

Title	高分子系の主成分解析と緩和モード解析(2004年度後期基礎物理学研究所研究会「モンテカルロ法の新展開3」, 研究会報告)
Author(s)	光武, 亜代理; 飯島, 寛允; 高野, 宏
Citation	物性研究 (2005), 85(3): 376-381
Issue Date	2005-12-20
URL	http://hdl.handle.net/2433/110377
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

高分子系の主成分解析と緩和モード解析

慶應義塾大学 理工学部 光武 亜代理, 飯島 寛允, 高野 宏¹

1 はじめに

高分子系の動的性質をシミュレーションのデータから解析するために用いられてきた方法に、緩和モード解析の方法がある。¹⁾⁻³⁾ この方法は、タンパク質の静的性質の解析に用いられる主成分解析の方法⁴⁾の動的拡張になっていると見ることが出来る。ここでは、複数の準安定状態を持つ小ペプチド鎖の系のモンテカルロ・シミュレーションを行い、その結果に主成分解析と緩和モード解析を適用し、これらの方法の比較検討を行う。

2 主成分解析 (PCA) の方法

最初に、主成分解析 (Principal Component Analysis) の方法について説明する。この方法は、多変数の統計データのある軸に射影したとき、射影したデータの分散が最大となる軸の方向を順次求めていく方法である。タンパク質の構造のゆらぎの静的性質をシミュレーションのデータから解析する方法として用いられている。具体的には、 i 番目の原子団の位置座標 $\{\mathbf{R}_i\}$ に関する分散共分散行列 $\langle (R_{(i,\alpha)} - \langle R_{(i,\alpha)} \rangle)(R_{(j,\beta)} - \langle R_{(j,\beta)} \rangle) \rangle$ に対する固有値問題を解くことにより、軸の方向が固有ベクトルとして、軸に射影したデータの分散が固有値として得られる。ここで、 $R_{(i,\alpha)}$ は \mathbf{R}_i の α 成分である。このことは、次のように導かれる。

今問題とする変数の組を $\mathbf{R} = {}^t(\dots, R_{(i,\alpha)}, \dots)$ のように縦ベクトルで表す。このベクトルの複数のサンプルが、多変数の統計データとして、シミュレーションから得られることになる。このデータを射影する軸方向の単位ベクトルを $\mathbf{v} = {}^t(\dots, v_{(i,\alpha)}, \dots)$; とする。単位ベクトルなので、 $\mathbf{v} \cdot \mathbf{v} = {}^t\mathbf{v}\mathbf{v} = 1$ を満たす。このとき、射影した変数は $\Phi = \mathbf{v} \cdot \mathbf{R} = {}^t\mathbf{v}\mathbf{R} = {}^t\mathbf{R}\mathbf{v}$ と表され、射影したデータの分散は $\langle (\Phi - \langle \Phi \rangle)^2 \rangle = \langle \{ {}^t\mathbf{v}(\mathbf{R} - \langle \mathbf{R} \rangle) \}^2 \rangle = \langle {}^t\mathbf{v}(\mathbf{R} - \langle \mathbf{R} \rangle) {}^t(\mathbf{R} - \langle \mathbf{R} \rangle)\mathbf{v} \rangle = {}^t\mathbf{v}\langle (\mathbf{R} - \langle \mathbf{R} \rangle) {}^t(\mathbf{R} - \langle \mathbf{R} \rangle)\mathbf{v} \rangle = {}^t\mathbf{v}\mathbf{C}\mathbf{v}$ と表される。ここで、 $\mathbf{C} = \langle (\mathbf{R} - \langle \mathbf{R} \rangle) {}^t(\mathbf{R} - \langle \mathbf{R} \rangle) \rangle$ は問題としている変数の分散共分散行列であり、その成分は $C_{(i,\alpha),(j,\beta)} = \langle (R_{(i,\alpha)} - \langle R_{(i,\alpha)} \rangle)(R_{(j,\beta)} - \langle R_{(j,\beta)} \rangle) \rangle$ と書ける。

射影したデータの分散を最大化する問題は、 ${}^t\mathbf{v}\mathbf{v} = 1$ の条件下で ${}^t\mathbf{v}\mathbf{C}\mathbf{v}$ を最大化する問題である。これは、 $\mathcal{R} = \frac{{}^t\mathbf{v}\mathbf{C}\mathbf{v}}{{}^t\mathbf{v}\mathbf{v}}$ に対する変分問題 $\delta\mathcal{R} = 0$ であり、その解となる \mathbf{v} が軸の方向を、 \mathcal{R} の極値が軸に射影したデータの分散を与える。この変分問題は、固有値問題 $\mathbf{C}\mathbf{v}_p = \Lambda_p\mathbf{v}_p$ を規格化条件 ${}^t\mathbf{v}_p\mathbf{v}_q = \delta_{p,q}$ の下で解くことに等価である。 p 番目に大きい固有値を Λ_p 、対応する固有ベクトルを \mathbf{v}_p とすると、 \mathbf{v}_p が p 番目の軸方向を与え、 Λ_p が p 番目の軸方向の分散を与える。

タンパク質のシミュレーションのデータの解析では、シミュレーションで得られた構造の複数のサンプルに対し、重心を揃える、平均構造からの 2 乗変位が最小になるように各構造を回転するという操作を行ってから、分散共分散行列を計算し、主成分解析を行う。分散の大きい順に第 1 主成分、第 2 主成分、... と呼ぶことにすると、例えば第 1 主成分と第 2 主成分に対して値の出現頻度を計算し、第 1、第 2 主成分の値を変数とした自由エネルギー曲面を求め、自由エネルギーの極小 (準安定状態) に対応する主成分の値から準安定状態の構造を同定するということが行われている。

¹E-mail: takano@rk.phys.keio.ac.jp

3 緩和モード解析 (RMA) の方法

次に緩和モード解析 (Relaxation Mode Analysis) の方法について説明する。この方法は、Master 方程式の時間発展演算子の固有値 (緩和率) と固有関数 (緩和モード) を変分法で評価する方法であり、高分子系の構造のゆらぎの動的性質をシミュレーションのデータから解析する方法として用いられている。具体的には、平衡状態での時間相関行列を $\langle (R_{(i,\alpha)}(t) - \langle R_{(i,\alpha)} \rangle)(R_{(j,\beta)}(0) - \langle R_{(j,\beta)} \rangle) \rangle$ とするとき、時間間隔 $t = t_0 + \tau$ と $t = t_0$ の場合の 2 つの時間相関行列に対する一般化固有値問題を解くことにより、緩和率が固有値として、緩和モードが固有ベクトルとして得られる。以下で、このことを導く。このように、緩和モード解析の方法は主成分解析の方法の動的拡張とも見ることができる。

最初に、時間発展演算子の固有関数としての緩和モードについて考える。マスター方程式を $\frac{\partial}{\partial t} P(Q; t) = -\sum_{Q'} \Gamma(Q|Q') P(Q'; t)$ とする。ここで、 $P(Q; t)$ は時刻 t に系が配置 Q をとる確率であり、 $\Gamma(Q|Q')$ は時間発展演算子 (遷移行列) である。以下では、時間発展演算子が平衡分布関数 $P_{\text{eq}}(Q)$ に対し、詳細釣り合いの条件 $\Gamma(Q|Q') P_{\text{eq}}(Q') = \Gamma(Q'|Q) P_{\text{eq}}(Q)$ を満たしている場合を考える。時間発展演算子に対する固有値方程式は、 $\sum_Q \phi_n(Q) \Gamma(Q|Q') = \lambda_n \phi_n(Q')$ および $\sum_{Q'} \Gamma(Q|Q') \phi_n(Q') P_{\text{eq}}(Q') = \lambda_n \phi_n(Q) P_{\text{eq}}(Q)$ で与えられる。ここで、 λ_n は固有値、 $\phi_n(Q)$ は左固有関数である。詳細釣り合いの条件より、右固有関数は $\phi_n(Q) P_{\text{eq}}(Q)$ で与えられる。左右の固有関数は、互いに規格直交関係 $\sum_Q \phi_m(Q) \phi_n(Q) P_{\text{eq}}(Q) = \langle \phi_m \phi_n \rangle = \delta_{m,n}$ を満たすように取る。このとき、固有関数の時間相関関数は $\langle \phi_m(t) \phi_n(0) \rangle = \sum_Q \sum_{Q'} \phi_m(Q) e^{-\Gamma t} (Q|Q') \phi_n(Q') P_{\text{eq}}(Q') = \delta_{m,n} \exp(-\lambda_n t)$ で与えられる。物理量 $A(Q)$ と $B(Q)$ がそれぞれ $A(Q) = \sum_n a_n \phi_n(Q)$ および $B(Q) = \sum_n b_n \phi_n(Q)$ のように展開されたとき、これらの平衡での時間相関関数は $\langle A(t) B(0) \rangle = \sum_n a_n b_n \exp(-\lambda_n t)$ のように与えられる。これより、時間発展演算子の左固有関数 ϕ_n と固有値 λ_n を、それぞれ、緩和モード、緩和率と呼ぶことができる。

上記の固有値問題を変分問題として考える。時刻 t で Q' 、時刻 $t + \tau$ で Q をとる条件付き確率を $T_\tau(Q|Q') = e^{-\Gamma \tau} (Q|Q')$ と書くと、上記の固有値方程式は $\sum_Q \phi_n(Q) T_\tau(Q|Q') = e^{-\lambda_n \tau} \phi_n(Q')$ および $\sum_{Q'} T_\tau(Q|Q') \phi_n(Q') P_{\text{eq}}(Q') = e^{-\lambda_n \tau} \phi_n(Q) P_{\text{eq}}(Q)$ の形で与えられる。この積分形の固有値方程式と等価な変分問題は、 $\mathcal{R} = \frac{\sum_Q \sum_{Q'} \phi_n(Q) T_\tau(Q|Q') \phi_n(Q') P_{\text{eq}}(Q')}{\sum_Q \phi_n(Q) \phi_n(Q) P_{\text{eq}}(Q)} = \frac{\langle \phi_n(\tau) \phi_n(0) \rangle}{\langle \phi_n(0) \phi_n(0) \rangle}$ に対する変分問題 $\delta \mathcal{R} = 0$ として与えられる。この解となる ϕ_n が固有関数であり、それに対する \mathcal{R} の極値が固有値 $\exp(-\lambda_n \tau)$ を与える。

固有値問題が変分問題になったので、試行関数を用いて固有関数 (緩和モード) と固有値 (緩和率) を近似的に計算する方法について説明する。試行関数として次のような近似的固有関数を考える。問題とする物理量を $\mathbf{R}(Q) = {}^t(\dots, R_{(i,\alpha)}(Q), \dots)$ とし、その平衡値からのずれを $\delta \mathbf{R}(Q) = \mathbf{R}(Q) - \langle \mathbf{R} \rangle$ とする。配置 Q から出発して時間 t 後の $\delta \mathbf{R}$ の期待値を $\delta \mathbf{R}(t; Q) = \sum_{Q'} \delta \mathbf{R}(Q') T_t(Q'|Q)$ と書くことにする。変分パラメータ $\mathbf{f} = {}^t(\dots, f_{(i,\alpha)}, \dots)$ を用いて、左固有関数の試行関数として $\phi(Q) = \mathbf{f} \cdot \delta \mathbf{R}(t_0/2; Q) = {}^t \mathbf{f} \delta \mathbf{R}(t_0/2; Q) = {}^t \delta \mathbf{R}(t_0/2; Q) \mathbf{f}$ とする。この試行関数の時間相関関数は、 $\langle \phi(\tau) \phi(0) \rangle = \sum_Q \sum_{Q'} {}^t \mathbf{f} \delta \mathbf{R}(t_0/2; Q) T_\tau(Q|Q') {}^t \delta \mathbf{R}(t_0/2; Q') \mathbf{f} P_{\text{eq}}(Q') = \sum_{Q_1} \sum_{Q_2} {}^t \mathbf{f} \delta \mathbf{R}(Q_1) T_{t_0+\tau}(Q_1|Q_2) {}^t \delta \mathbf{R}(Q_2) P_{\text{eq}}(Q_2) \mathbf{f} = {}^t \mathbf{f} (\delta \mathbf{R}(t_0 + \tau) {}^t \delta \mathbf{R}(0)) \mathbf{f} = {}^t \mathbf{f} C(t_0 + \tau) \mathbf{f}$ で与えられる。ここで、 $C(t) = \langle \delta \mathbf{R}(t) {}^t \delta \mathbf{R}(0) \rangle$ は時間相関行列で、その成分は $C_{(i,\alpha),(j,\beta)}(t) = \langle \delta R_{(i,\alpha)}(t) \delta R_{(j,\beta)}(0) \rangle$ で与えられる。変分関数 \mathcal{R} は $\mathcal{R} = \frac{\langle \phi(\tau) \phi(0) \rangle}{\langle \phi(0) \phi(0) \rangle} = \frac{{}^t \mathbf{f} C(t_0 + \tau) \mathbf{f}}{{}^t \mathbf{f} C(t_0) \mathbf{f}}$ で与えられる。この変分問題 $\delta \mathcal{R} = 0$ は一般化固有値問題 $C(t_0 + \tau) \mathbf{f}_n = e^{-\lambda_n \tau} C(t_0) \mathbf{f}_n$ と等価である。この一般化固有値問題を、規格直交関係 $\langle \phi_m \phi_n \rangle = \delta_{m,n}$ を表す条件 ${}^t \mathbf{f}_m C(t_0) \mathbf{f}_n = \delta_{m,n}$ の下で解くことにより、近似的固有関数と固有値が得られる。緩和モード解析は、シミュレーションから時間相関行列 $C(t)$ を計算し、上記の一般化固有値問題を解くことで、近似的に緩和モードと緩和率を評価す

る方法である.

このようにして求められた固有関数を用いて $\delta\mathbf{R}(Q)$ を $\delta\mathbf{R}(Q) \simeq \sum_n \tilde{\mathbf{g}}_n \phi_n(Q)$ のように展開する. その展開係数は $\tilde{\mathbf{g}}_n = e^{\lambda_n t_0/2} C(t_0) \mathbf{f}_n$ で与えられる. この展開係数 $\tilde{\mathbf{g}}_n$ と緩和率 λ_n を用いて時間相関行列 $C(t)$ を $C(t) = \langle \delta\mathbf{R}(t) \delta\mathbf{R}(0) \rangle \simeq \sum_n \tilde{\mathbf{g}}_n \tilde{\mathbf{g}}_n^t \exp(-\lambda_n t)$ のように再構成することができる. また, 配置 Q のときの緩和モード $\phi_n(Q)$ の値は $\phi_n(Q) \simeq e^{-\lambda_n t_0/2} \mathbf{f}_n^t \delta\mathbf{R}(Q)$ のように与えられる. 以下では, n 番目に遅い (小さい) 緩和率を λ_n とする. すなわち, $0 < \lambda_1 < \lambda_2 < \lambda_3 < \dots$ と選ぶ.

4 小ペプチド鎖のモンテカルロ・シミュレーション

真空中の小ペプチド鎖 Met-Enkephalin のモンテカルロ・シミュレーションを行い, 主成分解析の方法と緩和モード解析の方法を適用した結果について報告する. Met-Enkephalin は5残基から成るペプチドで, アミノ酸配列 Tyr-Gly-Gly-Phe-Met を持つ. この系はシミュレーションや解析のテストを行う系としてよく使われている. 真空中では, 水素結合を作りコンパクトな構造をとり, 複数の安定, 準安定な水素結合のパターンがあることが知られている.

力場に ECEEP/2 を使い, 2面角を変化させ Metropolis 法によるモンテカルロ・シミュレーションを行った. 1回の試行で2面角を ± 5 度の範囲で変化させた場合と, ランダムに2面角を選んだ場合を行った. 角度変化が ± 5 度の範囲の場合, 最初の 25M ステップを捨て, 次の 175M ステップを用いて計算を行った. 角度変化がランダムの場合, 最初の 25M ステップを捨て, 次の 240M ステップを用いた. 100 ステップ毎に構造のサンプルをとった.

タンパク質の場合に行われるように, 重心を一致させた平均構造からの2乗変位が最小となるように, 各構造を回転した後, 各残基の C_α の位置に関する分散共分散行列, 時間相関行列を計算し, 解析を行った. 緩和モード解析には, 角度変化が ± 5 度の範囲の場合, $t_0=6k$ ステップ, $\tau=1k$ ステップ, ランダムの場合, $t_0=1k$ ステップ, $\tau=1k$ ステップというパラメータを用いた.

問題とする物理量として, j 番目の C_α の重心からの位置ベクトルの β 成分 $R_{(j,\beta)}$ をとり, 時間相関行列 $C_{(i,\alpha),(j,\beta)}(t) = \langle (R_{(i,\alpha)}(t) - \langle R_{(i,\alpha)} \rangle) (R_{(j,\beta)}(0) - \langle R_{(j,\beta)} \rangle) \rangle$ を計算し, 解析を行った.

以下, 角度変化を ± 5 度の範囲に限定した場合の結果を示す. 主成分解析の固有値分布と緩和モード解析の緩和率分布を図1に示す. 主成分解析の固有値 (分散) が1番目が突出して他より大きいのに対し, 緩和モード解析の緩和率は, 1, 2, 3番目が他より小さくなっている.

図2に, $p=1, 2, 3$ に対し, 主成分解析の p 番目の固有ベクトルの (i, α) 成分 $v_{p,(i,\alpha)}$ と緩和モード解析の p 番目の緩和モードの $\tilde{\mathbf{g}}_{p,(i,\alpha)}$ を示す. ここで, $\tilde{\mathbf{g}}_{p,(i,\alpha)}$ は $R_{(i,\alpha)}(t) - \langle R_{(i,\alpha)} \rangle$ を緩和モード

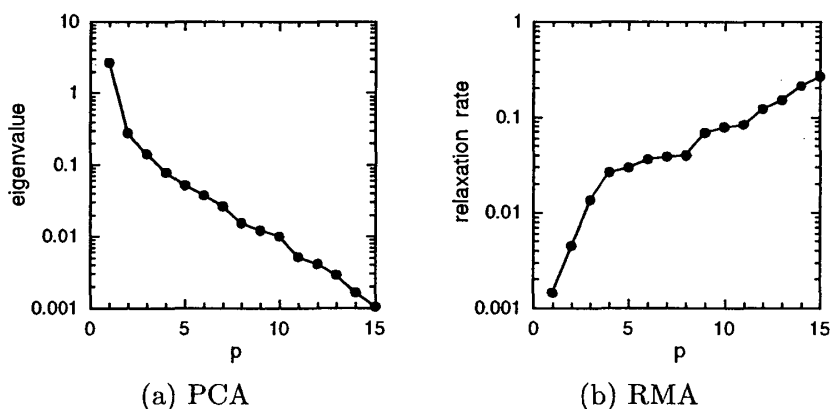


図1: 主成分解析の固有値 (分散) 分布 (a) と緩和モード解析の固有値 (緩和率) 分布 (b).

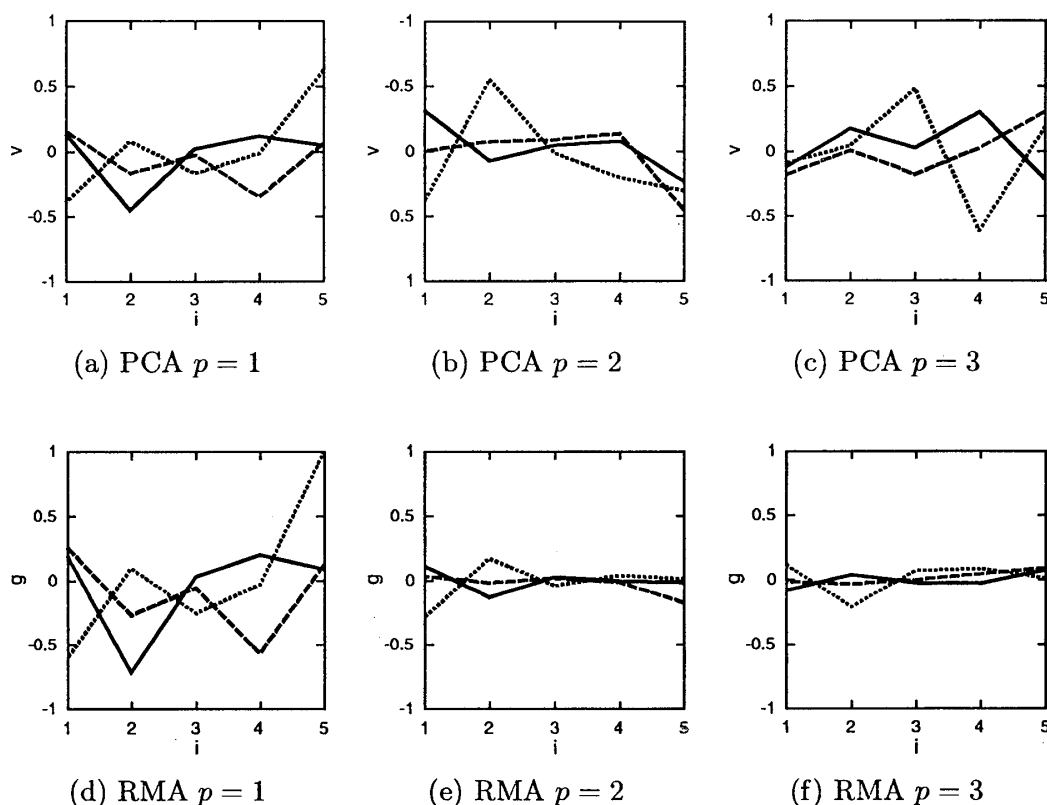


図 2: 主成分解析の p 番目の固有ベクトル $v_{p,(i,\alpha)}$ (a), (b), (c) と緩和モード解析の p 番目のモードの $\tilde{g}_{p,(i,\alpha)}$ (d), (e), (f). $p = 1, 2, 3$ の結果. 横軸は i , 3本の線が3つの α に対応する.

で展開したときの p 番目の緩和モードの係数で, p 番目の緩和モードの $R_{(i,\alpha)}(t) - \langle R_{(i,\alpha)} \rangle$ への寄与を表す. $p = 1$ に対しては, $v_{p,(i,\alpha)}$ と $\tilde{g}_{p,(i,\alpha)}$ はほとんど同じであるのに対し, $p = 2$ に対しては, 傾向は似ているものの異なっている. $p = 3$ に対しては, 全く異なっている.

図 3 に自己相関関数 $C_{(i,\alpha),(i,\alpha)}(t) = \langle (R_{(i,\alpha)}(t) - \langle R_{(i,\alpha)} \rangle)(R_{(i,\alpha)}(0) - \langle R_{(i,\alpha)} \rangle) \rangle$ を, シミュレーションから直接求めた場合と, 緩和モード解析の方法で評価した λ_p と $\tilde{g}_{p,(i,\alpha)}$ から再構成した場合の結果を示す. $t_0 = 6k$ ステップ, $t_0 + \tau = 7k$ ステップ (図中の $t = 6$ と 7 に対応) という短時間の相関行列から λ_p と $\tilde{g}_{p,(i,\alpha)}$ を評価したにもかかわらず, 再構成した自己相関関数はシミュレーションの結果と長時間迄良く合っており, 緩和モード解析による評価が良いことがわかる.

主成分解析の第 1 主成分と第 2 主成分の値の出現頻度を計算し, その頻度の対数を取り, 符号を変えることで, 第 1 主成分と第 2 主成分の関数としての自由エネルギーを評価することができる. 同様に, 緩和モード解析の第 1 および第 2 緩和モードの値の出現頻度を計算し, 第 1 および第 2 緩和モードの値の関数としての自由エネルギーを評価することができる. 図 4 にその結果を示す. いずれの場合も, 3 つの自由エネルギーの極小点がある. 第 1 主成分あるいは第 1 緩和モードの値の正の方に 1 つの深い極小点があり, 負の方に 2 つの比較的浅い極小点がある. 主成分解析の場合, 浅い方の極小点の第 1 主成分の値が 2 つで異なっているのに対し, 緩和モード解析では, 浅い方の極小点の第 1 緩和モードの値がほぼ同じになっている. 即ち, 第 1 緩和モードは, 安定状態 (深い極小点) と 2 つの準安定状態 (浅い極小点) の間の移り変わりを記述し, 第 2 緩和モードは, 2 つの準安定状態 (浅い極小点) の間の移り変わりを記述していることがわかる. 主成分解析より緩和モード解析の方が, 動的性質に関しては良い記述をしていると考えられる.

以下では, 2 面角の変化を ± 5 度の範囲にした場合の結果とランダムに 2 面角を選んだ場合の結

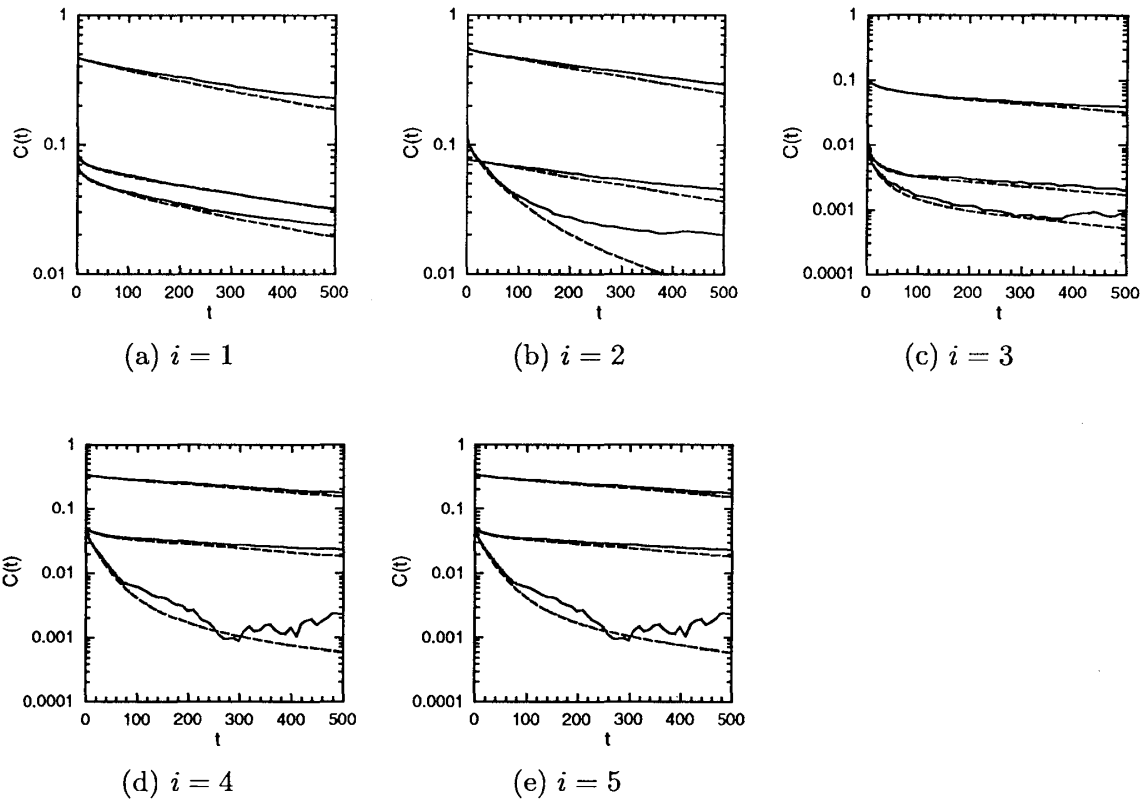


図 3: 自己相関関数 $C_{(i,\alpha),(i,\alpha)}(t)$ の t 依存性. (a), (b), (c), (d), (e) が, それぞれ, $i = 1, 2, 3, 4, 5$ の場合. 各図中の 3 組の線が 3 つの α に対応. 各組の線の内, 実線がシミュレーションから直接計算したもの. 破線が緩和モード解析から評価した λ_p と $\tilde{g}_{p,(i,\alpha)}$ から再構成したもの. 時間 t の単位は 1k ステップ.

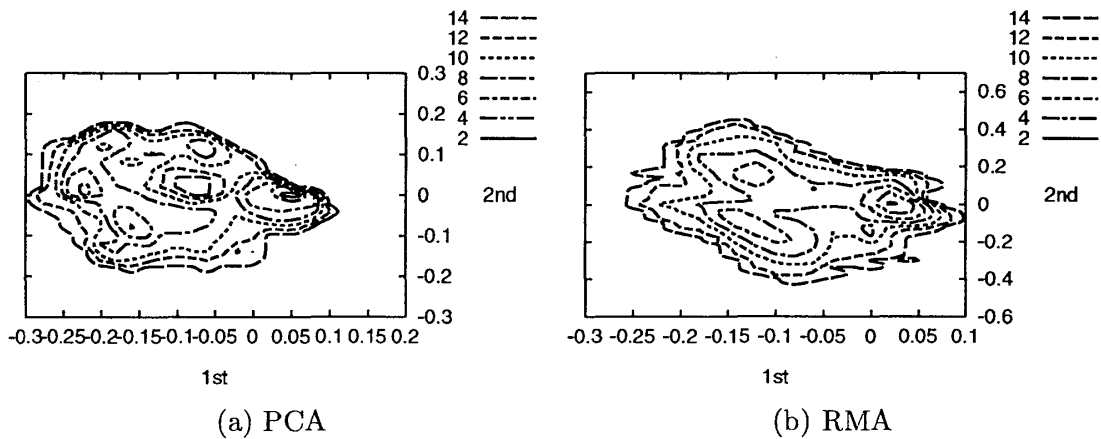


図 4: 主成分解析の第 1 主成分と第 2 主成分の関数としての自由エネルギー曲面の等高線 (a) と緩和モード解析の第 1, 第 2 緩和モードの値の関数としての自由エネルギー曲面の等高線 (b).

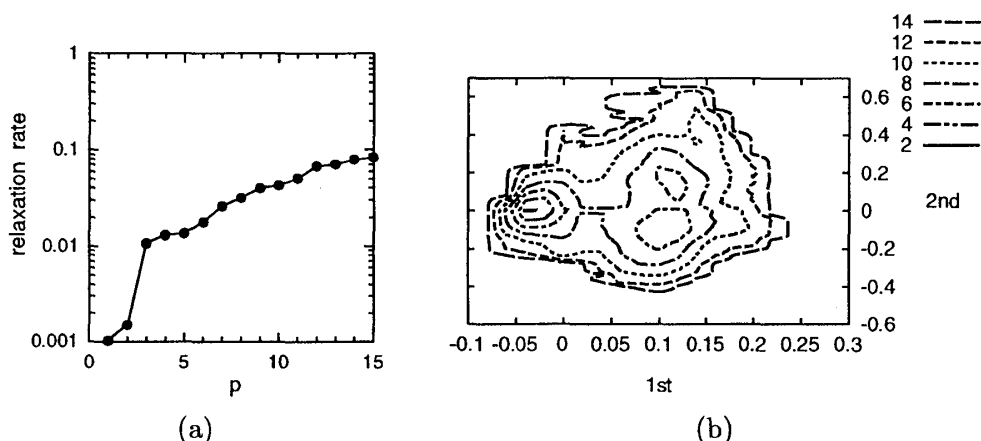


図 5: ランダムに 2 面角を選んだ場合の緩和モード解析の結果. 緩和率分布 (a) と第 1 および第 2 緩和モードの値の関数としての自由エネルギー曲面の等高線 (b).

果を比較する. 図 5 にランダムの場合の緩和率分布と, 第 1 および第 2 緩和モードの値の関数としての自由エネルギー曲面の等高線を示す. 図 1(b) と図 4(b) の 2 面角の変化を ± 5 度の範囲にした場合の結果と比較すると, 図 5(a) の緩和率分布は, 1, 2 番目の緩和率が互いに近く, 他よりかけ離れて小さいことが図 1(b) と異なっている. 図 5(b) の自由エネルギー曲面は, 浅い方の 2 つの極小点の第 1 緩和モードの値がほぼ同じになっているという特徴は図 4(b) と同じであるが, 地形は異なっている. これらは, ダイナミクスの違いの反映であると考えられる.

5 まとめ

小ペプチドの真空中のモンテカルロ・シミュレーションを行い, 主成分解析と緩和モード解析を行った. 第 1 緩和モードは主成分解析の第 1 固有ベクトルとほぼ同じで, 安定状態と準安定状態間の遷移を記述するモードであった. 第 2 緩和モードは主成分解析の第 2 固有ベクトルと異なっており, 2 つの準安定状態間の遷移を記述するモードであった. 本研究により, コンパクトな構造をとる高分子系に対しても緩和モード解析の方法が有効であり, 安定状態, 準安定状態間の遷移を記述できることが示された. また, 緩和モード解析の方法によって, モンテカルロ・シミュレーションの遷移確率の選び方, 即ちダイナミクスの違いの影響も見ることが示された.

今後, タンパク質の平衡状態のまわりの揺らぎを動的に解析するために, 水分子を陽に取り扱う水溶液中のタンパク質の分子動力学シミュレーションに対して, 緩和モード解析を適用していくことが考えられる.

参考文献

- 1) H. Takano and S. Miyashita, J. Phys. Soc. Jpn. **64** (1995), 3688.
- 2) S. Koseki, H. Hirao and H. Takano, J. Phys. Soc. Jpn. **66** (1997), 1631.
- 3) H. Hirao, S. Koseki and H. Takano, J. Phys. Soc. Jpn. **66** (1997), 3399.
- 4) A. Kitao, F. Hirata and N. Go, J. Chem. Phys. **158** (1991), 447.