると便利である．例えば  $n$  番目の原子のデカルト座標  $(s_n, t_n, u_n)$  は質量重み付きの座標では  $(x_{3n-2}, x_{3n-1}, x_{3n}) = (\sqrt{m_n} s_n, \sqrt{m_n} t_n, \sqrt{m_n} u_n)$  となる．蛋白質の全原子数を  $N$  とすると， $3N$  次元のベクトル  $(x_1, x_2, x_3, \dots, x_{3n-2}, x_{3n-1}, x_{3n}, \dots, x_{3N-2}, x_{3N-1}, x_{3N})$  は  $3N$  次元空間内でのあるひとつの蛋白質の構造を表わす．主成分分析でもちいる分散共分散行列  $A = \{a_{ij}\}$  は，

$$(2.3) \quad a_{ij} \equiv \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

と定義する．ただし， $\langle \dots \rangle$  は全データでの平均を意味する．分散共分散行列  $A$  は今の場合物理的な意味がはっきりするよう質量重み付きの座標の2次モーメント行列と呼ぶほうが適当である．基準振動解析では2次モーメント行列は，

$$(2.4) \quad A = k_B T F^{-1}$$

で与えられる．ここで  $k_B$  はボルツマン定数， $T$  は温度である．主成分分析では行列の対角化によって主成分軸をあらわす固有ベクトル  $w_l$  と固有値  $\zeta_l$  が標準固有値問題，

$$(2.5) \quad A w_l = \zeta_l w_l$$

を解くことで求められる．つまり基準振動解析は主成分分析の特殊な場合とみなすことができる．固有値  $\zeta_l$  は  $l$  番目の基準振動のゆらぎの分散をあらわす．基準振動解析の場合には固有値  $\zeta_l$  を

$$(2.6) \quad \zeta_l = k_B T / \omega_l^2$$

と書くと,  $\omega_l$  は固有振動数を表わすことになる. エネルギー極小点を原点として  $l$  番目の基準振動の座標を  $\sigma_l$  とすると, ポテンシャルエネルギーはパラボラとなり,

$$(2.7) \quad E = \frac{1}{2} \sum_{i=1}^{3N-6} \omega_i^2 \sigma_i^2$$

と表わすことができる. ただし, ここでは外力がないものとし, 分子の並進と回転に関わる 6 つの基準振動は除いた  $(3N - 6)$  個の基準振動のみを考慮した. ここで温度  $T$  のボルツマン分布を考えると,  $l$  番目の基準振動の方向で蛋白質が座標  $\sigma_l$  の構造をとる確率は,

$$(2.8) \quad P(\sigma_l) = \sqrt{\frac{k_B T}{2\pi\omega_l^2}} \exp\left(-\frac{\omega_l^2 \sigma_l^2}{2k_B T}\right) = \frac{1}{\sqrt{2\pi\zeta_l}} \exp\left(-\frac{\sigma_l^2}{2\zeta_l}\right)$$

となり, エネルギー極小点を中心とするガウス分布となる.

ここで典型的な蛋白質の振動数分布を図 1 に示す. 振動数分布は蛋白質の立体構造の違いや蛋白質の種類にほとんどよらないことが知られている(ただし, ここでは最低振動数付近の微妙な違いがわかるようにログ・プロットで示した). しかし, 具体的にどの部位がどのようにゆらいているかを知るためにはそれぞれの蛋白質について基準振動解析をおこなう必要がある.

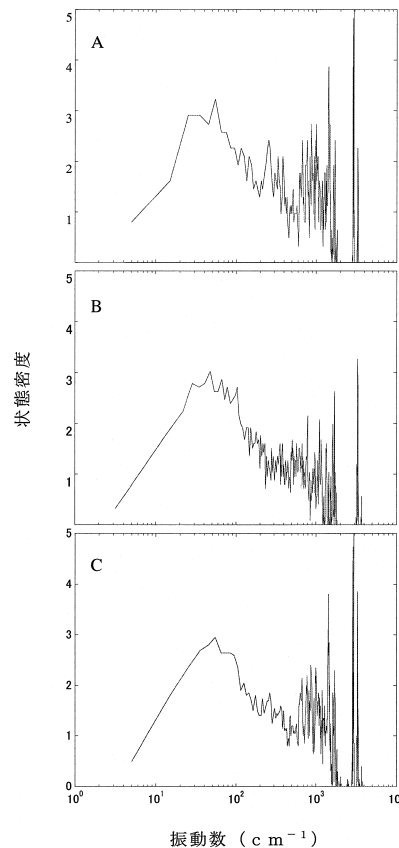


図 1. 様々な数のアミノ酸残基からなる蛋白質の基準振動の振動数毎の分布. A. クランビン(46 残基), B. エグリン  $\alpha$ (70 残基), C. ヒト・リゾチーム(130 残基).

蛋白質の高振動モード(振動数が  $1000\text{ cm}^{-1} \sim 4000\text{ cm}^{-1}$ )は化学結合長・結合角の伸縮運動などの蛋白質のごく一部がゆらぐ局所的な運動である。しかし、振動数が低くなるにつれてより大きな部分がかたまりとなって同時に相関をもって動く集団運動となる。これらを協調的運動 (Collective Motion) あるいは協奏的運動 (Concerted Motion) と呼ぶ。蛋白質の場合、 $150\text{ cm}^{-1}$  以下の振動数を持つモードは協調的運動であり、ゆらぎの範囲は蛋白質を構成するアミノ酸残基のうち数残基以上におよぶ。アミノ酸 100 残基程度からなる蛋白質の場合は、最低振動数は  $2 \sim 5\text{ cm}^{-1}$  程度(周期は 10 ピコ秒前後)である。最低振動数付近ではゆらぎの範囲は蛋白質分子全体におよぶ。また、式 (2.6) から低い振動数の基準振動モードはゆらぎの分散に大きく寄与することがわかる。蛋白質の低振動モードは蛋白質全体がゆらぐ協調的運動であって蛋白質のゆらぎの大きさを主に決めているのである。

大きくゆらぐ方向をあらゆる集団座標とゆらぎの大きさを大まかに決めるためにはより単純化した方法 (Tirion (1996), Bahar et al. (1997)) がある。この方法では、アミノ酸残基を相互作用の単位とするように粗視化する。残基間の距離が近い場合にはある一定の力の定数を持ったばねで繋がれており、残基間の距離が遠いときには相互作用をしないものとする。このように近似すると式 (2.2) の 2 次微分行列は解析的に求められる行列 (Kirchhoff 行列) になり、エネルギー極小化をすることなく即座に求めることができる。この近似法がある程度うまくいっているのは、既に 2.1 節で述べたように蛋白質は充填率が高い状態であり、パッキングの状態が蛋白質のゆらぎの異方性が決まっていることと関係しているのだろう。基準振動解析全般に関してのより詳しい解説には文献 Hayward and Go (1995) や文献 Hayward (2001) がある。

#### 2.4 調和性・非調和性を見る方法：主成分分析

モンテカルロ法や分子動力学法を用いると、蛋白質の立体構造変化を分子シミュレーションのステップに沿って連続的に得ることができる。計算された構造群は物理的な熱ゆらぎを反映している。分子シミュレーションで得られた原子座標をもちいれば直接 (2.3) 式の 2 次モーメント行列を計算することができる。更に (2.5) 式を用いれば主成分軸と各軸方向での分散を求めることができる。モンテカルロ法や分子動力学法に (2.1) 式で与えられる立体構造エネルギーを直接用いたとすると、非調和的なゆらぎもみることができる。水分子など、蛋白質以外の要素を含めた分子シミュレーションをおこなった場合にも、蛋白質の原子座標のみを用いて主成分分析をおこなうことが多い。また、蛋白質の全原子を使わずに、主鎖の原子座標のみを使ったり、各アミノ酸残基の重心座標・ $C^\alpha$  原子の座標や内部座標 (2 面角など) を用いてもよい。ゆらぎを主成分分析で見た場合どのような結果が得られるかについては次の節で詳しく述べる。主成分分析に関する最近の研究についての解説は文献 Kitao and Go (1999), Berendsen and Hayward (2000) をご覧いただきたい。

### 3. ゆらぎを主成分分析でみる

#### 3.1 どのように調和性・非調和性があらわれてくるか？

蛋白質の分子動力学法は基準振動解析よりも先に試みられたが、主成分分析を用いた分子動力学計算の結果の解析は基準振動解析より後でおこなわれた (Levy et al. (1984))。当時は分子動力学計算のシミュレーション時間と基準振動の最低振動モードの周期が同程度であった。従って分子動力学やモンテカルロ法を用いてもエネルギー極小間のジャンプはほとんど起こらず、ほぼひとつの極小点近傍でのゆらぎを観測していたと考えてよい (例としては文献 Horiuchi and Go (1991) がある)。この場合、非調和性の効果は分子動力学で得られたゆらぎが基準振動解析によるゆらぎより小さくなるという現象にあらわれてくる。エネルギーバリアを越えない

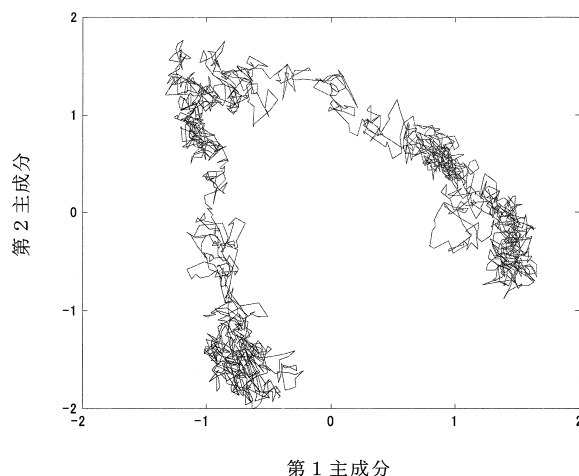


図 2. 蛋白質ヒト・リゾチームが水中で  $1.0 \times 10^{-9}$  秒間ゆらいでいく様子を分子動力学法で計算し、第 1・第 2 主成分軸で張られる空間内に射影したもの．第 1 主成分のゆらぎの大きさは  $0.76 \text{ \AA}$  で、第 2 主成分のゆらぎは  $0.5 \text{ \AA}$  でスケールしてある．

範囲のゆらぎでは、高次項は 2 次関数よりも急激に立ち上がるのでゆらぎを調和近似よりも小さくする．ただし、このような効果はほとんどのモードでは無視できる．最低振動数付近ではゆらぎに 30% 程度までの違いが出ることがあるが、以下述べるようなケースに比べると影響は小さいといってよい．

主成分分析が分子動力学法の解析に本格的に用いられるようになったのは 1990 年代に入ってからである (Kitao et al. (1991), Garcia (1992), Amadei et al. (1993))．分子動力学のシミュレーションにおける時間が基準振動の最低振動モードの周期より十分長くなってくると、それ以前とは異なった大きな非調和性が観測されるようになった．これはエネルギー極小近傍での非調和性ではなくてエネルギー極小間のジャンプによるものである．このようなジャンプによる調和性からのずれは非常に大きい．従って、十分長いシミュレーションをおこなった場合、ゆらぎの非調和性はおもにジャンプが原因である．図 2 には、第 1・第 2 主成分軸で張られる空間に蛋白質の構造変化の履歴を射影し、構造の時間変化を示したものである．このように構造群はいくつかのクラスターを形成している．クラスター内の変化はあるエネルギー極小点近傍のゆらぎであり、クラスター間の遷移はエネルギー極小間のジャンプと考えるのが妥当である．クラスター間の遷移には分子内の原子パッキングの大きな変化を伴うことが判っている．そうなればエネルギー障壁を越える必要があるからである．しかし、このような振る舞いを拡散過程として記述する考え方もある (5.2 節参照)．

### 3.2 調和性と非調和性を含めた蛋白質のゆらぎ

それでは調和性や非調和性をどのように評価すればいいのだろうか？ これに関しては様々な試みがある (Hayward et al. (1994, 1995), Brooks et al. (1995), Janezic and Brooks (1995), Janezic et al. (1995), Troyer and Cohen (1995), Garcia and Harman (1996))．ここでは非調和性因子を用いた評価法 (Hayward et al. (1995)) について説明する．非調和性因子は、分子動力学計算などで得られる非調和性を含むモデルから計算された各主成分のゆらぎと基準振動解析によって計算される同じ軸方向でのゆらぎの比として定義される．従って実際の蛋白質のゆらぎが調和的であれば非調和性因子は 1 となる．非調和性因子が 1 以上になる主成分は全体の数

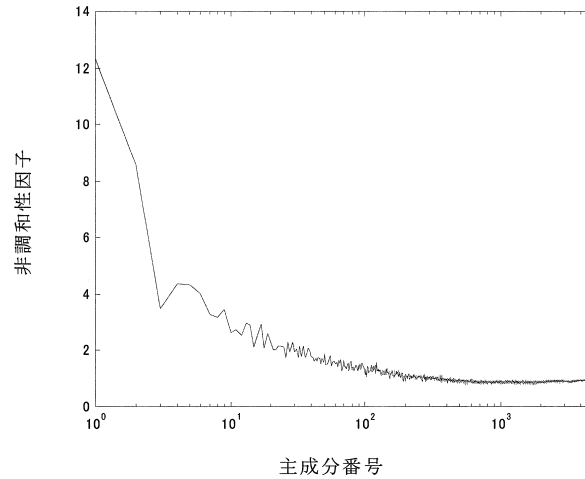


図 3. 図 2 と同じ分子動力学の結果．非調和性因子が主成分番号によってどのように変化するかを示したもの．300 番目以降では 1 に収束して非調和性がなくなっている．

パーセント程度でそれ以外の主成分ではほぼ 1 となる(図 3)．また最初の数個の主成分に関しては非調和性因子が極端に 1 より大きくなるのがいくつかの蛋白質について確かめられている．つまり，ほとんどの自由度で蛋白質のゆらぎは調和的であり，極少数の自由度が非調和的な振る舞いを示す．

#### 4. JAM モデルを用いてエネルギー地形を眺める

##### 4.1 蛋白質のゆらぎのモデル化

これまで述べてきたことから見えてきた蛋白質のエネルギー地形とゆらぎの特徴はどのようにまとめられるだろうか．それは JAM (Jumping-Among-Minima) モデル (Kitao et al. (1998)) という概念に凝縮される．既に見てきたように，蛋白質の天然状態にはたくさんのエネルギー極小があり，エネルギー的には区別しにくい．蛋白質の状態点はほとんどの時間，あるエネルギー極小点の近傍にエネルギー的にトラップされている．しかし，稀に大きな熱ゆらぎが起こり，周りのエネルギー障壁を越えて別のエネルギー極小点の近傍へと構造のジャンプが起こる．極小点近傍でのゆらぎと極小間のジャンプ，この 2 つのタイプのゆらぎは 2 次モーメント(分散共分散)行列において区別して考えることができる．まず 2 次モーメントを計算するために分子動力学やモンテカルロ法をもちいて蛋白質の立体構造を多数生成する．それから直接 2 次モーメントを計算するかわりに，立体構造群を  $M$  個のグループに分ける．このグループ分けの方法は任意だが物理的な意味が明らかになるようにするとよい．今の場合，状態点が属するエネルギー極小ごとにグループに分けるのが望ましい．極小間を隔てるエネルギー障壁には様々な高さのものがあるので，見たいエネルギー解像度に応じて分類の方法を考える必要がある．この分類方法の仕方には物理的な概念が入ってくるのである．何らかの物理的描像に基づいて分類をおこなったとすると，式 (2.3) の 2 次モーメントは，

$$(4.1) \quad a_{ij} = \sum_{k=1}^M f_k (\langle x_i \rangle_k - \langle x_i \rangle) (\langle x_j \rangle_k - \langle x_j \rangle) + \sum_{k=1}^M f_k \langle (x_i - \langle x_i \rangle_k) (x_j - \langle x_j \rangle_k) \rangle_k \\ = s_{ij} + \sum_{k=1}^M f_k (a_{ij})_k$$

のように 2 つの項に分けることができる。ただし、 $\langle \dots \rangle_k$  は  $k$  番目のグループ(構造群)の中での平均を取ることを意味し、 $f_k$  はグループに属するデータ数の総データ数に対する割合を表わす。 $(a_{ij})_k$  は  $k$  番目のグループ内で計算した 2 次モーメントである。式 (4.1) の第 1 項はそれぞれのグループの平均が全体の平均からどの程度ばらついているかを表わし、第 2 項はそれぞれの状態でのゆらぎの荷重平均をあらわす。このように蛋白質のゆらぎを分類することが JAM モデルの描像である。式 (4.1) の第 1 項・第 2 項のゆらぎをあらわすためにそれぞれどのくらいの数の自由度が必要なのかは、第 1 項・第 2 項の行列の固有値問題を解けば明らかになる。ここで述べたような分割は更に  $(a_{ij})_k$  を分割することで階層的におこなうこともできるがここでは省略する。JAM モデルの描像は多くの物質にも当てはまる一般性の高い概念であることがわかるだろう。それでは、蛋白質では更にどのような特徴があるのか。まず重要な第 1 番の点は、極小点近傍のエネルギー曲面の形状はどのエネルギー極小点近傍でもほとんど変わらないことである。これは  $(a_{ij})_k$  が  $k$  に依存しないことから示すことができる。更に極小点近傍のエネルギー曲面の形状はパラボラ(2 次曲面)形であって非調和性はほとんどない。これは既に 3.1 節で述べたことである。第 2 の点は、非調和的なエネルギー極小点間のジャンプ (JAM) は少数の自由度で張られる部分空間内で起こる事である。JAM は確率的に発生するまれな現象である。これについては既に 3.2 節で述べた。




#### 4.2 エネルギー地形はどうなっているのか?

それでは JAM モデルの観点からみた蛋白質のエネルギー地形の特徴を更に詳しく見ていこう。分子シミュレーションや実際のグループ分割の方法についてはここでは述べないが、様々な試みの結果、最終的に JAM モデルの描像による蛋白質のエネルギー地形とゆらぎの特徴は表 1 のように 3 つのタイプの自由度にまとめられることが明らかになった。ほとんどの自由度は、ジャンプと関係せず、調和的なすなわち 2 次曲線のエネルギー面をもっている。計算から求められた形はシミュレーションの長さには依存しないし、基準振動解析による予想と一致するので、エネルギー地形を直接反映していると考えてよい(非調和性因子も 1 である)。実際のヒトリゾチームの場合、調和モードの数は全体の自由度の 95% を占める。非調和な自由度(多重階層モードと単階層モード)はごく少数であり、エネルギー極小間のジャンプに関わるモード(JAM モード)である。ゆらぎが遅くなるほど、またゆらぎが大きくなるほど、それにかかわる自由度は少なくなることがわかる。単階層モードのエネルギー面の形状も調和モードと同じようにパラボラ状であるが非調和性因子は 1 より大きい。これらのモードの方向では表に示したように微細な極小点多数が更に上の階層の 2 次曲面上に分布しているからだと考えられる。多重階層モードには、さまざまな高さのエネルギー障壁がある。この自由度のゆらぎはドメイン運動などの非局所的な大きなゆらぎで、リガンドとの結合に伴う構造変化など、機能と密接に関係するモードである。

ここで示した JAM モデルによるダイナミクスの描像は、多くの蛋白質に共通する物性を示したものである。蛋白質の個性は、例えば機能に関する多重階層モードが具体的にどのようなゆらぎであるかにあらわれてくる。大きな協調的 JAM モードと機能の関係については文献 Berendsen and Hayward (2000) を参照していただきたい。



表 1. 蛋白質のエネルギー地形とゆらぎ .

	多重階層モード	単階層モード	調和モード
エネルギー地形の断面図			
どのようなゆらぎか?	ドメイン運動などの大きなゆらぎ・リガンド結合などの機能と関係	局所的なゆらぎ	局所的なゆらぎ
非調和性因子原子パッキングの変化	2以上大	2から1なし	1なし
二面角の変化	主鎖および側鎖に大きな変化	主に表面の側鎖に変化	微小
ゆらぎの時間領域	$10^{-13}$ 秒以上 (いろいろな時間スケールの運動が共存)	$10^{-13} \sim 10^{-11}$ 秒	$10^{-14} \sim 10^{-12}$ 秒
ヒトリゾチームの場合			
モード数	30	270	5817
全体に占めるモード数の割合	0.5%	4.5%	95%
ゆらぎへの寄与率	82%	15%	3%
ゆらぎの振幅	0.08 Å以上	0.08 ~ 0.015 Å	0.015 Å以下

## 5. サンプリングの問題

### 5.1 ゆらぎの時空とサンプリング

蛋白質のゆらぎの時空間は残念ながらまだ分子動力学法などの分子シミュレーションでは完全にカバーできない．通常の分子動力学法の時間領域は  $10^{-15}$  秒から  $10^{-9}$  秒前後で， $10^{-6}$  秒までの計算 (Duan and Kollman (1998)) はまだ特殊な例とわいていい．一方，蛋白質の特徴的なゆらぎは  $10^{-15}$  秒から秒程度までであるから，分子シミュレーションでは十分なゆらぎのサンプリングをすることができないのである．従って，以前より長いシミュレーションをおこなえばおこなうほど今までは観測できなかったまれに起こるイベントを観測する可能性が高まる．しかし，むやみに長い計算時間を使って計算するのはあまり効率的な方法ではない．1つの長いシミュレーションをおこなうよりもたくさんの短いシミュレーションを異なった初期状態から出発して多数おこなう方がサンプリングの効率がいいこともわかってきており (Caves et al. (1998))，今後は非ボルツマン・サンプリング法などと組み合わせてより効率的な計算がおこなわれることが期待されている．

### 5.2 ジャンプと拡散：2つのモデル

JAM モデルの描像は多くの実験事実 (例えば Frauenfelder et al. (1991)) や分子動力学の結果をよく説明することができる．しかし前節で述べたサンプリングの問題があるので，エネルギー地形の詳細，たとえばエネルギーのくぼみは天然状態でいくつぐらいあるのかといった問題を解決することはできていない．蛋白質のゆらぎは極少数 (数個程度) の自由度では非常に大きいという特徴がある．これらの構造変化にともなうエネルギー障壁はある程度低いと考えてよい．そのため，これらのゆらぎを拡散運動として記述する見方もある (Amadei et al. (1999))．実際，エネルギー障壁がまったくない単純な拡散運動と考えても，ゆらぎの特に大きな主成分軸方向でのゆらぎの時間依存性の特徴をある程度説明することが可能である．しかし Caves らは異なった初期条件から多数の分子動力学計算をおこない，それによって得られた蛋白質の立

体構造は少数の決まった状態にトラップされることを示している (Caves et al. (1998)). この結果から考えるとゆらぎの大きな方向に関して有意な高さのエネルギー障壁があると考えられるほうが自然である. 今後, エネルギー極小点の分布とエネルギー障壁の高さをより詳細に見ていくことが必要になっていだろう. そのためには構造空間のサンプリング方法の改良が必要である.

## 6. JAM モデルをもちいた実験データの解釈

### 6.1 複数の手法をもちいて蛋白質のエネルギー地形を眺める方法

さて, これまで述べたように, 現在, 分子シミュレーションだけでは蛋白質のゆらぎの時空をすべて記述することはできない. 蛋白質のダイナミクスは複雑で, ひとつの実験手段ですべてを明らかにする事は困難なのだ. しかし, 複数の手段から得られるデータを総合的に解釈すれば, より広い時空での蛋白質の振る舞いが明らかになる. 蛋白質の比較的スローなゆらぎを観測するためには, 分子動力学やいろいろな実験データをどのように接続して総合的に解釈すればいいのか? そのために JAM モデルの描像が役立つ. この描像を応用するためのヒントは既にこれまで見てきた以下の点である. I) 蛋白質は多数の調和的な自由度とごく少数の非調和的な自由度をもつ. II) エネルギー極小点近傍のゆらぎは比較的速いゆらぎである. III) エネルギー極小点近傍のゆらぎは調和的でどの極小点にいるかにほとんど依存しない. IV) 非調和なゆらぎは確率的な運動である. V) 分子シミュレーションをもちいると, 比較的速いゆらぎを高解像度で観測できる. VI) 分子シミュレーションより遅い時間領域を観測できる実験手段は存在するが, 実験データには複雑なダイナミクスを反映しているので解釈が難しい.

これらのヒントから, 分子シミュレーションからはエネルギー極小点近傍の速い動きを決定し, 非調和なゆらぎを実験から決定するというアイデアが生まれてくる. 非調和なゆらぎを記述するには少数の自由度のゆらぎを考えればよいのだから, 実験から決定すべきパラメタの数は少数ですむ. これが JAM モデルに基づいて実験データを解析する方法の基本的なアイデアである.

それでは具体的に実験と理論をどのように接続すればよいのか? 例えば実験からある物理量の時空相関関数  $C(r, t)$  が求まったとしよう ( $r$  と  $t$  はそれぞれ座標と時間を表す). 分子シミュレーションからはエネルギー極小点近傍での時空相関関数  $C^{min}(r, t)$  を決めておく. ヒント III) を仮定し外部運動から寄与  $C^{ext}(r, t)$  ・調和的なゆらぎ・非調和なゆらぎがそれぞれ独立であれば,

$$(6.1) \quad C(r, t) = C^{ext}(r, t)C^{JAM}(r, t)C^{min}(r, t)$$

となる. 外部運動が単純なモデルで与えられれば,  $C^{JAM}(r, t)$  を実験と理論のハイブリッドから決める事ができるのである.  $C^{JAM}(r, t)$  は機能的に重要であると期待される比較的ゆっくりとしたゆらぎの情報を与えてくれる. このアイデアはいろいろな実験データとのハイブリッドを可能とする.

非干渉性中性子非弾性散乱では, 動的構造因子  $S_{inc}(Q, \omega)$  が観測される ( $Q$  は波数,  $\omega$  は振動数).  $\omega$  空間では  $S_{inc}(Q, \omega)$  は各成分の畳み込み積分で,

$$(6.2) \quad S_{inc}(Q, \omega) = S_{inc}^{ext}(Q, \omega) \otimes S_{inc}^{JAM}(Q, \omega) \otimes S_{inc}^{min}(Q, \omega)$$

のように与えられる.  $S_{inc}^{min}(Q, \omega)$  を基準振動解析などから計算し,  $S_{inc}^{ext}(Q, \omega)$  のモデルが与えられれば,  $S_{inc}^{JAM}(Q, \omega)$  を決める事が可能となる ( $\otimes$  は振動数に関する畳み込み積分) (Kitao and Go (1999)).  $S_{inc}^{JAM}(Q, \omega)$  がまったく拡散的な運動であれば特徴的な振動数をもたないのでこの

項は  $S_{inc}^{min}(Q, \omega)$  の振動ピークの線幅を広げる効果しかないが,  $S_{inc}^{JAM}(Q, \omega)$  が特徴的な振動数を持つ場合には,  $S_{inc}^{min}(Q, \omega)$  の振動ピークをシフトさせることになる.

実験から時間依存性ではなく物理量  $A$  のアンサンブル平均  $\langle A \rangle$  が得られる場合には, JAM モデルの描像からアンサンブル平均を

$$(6.3) \quad \langle A \rangle = \sum_k f_k \langle A \rangle_k$$

のように, 各エネルギー極小点  $k$  の近傍での平均  $\langle A \rangle_k$  を更にその統計重率  $f_k$  で平均したもので置き換えることができる (Kitao and Wagner (2000)). ヒント II) を考慮すれば, 平均  $\langle A \rangle_k$  は分子シミュレーションから十分計算する事ができる. あとは重率  $f_k$  を実験データの  $\langle A \rangle$  を再現するように決定すればよい. 決めるべきパラメタの数はエネルギー極小点の数であるので,  $\langle A \rangle$  のセットがそれより十分多ければこの方法が使える. この方法は様々な実験データの解析に使えるが, X 線結晶解析や NMR などが最適な方法であると考えられる.

## 6.2 NMR と分子シミュレーションをもちいて得られたエネルギー地形・機能との関係

それでは実際に前節で述べた方法がどのように使えるかをみてみよう. NMR の緩和データからは, ピコ ( $10^{-12}$ ) 秒からミリ ( $10^{-3}$ ) 秒程度までのダイナミクスの情報が得られる. その情報と分子シミュレーションを組み合わせ, 蛋白質のスローなゆらぎを詳細に調べることができる. 以下にはそれらのゆらぎが機能にどのような役割を果たしているかが明らかにされた例をあげる.

蛋白質 CD2 は T 細胞や Natural Killer 細胞の表面に存在し, 別の細胞表面にある蛋白質 CD58 と結合することで 2 つの細胞を接着し, 免疫作用を補助する. CD2 と CD58 の結合が阻害さ

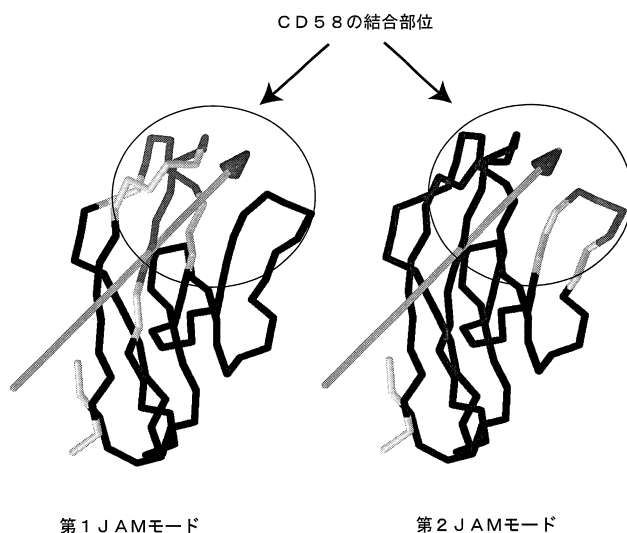


図 4. 蛋白質 CD2 が持つ, CD58 への結合に関係する 2 種類のゆらぎ, 第 1 JAM モードと第 2 JAM モードを示した図. 黒い部分と濃い灰色の部分はそれぞれ固いユニットとして協調的に動く部分 (ダイナミックドメイン) であり, 薄い灰色の部分はそれらをつなぐ部位である. 黒い部分を固定して考えた場合, 濃い灰色の部分が矢印を軸として回転する. どちらのモードも CD58 との結合部位が表面に露出するようにゆらぐモードである. プログラム DynDom (Hayward and Berendsen (1998)) を用いて解析したもの.

れると細胞間の相互作用によって引き起こされるべき免疫反応が阻害されることが知られている。この研究では分子動力学法をもちいてエネルギー極小近傍でのゆらぎを計算し、NMR(核磁気共鳴)の緩和時間のデータ(Wyss et al. (1997))を再現するように重率  $f_k$  を求め、CD2のゆらぎと機能のかかわりを明らかにしている(Kitao and Wagner (2000))。

これによってCD2がどのような構造変化をするかが明らかになった。CD2は主に2つの準安定な状態をもつ。それぞれの状態の統計重率  $f_k$  は0.42と0.24であるから、CD2の構造は42%および24%の確率でそれぞれの状態をとることがわかった。これ以外にもより重率の低い状態が存在する。このようにNMRの実験からもシミュレーションで見られるように蛋白質が少数の準安定な状態間をジャンプしていることが明らかになった。CD2がCD58と結合するには結合面がより外部に露出された形に構造変化する必要がある。CD2が単独で存在する時に起こっている準安定な状態間のジャンプは実はこの構造変化と深く関係している。JAMによって起こる構造変化のモードを図4に示す。CD2は単独では結合面が比較的閉じた形で存在しやすい。しかし、ナノ( $10^{-9}$ )秒程度時間スケールで確率的なジャンプがおり、しばしば結合面がより露出された形に構造変化している。CD2はCD58の存在しない状態でも、CD58との結合するのに適した構造へ変形するモードを持っているのである。このように実験データをJAMモデルの描像を用いて解析することで、構造空間の中でエネルギー極小がどのように分布しているかを調べ、機能のメカニズムに迫ることが可能である。

## 7. おわりに

これまで見てきたように主成分分析は蛋白質のエネルギー地形とゆらぎを解析するためにとても有用な方法である。しかし、蛋白質という物質の特性、すなわちゆらぎの時間領域が広いことによって生じるサンプリングの困難さのため、エネルギー地形の詳細はまだ十分わかっていない。これらの問題は更なる分子シミュレーション法の発展により今後次第に明らかになっていくものと期待される。それに伴って、これまではあまり広い範囲で変えることが難しかった温度や圧力などの状態を変えた解析を更におこなうことで蛋白質の物性と機能のメカニズムが解明されることだろう。およそ1000程度であると思われる機能のパターンごとにエネルギー地形を明らかにしていくことで、蛋白質の機能のメカニズムが解明される日が訪れることが期待される。

## 参 考 文 献

- Amadei, A., Linssen, A. B. M. and Berendsen, H. J. C. (1993). Essential dynamics of proteins, *Proteins*, **17**, 412–425.
- Amadei, A., de Groot, B. L., Ceruso, M. A., Paci, M., Di Nola, A. and Berendsen, H. J. (1999). A kinetic model for the internal motions of proteins: Diffusion between multiple harmonic wells, *Proteins*, **35**, 283–292.
- Bahar, I., Atilgan, A. R. and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Folding & Design*, **2**, 173–181.
- Berendsen, H. J. and Hayward, S. (2000). Collective protein dynamics in relation to function, *Current Opinion in Structural Biology*, **10**, 165–169.
- Brooks, B. R., Janezic, D. and Karplus, M. (1995). Harmonic analysis of large systems. I. Methodology, *Journal of Computational Chemistry*, **16**, 1522–1542.
- Caves, L. S. D., Evanseck, J. D. and Karplus, M. (1998). Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin, *Protein Science*, **7**, 649–666.

- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of American Chemical Society*, **117**, 5179–5197.
- Duan, Y. and Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science*, **282**, 740–744.
- Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991). The energy landscapes and motions of proteins, *Science*, **254**, 1598–1603.
- Garcia, A. E. (1992). Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett.*, **68**, 2696–2699.
- Garcia, A. E. and Harman, J. G. (1996). Simulations of CRP:(cAMP)<sub>2</sub> in noncrystalline environments show a subunit transition from the open to the closed conformation, *Protein Science*, **5**, 62–71.
- Hayward, S. (2001). Normal mode analysis of biological molecules, *Computational Biochemistry and Biophysics* (eds. O. M. Becker, A. D. J. MacKerell, B. Roux and M. Watanabe), Marcel Dekker, New York.
- Hayward, S. and Berendsen, H. J. C. (1998). Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme, *Proteins*, **30**, 144–154.
- Hayward, S. and Go, N. (1995). Collective variable description of native protein dynamics, *Annual Review of Physical Chemistry*, **46**, 223–250.
- Hayward, S., Kitao, A. and Go, N. (1994). Harmonic and anharmonic effects in the dynamics of BPTI: A normal mode analysis and principal component analysis, *Protein Science*, **3**, 936–943.
- Hayward, S., Kitao, A. and Go, N. (1995). Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis, *Proteins*, **23**, 177–186.
- Horiuchi, T. and Go, N. (1991). Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: Human lysozyme, *Proteins*, **10**, 106–116.
- Janezic, D. and Brooks, B. R. (1995). Harmonic analysis of large systems. II. Comparison of different protein models, *Journal of Computational Chemistry*, **16**, 1543–1553.
- Janezic, D., Venable, R. M. and Brooks, B. R. (1995). Harmonic analysis of large systems. III. Comparison with molecular dynamics, *Journal of Computational Chemistry*, **16**, 1554–1566.
- Kitao, A. and Go, N. (1999). Investigating protein dynamics in collective coordinate space, *Current Opinion in Structural Biology*, **9**, 164–169.
- Kitao, A. and Wagner, G. (2000). Structure refinement of human CD2 versus geometric and dynamics constraints, *Proc. Nat. Acad. Sci. U. S. A.*, **97**, 2064–2068.
- Kitao, A., Hirata, F. and Go, N. (1991). The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulation of melittin in water and in vacuum, *Chem. Phys.*, **158**, 447–472.
- Kitao, A., Hayward, S. and Go, N. (1998). Energy landscape of a native protein: Jumping-among-minima model, *Proteins*, **33**, 496–517.
- Levy, R. M., Srinivasan, A. R., Olson, W. K. and McCammon, J. A. (1984). Quasi-harmonic method for studying very low frequency modes in proteins, *Biopolymers*, **23**, 1099–1112.
- Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, *Phys. Rev. Lett.*, **77**, 1905–1908.
- Troyer, J. M. and Cohen, F. E. (1995). Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory, *Proteins*, **23**, 97–110.
- Wyss, D. F., Dayie, K. T. and Wagner, G. (1997). The counterreceptor binding site of human CD2 exhibits an extended surface patch with multiple conformations fluctuating with millisecond to microsecond motions, *Protein Science*, **6**, 534–542.

## Principal Component Analysis to Elucidate Protein Energy Landscape

Akio Kitao

(Graduate School of Science, Department of Chemistry, Kyoto University)

Principal component analysis is a powerful tool to investigate the protein energy landscape since protein fluctuations are highly anisotropic in nature. Because of the extreme complexity of protein three-dimensional structures, a large number of substates, which are energetically comparable with one another and distinguishable by their conformations, exist on the protein energy surface in the native state. Jumping-Among-Minima (JAM) model is proposed in order to explore the protein energy landscape. A novel picture of the protein energy landscape is discussed.

---

Key words: Protein, energy landscape, JAM model, harmonicity, anharmonicity.