

**修士論文** (2010 年度)**混雑効果に誘導される高速なフォールディング過程**

東京大学大学院総合文化研究科広域科学専攻

関連基礎科学系 大竹 亮介

**目次**

|          |                                  |            |
|----------|----------------------------------|------------|
| <b>1</b> | <b>研究の背景</b>                     | <b>658</b> |
| 1.1      | 細胞内の分子混雑状態 . . . . .             | 658        |
| 1.2      | タンパク質のフォールディング過程 . . . . .       | 659        |
| 1.3      | 近年のタンパク質計算科学の発展と本研究の立場 . . . . . | 661        |
| 1.4      | タンパク質に対する混雑効果の研究 . . . . .       | 662        |
| 1.5      | 本稿では何を論じるか . . . . .             | 663        |
| <b>2</b> | <b>モデルの設定と方法</b>                 | <b>664</b> |
| 2.1      | タンパク質の HP モデル . . . . .          | 664        |
| 2.2      | 本研究で対象とするモデルの設定 . . . . .        | 665        |
| 2.3      | velocity Verlet 法 . . . . .      | 667        |
| 2.4      | グラフィックスツールの作成 . . . . .          | 668        |
| 2.5      | Shape Parameter の設定 . . . . .    | 668        |
| 2.6      | 配列の決定 . . . . .                  | 669        |
| <b>3</b> | <b>通常の分子動力学計算によるシミュレーション</b>     | <b>673</b> |
| 3.1      | MD 計算の結果 . . . . .               | 673        |
| 3.2      | ランダムな初期状態からの緩和 . . . . .         | 676        |
| <b>4</b> | <b>レプリカ交換法による解析</b>              | <b>678</b> |
| 4.1      | レプリカ交換分子動力学法の概要 . . . . .        | 678        |
| 4.2      | レプリカ交換分子動力学法のアルゴリズム . . . . .    | 678        |
| 4.3      | アルゴリズムの導出・確認 . . . . .           | 679        |
| 4.4      | REMD の結果 . . . . .               | 681        |
| 4.5      | 自己相関関数の転移 . . . . .              | 684        |
| 4.6      | 二次元ヒストグラムの温度変化に伴う遷移 . . . . .    | 685        |
| 4.7      | 安定構造の可視化 . . . . .               | 685        |
| <b>5</b> | <b>Confined 系</b>                | <b>692</b> |
| 5.1      | 束縛ポテンシャルの導入 . . . . .            | 692        |
| 5.2      | Confined 系の REMD による解析 . . . . . | 693        |

|     |                                 |     |
|-----|---------------------------------|-----|
| 5.3 | Confined 系における安定構造 . . . . .    | 694 |
| 5.4 | 自己相関関数のピークの変化 . . . . .         | 695 |
| 5.5 | 閉じ込めによる高速なフォールディングの発見 . . . . . | 695 |
| 6   | 自由エネルギー地形による議論 . . . . .        | 704 |
| 6.1 | アンブレラサンプリング法 . . . . .          | 704 |
| 6.2 | 温度アンブレラサンプリング法 . . . . .        | 706 |
| 6.3 | シミュレーションの方法 . . . . .           | 707 |
| 6.4 | 自由エネルギー地形の概形 . . . . .          | 708 |
| 7   | Discussion . . . . .            | 712 |
| 7.1 | 示された結果のまとめ . . . . .            | 712 |
| 7.2 | 高速なフォールディングはなぜ起こるか . . . . .    | 712 |
| 7.3 | 本研究から提案されること . . . . .          | 713 |
|     | 謝辞 . . . . .                    | 714 |
|     | 引用文献 . . . . .                  | 714 |

## 1 研究の背景

### 1.1 細胞内の分子混雑状態

生物の細胞内では、生体由来の各種タンパク質や RNA、脂質膜などの巨大分子が極めて高い濃度で水溶している。体積にして細胞の 20-30% は巨大分子で占められていると考えられており [1]、細胞内における巨大分子間の距離は *in vitro* 実験で扱われる希薄溶液に比べ非常に短い (fig.1(a))。このような環境は「分子混雑状態」「分子クラウディング状態」と呼ばれ、溶液中の分子の振舞いが希薄溶液の場合と異なる例が多数観察されている [2][3][4]。分子混雑状態の下で分子の振舞いに違いが起こるこの効果は「混雑効果」あるいは「分子クラウディング効果」と呼ばれ、生物物理学の分野で特に関心を持たれている現象の一つである。分子混雑状態下では、1 分子あたりに割り当てられる体積が減じられるため、分子の配位エントロピーが減少し、結果、溶液全体の自由エネルギーは増加し、分子が存在することによる化学ポテンシャルも上がる [5]。このような排除体積効果により、実験的に観察される分子混雑効果は起こされていると考えられており、実際、多くの実験・理論研究がこれまでに行われてきた。

空間中の体積を広く占有しながらプロセスが進む反応に対し、分子混雑状態は特に影響する。特異な立体構造をとり安定化するタンパク質や DNA に対する混雑効果が多く報告されており、特に固有の立体構造へとフォールディングされることで初めて機能を発現する球状タンパク質への混雑効果は特に顕著である [6][7]。本稿は、このタンパク質に対する混雑効果に関し筆者が粗視化モデルを設定し、数値実験研究を行った結果を報告するものである\*1。

\*1 本論文には多数の図が掲載されているが、図のカラー版は次のホームページ上で閲覧可能である：  
<http://chaos.c.u-tokyo.ac.jp/~ohtake/>

## 1.2 タンパク質のフォールディング過程

ここで、タンパク質のフォールディングに関して説明する。タンパク質は、20 種類のアミノ酸を単位とする直鎖状のポリマーであり、タンパク質中で連なっているそれぞれのアミノ酸は残基と呼ばれる。タンパク質が生体内で合成されると、アミノ酸残基同士の親水・疎水相互作用、静電的相互作用、あるいは架橋形成などにより立体的に折畳まれる<sup>\*2</sup>。タンパク質のフォールディング過程は、分子中の異なる成分間で働く相互作用エネルギーと分子鎖が持つコンフォメーションエントロピーに基づく自由エネルギーの和で表現される全自由エネルギーを最小にする構造を得ようと自己組織化していく過程である [9]。生体内で機能しているタンパク質は、アミノ酸配列により決まる自由エネルギー地形のうちで最安定な状態をとっていると考えられており、これを天然構造と呼ぶ。しかしながら、生物が持つタンパク質の平均残基数は 300 残基ほどと見積もられており、一つのタンパク質がとり得る立体構造は極めて膨大な数となる。このことに関連し、熱・統計力学の思考実験から重大なパラドックスが指摘された。

### 1.2.1 Levinthal のパラドックス

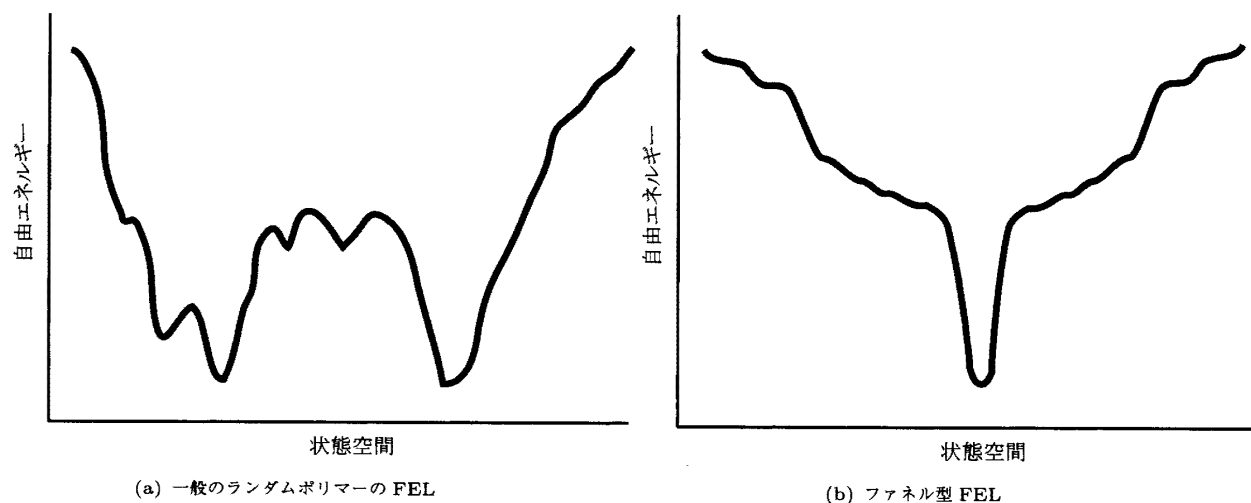
1968～69 年、Levinthal はタンパク質のフォールディングに関し、発表した論文および学会講演の中で次のような指摘を行っている [10][11]：

「タンパク質がランダムコイルの状態から折畳まれていき、天然状態と呼ばれる特有の立体構造をとるようになるという実験事実がある。この結果を説明する一つの仮説は、天然状態は自由エネルギー最小の状態であり、熱平衡状態下では天然状態をとる分子の数が最も多くなるという見方である。別の見方としては、天然状態は特異に選ばれた準安定状態の一つに過ぎないとする立場で、この場合、熱平衡状態は実現されていないということになる。」

「タンパク質 1 分子にはおおよそ 2000 の原子が含まれている。空間的な自由度が結合様式により制限されていることを考慮すると、原子間の結合角に関し 300 ほどの自由度がある。仮に原子間の結合角が 10 の状態しかとれないとしても、可能な構造は  $10^{300}$  個あることになる。天然のタンパク質は数秒以内に折畳まれるのでタンパク質が立体構造としてとることが可能な状態をすべてサンプリングしているとは考えられない。」

一つの立体構造をサンプリングする時間が 1 フェムト秒 ( $10^{-15}$  秒) であると仮定しても、 $10^{300}$  という状態数がすべてサンプリングされるまでには天文学的な時間が掛かることになる。これは「Levinthal のパラドックス」と呼ばれる問題として生物物理学の分野では広く認知されている。タンパク質フォールディングは、熱平衡により実現されているものとは考えにくく、合成された時点のランダムな状態から天然状態へと効率的に折畳まれていると考えられる。生体内のタンパク質には、速やかに折畳まれるようにフォールディングの経路が与えられているとする仮説が、次節で説明する「フォールディング・ファネル・モデル」である。

<sup>\*2</sup> 近年、立体構造を持たずに生体内で機能するタンパク質である IDP が報告され、生物が持つタンパク質のうち、種類にして 50% 程度が IDP であるという知見が得られた [8]。しかし、酵素活性を示す球状タンパク質や細胞骨格を作る構造タンパク質は特定の立体構造をとることでそれぞれの機能が初めて発現される場合が多く、従来通りの「タンパク質は立体構造をとる」という知見は、IDP でなければ正しい。



**fig.1** : ランダムポリマーの FEL の模式図 一般にランダムヘテロポリマーの FEL は (a) 図のようになるが、天然タンパク質の FEL は進化の過程で (b) 図のように効率的なフォールディングが起きるようデザインされているという仮説がフォールディング・ファネル・モデルである。

### 1.2.2 タンパク質のフォールディング・ファネル・モデル

統計力学に基づいたタンパク質の理論モデルである格子モデルやヘテロ高分子鎖モデルなどの計算機シミュレーション研究から、タンパク質のフォールディングに関し、「フォールディング・ファネル」という概念が生まれた。ファネル (funnel) とは「漏斗」という意味である。

多次元の相空間中の点によりタンパク質の立体構造を記述し、それぞれの点において定義される自由エネルギーを別の軸にとり、曲面としたものを自由エネルギー地形、FEL (Free Energy Landscape の略) と呼ぶが、フォールディングするランダムヘテロポリマーの FEL は、模式的には一般に fig.1(a) のように描けることが知られている。自由エネルギーが最低となる構造が天然構造であるとされるが、fig.1(a) のような FEL を持つヘテロポリマーの場合、天然構造に近い自由エネルギーを持つ構造は全く似ていない別の状態が多数存在し、ランダムな状態からフォールディングを出発させると高い確率でミスフォールド状態 (天然構造ではないフォールディング構造に落ち着いた状態) に陥る。いったんミスフォールド状態に陥ると、天然構造へ至るには FEL 中の高いエネルギー状態を経なければならず、この FEL 中の山を越えて尚且つ最安定状態へ辿り着き、状態が落ち着くという可能性は極めて低い。つまり、このような FEL を持つ分子の場合、分子同士で立体構造が揃う場合はほとんどないと言ってよい。

一方で、生体内で機能するタンパク質の FEL は fig.1(b) のような形になっており、フォールディングはこの漏斗状の FEL の上を滑るようにして進行するとするモデルが、「タンパク質のフォールディング・ファネル・モデル」である。タンパク質の状態が自由エネルギーのより低い方向へと遷移していくことで、ランダムな立体構造から出発しても天然状態へ高速に至るとしている。

FEL はタンパク質のアミノ酸配列と一対一に対応すると考えられるが、アミノ酸配列の進化の過程でフォールディングが速く完了するよう FEL は洗練化されていき、滑らかなファネル状となっていったとこの仮説は説明する。すなわち、生体内で働くタンパク質の FEL は進化的にデザインされており、天然構造にあたる FEL 中の領域が際立ったエネルギー谷であるというだけでなく、フォールディングが階層的に、滑らかに進行するように自由エネルギー地形が広い範囲にわたってうまく設計されているとする仮説である。フォールディング・ファネルにも準安定状態にあたる凹部は多数存在しているが、これが実験的に観察されるフォールディング中間状態に相当するものであり、また、この準安定状態を囲む自由エネルギー地形中の

峰は状態遷移の律速の原因となるとはいえ、準安定状態からの脱出は高確率で起こると考えている。

現在のところ、フォールディング・ファネルは概念上のモデルであり、実験的に、あるいは高精度のシミュレーション上で確認された経験は、巨大タンパク質に関しては、現在のところはない（ペプチド程度の低分子鎖ならば報告がある [14]）。しかしながら Levinthal のパラドックスを回避する考え方であり、また、合理的なモデルでもあるので、フォールディング・ファネル・モデルは多くの研究者に受け入れられ、また、この仮説を検証するための理論的、実験的な研究がこれまで活発に行われてきた。ランダムな状態から出発し、高確率で最安定構造へ至る FEL を持つモデル鎖をデザインすることにも成功しており [15]、生体内でも同じ状況が実現されているという可能性は十分にある。だが、大規模コンピュータを用いた近年の高精度タンパク質フォールディングシミュレーションは、天然状態へ向かう高速なパスの存在に肯定的な結果をもたらしてはいない。注目するタンパク質に含まれる全ての原子を考慮する高精度のモデルによりフォールディング・シミュレーションを行っても、やはり準安定状態に引っ掛かってしまい、天然構造が実現されない場合があることがわかってきた。

近年、スーパーコンピュータを用いた手法をとる計算化学の研究者から、タンパク質フォールディングに関する多くの報告が行われている。これらの研究で用いられているシミュレーション法は、ミクロなスケールで起こっているタンパク質フォールディングダイナミクスをほぼ狂いなく再現する手法であると考えてよいが、天然のタンパク質分子をコンピュータ上に再現し、ランダムな状態から分子動力学計算を開始しても、フォールディング・ファネル・モデルで唱えられたような高速なフォールディング過程が観察されないと報告されている [16]。FEL 中でフォールディング経路がデザインされており、タンパク質はこの経路の存在のために効率よくフォールディングされているという仮説に多くの例外が出てしまうことになる。

これに関連し、本稿は、混雑効果により準安定状態への滞留が解消され、高速なフォールディング過程が形成される場合があることを報告するものである。

### 1.3 近年のタンパク質計算科学の発展と本研究の立場

2010 年現在、大規模並列分散計算処理技術の発展により、タンパク質フォールディングの高精度シミュレーションが実際的な時間内で実行可能となりつつある。ここで高精度シミュレーションと言ったのは、タンパク質分子と溶媒となる水分子を構成する全ての原子を考慮する大規模計算のことであり、以下の例がある：

- ・ 全原子分子動力学法：原子間の相互作用を古典力学で表される力場により経験的に近似しニュートン運動方程式を数値的に解く方法
- ・ 第一原理分子動力学法：分子間ポテンシャルを量子力学的に（第一原理的に）求めるシミュレーション法

ところで、このような高精度のシミュレーション方法が開発され、実行されている中では、本稿にて扱うような抽象化したモデルは研究対象としての意義が薄くあるようにも思える。しかし、現在の高精度シミュレーションには扱える問題と扱えない問題とがあり、この区別は主に高精度シミュレーションの計算量の膨大さに起因する。抽象化したモデルでまず議論すべき問題は依然として存在しており、その一つが混雑効果についての研究であると筆者は考えているが、まず、現在の高精度シミュレーションの技術的限界はどこにあるのかについてここで簡単に説明を行う。

先に紹介した分散コンピューティングシステムは、数千～数十万の CPU コアを持つ巨大な計算機環境であるが、これらの計算能力をもってしても、タンパク質のシミュレーションは重い負荷の掛かる問題であり、計算量の限界がすぐに来る。タンパク質は数千の原子を内在する、自由度が極めて高い系であって、タンバ

ク質と溶媒が持つ全原子を考慮する分子動力学シミュレーションを行う場合、その計算が考慮しなければならない変数の数は莫大な数字となる。更にこれらの原子間で考えるべきポテンシャルを計算することになるから、いかにフォールディングシミュレーションが重い計算であるかがわかる。

計算量の膨大さを更に上乗せしているのは、シミュレーションを実行すべき時間の長さである。タンパク質は、フォールディングや「開閉」などの大規模な構造変化を起こすことで生体内で機能を示すとされるが、そのような構造変化が起きる時間スケールはマイクロ秒からミリ秒、あるいは秒のスケールである。一方で、高精度分子動力学計算が1ステップ進む際、進む時間はフェムト秒 ( $10^{-15}$ ) のスケールであるから  $10^{10} \sim 10^{15}$  倍の時間スケールの差があり、フォールディングシミュレーションを実行するには膨大な計算量が必要となる。

計算量は極めて巨大だが、これらの方法で行われるシミュレーションの結果は非常に精度が高く、ミクロスケールで起こっているタンパク質の動的振舞いをほぼ狂いなく再現するものであると考えてよい。この高精度シミュレーションにより、先のフォールディング・ファネルの仮説は検証ができるはずだ。高精度の分子動力学計算を天然構造から外れた状態から出発させるとどうなるだろうか？天然タンパク質の FEL は巧妙にデザインされており、漏斗状であると仮定される FEL 上を滑っていき、漏斗の底にあたる天然構造へ自然に至るというフォールディング過程は観察されるだろうか。実際はそうではなく、天然構造とは別の準安定状態に引っかかり、フォールディングに失敗する例がほとんどであるという報告が行われている [16]。大域的な自由エネルギー地形は、フォールディング・ファネルへの落ち込みを誘導させる経路を与えるものではない、あるいは、仮に FEL が漏斗状であったとしても、FEL の斜面に多くの引っ掛かりがあり、ランダムな状態のタンパク質を溶液中に置いただけではフォールディングは完了されないと推測される。本研究では、タンパク質と水だけではなく、更に分子混雑がないとフォールディングは完了できないという仮説を唱えるものであり、その検証を、独自に設定した粗視化モデルにより行う。

ところで、本節で述べたように、タンパク質のシミュレーションは、精度を上げることは可能であるが、それに連れて計算量が膨大化し、シミュレーションが行えなくなるというジレンマがある対象である。1分子のシミュレーションでさえ現在の最も高性能なスーパーコンピュータの能力限界を越えるのだから、タンパク質の多分子シミュレーションは、現時点では到底不可能であり、生体内の分子混雑状態を高精度に再現することは技術的限界を大きく越えている。そして本研究テーマにて課題としたいことは、具体的なタンパク質のフォールディングに対する混雑効果ではなく、フォールディングするヘテロポリマーに空間的な束縛、周辺の障害物との衝突を導入し、結果、フォールディングに対しどのような影響が出るのか検証するということであり、物質に関しては具体性のない興味である。精度の追求ではなく、対象の詳細にはとらわれない定性的な議論を行うとする立場を本稿はとる。どのような単純化したモデルを立てて研究を行うかについては次章を参照して頂きたい。また、単純化したモデルでないと本稿の研究対象は解析が遂行できないという現実も、いま述べた立場を筆者がとる主な理由の一つである。

## 1.4 タンパク質に対する混雑効果の研究

タンパク質に対する混雑効果の研究は、実験的にも理論的にもこれまで多く行われてきた。実験研究では、デキストランなどのクラウディング剤により大腸菌由来の DNA 結合性タンパク質の DNA 結合確率の増加を観察した Murphy *et al.* 1994 [2] や、やはりクラウディング剤の付与により人工的に分子混雑状態を作り出し、ニワトリリゾチームの転移温度の上昇を報告した Sasahara *et al.* 2003 [6] などが挙げられる。また理論方面からも、混雑効果によりタンパク質のフォールディング確率が上がる、あるいはフォールディング転移温度が上がるという現象は、物理モデルを使った研究により解析されてきた。Takagi *et al.* 2003 [17] は、

郷モデルと呼ばれるタンパク質の粗視化モデルに、シャペロニンタンパク質による閉じ込めを想定した束縛ポテンシャルを追加し、その内部でのタンパク質の振舞いを研究した。同様の研究結果はいくつか発表されており、Mittal and Best 2008 [18] による、やはり類似の束縛ポテンシャルによるフォールディング確率への影響を論じた論文がある。これらの理論解析で用いられた束縛ポテンシャルは、細胞内の分子混雑による空間的束縛ともとれるので、本稿でも同様のポテンシャルを後に導入し、解析を行う。

## 1.5 本稿では何を論じるか

本稿の主題とは次である：

分子混雑はタンパク質のフォールディング過程に対していかなる影響を与えているのか、また、有益に働く場合があるのかどうか、研究を行う。高精度シミュレーションによるフォールディングシミュレーションは、準安定状態への滞留により成功しない場合があることがわかってきたが、現行の高精度シミュレーションが無視している混雑効果が実はこの滞留の解決のために重要な影響を与えているのではないかという仮説をここに立て、本研究を通じ検証した結果を示す。

研究目標の遂行のため、次の方法をとる：

タンパク質フォールディングを模擬する抽象化したヘテロポリマー鎖モデルを設定し、定性的な振舞いを解析する目的で研究を行う。具体的には、モデル鎖を分子動力学法により動かし、統計力学分野の各種手法によりアンサンブル平均や FEL を調べ、また、それらの物理量に表れるマクロな性質が混雑効果によりどのように変化するか研究を行う。扱った主な手法は、分子動力学シミュレーション (MD)、レプリカ交換法 (REM)、アンブレラサンプリング法などである。

本研究における新規性は以下である：

最安定構造を *a priori* に決めないヘテロポリマー鎖モデルを使い、同じ立体構造へと状態が収束するフォールディング過程が混雑効果により誘導されているという可能性について論じたこと。天然のタンパク質が生体内で同じ構造へ再現性よく落ち着くという現象は、高精度シミュレーションでは再現できていない。高精度シミュレーションが考慮していない分子混雑が、この天然タンパク質の特異な物性の原因ではないかという検証を、粗視化モデルを用いて定性的な研究により行った。分子混雑によりタンパク質のフォールディング転移温度が上昇するという現象は、粗視化モデルの研究により行われてきたが、これらの先行研究では最安定構造を *a priori* に決めているために別の準安定構造との競合については議論できずにいた。モデル鎖のポテンシャルの設定に際し、最安定構造を決めないという立場をとったことで、高精度シミュレーションから報告されている、準安定構造への滞留の問題を再現し、滞留の解消に混雑効果が効いているという議論と検証を行った点は、タンパク質研究において新規な点である。

次章より本論に入る。

## 2 モデルの設定と方法

### 2.1 タンパク質の HP モデル

生体内のタンパク質は、平均して 300 アミノ酸残基ほどで構成されており、残基間の相互作用により立体的に折畳まれ、機能を発現する。この相互作用を抽象化し、単純化してタンパク質フォールディングを再現する粗視化モデルの一つが、HP モデルである。本研究では HP モデルを用いた解析を行っているが、まずは HP モデル全般について概説を行いたい。

「HP モデル」の「HP」とは「Hydrophobic-Polar」の略であり、20 種類のアミノ酸残基を親水性残基と疎水性残基の 2 種類に大胆に分け、配列を 1、0 で記述するモデルである。残基間相互作用は、考えている残基対の組み合わせにより決定する。このモデルと実際のタンパク質との対応がどのようにつくのかという点であるが、HP モデルの一つの残基が実際のタンパク質のアミノ酸数個と対応するものと考えれば良い。天然のタンパク質配列中では、親水性残基、疎水性残基はそれぞれ連続する場合が多く、ゆえにタンパク質鎖の中には親水性の領域と疎水性の領域が存在している。水溶液中でタンパク質が折畳まれる際には親水-疎水相互作用の影響を受け、疎水性の領域を内側に、親水性の領域を外側に配置させる形で自発的にフォールディングする。これら親水基・疎水基が連続した領域を HP モデル中の一つの残基として抽象化するのである。タンパク質のフォールディングには階層性があり、小さいスケールでは鎖中の局所部位における水素結合による二次構造の形成から、大きなスケールでは立体構造全体を大まかに決める親水-疎水相互作用による立体配置の決定までである。HP モデルは大きなスケールのフォールディングに注目し、二次構造などの小さなスケールのフォールディングに関しては記述を行わない。

タンパク質の HP モデルにも様々な形が提案され、研究されている。主な分類として、格子型か非格子型かという点が重要であろう。

■**格子モデル** 格子モデルは、残基の位置を 3 次元格子上に配置させ、鎖を self-avoiding walk させるように動かすモデルである。隣接格子に関してエネルギーの計算を行い、格子上の配置から系全体のエネルギーを計算し、モンテカルロ法により振舞いを解析する。タンパク質のエネルギー最小化の問題や、自由エネルギー地形についての一般の性質を議論するために使われるモデルであり、計算量は軽い。この模型では、状態数が限られるためにその振舞いを徹底的に調べ上げることが可能であり、タンパク質フォールディングに関する描像が曖昧であった時代に、理論的な予想を立てる場合に頻繁に使用された。ただ、ダイナミクスを記述する必要がある場合には弱いモデルであると言える。

■**非格子モデル** 残基を空間中に自由に配置するモデルで、相互作用は残基間の距離により決まる。分子動力学法により時間発展が記述されるが、格子モデルと比べて自由度が大きいため計算量は大きくなる。速度や力に関する振舞いを調べる場合には有効なモデルである。

一般に、これらの HP モデルの振舞いは、立体構造に関しガラス的な振舞いを示すものとなる。すなわち、高温域では立体構造の揺らぎが大きい無秩序な状態となり、構造は定まらないが、低温域では揺らぎが小さくなり、固まった構造をとる。しかし、どの構造をとって動かなくなるかは試行により異なっており、これは低温域においては FEL 中に多数の準安定状態が存在しているためで、このとき系は立体構造についてガラス状態にあるものと言える。本来は「ガラス」という言葉は有限な系に対し使われるべきではないが、無秩序ではなく、遅い緩和を起こす状態を表現する際に便宜上本稿では使用することにする。

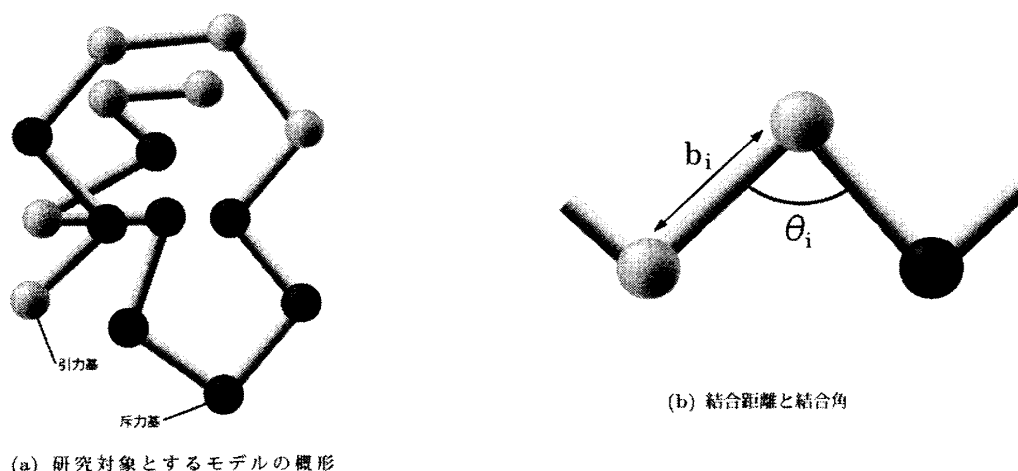


fig.2 : モデルの模式図

## 2.2 本研究で対象とするモデルの設定

本研究においても、タンパク質を抽象化した物理モデルである HP モデルを設定するが、モデル鎖への衝突が重要になるであろうという予想から、非格子型の HP モデルを採用し、分子動力学法により解析を行うことにする。実際のタンパク質が持つ性質を模擬する必要から、次の要請をする：

1. フォールディング転移を起こし、その様子が認識できる
2. 多くの準安定状態を持つ
3. 自由度が少なく計算量が少ないモデルである

1、2 番目に掲げた要請は、HP モデルを設定するとほぼ自動的に満足される。3 番目に掲げた要請の理由は次の通りである：自由度が高い、ゆえに計算量が膨大となるモデルである必要はなく、複雑な自由エネルギー地形を持ち、フォールディングするという条件を満たすモデルであれば、なるべく単純なものがよい。また、計算量が大きいモデルは現在筆者が利用できる計算機環境から考えても適切ではない。

設定するモデル鎖のポテンシャル  $V$  を次のように設定する：

$$\begin{aligned}
 V = & \sum_{bonds} K_r (b_i - b_0)^2 + \sum_{angles} K_\theta (\theta_i - \theta_0)^2 \\
 & + \sum_{i,j} \left( \epsilon_1 \left( \frac{C}{r_{i,j}} \right)^6 + 2\epsilon_2 \left( \frac{C}{r_{i,j}} \right)^{12} \right)
 \end{aligned} \tag{1}$$

この関数により表されるモデル鎖の模式図を fig.2 に示したので、この図を参考にしながら (1) 式の内容を説明することにする。 $i, j$  は残基あるいはその間にある結合などの番号を表し、残基には引力基と斥力基の 2 種類があると考え、残基の番号によりどちらの種類であるかが決定されている (2.6 節でこの配列の設定を行う)。(1) 式右辺第 1 項は隣接残基間の結合距離、すなわち fig.2(b) における  $b_i$  に関するポテンシャルであり、最適値を  $b_0$  ( $= 4.0$ ) としている。 $K_r = 100.0$  である。第 2 項は、fig.2(b) の  $\theta_i$  で表される、隣接残基間の結合角に関するポテンシャルであり、最適値を  $\theta_0$  ( $= 2.09$ ) としている。 $K_\theta = 20.0$  である。第 3 項が

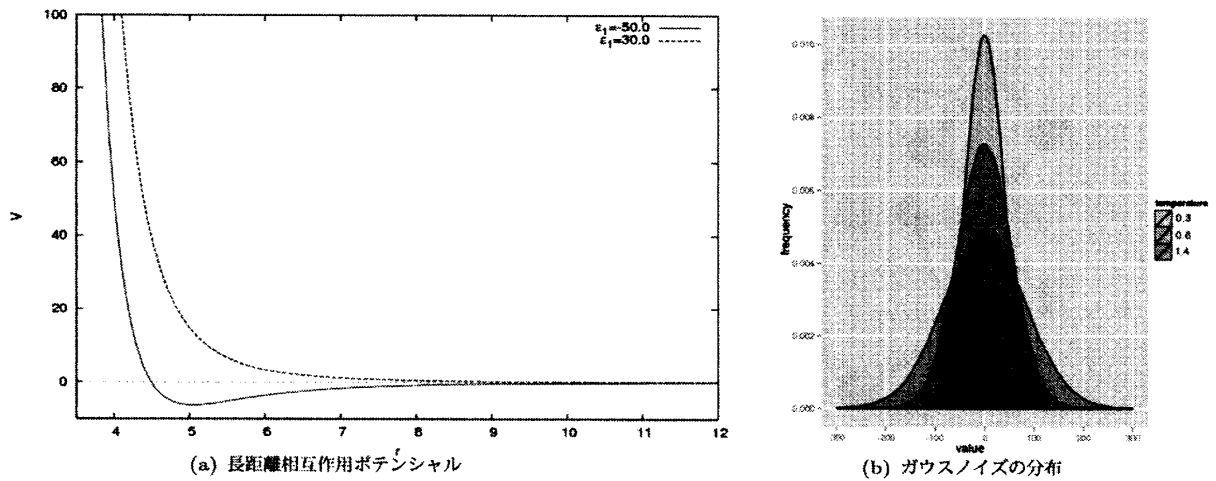


fig.3 : 長距離相互作用ポテンシャルとガウスノイズ

長距離残基間相互作用に関するポテンシャルであり、考える残基対により  $\epsilon_1$  が異なっている。残基同士の距離  $r_{i,j}$  に依存する関数であり、式中のパラメータ  $C$  は 4.0 である。 $i$  番目の残基と  $j$  番目の残基の対に注目したとき、それぞれが引力基であるか、それとも斥力基であるかでポテンシャルが異なる。双方が引力基である、もしくは引力基と斥力基の組み合わせである場合、 $\epsilon_1 = -50.0$  となり、fig.3(a) の赤線で示したポテンシャルを持つことになり、互いに引き合う。双方が斥力基である場合、fig.3(a) の青線で示したポテンシャルを持つことになり、互いに反発する。また、このモデルは理想鎖であり、上記ポテンシャルが許す限り残基の質点は自由に動くものとしている。

そして、次のランジュヴァン方程式に従い時間発展するとした。

$$\begin{aligned} m\dot{\mathbf{v}}_i &= \mathbf{f}_i - \gamma\mathbf{v}_i + \xi_i \\ \langle \xi_i(t)\xi_j(t') \rangle &= 2\gamma k_B T \delta_{i,j} \delta(t-t') \end{aligned} \quad (2)$$

$m$  は残基一つの質量であり、 $m = 10.0$  である。 $\mathbf{v}_i$  が残基  $i$  の速度であり、 $\mathbf{f}_i$  が残基  $i$  に働く力を表す。 $\gamma$  は粘性係数であり、 $\gamma = 0.25$  である。簡単のため、ボルツマン定数  $k_B$  は 1.0 とした。 $\delta_{i,j}$  および  $\delta(t-t')$  はディラックのデルタ関数であり、 $\xi(t)$  で表されるノイズに残基間および時間に関して相関がないことを示している。このノイズは白色ガウスノイズであり、温度  $T$  により fig.3(b) のように分布型が変化する。

このポテンシャルの形、およびランジュヴァン方程式中のパラメータは Nakamura 2004 [19] に従う。先に設定したポテンシャル形は、Nakamura 2004 において設定されているポテンシャル形を、二面角と残基間相互作用のポテンシャルの部分に関して改変したものである。Nakamura 2004 では残基数 50 程度のタンパク質ドメインに対する詳細な近似を行う非格子型郷モデルを採用しているので、上で設定したモデルとは考えている状況が異なるが、本研究では混雑効果に関する定性的な議論を行いたいという理由から、実際のタンパク質が持つ各種ポテンシャル、残基の質量等それぞれのパラメータの相対比に関してはモデルとの対応を無視している。設定したパラメータでランジュヴァン方程式による MD シミュレーションを行うと、残基間の距離がほぼ等距離となり、残基間の引力-斥力相互作用により結合角が若干揺らぐという振舞いが観察されるので、上記モデルで考慮するいずれかの力が異常に強いということはないと判断し、期待する運動は行われていると考え、(1) 式で示したモデルの形を採用した。

## 2.3 velocity Verlet 法

MD シミュレーションのアルゴリズムには、計算の安定性や精度の面で優れているとされる velocity Verlet 法を採用した。本節ではこの velocity Verlet 法のアルゴリズムの導出を確認しておく。

$\mathbf{r}_i$  を  $i$  番目の残基の座標として、モデル鎖の運動は次のニュートン運動方程式により規定される。

$$m \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{f}_i \quad (3)$$

ポテンシャルは既に前節で与えた。 $\mathbf{f}_i$  はポテンシャルの座標微分となり、以下で与えられる：

$$\mathbf{f}_i = -\nabla V(\mathbf{r}_i) \quad (4)$$

座標を  $\mathbf{r}_i(t)$  で表し、微小時間  $h$  進んだ時点での座標を  $\mathbf{r}_i(t+h)$  で表すと、テイラー展開から次の代数式が成立する：

$$\mathbf{r}_i(t+h) = \mathbf{r}_i(t) + h \frac{d\mathbf{r}_i(t)}{dt} + \frac{1}{2!} h^2 \frac{d^2 \mathbf{r}_i}{dt^2} + \frac{1}{3!} h^3 \frac{d^3 \mathbf{r}_i(t)}{dt^3} + \dots \quad (5)$$

ここで、

$$\frac{d\mathbf{r}_i(t)}{dt} = \mathbf{v}_i(t) \quad (6)$$

であるから、(5) 式の 3 階微分以上を無視すると次式が得られる。

$$\mathbf{r}_i(t+h) = \mathbf{r}_i(t) + h \mathbf{v}_i(t) + \frac{h^2}{2m} \mathbf{f}_i(t) \quad (7)$$

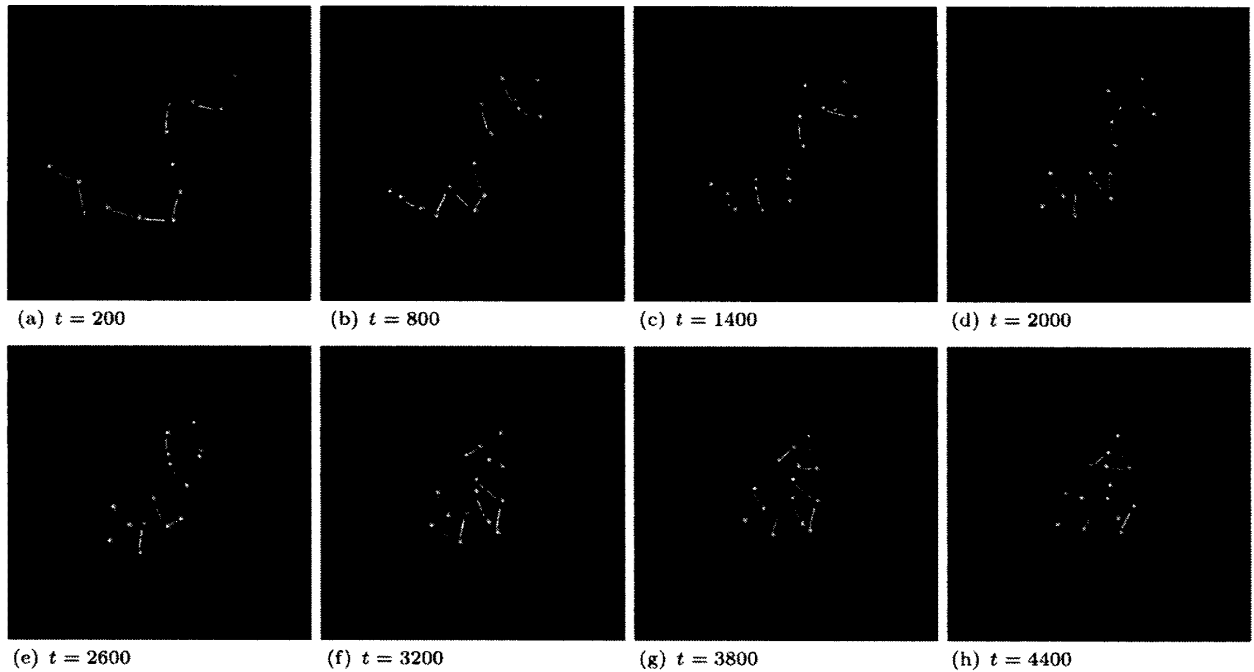
速度の項に関して新たに方程式が必要となる。速度を一階微分すると加速度となる。精度を高めるため次のように設定する。

$$\mathbf{v}_i(t+h) = \mathbf{v}_i(t) + \frac{h}{2m} (\mathbf{f}_i(t) + \mathbf{f}_i(t+h)) \quad (8)$$

式 (7)、(8) から、分子の運動を追跡するアルゴリズム (velocity Verlet 法) が得られた。時間ステップ  $h$  は 0.01 としてシミュレーションを行う。

また、初期条件を以下の手順の通り設定する：

1. 原点座標 (0,0,0) に一端の残基  $R_0$  を置く
2. 中心を (0,0,0) にとり、原点残基間結合距離の最適値  $b_0$  を半径とする球の球面状に  $R_0$  の隣の残基  $R_1$  をランダムに配置する
3.  $R_n (n \geq 2)$  に関しては、 $R_{n-1}$ 、 $R_n$ 、 $R_{n+1}$  が成す角度  $\theta_n$  が結合角の最適値  $\theta_0 \pm \delta_{\theta(n)}$  となるように、また、残基間結合距離がやはり  $b_0$  となるように、更に二面角  $\phi$  をランダムに配置し、設定する ( $\delta_{\theta(n)}$  は大きさ 1 の白色ガウスノイズ)
4. マクスウェルの速度分布に従うよう各残基の速度を乱数により決定する



**fig.4** : モデル鎖の 3-D 描画 引力基を赤色、斥力基を青色のボールで示している。副題は、??節で示す通り、あるランダムな初期状態から MD シミュレーションにより発展させた時間数を示す。

## 2.4 グラフィックスツールの作成

本研究は、タンパク質のフォールディングを研究対象としている。モデル鎖の形がシミュレーション中に 3 次元座標中でどのような状態にあるかを簡便に知る手段が求められる。この理由により、モデル鎖の残基の座標が情報として得られたとき、そのモデル鎖の形を 3-D グラフィックスで簡単に映し出す可視化ソフトウェアを作成することにした。簡便に 3-D 画像を作成できる環境を提供する API である OpenGL を使い、目的の可視化ソフトウェアを作成した。モデル鎖中の 15 個の残基座標データをファイル入力すると、引力基は赤色のボールで、斥力基は青色のボールで表現され、残基を表すボールの間は円柱状のオブジェクトで繋いでいる。出力例を fig.4 に示した。後の章においてもこのグラフィックスツールを用いて作成した画像を使用する。

## 2.5 Shape Parameter の設定

モデル中の質点座標は 15 個という小さな数ではあるが、それでも自由度は高い。しかし本研究では立体構造に注目したいので、座標に関する自由度を潰し、なおかつどのような構造にあるのか判定できるパラメータを設定しておきたい。これらのパラメータを以降 Shape Parameter と呼称することにする。本研究では、次の数を Shape Parameter として設定した：

- ・  $R_g^2$  : 残基の座標の分散

$$R_g^2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_0)^2$$

- ・  $\Sigma\theta$ : 残基間の結合における結合角の和

$$\Sigma\theta = \sum_{i=1}^{N-1} \theta_i$$

- ・  $P_g^2$ : 斥力基の座標の分散

$$P_g^2 = \frac{1}{N_P} \sum_{\text{HP}[i]=1} (\mathbf{r}_i - \mathbf{r}_P)^2$$

$R_g^2$  は一般に慣性半径と呼ばれる量で、分子鎖の拡がり具合を表す。これに加え、 $\Sigma\theta$ 、 $P_g^2$  というパラメータをとることで、シミュレーション中に現れるほとんどの立体構造を区別できることになる。 $P_g^2$  の定義の中で表れる  $\sum$  は「斥力基の和」を意味しており、また、 $\mathbf{r}_P$  は斥力基の座標の平均を意味する。これらの Shape Parameter による 3 軸の空間をとれば強力に構造を分類できるが、本稿では特別必要がない限り、 $R_g^2$  と  $\Sigma\theta$  を用いるのみとし、 $P_g^2$  はほとんど解析に用いていない。Shape Parameter が二つあれば、ほとんどの場合は構造を区別できるためである。

## 2.6 配列の決定

本研究が対象とする HP モデルはヘテロポリマーであるが、これは残基の空間中の配置、すなわち立体構造によりポテンシャルが異なり、この事情は実際のタンパク質が持つ複雑な自由エネルギー地形を模擬する。要請として、フォールディング転移温度において、MD シミュレーション中に状態間遷移が観察できるモデルであるべきだとする。自由エネルギー地形の形状を特徴づけているパラメータは HP モデルの引力基と斥力基がどのような配列で並ぶかという配列情報であるので、上記の要請を満たすような配列を探すことにする。本節では、この配列の探索作業について説明する。

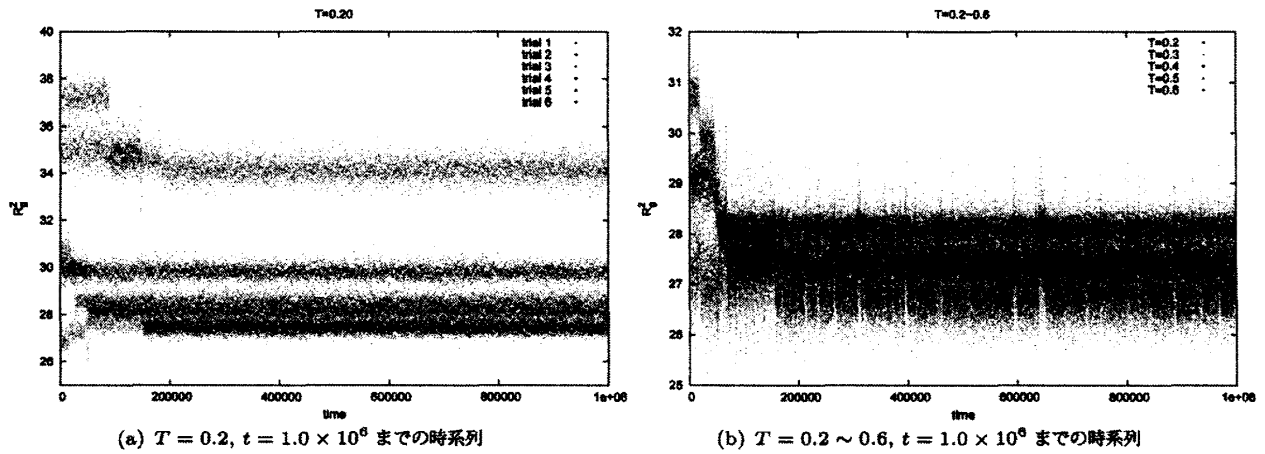
残基数  $N=15$  として配列をランダムに設定し、MD 計算を行い、その結果として出力される残基座標の分散  $R_g^2$  を時系列プロットすることで多数の準安定構造を持つと期待される配列を網羅的に探すことにした。以下、ヘテロポリマーモデルの配列を、斥力基を 1、引力基を 0 として、例えば  $\{1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1\}$  などと表すことにする。まず、いま示した配列をサンプルとして、2.4 節で紹介したグラフィックスツールを用いて温度を変化させながら、フォールディング転移が起きていると思われる振舞いを示す温度を網羅的に探した。その結果、転移温度は  $T = 0.6$  付近であると考えられた。根拠は次の通りである：

- ・  $T > 0.7$  の領域では、 $R_g^2$  の時系列は無秩序になり、アンフォールド状態と考えられる揺らぎの大きい振舞いが観察され、グラフィックスツールでも同様の振舞いが観察される。
- ・  $T < 0.5$  の領域では、 $R_g^2$  の時系列は fig.5(a) のようになり、揺らぎの小さい振舞いが常に見られ、また、その値は同じ温度でも試行により異なる。これは、準安定状態が多数生じており、どの状態に落ち着くかにより  $R_g^2$  がとる値が異なるためであると考えられる。
- ・  $T = 0.6$  における  $R_g^2$  の時系列を見ると fig.5(b) 中の赤色のプロットで示したような振舞いが観察され、様々な状態間を行き来していると考えるべき振舞いが観察される。

先のサンプル配列の転移点は  $T = 0.6$  付近であると考え、また同時に、他の配列においても転移温度は  $T = 0.6$  付近であると考えた。

疑似乱数により  $N = 15$  の HP モデルの配列を多数得て、 $T = 0.6$  において MD 計算をし、 $R_g^2$  のヒストグラムを描いたとき、その形が最も複雑なものを研究対象として選ぶことにした。

疑似乱数から次の 8 つの配列を得た。それぞれの配列を Seq-1 などと名付けることにする：

fig.5 :  $T = 0.2 \sim 0.6$  での  $R_g^2$  の時系列

$$\begin{aligned}
 \text{Seq-1} &= \{1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0\}, & \text{Seq-2} &= \{1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0\} \\
 \text{Seq-3} &= \{1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0\}, & \text{Seq-4} &= \{0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1\} \\
 \text{Seq-5} &= \{1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1\}, & \text{Seq-6} &= \{0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0\} \\
 \text{Seq-7} &= \{1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0\}, & \text{Seq-8} &= \{0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0\}
 \end{aligned} \tag{9}$$

この中から今後の解析に用いる配列を決める。MD による  $2.5 \times 10^9$  ステップ ( $t = 0 \sim 2.5 \times 10^7$ ) のテストシミュレーションを行い、ヒストグラムを作成する。温度  $T = 0.5$  としてテストシミュレーションを行い、なるべく多数の状態が結果の出力となる時系列中に現れるものを選び、また、その状態をヒストグラムの形から認識しやすいものを選ぶことにする。注目する変数には  $R_g^2$  を選んだ。

結果は fig.6~7 に示したようになった。fig.6~7 左側の図では  $R_g^2$  の時系列頻度を表すヒストグラム、右側の図では  $R_g^2$  の時系列を示した。右側の時系列プロットを見ると、構造転移がほとんど起きていないもの (Seq-1,2,4,6) と頻繁に起きるもの (Seq-3,5,7,8) があることがわかる。前者については、 $T = 0.5$  という温度でも、一つの安定構造に留まりやすい配列であることから、タンパク質の複雑な自由エネルギー構造を考える上ではモデルとしては適当でないと判断し、後者の中から研究対象とする配列を選ぶことにする。Seq-3、Seq-5、Seq-7 のヒストグラムを見ると、構造転移は起きているけれどもヒストグラムを描くとその様子が見られず、凹凸の様子が潰れてしまっている場合がある。特に Seq-7 で顕著である。時系列よりもヒストグラムから準安定構造の出現の様子を観察したいという事情から、構造転移が示される場合の時系列から得たヒストグラムは凹凸がなるべく激しい方が都合がよい。その要請を満たす配列は、今の 8 個の配列の候補の中では Seq-8 であるから、この Seq-8 を採用し、研究を進めることにする。

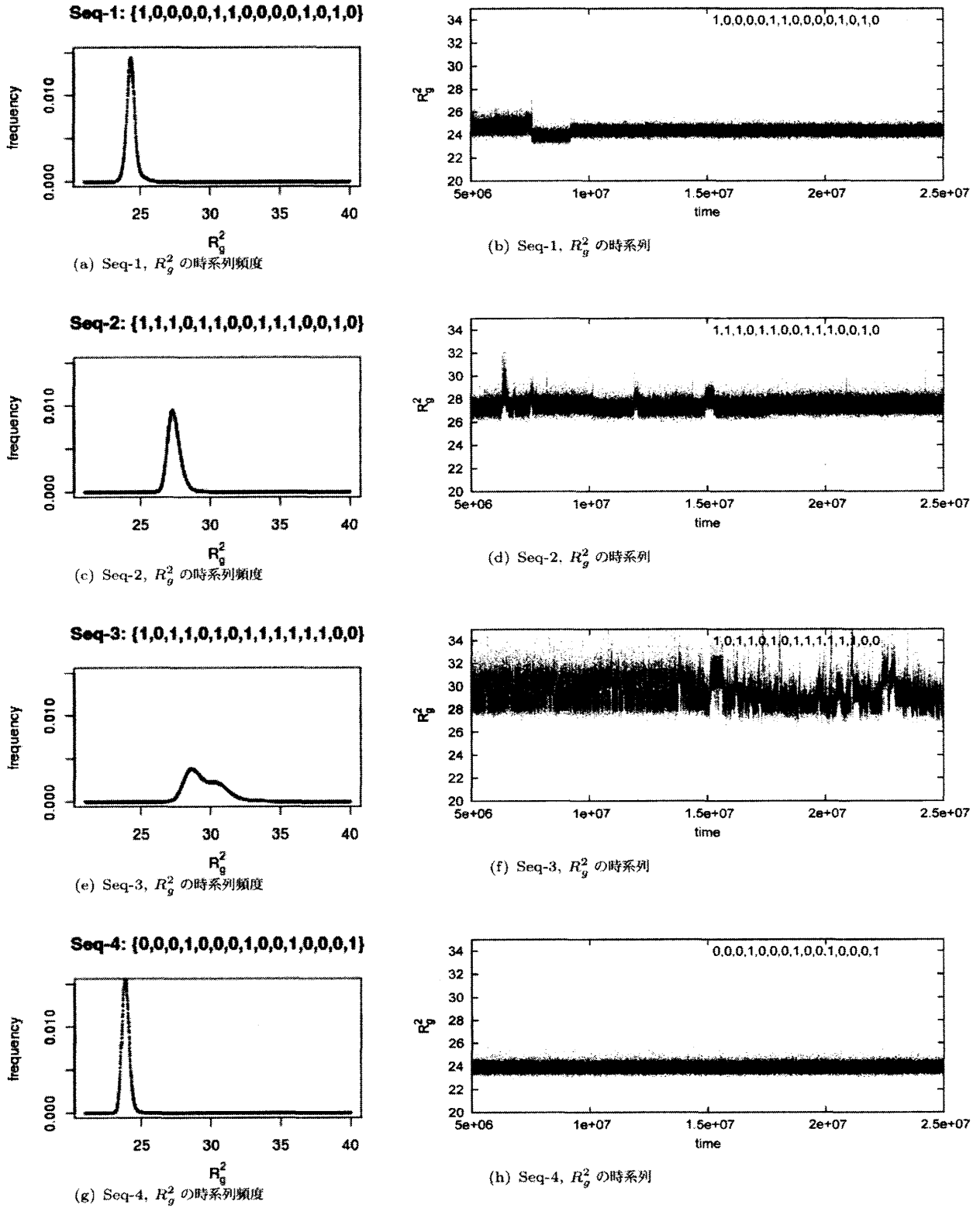


fig.6 : サンプルとして選んだ配列のテストシミュレーション (I)

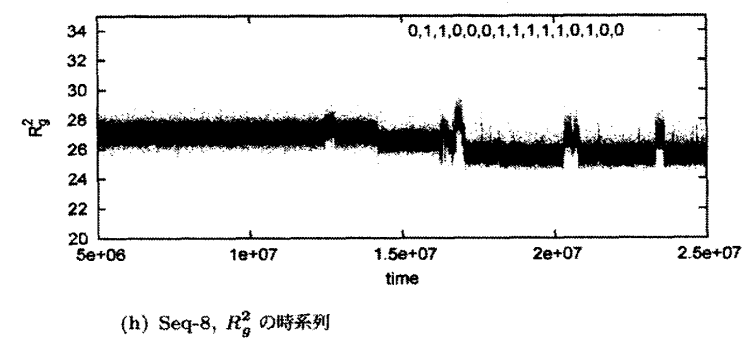
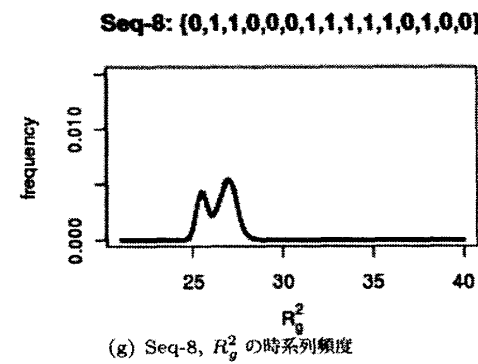
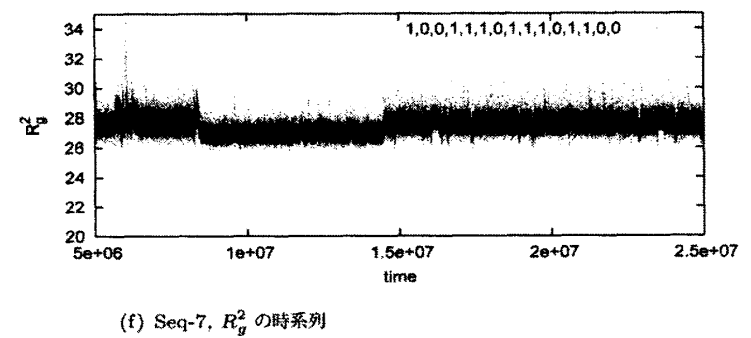
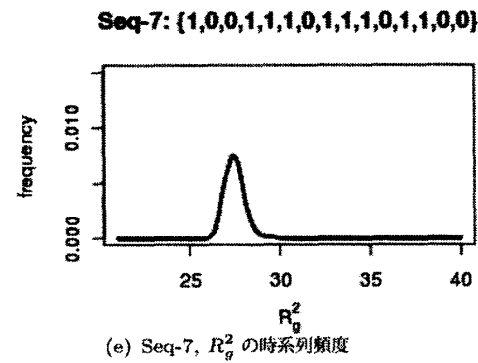
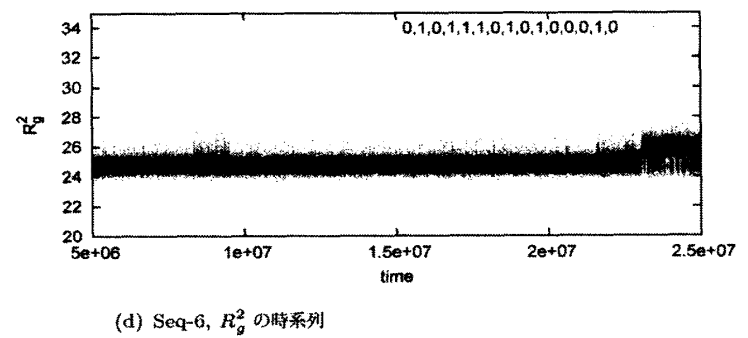
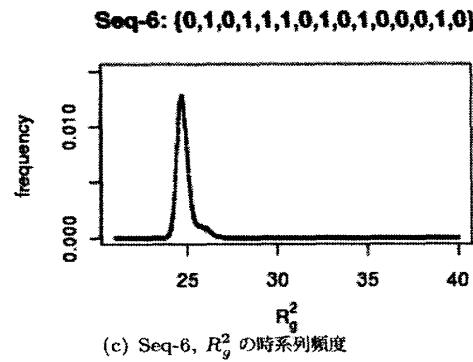
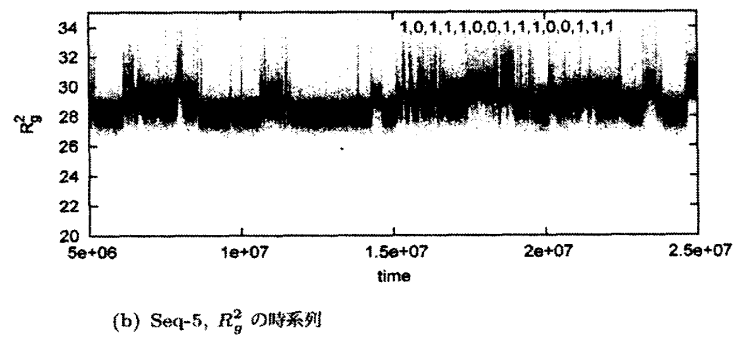
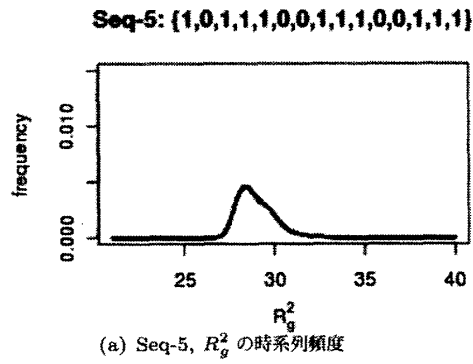


fig.7: サンプルとして選んだ配列のテストシミュレーション (II)

### 3 通常の分子動力学計算によるシミュレーション

#### 3.1 MD 計算の結果

$T = 0.1 \sim 1.4$  の温度領域で MD 計算を行い、それぞれの時系列データから時間平均を算出する試みを行った。これについては、成功した点と失敗した点があるが、双方ともに本稿にとっては有益であるため報告する。

次の変数の時系列平均を温度ごとにプロットしたい：

- ・  $R_g^2$  の時系列平均
- ・  $\Sigma\theta$  の時系列平均
- ・ 分子鎖のエネルギー  $E$  の時系列平均

先行研究 [19] を踏まえて考えると、モデル鎖はある温度  $T_f$  でフォールディング転移が起きると考えられるが、このフォールディング転移温度  $T_f$  は混雑効果により変化することが予想され、また、同じく混雑効果により各パラメータ時系列平均の温度プロットの概形も変化すると予想される。すなわち、本研究にとってこのプロットは非常に重要なデータとなる。

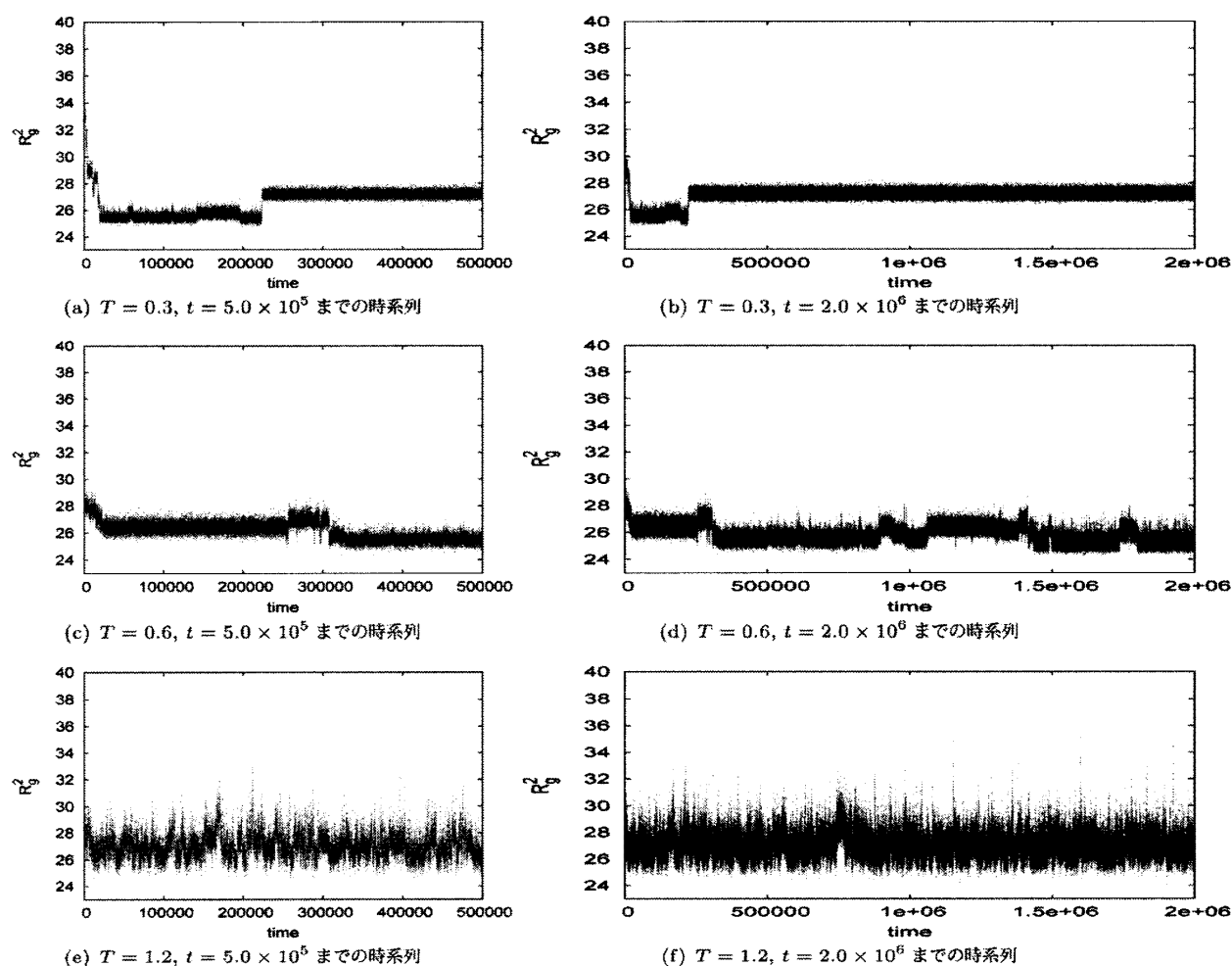


fig.8 : 各温度における初期緩和の様子

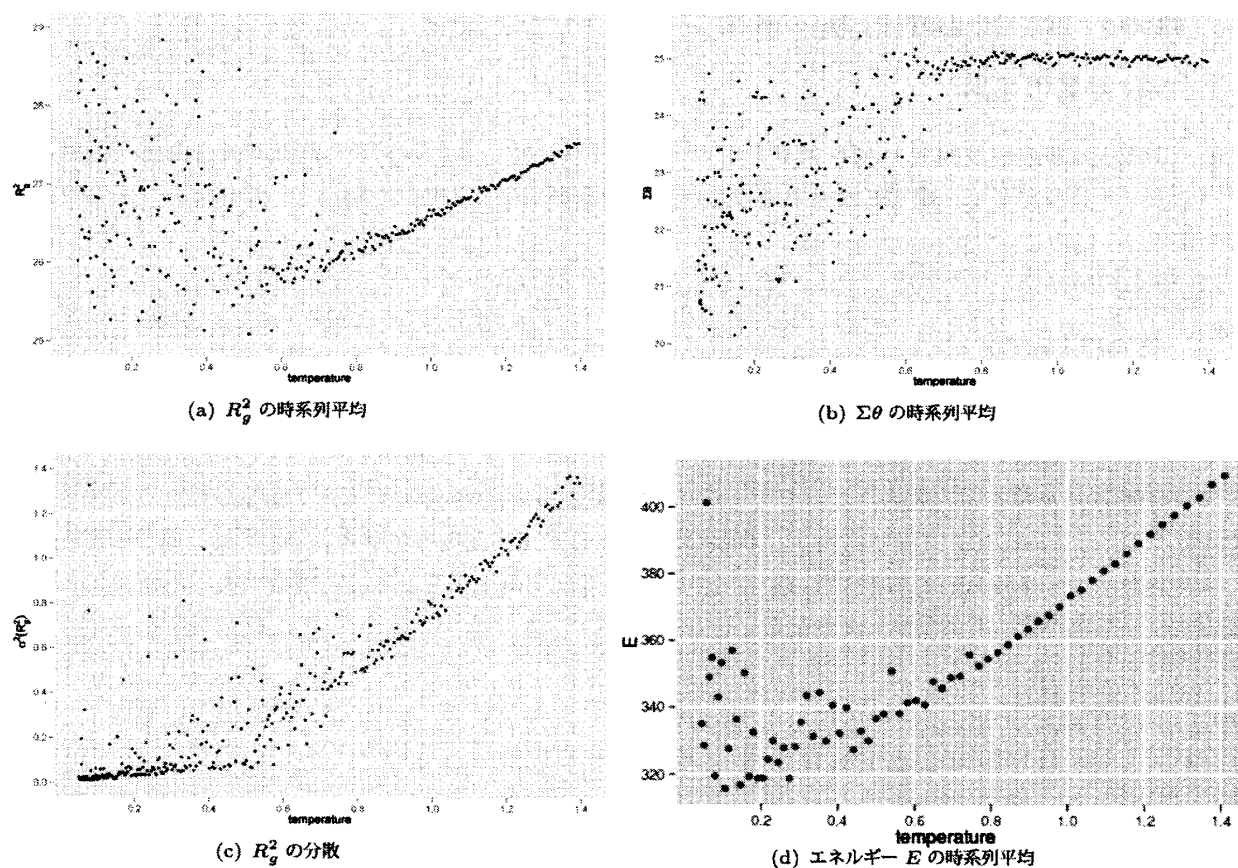


fig.9 : MD シミュレーションにより得た各物理量の時系列平均

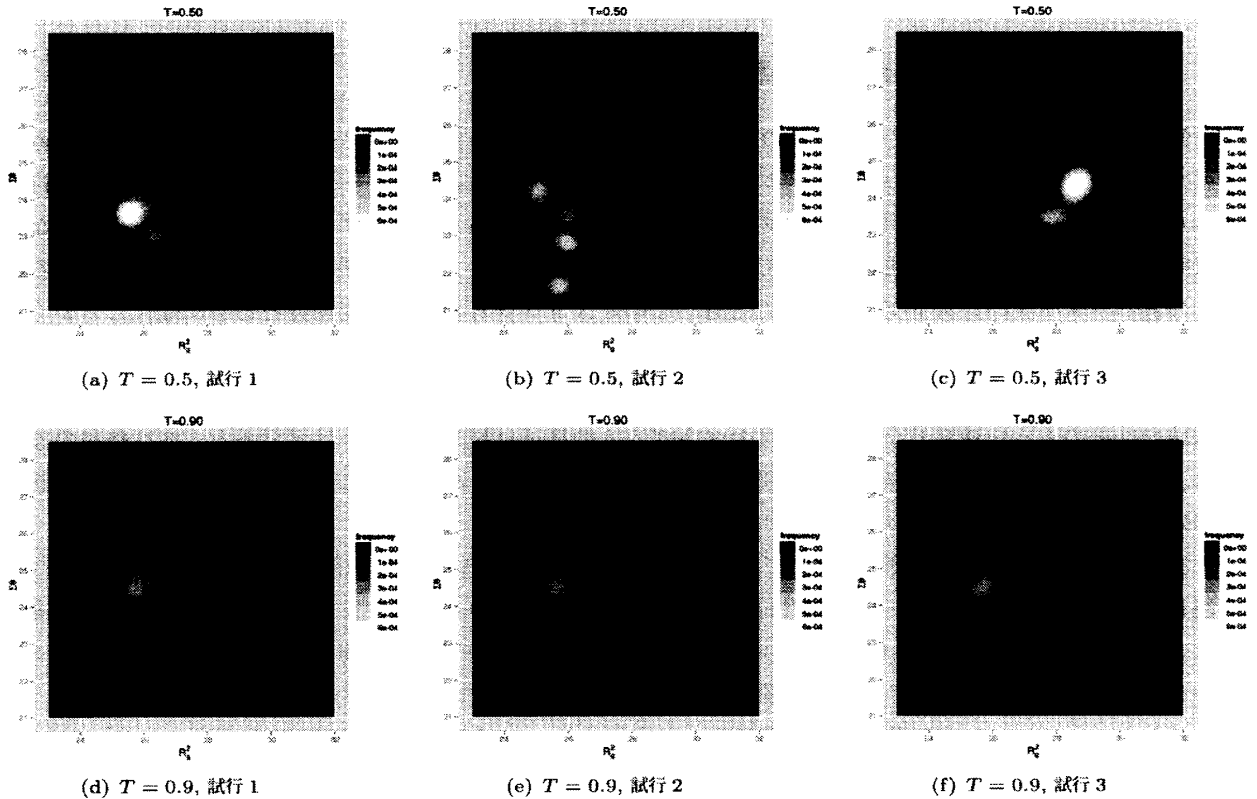
前章にて示した方法に従って、各温度において  $t = 0 \sim 5.0 \times 10^6$ 、時間ステップ数にして  $5.0 \times 10^8$  ステップの MD シミュレーションを行った。ランダムな状態からシミュレーションを始めると設定しているため、時系列の初期には初期緩和が起こるが、この初期緩和の状態にある時系列データは平均の算出根拠からは外し、その後の時系列データを使用する。それぞれのパラメータの初期緩和の様子を fig.8 に示した。fig.8 には温度の代表点として  $T = 0.3, 0.6, 1.2$  をとり、その時系列を示している。この様子を見ると、それぞれの温度で  $t = 1.0 \times 10^6$  程度では既に十分に緩和は起きていると考えられる。 $t = 0 \sim 1.0 \times 10^6$  までは切り捨て、 $t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  を時系列データとして採用し、目的のパラメータの算出に用いた。そして、その結果を示した図が fig.9 である。結果として、カノニカル平均を出すという目的で行ったこのシミュレーションは失敗している。低温域において系はガラス状態となるためである。

fig.9(a) には  $R_g^2$  の時系列平均、fig.9(b) には  $\Sigma\theta$  の温度プロットを示した。それぞれを見ると、高温側ではプロットが収束しているが、低温側では平均値にバラつきが目立っており、今回の MD 計算では有効なサンプリングは全く行われていないと考えてよい。プロットにバラつきが目立ち始める温度は  $T = 0.8 \sim 1.0$  であると考えられる。この温度を、注目する系のガラス転移点  $T_g$  と考えた。

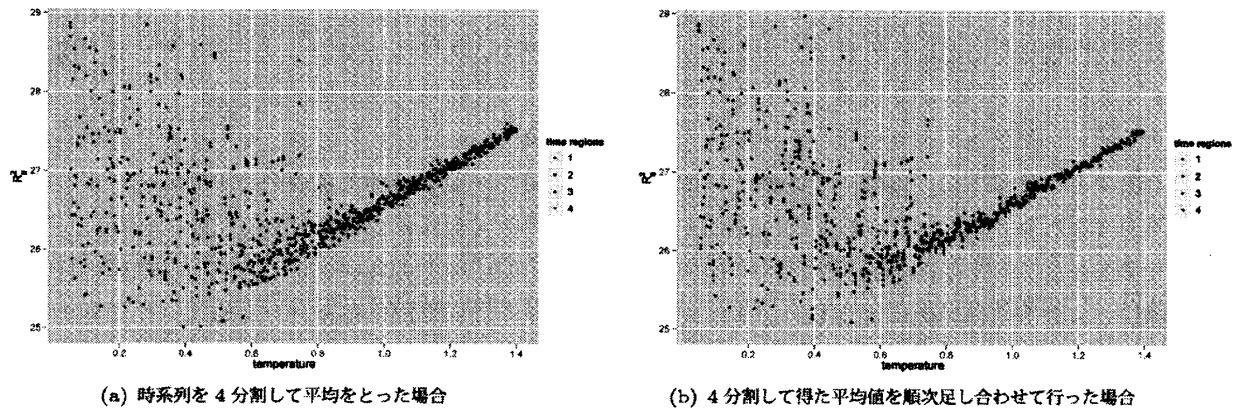
時系列中で  $R_g^2$ 、 $\Sigma\theta$  それぞれの変数がどのような変化を起こしているのかを視覚化するため、得られた時系列データから二次元ヒストグラムを描き、それを示すことにする。fig.10 がそのヒストグラムである。fig.10(a)、(b)、(c) はそれぞれ  $T = 0.5$  という同じ温度で MD 計算を行った結果得た時系列データから作成したヒストグラムであるが、これらの図を見てわかるように、試行ごとに落ち着く状態が異なっている。fig.10(d)、(e)、(f) はそれぞれ  $T = 0.9$  という温度で独立に行った 3 回の MD 計算の結果から得たヒストグラムであるが、この温度ではヒストグラムの様子は違いはほぼ見られない。低温側で時系列平均のプロット

が全く行えない理由は、転移温度以下では多数の準安定状態が生じ、MD 計算を行うとその局所安定状態へ落ち込み、抜け出せなくなるためであると考えられる（遅い緩和）。モデルの安定なフォールディング構造は多数存在するが、フォールディングしてしまうとポリマー内の残基間相互作用により安定化してしまい、別のフォールディング構造へ移る確率は極めて低くなる。逆に構造の解けたアンフォールド状態では様々な状態を行き来することが出来、結果、サンプリングが実行した MD 計算の時間の範囲内で十分に完了し、試行ごとに同様のヒストグラムが得られ、その結果が fig.9 のプロットの収束に現れている。

本研究を進めるにあたり、系がガラス相に陥ることでカノニカル平均がわからなくなることは厄介である。ガラス転移点よりも低温にあるとされるフォールディング転移点付近での系の振舞いを知ることができなくなるためである。ガラス状態の緩和に混雑効果は影響を与えるかもしれないが（これについては本稿後半で考察する）、まずは混雑効果なしの条件下でのモデル鎖のフォールディング転移温度を調べたいと考えており、今の状態だとその転移温度を知ることができない。時系列データを長くとることによってガラス状態は緩和されるだろうか。時系列中で緩和が徐々にでも起こっているならば、単純に時系列を長くすればサンプリングは可能である。 $t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  まである時系列データを 4 分割し、それぞれの時間領域で時系列平均を算出し、先と同様のプロットを得た。それが fig.11 である。時系列の後ろの方で緩和が起きているのならば、時間領域がより後になるにつれプロットに収束が起こるはずだが、fig.11 からはそうは読み取れない（fig.11(a)）。また、それぞれの時間領域の時系列平均を次々足しては平均をとり、その様子をプロットしてみても、高温域のアンフォールド相の収束が見られるだけで、 $T < T_g$  の低温域においては平均の収束



**fig.10** :  $T = 0.5$ ,  $T = 0.9$  における  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラム  $t = 5.0 \times 10^6$  までの MD シミュレーションを行って得た時系列から、 $t > 1.0 \times 10^6$  の時系列について  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラムを得た。Shape Parameter である  $R_g^2$  を横軸に、 $\Sigma\theta$  を縦軸にとっている。平面上の配色が時系列頻度を表す。上段の (a), (b), (c) が  $T = 0.5$  として 3 回の MD シミュレーションを行って得たヒストグラムであり、下段 (d), (e), (f) が  $T = 0.9$  の場合である。 $T = 0.5$  の場合、各試行により現れるパターンが異なっているが、一方で  $T = 0.9$  の場合は一様である。



**fig.11** : 時系列を分けて平均をプロットした場合  $t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  までの時系列を  $1.0 \times 10^6$  ごとに4分割し、それぞれの時間領域ごとに  $R_g^2$  の時系列平均を得た場合 (a) と、それぞれの時間領域で得た平均値を順次足してプロットした場合 (b) の図を示す。点の色は time regions として分けているが、(a) では  $t = 1.0 \times 10^6 \sim 2.0 \times 10^6$  の time region を1,  $t = 2.0 \times 10^6 \sim 3.0 \times 10^6$  の time region を2などとして配色しており、(b) では  $t = 1.0 \times 10^6 \sim 2.0 \times 10^6$  の time region を1,  $t = 1.0 \times 10^6 \sim 3.0 \times 10^6$  の time region を2などとして配色している。時系列が後になれば、あるいは字形列を長くすれば低温域で平均値の収束が見られるということはない。

は全く見られず、分散したままである (fig.11(b))。現在の計算機資源から計算可能な時間スケールでは低温域のカノニカル平均を求めることができず、議論ができない。効率的なサンプリングを実行するための手法がここで必要となった。次章から、使用するアルゴリズムを示すが、その前に、通常の MD 計算で観察される初期緩和について次の節でまとめておきたい。

### 3.2 ランダムな初期状態からの緩和

予備的なデータではあるが、モデル鎖の Shape Parameter、そしてポテンシャル  $V$  が、初期のランダムな状態からどのように緩和していくのかについて調べることにした。各パラメータは、時間発展とともに自由エネルギー地形中で安定な状態となる値の周辺に落ち着く確率が高くなるはずである。別々のランダムな初期状態から始まる、それぞれ独立した MD 計算を十分に多数回行い、この時系列を標本集団として、ある時点での各パラメータのヒストグラムを得たとき、このヒストグラムの形は時間発展とともに変化していきだろうと考えられる。フォールド相となる温度領域では時間が経過するにつれ秩序的に折畳まれていくはずなので、ヒストグラムの外形は凹凸の激しい形に近づいていくと考えられ、また、アンフォールド相となる高温領域ではそのような秩序的な構造を示すヒストグラムの様子は観察されないはずである。初期緩和の振舞いは、他節で論じる熱平衡状態におけるモデル鎖の振舞いを考える中では論じることができないので、初期緩和についてはこの節にまとめておくことにする。

上記の目的、つまり、初期緩和において MD による時間発展に伴い秩序構造が現れていく（フォールド相）／現れていかない（アンフォールド相）様子を観察するべく、次のようなシミュレーションを設定し、実行した。すなわち、 $T = 0.4, 0.6, 0.8, 1.0$  というそれぞれの温度において、毎回別の乱数を振り 600 回のシミュレーションを行うという操作である。時系列は  $t = 0 \sim 500000$ 、すなわち  $5.0 \times 10^7$  ステップである。

このシミュレーションから得られた時系列データ標本から、 $t = 500, 10000, 500000$  において、 $R_g^2$ 、 $\Sigma\theta$ 、 $V$  それぞれのパラメータのヒストグラムを得た。この様子を fig.12 に示す。fig.12 から以下のことがわかる：

1.  $t = 500$  の、初期状態から始めてほとんど緩和が起きていない時点では、ヒストグラムは対象となる物理量のパラメータに関わらずベル型を示し、フォールディング構造は見られない。
2.  $t = 500000$  の時点でのヒストグラムを見ると、 $T < 1.0$  の領域では  $R_g^2$  と  $\Sigma\theta$  のヒストグラムに凹凸

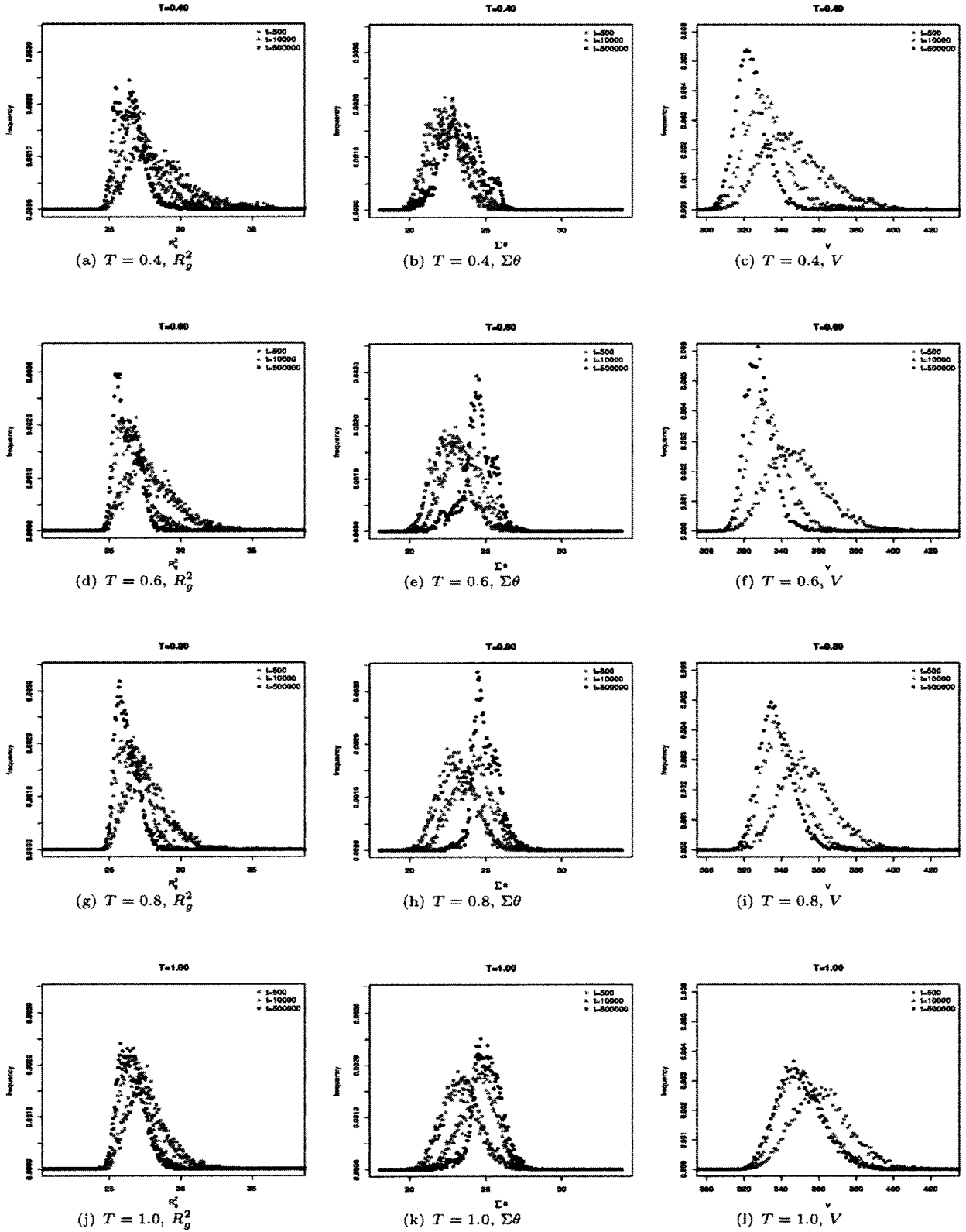


fig.12 : 異なる初期条件から出発した場合の各物理量の時間別頻度分布  $T = 0.4, T = 0.6, T = 0.8, T = 1.0$  の場合に関し、それぞれの温度において  $t = 500000$  の長さで 600 回の独立な MD シミュレーションを行った。それぞれのシミュレーションは初期条件を毎回ランダムに振っている。 $t = 500, 10000, 500000$  の時点で物理量  $R_g^2, \Sigma\theta, V$  をとり、600 回の試行をもとにそれぞれの時点での各物理量の頻度分布をとりヒストグラムとして示したものが上図である。上段から、 $T = 0.4, 0.6, 0.8, 1.0$  と温度で分かれており、また、左 1 列目が  $R_g^2$ 、2 列目が  $\Sigma\theta$ 、3 列目が  $V$  のヒストグラムを示している。それぞれの図中では、桃色で  $t = 500$  の時点、水色で  $t = 10000$  の時点、藍色で  $t = 500000$  の時点でのヒストグラムを示している。

のある分布が現れており、時間発展に連れヒストグラムはこの状態へと近づいている。フォールド相における秩序的な状態への緩和が観察できているものと考えられる。

3.  $t = 500000$  という緩和が起こっていると考えられる時点においても、 $T > 1.0$  の高温領域ではいずれのパラメータもヒストグラムはベル型を示し、この温度領域がアンフォールド相であることを裏付けている。
4. 特に  $\Sigma\theta$  のヒストグラムで顕著だが、 $T = 0.4$  の低温領域ではヒストグラムは複雑な多峰型を描いており、多数の準安定状態に簡単に陥りやすいことを示している。

フォールド相であり、かつ MD 計算が十分に緩和していれば凹凸の激しいヒストグラムになるはずであり、アンフォールド相もしくは緩和していない状態であればヒストグラムはベル型に近い形をとるはずである。 $T = 0.4$  などの低温域で、500000 以上の時間が経てばフォールディング構造は現れてくるものと考えられる。

## 4 レプリカ交換法による解析

### 4.1 レプリカ交換分子動力学法の概要

前々節で示した結果から、通常の MD 計算ではカノニカル平均を算出するためのサンプリングが現実的に可能な時間スケールのシミュレーションからは実行できないことがわかった。このような遅い緩和の問題を解決するために、現在までに多くの手法が開発されている。一般に「拡張アンサンブル法」と呼ばれるが、よく知られている拡張アンサンブル法として、マルチカノニカル法、シミュレーテッド・テンパリング法（焼き戻し法）、レプリカ交換法などがある。本研究では、レプリカ交換法を MD 計算に使用できるように改変したアルゴリズムであるレプリカ交換 MD 法（REMD）を採用し、シミュレーションに用いることにした。レプリカ交換法は、自由エネルギー構造が複雑な系についてサンプリングを行う場合、しばしば陥る準安定状態から効率的に抜け出し、なおかつ詳細釣り合いの条件を満たすサンプリングを行うアルゴリズムである。MD とは名がついているが、サンプリングの方法に MD シミュレーションを用いているだけであって、REMD は動力学を記述するアルゴリズムではないことに注意したい。

REMD のアルゴリズムの概略をここに示しておく。内容は Sugita 1999 [20] において示されている内容に従う。

### 4.2 レプリカ交換分子動力学法のアルゴリズム

REMD では、MD シミュレーションを多数独立して並行させながら走らせる。ここで「独立」とは、MD 計算の上では相互作用をしない、という意味である。それぞれの MD 計算を行っている系は「レプリカ」と呼ばれ、また、この独立した多数のシミュレーションを行わせている系は「レプリカ系」と呼ばれる。各レプリカでは MD 計算を行っているが、MD 計算を行わせている温度が異なっており、レプリカ系とは低温のレプリカから高温のレプリカまでを順に並べたものである。MD 計算を行う各レプリカに番号をふることにする。番号を  $q$  で表すことにし、 $1 \leq q \leq Q$  であるとする。各レプリカの温度は、このレプリカの番号  $q$  により決まっており、シミュレーション前に定義する。それぞれのレプリカの温度は、 $T_1 < T_2 < \dots < T_q < \dots < T_Q$  のように、レプリカの番号に対応して低温から高温へと並んでいる。同時に、各レプリカで考慮する全座標および全運動量にも番号をふり、 $j$  で表す。 $q$  番目のレプリカ内の全座標を  $\mathbf{r}_q^j$ 、全運動量を  $\mathbf{q}_q^j$  という記号により表すことにする。番号  $q$  のレプリカ中で MD 計算されている状態ベクトルに対し  $j$  という番号を振っているが、この  $j$  は交換操作によりレプリカ間で交換されるために変化し、1 から  $Q$  までの整数をとる。ここで

言う「交換」とはレプリカ交換と呼ばれる操作を指す。レプリカ交換とは、あるステップ数の MD 計算を各レプリカでそれぞれ独立して行った後、レプリカ系の中で隣接するレプリカについて、その交換対象対を一つだけ等確率で選び、レプリカの温度は変えずに、座標、速度を交換する操作である。実は隣接したレプリカ間でなくともアルゴリズムは成立するが、隣接したレプリカ間である方が効率的な交換が行えるため、交換対の選び方は隣接したもの同士に限ることにする。この交換は、詳細釣り合いの条件を満たすよう確率的に行い、本稿ではメトロポリス遷移を採用している。

レプリカ交換は温度を徐々に変化させる操作であり、結果、疑似アニーリングのような操作を行っていることになる。このために REMD ではカノニカル分布に従う効率的なサンプリングが行えることになる。

次節においてアルゴリズムの導出・確認を行う。

### 4.3 アルゴリズムの導出・確認

$q$  はレプリカの温度を表し、レプリカそれぞれに固有の値が与えられているが、一方で  $j$  は系を記述する座標と速度のベクトルを表す番号である。こうして二つの番号  $q$  と  $j$  を独立してつけた理由は、交換操作の説明を行う際の便宜を図るためである。REMD では、 $q$  方向の軸の中に割り振られた  $j$  を、 $q$  の軸の中で交換していくという操作を行う。 $j$  を固定して考えれば、状態を記述する  $\mathbf{r}_q^j, \mathbf{v}_q^j$  は  $j$  が  $q$  軸中で交換されていく操作に従い、シミュレーションを行う温度  $T_q$  が徐々に変化させられるという状況である。

以上をまとめると、各レプリカで考慮する質点の数を  $N$  として、それぞれのレプリカは次の変数により特徴づけられることになる：

- ・ 各レプリカの温度  $T_q$  ( $T_1 < T_2 < \dots < T_q \dots < T_Q$ )
- ・ 各レプリカ中で考慮する全ての質点 ( $N$  個) の座標  $\mathbf{r}_q^j$
- ・ 各レプリカ中で考慮する全ての質点 ( $N$  個) の運動量  $\mathbf{p}_q^j$

各レプリカにおいて考慮する位置、速度ベクトルを、レプリカ系に含まれる全てのレプリカについて合わせてまとめたベクトルを考え、それを  $\{\mathbf{r}\}, \{\mathbf{p}\}$  とする。更に便宜のため、各レプリカの状態を記述するベクトルを全て表す  $\chi_q^j$  を設定する。 $\chi_q^j = (\mathbf{r}_q^j, \mathbf{p}_q^j)$  である。また、 $\{\mathbf{r}\}, \{\mathbf{p}\}$  をまとめて  $\{\chi\}$  とする。REMD の場合、REMC の場合とは異なり、交換の際にマクスウェル分布に基づく運動量の補正を行うが、この事情を解決する便宜のため、 $\chi_q^{j'}$  を導入し、 $\chi_q^{j'} = (\mathbf{r}_q^j, \sqrt{\frac{T_{q\pm 1}}{T_q}} \mathbf{p}_q^j)$  と定義する。 $T_{q\pm 1}$  の符号は、元のレプリカが交換により高温側へ移るか（この場合  $T_{q+1}$ ）低温側へ移るか（この場合  $T_{q-1}$ ）に依る。交換とは次の操作を言う：

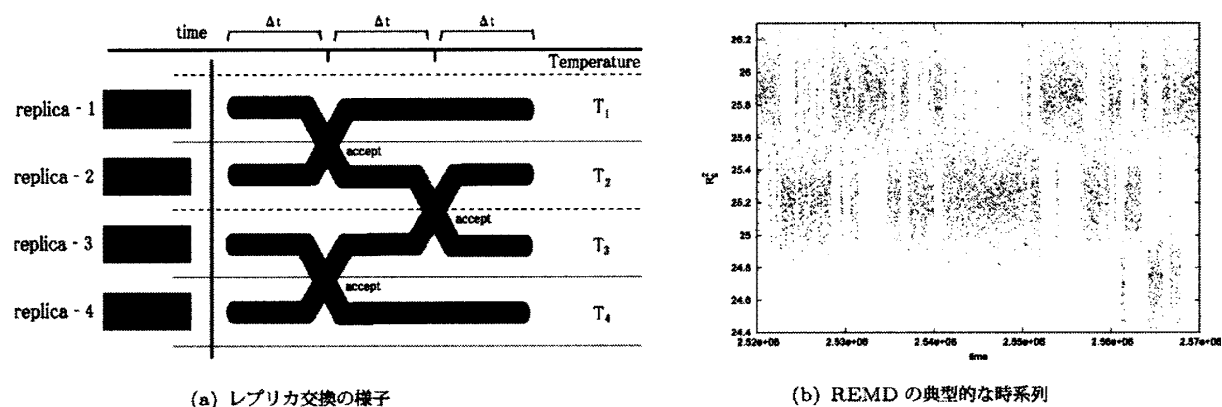


fig.13 : REMD の模式図と時系列

$$\{\dots, \chi_q^{j_a}, \chi_{q+1}^{j_b}, \dots\} \rightarrow \{\dots, \chi_q^{j_b}, \chi_{q+1}^{j_a}, \dots\} \quad (10)$$

ここで、 $j_a$  と  $j_b$  は、交換操作の時点でレプリカ  $q$  と  $q+1$  がそれぞれ持っている状態ベクトルの組の番号である。これにより、レプリカ系全体を記述するハミルトニアンは  $\mathcal{H}_q(\{\chi\})$  と表されるが、 $\mathcal{H}(\{\chi\})$  はレプリカそれぞれの運動エネルギーをレプリカ系全体で和をとった量  $K(\{\mathbf{p}\})$  と、それに対応するポテンシャルエネルギー  $V(\{\mathbf{r}\})$  の和で表される：

$$\mathcal{H}(\{\chi\}) = K(\{\mathbf{q}\}) + V(\{\mathbf{r}\}) \quad (11)$$

すなわち、レプリカ系がある状態  $(\{\chi\})$  をとるとき、ハミルトニアン  $\mathcal{H}(\{\chi\})$  は対応する値をとるが、各レプリカが持つハミルトニアンを  $\mathcal{H}_q(\chi_q^j)$  として表すと、温度一定条件下におけるその状態の出現確率  $W(\{\chi\})$  は次のボルツマン因子により決まる：

$$W(\{\chi\}) = \prod_{q=1}^Q \exp(-\beta_q \mathcal{H}_q(\chi_q^j))^{*3} \quad (12)$$

ここで、 $\beta_q$  とは  $q$  番目のレプリカの逆温度であり、 $\beta_q = 1/k_B T_q$  である。

以下、レプリカを交換する操作を説明する。既述の通り、これは  $j$  に番号付けされる状態  $\chi_q^j$  を  $q$  軸方向で交換する操作である。

レプリカ系がとる状態出現頻度はカノニカル分布に従うとしたので、レプリカ  $q$  と  $q+1$  ( $1 \leq q \leq Q-1$ ) の状態  $\{\chi_q^{j_a}\}$  と  $\{\chi_{q+1}^{j_b}\}$  が交換される操作が実行される確率は、次の詳細釣り合いの条件を満たさなければならない：

$$\begin{aligned} W(\{\dots, \chi_q^{j_a}, \chi_{q+1}^{j_b}, \dots\}) w(\chi_q^{j_a}, \chi_{q+1}^{j_b} \rightarrow \chi_q^{j_b}, \chi_{q+1}^{j_a}) \\ = W(\{\dots, \chi_q^{j_b}, \chi_{q+1}^{j_a}, \dots\}) w(\chi_q^{j_b}, \chi_{q+1}^{j_a} \rightarrow \chi_q^{j_a}, \chi_{q+1}^{j_b}) \end{aligned} \quad (13)$$

ここで  $W(\{\dots, \chi_q^{j_a}, \chi_{q+1}^{j_b}, \dots\})$  と書いた値は、レプリカ系が状態  $\{\dots, \chi_q^{j_a}, \chi_{q+1}^{j_b}, \dots\}$  にある場合が起こる確率である。 $w(\chi_q^{j_a}, \chi_{q+1}^{j_b} \rightarrow \chi_q^{j_b}, \chi_{q+1}^{j_a})$  は、状態  $\chi_q^{j_a}$  を持つレプリカ  $q$  と、状態  $\chi_{q+1}^{j_b}$  を持つレプリカ  $q+1$  の間で詳細釣り合いの条件を満たす交換が起こる確率を示す。逆も同様である。式変形すると以下ようになる：

$$\frac{w(\chi_q^{j_a}, \chi_{q+1}^{j_b} \rightarrow \chi_q^{j_b}, \chi_{q+1}^{j_a})}{w(\chi_q^{j_b}, \chi_{q+1}^{j_a} \rightarrow \chi_q^{j_a}, \chi_{q+1}^{j_b})} = \frac{W(\{\dots, \chi_q^{j_b}, \chi_{q+1}^{j_a}, \dots\})}{W(\{\dots, \chi_q^{j_a}, \chi_{q+1}^{j_b}, \dots\})} \quad (14)$$

(12) 式から次のようになる：

$$\begin{aligned} \frac{w(\chi_q^{j_a}, \chi_{q+1}^{j_b} \rightarrow \chi_q^{j_b}, \chi_{q+1}^{j_a})}{w(\chi_q^{j_b}, \chi_{q+1}^{j_a} \rightarrow \chi_q^{j_a}, \chi_{q+1}^{j_b})} = \exp(-\Delta) \\ \Delta = \{\beta_q \mathcal{H}_q(\chi_q^{j_b}) + \beta_{q+1} \mathcal{H}_{q+1}(\chi_{q+1}^{j_a})\} - \{\beta_q \mathcal{H}_q(\chi_q^{j_a}) + \beta_{q+1} \mathcal{H}_{q+1}(\chi_{q+1}^{j_b})\} \end{aligned} \quad (15)$$

ここで、各レプリカのハミルトニアン  $\mathcal{H}_q(\chi_q^j)$  は、各レプリカのポテンシャルと運動エネルギーより成り、次のように表される：

---

\*3 ここでは、 $j$  は  $q$  により一意に決まる。

$$\mathcal{H}_q(\chi_q^j) = K(\mathbf{q}_q^j) + V(\mathbf{r}_q^j) \quad (16)$$

そして、交換を起こした場合に計算されるハミルトニアン  $\mathcal{H}_q(\chi_q^j)'$  は、マクスウェル分布に基づく補正を施されるため次のように表される：

$$\mathcal{H}_q(\chi_q^j)' = \frac{\beta_{q+1}}{\beta_q} K(\mathbf{q}_q^j) + V(\mathbf{r}_q^j) \quad (17)$$

よって (15) 式中の  $\Delta$  は次のようになる：

$$\begin{aligned} \Delta = & [\beta_q \{ \frac{\beta_{q+1}}{\beta_q} K(\mathbf{q}_q^{j_b}) + V(\mathbf{r}_q^{j_b}) \} + \beta_{q+1} \{ \frac{\beta_q}{\beta_{q+1}} K(\mathbf{q}_{q+1}^{j_a}) + V(\mathbf{r}_{q+1}^{j_a}) \}] \\ & - [\beta_q \{ K(\mathbf{q}_q^{j_a}) + V(\mathbf{r}_q^{j_a}) \} + \beta_{q+1} \{ K(\mathbf{q}_{q+1}^{j_b}) + V(\mathbf{r}_{q+1}^{j_b}) \}] \end{aligned} \quad (18)$$

ポテンシャル  $V$  は温度に依存しないことに注意し、式を整理すると

$$\Delta = (\beta_{q+1} - \beta_q) \{ V(\mathbf{r}_{q+1}^{j_a}) - V(\mathbf{r}_q^{j_b}) \} \quad (19)$$

となる。つまり、メトロポリス遷移を行う際には、交換の対象となるレプリカ対のポテンシャルを計算し、温度差とポテンシャル差に基づいた数  $\Delta$  を求めて確率的な交換を起こさせれば、詳細釣り合いの条件を満たす交換が行えるということがわかる。次のようなメトロポリス遷移を行わせれば (13) 式が満足されるサンプリングが行われる：

$$w(\chi_q^{j_a}, \chi_{q+1}^{j_b} \rightarrow \chi_q^{j_b}, \chi_{q+1}^{j_a}) = \begin{cases} 1 & \text{for } \Delta \leq 0 \\ \exp(-\Delta) & \text{for } \Delta > 0 \end{cases} \quad (20)$$

先述した通り、交換には方向がある。上のアルゴリズムにおいて、 $q+1$  を  $q-1$  と読み替えれば、逆方向の交換の説明となる。実際のシミュレーションにおいては両方向の交換を同じ回数行う必要がある。

以上のアルゴリズムが、本研究で用いたレプリカ交換分子動力学法である。

レプリカの数、各レプリカに割り振る温度、および交換頻度をうまく調整しないとレプリカ交換法は意味のある結果を出さない。試行錯誤の結果、次のような条件で安定した平均物理量を出力することがわかった：

- ・ レプリカ数：70
- ・ 各レプリカの温度： $T_q = 2 \times (0.15 + q/100)^2$  ( $0 \leq q < 70$ )
- ・ 時系列の長さ： $5.0 \times 10^8$  ステップ ( $t = 0 \sim 5.0 \times 10^6$ )
- ・ メトロポリス遷移を起こす頻度\*4：1000 ステップに 2 回

#### 4.4 REMD の結果

前節までに示した方法の通りに、モデル鎖の各パラメータの時系列平均を REMD により計算した。結果を次のように分けて説明する。

---

\*4 棄却する場合も含む。

1. シミュレーションの進行に伴うガラス状態の緩和
2. 各パラメータの時系列平均
3.  $R_g^2$  と  $\sum \theta$  による二次元ヒストグラムの温度変化に伴う遷移

上記の内容を小節に分け、それぞれを説明する。

#### 4.4.1 シミュレーションの進行に伴うガラス状態の緩和

REMD は動力学計算を行う系にアニーリングを起こし、緩和を促進させるアルゴリズムであるが、まずはその緩和の様子についてである。 $5.0 \times 10^6 t$  の長さのシミュレーションを行い、 $1.0 \times 10^6 t$  ごとに時系列を 5 つにわけてそれぞれの時間領域において平均値を算出してみたが、 $1.0 \times 10^6 t$  以上であれば結果は同一であることが分かったので、 $1.0 \times 10^6 t$  までの時系列を捨てて以降平均値の算出の根拠とすることにする。

#### 4.4.2 各パラメータの時系列平均

$t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  での時系列を用いて、 $R_g^2$  の時系列平均をプロットしたものを fig.14(a) に、 $\sum \theta$  の時系列平均をプロットしたものを fig.14(b) に示す。20 回の試行を行い、それぞれの温度における平均値と標準偏差を算出した。20 回の試行で計算される平均値も大きい赤点で fig.14(a)、fig.14(b) には示しており、さらに標準偏差の 2 倍の幅でエラーバーとなる灰色のリボンを付加して示している。また、同時にそれぞれのパラメータを NMD で計算した場合の結果 (3.1 節参照) も青点で重ねて示している。fig.14(c)、(d) には、レプリカ系で設定したそれぞれの温度における  $R_g^2$ 、 $\sum \theta$  の時系列頻度を示すカラーマップを示した。エネルギー  $E$  については次章で紹介することにする。

fig.14 のエラーバーの様子から、結果は収束していると判断した。使用した REMD のアルゴリズムは実際の分配関数に従う重率に十分によく従うサンプリングを達成していると判断する。ただ、シミュレーション対象とした範囲のうち、最も低温の領域では試行に収束が見られていない。これは、ごく低温の領域では熱揺らぎがほとんどないためにレプリカ交換がうまく働かず、それぞれの準安定状態に陥ったとき、レプリカ同士での相対的なエネルギー差に揺らぎがなくなってしまうためであると考えられる。しかしながら、この領域の温度を本研究では対象とせず、本研究の全過程を通じて転移温度付近に関するモデル鎖の振舞いに注意をするので、この点の不備は無視することにした。低温域の揺らぎを収束させるべく、この温度領域でのレプリカ数を増やすという操作も行ったのだが、結局は低温域における揺らぎを解決することはできなかった。また、 $T > 0.3$  の領域では結果は収束していることから、本節最初で示した結果を出す程度の長さの時系列を対象に今後も議論を行う。

#### 4.4.3 $R_g^2$ と $\sum \theta$ による二次元ヒストグラムの温度変化に伴う遷移

$R_g^2$  の様子からまずは説明する。fig.14(a) を見ると、その概形は以下のようにまとめられる：

- ・ 低温の  $T = 0.15$  付近では  $R_g^2 = 25.2$  程度で、そこから  $R_g^2$  は温度上昇と共に増加していく
- ・  $T = 0.25$  付近を極大として、 $R_g^2$  は  $T = 0.6$  付近まで減少していく
- ・  $T = 0.6$  付近を極小として、それ以上の高温領域では、 $R_g^2$  は温度  $T$  の上昇と共にほぼ線形関係の形で単調に増加していく

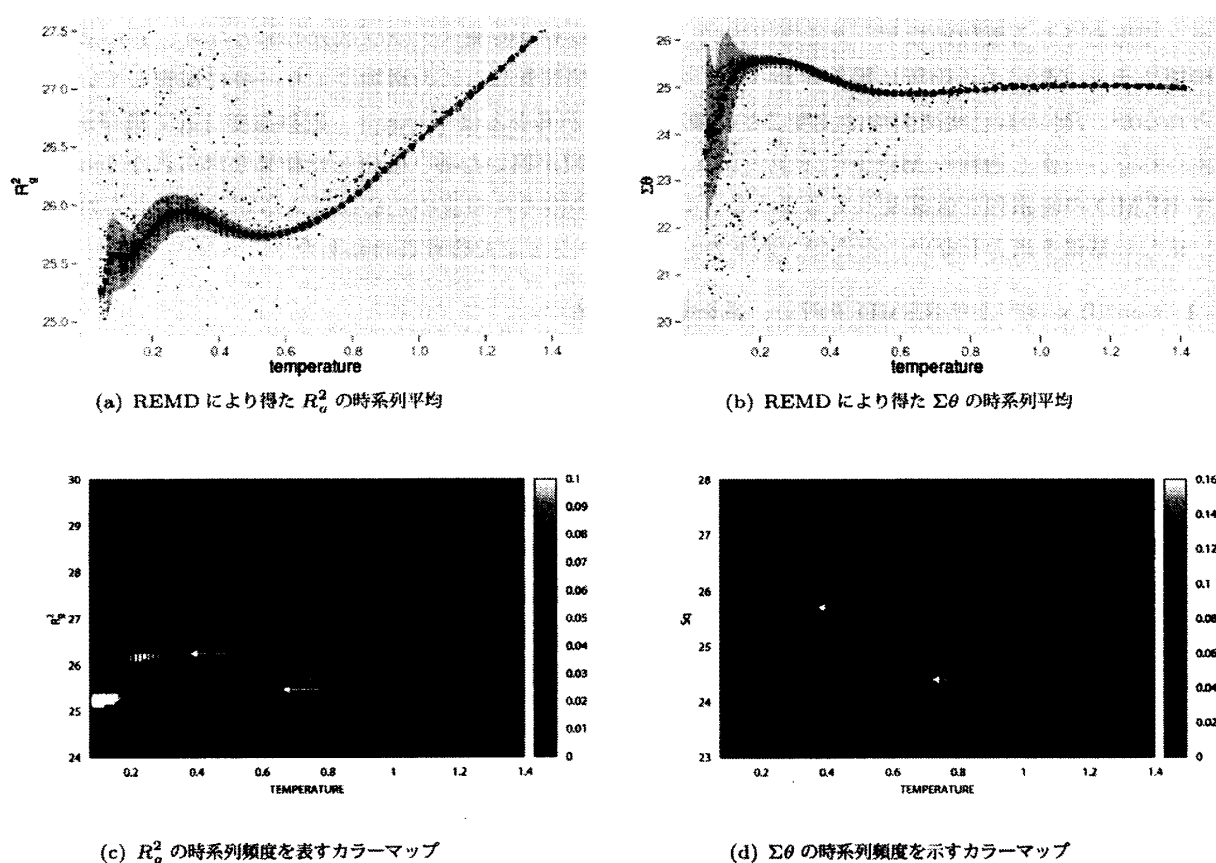
温度上昇に伴う揺動力の増加により、 $R_g^2$  は温度上昇に従い単調に増加していくものと一見は考えられるが、一方で REMD の結果は上記の通り、 $T < 0.6$  の領域では増減がある。3.1 節で示されたように、 $T < 0.6$  の領域はフォールド相にあり、構造がアンフォールドするために  $R_g^2$  が増減するというのではない。後で詳

しく紹介するが、これは  $R_g^2$  に差がある複数の構造が同時に現れているために起こる増減であり、温度によりどの構造がより安定であるかが異なるためにこのような平均値の増減が起きる。その様子は fig.14(c) のカラーマップにも現れており、 $T < 0.6$  の領域で明るくなっている領域がそれらの安定構造の分布を示しているが、どのような立体構造にあるのかについては本章 7 節で詳述する。

$\Sigma\theta$  は温度変化に伴い以下のような変化をする：

- ・ 低温の  $T = 0.2$  付近では  $\Sigma\theta$  は 25.5 付近の値をとり、そこから温度上昇と共に減少していく
- ・  $T = 0.5$  付近で  $\Sigma\theta$  は 25.0 付近の値をとって、それより高温では変化しなくなる

fig.17(d) のカラーマップの中には少なくとも二つの構造が競合関係にあると推測されるが、これについてもやはり後に詳述することにする。



**fig.14** : REMD により得た各 Shape Parameter の平均と時系列頻度 (a)  $t = 5.0 \times 10^6$  までの REMD シミュレーションを行い、 $t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  までの時系列を用いて  $R_g^2$  の平均を算出し温度ごとにプロットした。20 回のシミュレーションを行っており、それぞれの結果を細かい赤点で示している。20 回すべての平均値を更に平均した値を大きい赤点で示し、標準偏差の 2 倍をグレーのリボンとして示している。図中の青点は、??において示した通常の MD シミュレーションによる時系列平均である（時系列の長さ・領域は同じである）。(b) (a) と同様に  $\Sigma\theta$  の平均を示している。(a) と同じく標準偏差の 2 倍を示すリボン、通常の MD シミュレーションによる結果を示した。(c) REMD による  $R_g^2$  の時系列中に現れる値の頻度分布を温度ごとにとり、カラーマップとして示している。(d) REMD による  $\Sigma\theta$  の時系列中に現れる値の頻度分布を温度ごとにとり、カラーマップとして示している。

#### 4.5 自己相関関数の転移

REMDにより高速に緩和が起こることを利用し、自己相関関数を得て温度ごとにプロットすることを目的としたシミュレーションを行ったのでその結果を報告する。フォールディング転移が起こるのであれば、ある温度を境にして各パラメータの自己相関に飛びが起こるはずである。ここでは  $R_g^2$  に注目し自己相関関数による議論を行う。

$f(t)$  の自己相関関数  $C(\tau)$  は次式により定義される：

$$C(\tau) = \frac{1}{\sigma^2 T} \int_0^T (f(t) - \bar{f}) \cdot (f(t+\tau) - \bar{f}) dt \quad (21)$$

ここで  $\sigma^2$  は関数  $f(t)$  の時系列中の分散であり、 $\bar{f}$  は  $f$  の時系列平均、 $T$  は時系列の長さである。時系列があったとき、 $\tau$  だけ時間が進んだ時点での量  $f(t+\tau)$  との相関を示す量が自己相関関数である。シミュレーションで出力される離散的なデータを扱うため、上式の積分は時間ステップ数分の和となり、 $T$  による除算は時間ステップ数による除算に置き換わる。この自己相関関数は、 $\tau$  の増加により一般に減衰していくと考えられるが、 $R_g^2$  の自己相関関数は温度ごとに減衰の仕方が異なるはずであり、転移温度付近で飛びがあると考えられる。しかし通常の MD シミュレーションでは緩和が遅いため、安定した結果が得られなかった。そして REMD の時系列を直接使ってしまうと、レプリカ交換それ自身のために揺らぎが大きくとられてしまい、正しい結果を出力しない。そこで、以下のシミュレーションを行うことにした：

1.  $t = 1.0 \times 10^6$  まで REMD を行い、系を緩和させる
2.  $t = 1.0 \times 10^6$  からは交換操作を止め、通常の MD シミュレーションを各温度で行わせる
3.  $t = 1.0 \times 10^6 \sim 2.0 \times 10^6$  まで通常の MD シミュレーションを行い、時系列を得て (21) 式の  $\tau$  を動かし、それぞれの  $\tau$  において時系列データから自己相関関数を得る

以上のようなシミュレーションを行うと、fig.15 に示す結果が得られた。fig.15(a) には、温度を横軸、 $\tau$  を縦軸、配色を自己相関関数の値としてカラーマップにより自己相関関数の様子を表示している。 $\tau$  の上昇とともに自己相関関数の値は減衰していくが、 $T = 0.6$  に特殊な振舞いが見られ、この温度で明らかに自己相関関数は大きな値をとっている。 $\tau = 20000$  として  $R_g^2$  の自己相関関数を温度ごとに示したものが fig.15(b) であるが、 $T = 0.6$  に明白な飛びがあることが分かる。この転移はフォールディング転移であると考えられる。

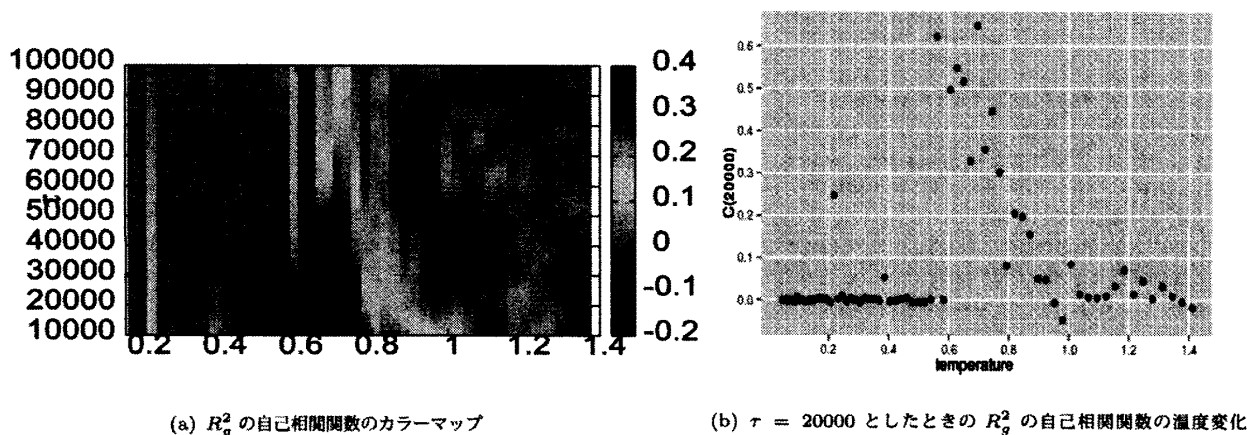


fig.15 :  $R_g^2$  の自己相関関数

フォールディング転移温度  $T_f$  で自己相関関数にピークが示される理由についてであるが、 $T = 0.6$  においては、 $R_g^2$  の時系列は fig.8(c), (d) に示すようなパターンを描くため、時系列平均と各時点での値とのずれが大きくなり、結果として自己相関関数が大きくなるのだと考えられる。逆にフォールド相では揺らぎが全くないために自己相関関数は 0 に近い値をとり、また、アンフォールド相においては高温になるに連れ無秩序さが増していくため、温度上昇とともに自己相関関数は減衰していく。

ここで示されたフォールディング転移点  $T_f$  以下の温度では、シミュレーションを行う時間幅の範囲においては系はまったく状態間遷移を起こさないが、かといって  $T_f$  以上の高温になると無秩序相になるかということそうではない点に注意したい。自己相関関数が大きい値を持つということは準安定状態に長く滞留し、観察した時系列幅の中で状態間遷移を起こしているということなので、アンフォールドされた状態であるとはいえない。自己相関関数が 0 となる点がガラス転移点  $T_g$  であると考えられるが、本節の結果からはその温度は  $T = 1.0$  付近であると考えられる。これは、3.1 節で示された、通常の MD シミュレーションによる平均がバラつき始める温度とも一致する。 $T_f = 0.6$ 、 $T_g \simeq 1.0$  として以降議論を進める。

## 4.6 二次元ヒストグラムの温度変化に伴う遷移

温度変化に伴い、モデル鎖の安定状態はどのように遷移していくのだろうか。それは、 $R_g^2$  と  $\Sigma\theta$  を両軸にとった二次元ヒストグラムを REMD の時系列データから得ることで観察できる。前節の結果を得た REMD 時系列データを用い、二次元ヒストグラムを得た。その様子を示した図が fig.16 である。レプリカとして分けた全ての温度で二次元ヒストグラムを得たが、特に注目すべき変化を表している温度を選び、fig.16 に示した。時系列中で特に高い頻度で現れる状態を示す領域は白色で示されているが、色によるヒストグラムではその高低を判別しにくいので、同じデータからポリゴンによるヒストグラムを得、fig.17 に示した。

$T = 0.08, 0.10$  のヒストグラムでは、7 箇所以上の準安定状態が確認できる。本シミュレーションで考慮した温度領域の中で最も低温の状態であるが、この温度では準安定状態が多数出現し、入れ替わりその状態の分布も変化し、温度が高くなると新たに出現したり、また、消失したりする。このことから、この温度領域の自由エネルギー地形は非常に複雑な構造を示すことが予想される。fig.15 の相関関数の時間相関係数の値から、 $T < 0.6$  の領域ではフォールド相であると考えられるが、フォールド相ではヒストグラムの様子もピークは鋭く、揺らぎはほとんどないと見える。

$T = 0.15$  となると準安定な構造は急激に減り、安定な構造は 3 つほどとなる。 $(R_g^2, \Sigma\theta) = (25.5, 24.5), (25.5, 25.5), (26.5, 25.7)$  にそれぞれの準安定構造の分布のピークがあると考えられる。3 つの準安定状態はそれぞれの相対的な頻度分布を変えながら転移温度である  $T = 0.60$  付近まで維持される。その様子は fig.14(c), (d) に示したヒストグラムの変化と対応している。

## 4.7 安定構造の可視化

前節で示した二次元ヒストグラムの中に現れたそれぞれのパッチは、どのような立体構造に対応するのだろうか。2.4 節で説明したグラフィックス・ツールを用いてそれらを可視化することにした。温度の代表点として  $T = 0.3, T = 0.6, T = 1.2$  を選び、それぞれのヒストグラム中のパッチにあたる立体構造を調べていくことにする。

$T = 0.3$  の場合を fig.19 に示した。図の見方の概略は次の通りである。fig.18 を参考にしながら説明する：

- ・ fig.18 中の A にあたる図では、各パッチにあたる安定構造のある時点での座標と速度、力を REMD の時系列から抜き出し、それらのベクトルを初期条件として、同じ温度において  $1.0 \times 10^5$  ステップ

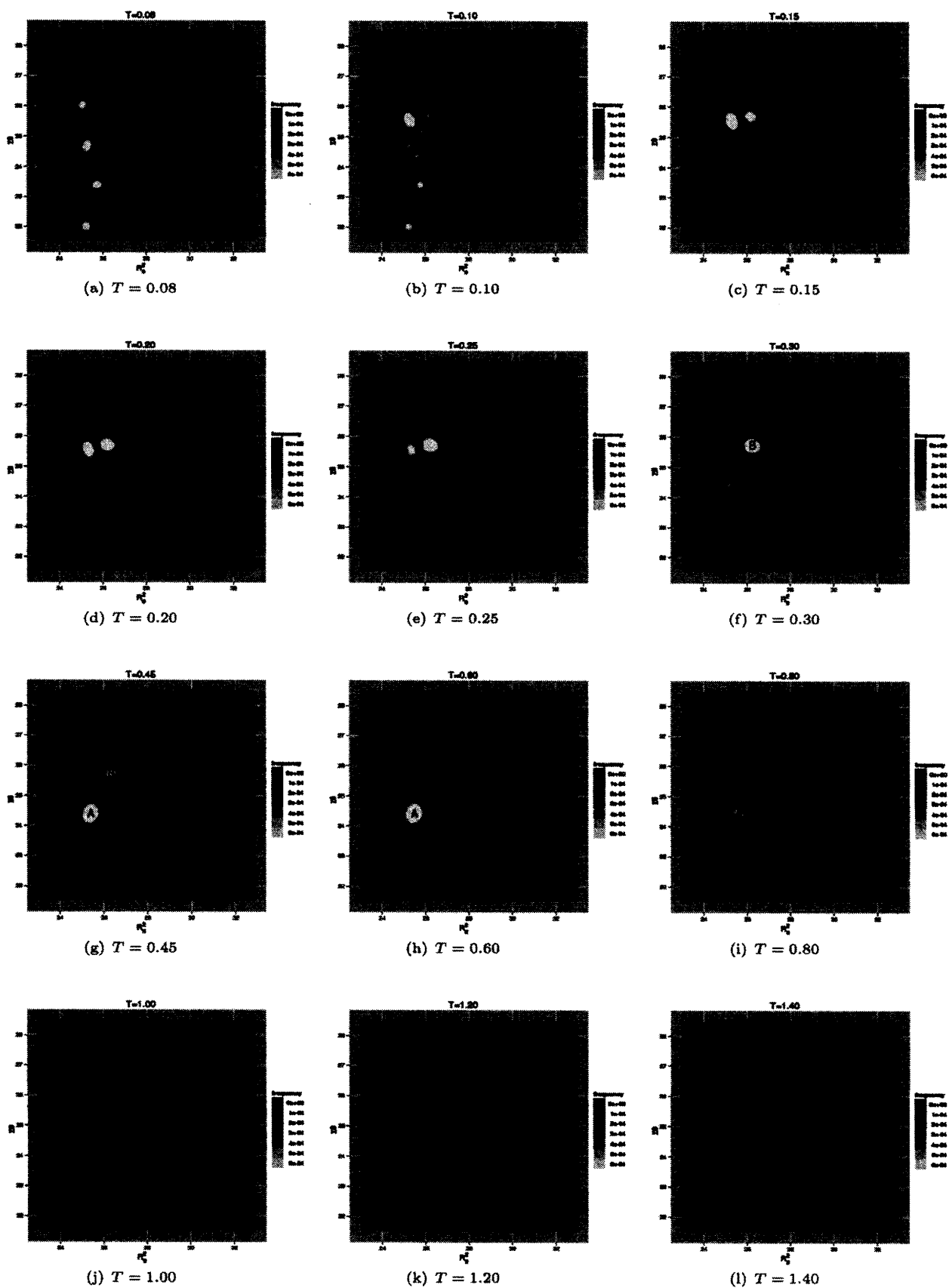


fig.16 : REMD シミュレーションから得た  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラム

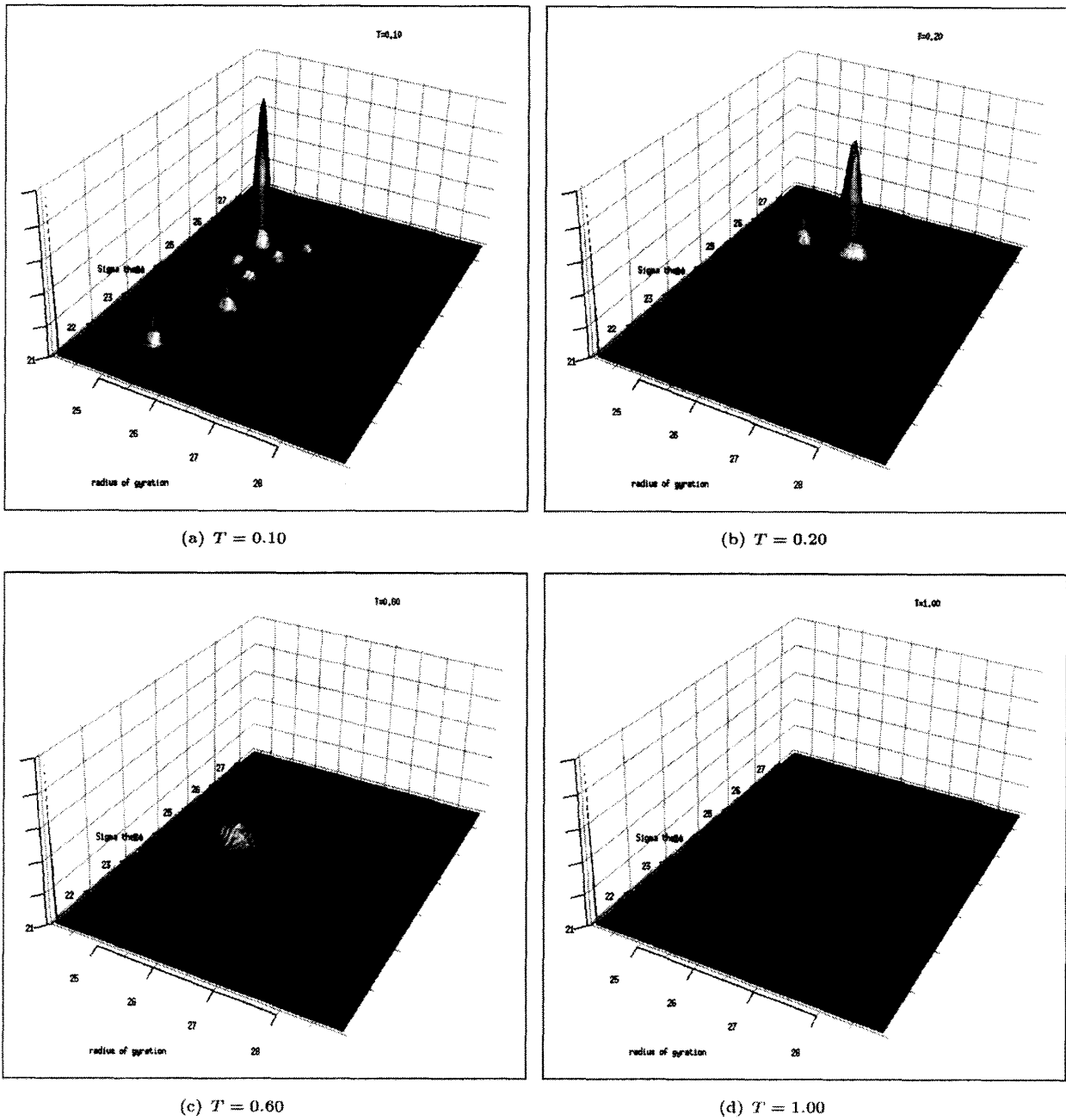


fig.17 : REMD シミュレーションから得た  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラム (3-D 表示)

(1000t) の NMD を走らせ、 $R_g^2$ - $\Sigma\theta$  平面内での軌道を描いている。 $R_g^2$ - $\Sigma\theta$  平面には REMD の結果より得た時系列頻度の等高線が描かれており、また、初期条件の違いにより軌道も色分けしてある。 $T = 0.3$ 、 $T = 0.6$  では、頻繁に現れる構造を初期条件とした軌道を描いており、また、 $T = 1.2$  の場合には初期条件は REMD の時系列中からランダムに数回選んで NMD を走らせ、その軌道をそれぞれ描いている。

- ・ fig.18 中の B にあたる図では、A で重ね書きされる前の  $R_g^2$ - $\Sigma\theta$  平面内での等高線のプロットを示している。
- ・ fig.18 中の C にあたる図では、 $T = 0.3$  と  $T = 0.6$  の場合には、初期条件となる構造をグラフィック

ス・ツールで示したものを、 $T = 1.2$  の場合には、NMD の時系列中で現れる不規則な構造を示している。A、B、C などの色分けしたラベルをそれぞれの図の左下に貼付しているが、これは図 A 中のパッチにも同じく貼付されており、対応がある。図中には残基番号を白字で示してある。

- ・ fig.18 中の D にあたる図では、図 A 中の  $1.0 \times 10^5$  ステップの NMD を行っている間の分子鎖中の結合角  $\theta$  を、残基番号に対応させながら折れ線グラフでプロットしている。1000 ステップ (10t) おきに結合角の系列を出力しており、それぞれの時間での出力を線で繋いでいる。

この図の形式は、fig.20、fig.21、fig.28、fig.29、fig.30 についても同様である。

低温域における REMD の時系列頻度二次元ヒストグラム中に現れるパッチは、やはりそれぞれ異なる構造に相当することが fig.19、fig.20 により示された。パッチと NMD による軌道がきれいに対応しているためである。しかし、fig.21 で示される  $T = 1.2$  のアンフォールド相では規則的な軌道は NMD では描けず、揺らぎが非常に大きい。図 D で示される結合角も、1000t という短い時系列の中でも大きく揺らいでいることがわかる。

赤色の A でラベルした構造は、広い温度領域で観察される構造であり、末端の残基が互いに向き合っているという特徴がある。 $R_g^2$  も小さく、次章で述べる Confined 系においてもやはり広い温度領域で観察される。

橙色の B でラベルした構造は、1 側の末端が螺旋状の配置となっていることが特徴であり、 $R_g^2$  が大きい。いま行っている、束縛のない状態でのシミュレーションでは広い温度領域で観察される構造であるが、Confined 系では観察されにくくなる。

緑色の C でラベルした構造は、1 側の末端と 15 側の末端が口金のように向き合う配置となっており、 $R_g^2$  は小さい。いまの束縛のない状態でのシミュレーションでは狭い温度範囲で観察される構造であるが、Confined 系では広い温度領域で観察される。

この他にも、レアイベントとなる構造が観察される場合があるが、ヒストグラムにも現れないため割愛した。

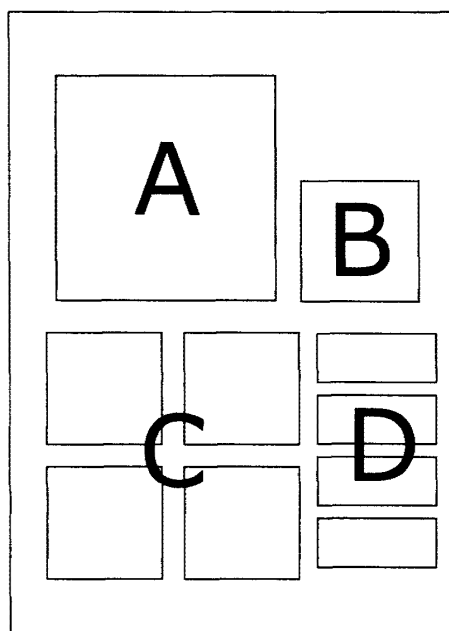


fig.18 : 安定構造の図の見方

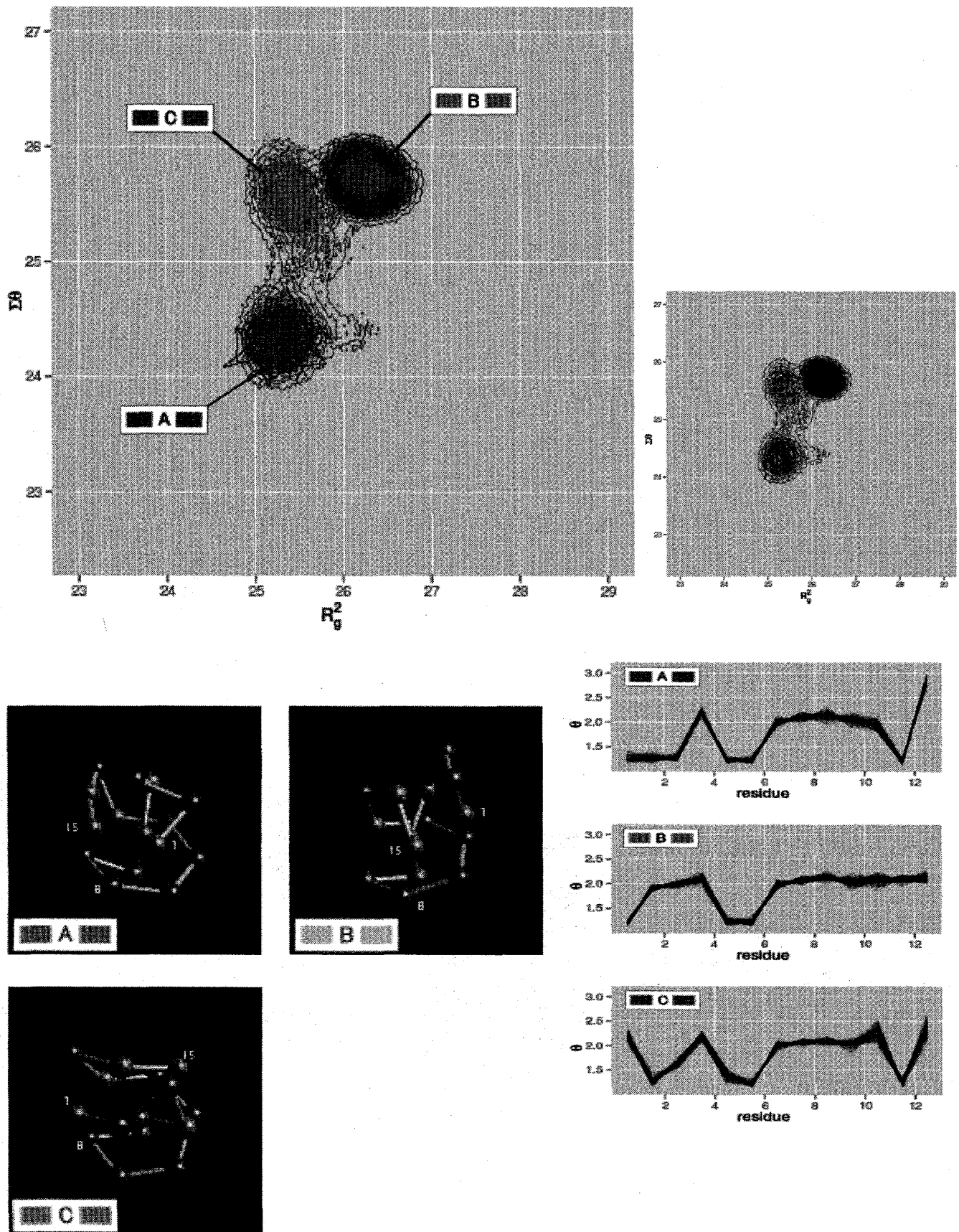


fig.19 :  $T = 0.3$  の場合

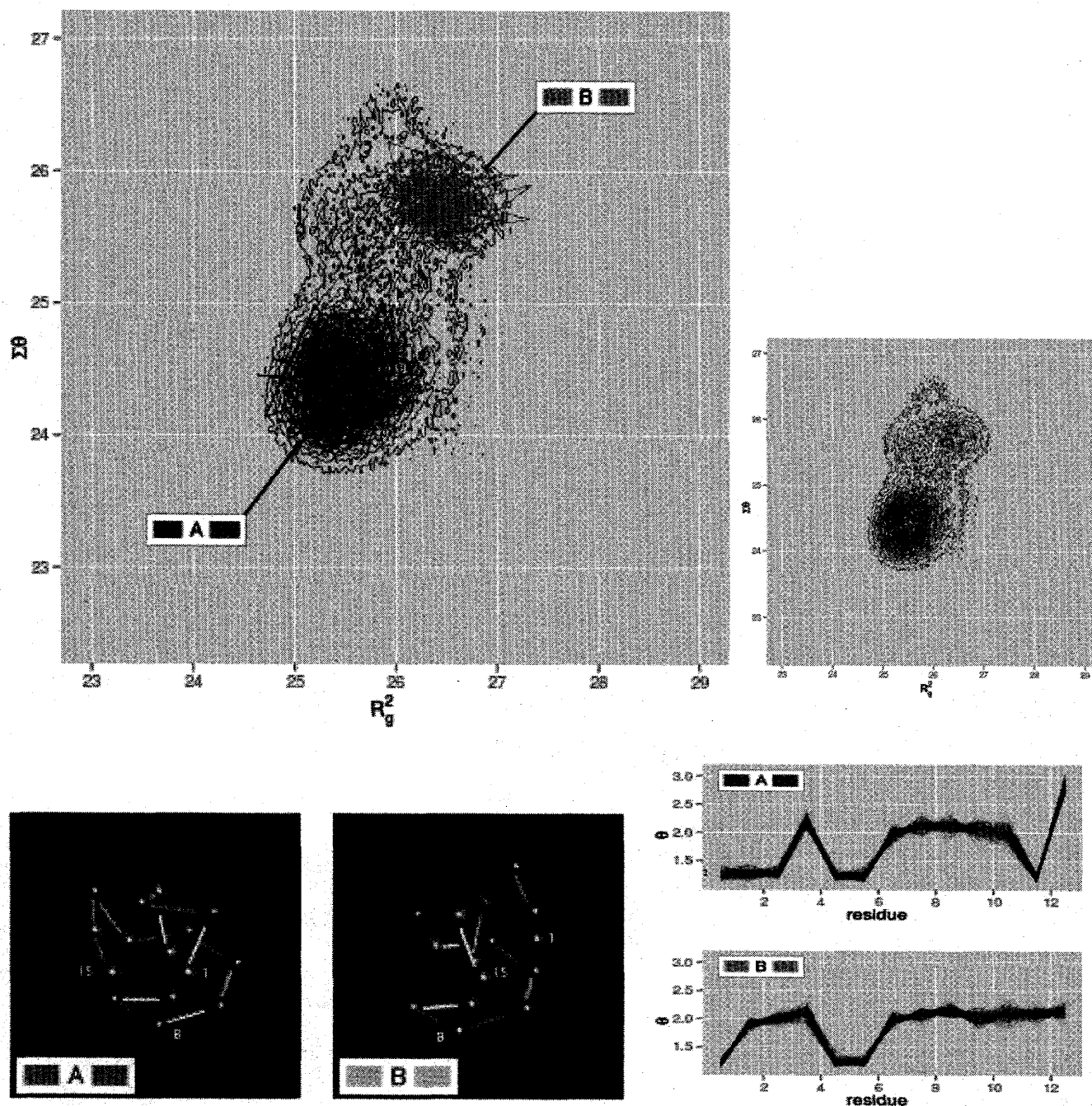


fig.20 :  $T = 0.6$  の場合

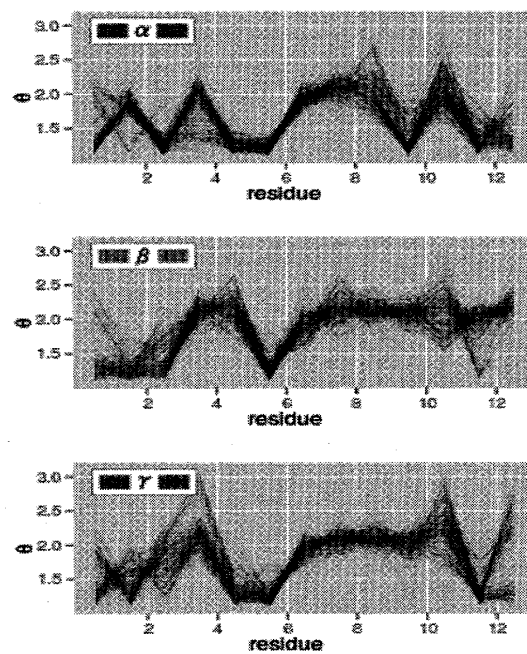
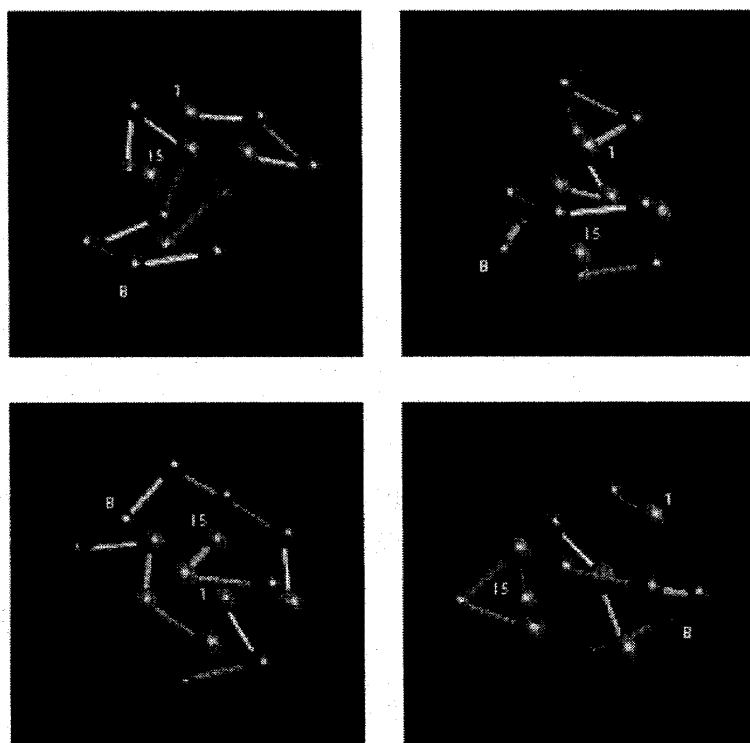
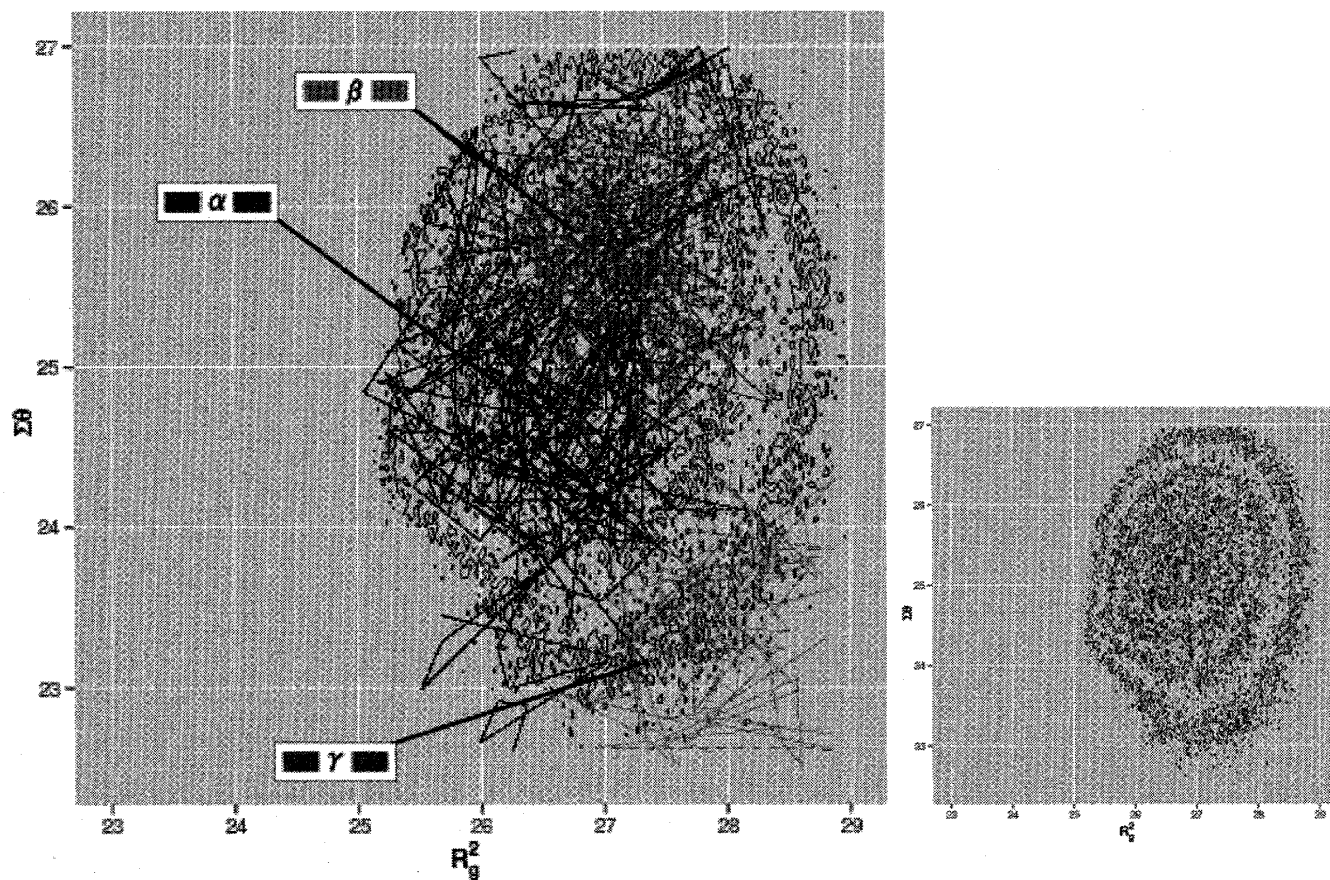


fig.21 :  $T = 1.2$  の場合

## 5 Confined 系

### 5.1 束縛ポテンシャルの導入

F.Takagi 2003 [17] は郷モデル\*5を用い、タンパク質を円柱状のポテンシャルの中に閉じ込めた場合の転移温度の変化について観察した。半径  $L$ 、高さ  $2L$  の円柱状のケージを設定し、その中にタンパク質を閉じ込め MD 計算を行うという操作である。郷モデルのポテンシャルに次のようなポテンシャルを追加し、タンパク質を束縛した状態での MD シミュレーションを行い、モデル鎖のフォールディングの振舞いの変化を観察した。

$$\sum_i K_{cage} \left[ \left( \frac{C}{2d_i} \right)^4 - 2 \left( \frac{C}{2d_i} \right)^2 + 1 \right] H(C/2 - d_i) \quad (22)$$

ここで、 $d_i$  としたのは、円柱状のポテンシャル壁との距離であり、また、 $H(x)$  はヘヴィサイド関数であり、 $H(x) = 1$  for  $x > 0$  かつ  $H(x) = 0$  for  $x < 0$  である。

その結果、フォールディング転移温度が高温側に移動するという現象を観察し、報告している。また、J.Mittal and R.B.Best 2008 [18] は、円柱状のポテンシャルの他に、平行板状、球状のポテンシャルを設定し、その束縛ポテンシャルの大きさと転移温度との関係を、やはり郷モデルの MD 計算により詳細に調べた。その結果、束縛ポテンシャルの大きさと転移温度は負の相関があり、束縛ポテンシャルの大きさの減少（つまり、束縛条件をより厳しくしていく）とともに転移温度は線形に上昇するということを観察した。本研究で設定したモデル鎖にも同様のポテンシャルを与え、これら先行研究で報告された内容と同様に、転移温度が上がるかどうか、また、その他特異な振舞いが現れるか調べた。束縛ポテンシャルの導入により、4 章 fig.16 に示したような状態頻度分布は変化すると予想される。これについても、4.3 節で導入した REMD 計算を用い、束縛ポテンシャルの下で REMD 計算を行うとどのような形になるのか調べた。

本稿では次の束縛ポテンシャルを与える：

$$\sum_i K_{cage} \left[ \left( \frac{C}{2d_i} \right)^4 - 2 \left( \frac{C}{2d_i} \right)^2 \right] H(C/2 - d_i) \quad (23)$$

ここで、 $K_{cage} = 1.0$ 、 $C = 4.0$  である。束縛ポテンシャルの概形を fig.22(a) に示した。ポテンシャル壁の形は球形となっている (fig.22(b))。以下、本稿では (23) 式のポテンシャルによりモデル鎖の閉じ込めを施した系を Confined 系と呼称することにする。

Confined 系の MD シミュレーションを行う場合、閉じ込めを行うポテンシャルの半径が重要になる。モデル鎖にとって適切な半径があり、あまりに半径を大きくとり過ぎても束縛による効果は出ないし、小さくとり過ぎるとポテンシャル壁との衝突が激しくなり、モデル鎖中の残基間相互作用を無視する形でアンフォールド状態となり、フォールディングそのものが現れなくなる。そしてモデル鎖の緩和も遅いため、束縛の効果が現れているかどうかを即座に判定することも困難である。最適な半径は REMD を用いながら試行錯誤して設定する外なく、実際、本研究ではそうした。以下、束縛ポテンシャルの半径を CageSize と呼ぶことにする。

\*5 郷モデルとは、タンパク質を MD 計算するための粗視化モデルの一つであり、本研究において扱っている HP モデルもこの郷モデルのポテンシャルを転用し使用している。

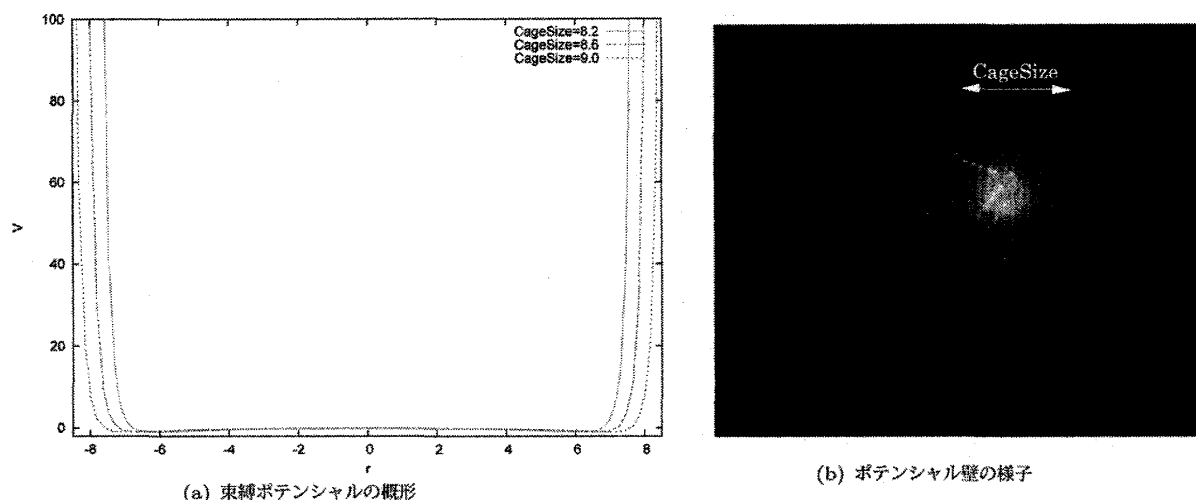


fig.22 : Confined 系とは

このような試行錯誤の結果、束縛ポテンシャルによるフォールディングへの影響が現れ、転移温度が上昇する井戸型ポテンシャルの半径が特定された。CageSize = 8.5 付近にそのような振舞いを起こす値があり、8.0 ~ 9.0 の範囲で期待する効果が現れる。この様子について、次節にて詳しく紹介する。

## 5.2 Confined 系の REMD による解析

Confined 系での温度- $R_g^2$  のプロットを REMD により得た結果を fig.23(a) に、温度- $\Sigma\theta$  のプロットを fig.23(b) に示す。4.4 節で得た、束縛のない場合での REMD の結果も同時にプロットしている。行った REMD の条件は、束縛ポテンシャルが導入されている点を除いて、4 章で行ったシミュレーションと同じであり、時系列の長さ、算出根拠とする時系列範囲、交換頻度等も含めすべて同じである。

fig.23 を見ると、CageSize により温度に伴う変化が異なっている。特に  $R_g^2$  は、 $T > 0.6$  の領域に関しては CageSize と対応して順に平均値が下がっていつている。これは、高温域でアンフォールド状態となっている際に束縛を受けていると、ポテンシャル壁との衝突により残基の運動が抑えられ、結果として  $R_g^2$  の平均が下がるためであると考えられる。一方で低温域においては順が崩れており、CageSize = 8.2 のプロットが上となっている。これは、フォールド相においては準安定構造の頻度分布に CageSize により違いが出るためであると考えられる。その様子は  $\Sigma\theta$  のプロットにもよく現れており、CageSize  $\geq 8.4$  では束縛のない状態と概形は同様であるが、CageSize = 8.2 の場合では、極大をとる温度やその値が他の場合と比べて大きく異なっている。

REMD の結果、系のエネルギー  $E$  の時系列平均は fig.24 に示すような値をとった。3.1 節で先送りしていた、束縛のない系での結果も示している。大まかにはエネルギー  $E$  の平均は温度と比例関係にあり、低温においても同様に比例関係にあることが REMD によりわかった。

fig.25 には、 $R_g^2$  と  $\Sigma\theta$  の時系列頻度の温度に伴う変化を示すカラーマップを示した。また、fig.26~27 には、温度ごとに  $R_g^2$ - $\Sigma\theta$  の二次元ヒストグラムを描いたものを、CageSize を変えて示した。ここでは特に CageSize 方向での変化について注目したい。

fig.25 のカラーマップ中のパターンには、特に  $T < 0.6$  の低温域において、CageSize により変化が見られる。安定構造の出現はカラーマップ中のピーク（色が明るい部分）として認識できる。

fig.26~27 にはパッチ状のパターンが示されているが、このそれぞれのパッチが安定構造に対応している。それぞれのパッチに「A」、「B」、「C」、「D」という番号を振る。CageSize の変化によりそれぞれのパッチ

がどう変化するか、パッチの番号ごとに説明する：

- ・ A :  $(R_g^2, \Sigma\theta) = (25, 24)$  付近に現れる構造である。CageSize = 8.2 (fig.27a) では  $T = 0.4$  付近から現れ始めるが、CageSize = 8.5, 8.6 (fig.26e,g) ではより低温の  $T = 0.2$  付近から現れており、しかし束縛の緩い状態 (fig.26k) となると再びごく低温域では現れなくなり、 $T = 0.4$  からの頻度が高くなるようになる。
- ・ B :  $(R_g^2, \Sigma\theta) = (26.5, 26)$  付近に現れる構造である。CageSize が 8.6 以上 (fig.26i,k) において観察された。 $T = 0.1$  付近から現れ、高温になると消失する。 $R_g^2$  が大きいいためか、束縛が緩いほど安定になる傾向があり、束縛のない状態では  $T = 0.6$  の場合でも頻度が高くなっている (fig.26x)。
- ・ C :  $(R_g^2, \Sigma\theta) = (25.5, 25.5)$  付近に現れる構造である。B とは逆に、 $R_g^2$  が小さいためか束縛が強いと安定になる傾向がある。CageSize = 8.2 では  $T = 1.4$  となってもまだパッチとして残存している。束縛が強い場合では広い温度領域で見られるが、束縛のない状態では  $T \geq 0.6$  の低温域で現れるのみで (fig.26v)、高温になると全く現れなくなる。
- ・ D :  $(R_g^2, \Sigma\theta) = (26, 25.5)$  付近に現れる構造である。CageSize = 8.2 ~ 8.5、 $T \geq 0.45$  の範囲で現れ、束縛が強い場合により安定となる傾向がある。高温になると消失する。

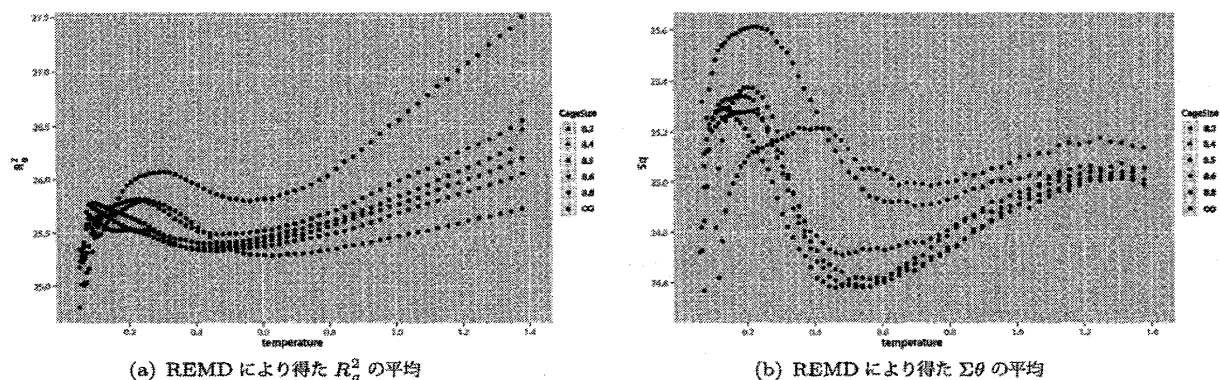
それぞれのパッチがどのような構造に対応するのか、次節において説明する。

### 5.3 Confined 系における安定構造

前節において示したパッチは、それぞれどのような構造に対応するのであろうか。CageSize = 8.5 の場合に注目し、4.7 節で行った説明と同様に、グラフィックス・ツールを用いて可視化したものを fig.28 ~ 30 に示す。fig.28 に示した図では、 $T=0.3$  におけるそれぞれの安定構造を示している。3-D 画像によるスナップショット、そして結合角  $\theta$  の折れ線グラフから、4.7 節の束縛のない状態でラベルした A と C の構造は対応していると考えられる。束縛のない状態では現れていた B の構造が無くなっており、C の構造がより安定となり、また、束縛のない状態ではほとんど観察されなかった D や E の構造が現れている。この他にも、頻度の低い多数の状態が観察されたが、ここでは割愛する。

D の構造は、束縛のない状態で現れていた B の構造とよく似るが、B に比べ、残基番号 1 側の末端が残基 8 番側とより離れる形となっており、15 番側の末端における結合角もパターンが違っている。

E の構造は、現れる頻度は低く、高温になると消失する。C の構造に見られる口金のような部分が互い違



**fig.23** : 各 Shape Parameter 平均値の CageSize による変化 CageSize の異なる束縛系の REMD シミュレーションを  $t = 5.0 \times 10^6$  まで行い、 $t = 1.0 \times 10^6 \sim 5.0 \times 10^6$  までの時系列を用いて  $R_g^2$ (a)、 $\Sigma\theta$ (b) の平均を算出し温度ごとにプロットした。点の配色が CageSize の違いを示している。

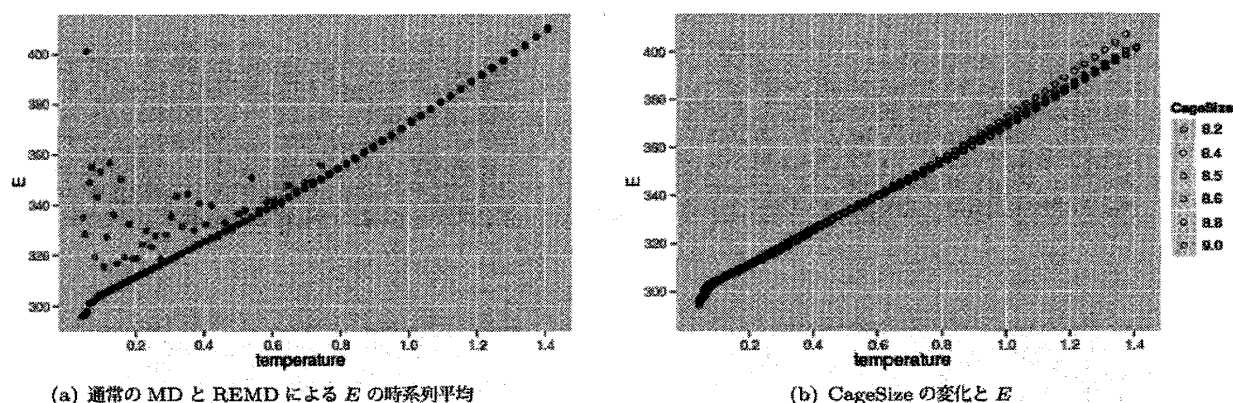


fig.24 : 各シミュレーションによるエネルギー  $E$  の時系列平均 (a) 束縛のない状態で、通常の MD シミュレーションの場合および REMD を用いた場合で算出される  $E$  を温度ごとにプロットした。橙色の点が通常の MD、紫色の点が REMD による結果を示す。 (b) CageSize を変えて REMD により  $E$  の時系列平均を算出し温度ごとにプロットした。配色の違いが CageSize の違いを表す。CageSize による変化はほとんど見られない。

いになったような構造を持っている。

fig.29 では  $T = 0.6$  の場合を、fig.30 では  $T = 1.2$  の場合を示しているが、温度上昇に伴い A と C の構造のみが現れるようになる。束縛のない状態と異なり、 $T = 1.2$  では結合角の折れ線グラフにまとまりがあり、アンフォールド状態であっても時系列中の自己相関が比較的高くなっていることを伺わせる。次節においては、この自己相関係数による議論を行う。

## 5.4 自己相関関数のピークの変化

先行研究によれば、束縛条件を厳しくするほど転移温度は上がるという。そのような振舞いが本モデルでも観察されるだろうか。4.5 節で行ったシミュレーションを Confined 系においても同様に行った。CageSize を変えてシミュレーションを行い、得られた結果を示した図が fig.31 である。(a) は束縛のない状態、(b) は CageSize = 8.5 の状態、(c) は CageSize = 8.2 の状態で得た自己相関関数のカラーマップであるが、束縛のある状態ではピークとなる赤色の領域が高温側に移動していることがわかる。しかし束縛の条件を強めていくとよりピークが高温側へ移るかというところでもなく、CageSize = 8.2 の場合ではピークは  $T = 0.7$  付近に出るが (a)、CageSize = 8.5 の場合では  $T = 1.0$  付近に移っており (b)、しかし束縛をなくすと  $T = 0.6$  付近となる (c)。先行研究で報告されている内容と同様に、束縛により転移温度が上昇したとも捉えられるが、 $\tau = 20000$  として温度-自己相関関数のプロットを行うと fig.34(d) のようになり、 $T = 0.6$  付近から自己相関関数に揺らぎが生じ始めることはどの CageSize においても変わらない。フォールディング転移温度の判定はこれでは困難なので、判定は行わないことにした。だが、自己相関関数のピークが、適当な CageSize をとることにより大きく高温側へ移る現象が観察されたことには注意を払いたい。

## 5.5 閉じ込めによる高速なフォールディングの発見

ガラス転移点以下で MD シミュレーションを行うと、系が準安定状態に長く滞留するという現象が起きるが、Confined 系において MD シミュレーションを行った際、CageSize を適切にとることで最安定状態に系が落ち着く場合が多くなるという現象が観察されたので報告したい。この結果は、「多数の準安定状態を持つタンパク質がなぜ短い時間の中で最安定構造へフォールディングするのか」という疑問に一つの回答の可能性を提示するものであると考えている。

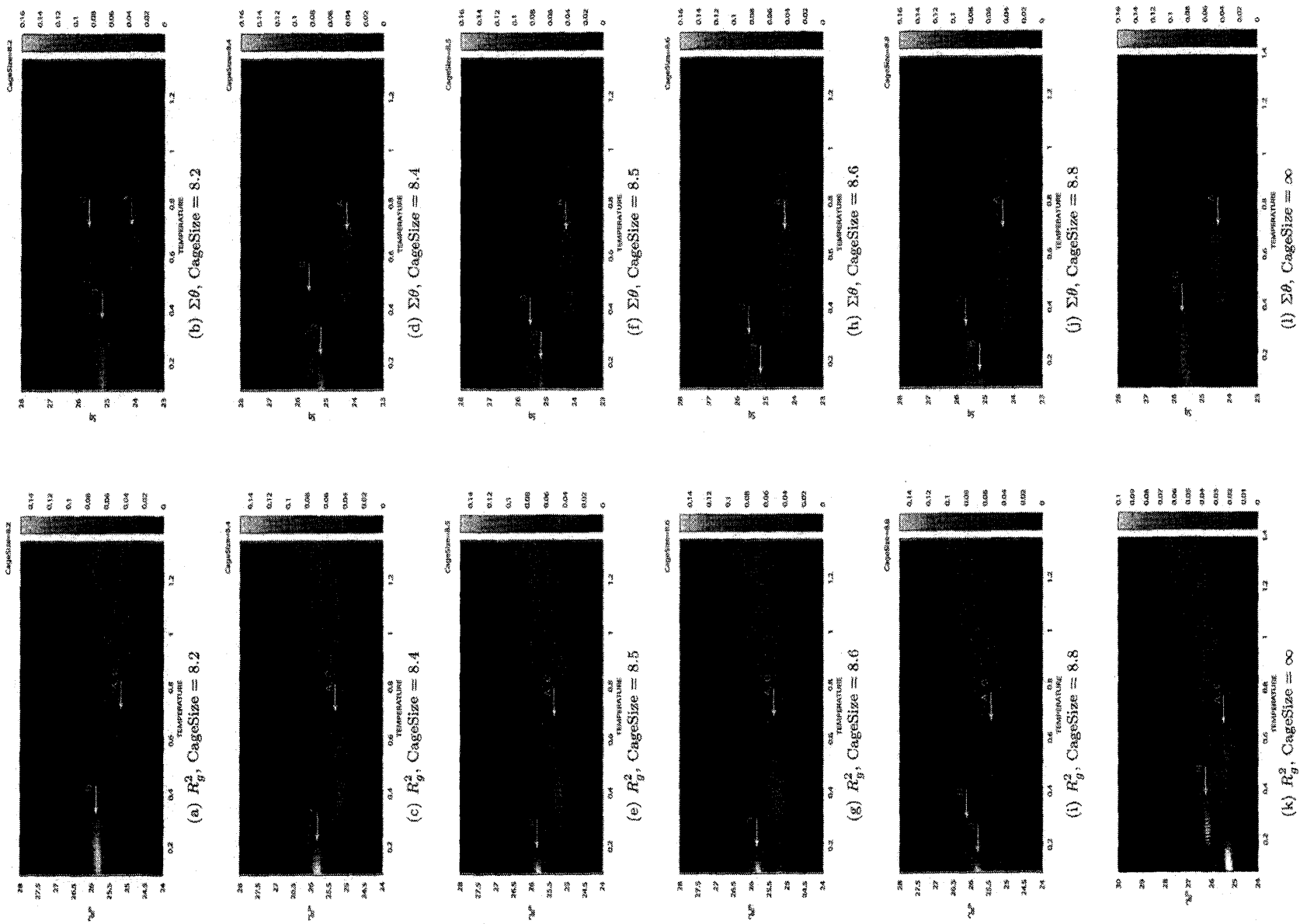


fig.25  $R_g^2$ ,  $\Sigma\theta$  の時系列頻度を表すカラーマップ 各 Shape Parameter について, REMD の時系列中に現れる値の頻度をカラーマップによるヒストグラムとして表した. 縦軸にそれぞれの Shape Parameter の値, 横軸に温度をとっている. 左側に  $R_g^2$  のカラーマップを, 右側に  $\Sigma\theta$  のカラーマップを配置させている.

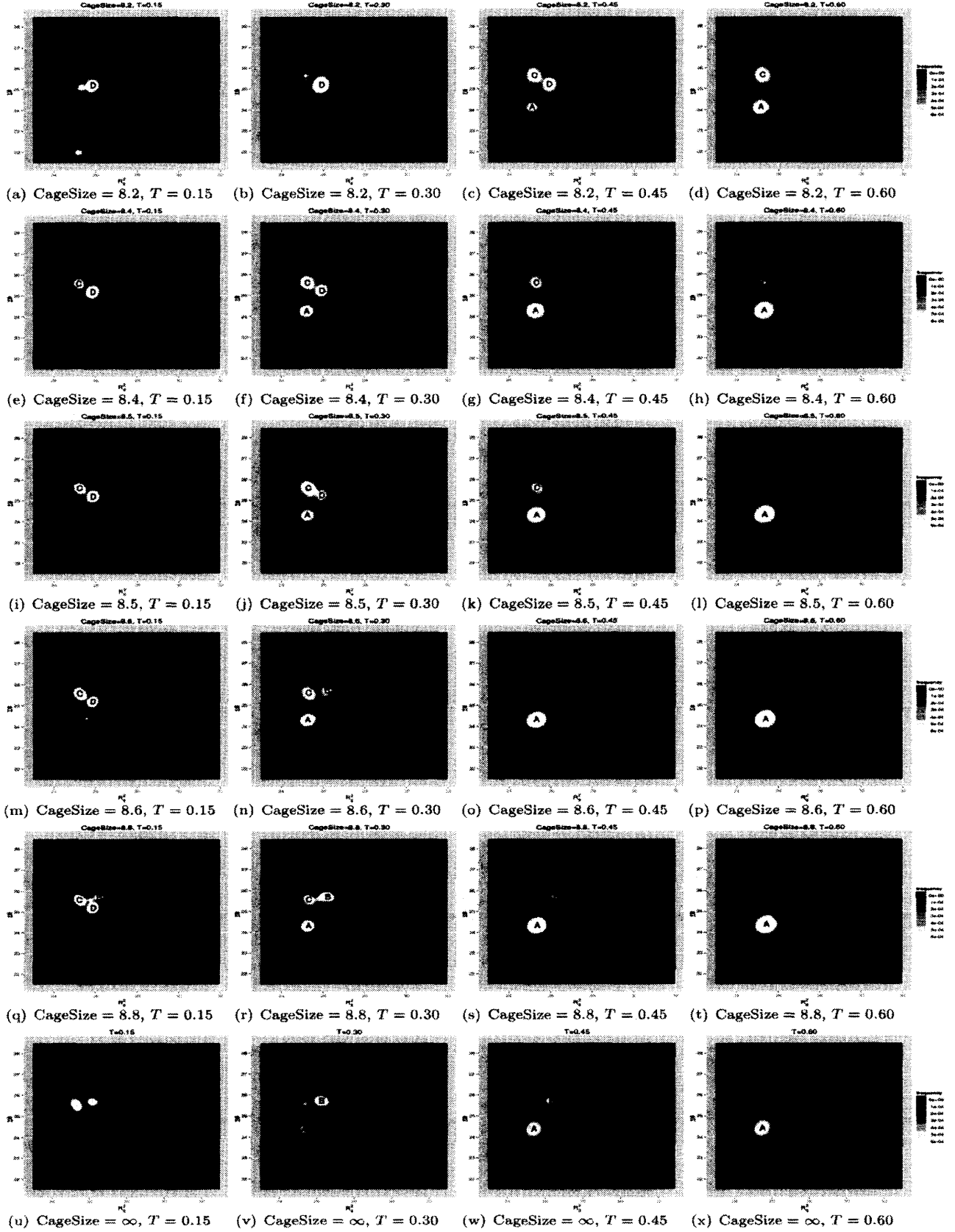


fig.26 : REMD より得た Confined 系の  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラム (I) CageSize = 8.2, 8.4, 8.5, 8.6, 8.8,  $\infty$  の場合に関し,  $t = 5.0 \times 10^6$  の長さの REMD シミュレーションを行い,  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラムを得, カラーマップで示した. 配色が時系列頻度を表す. ここでは  $T = 0.15, 0.30, 0.45, 0.60$  の温度で得たヒストグラムを示している. 左から右へ向かい高温となっており, 上から下へ向かい束縛が緩い方向に並んでいる.

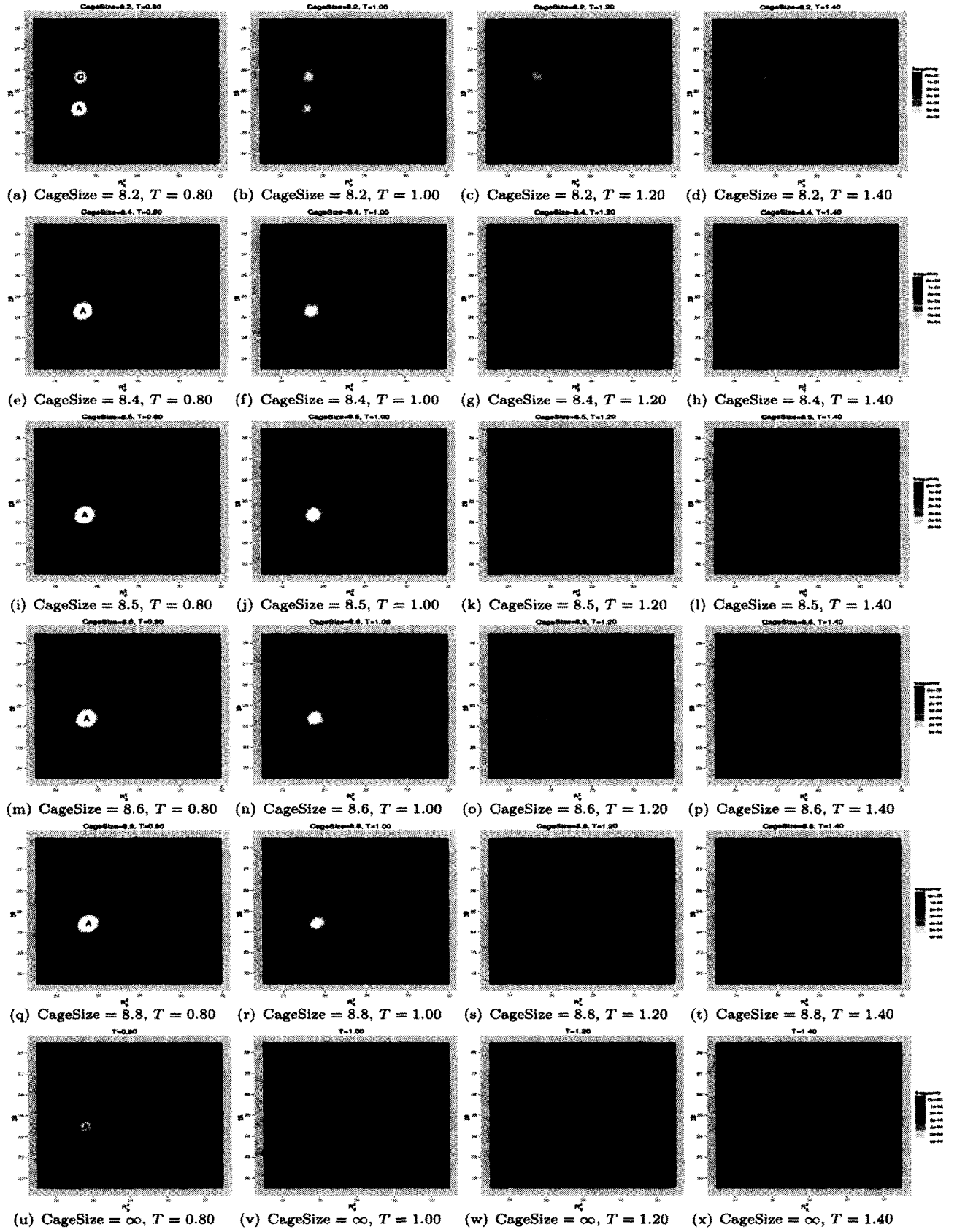


fig.27 : REMD より得た Confined 系の  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラム (II) CageSize = 8.2, 8.4, 8.5, 8.6, 8.8,  $\infty$  の場合に関し,  $t = 5.0 \times 10^6$  の長さの REMD シミュレーションを行い,  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラムを得, カラーマップで示した. 配色が時系列頻度を表す. ここでは  $T = 0.80, 1.00, 1.20, 1.40$  の温度で得たヒストグラムを示している. 左から右へ向かい高温となっており, 上から下へ向かい束縛が緩い方向に並んでいる.

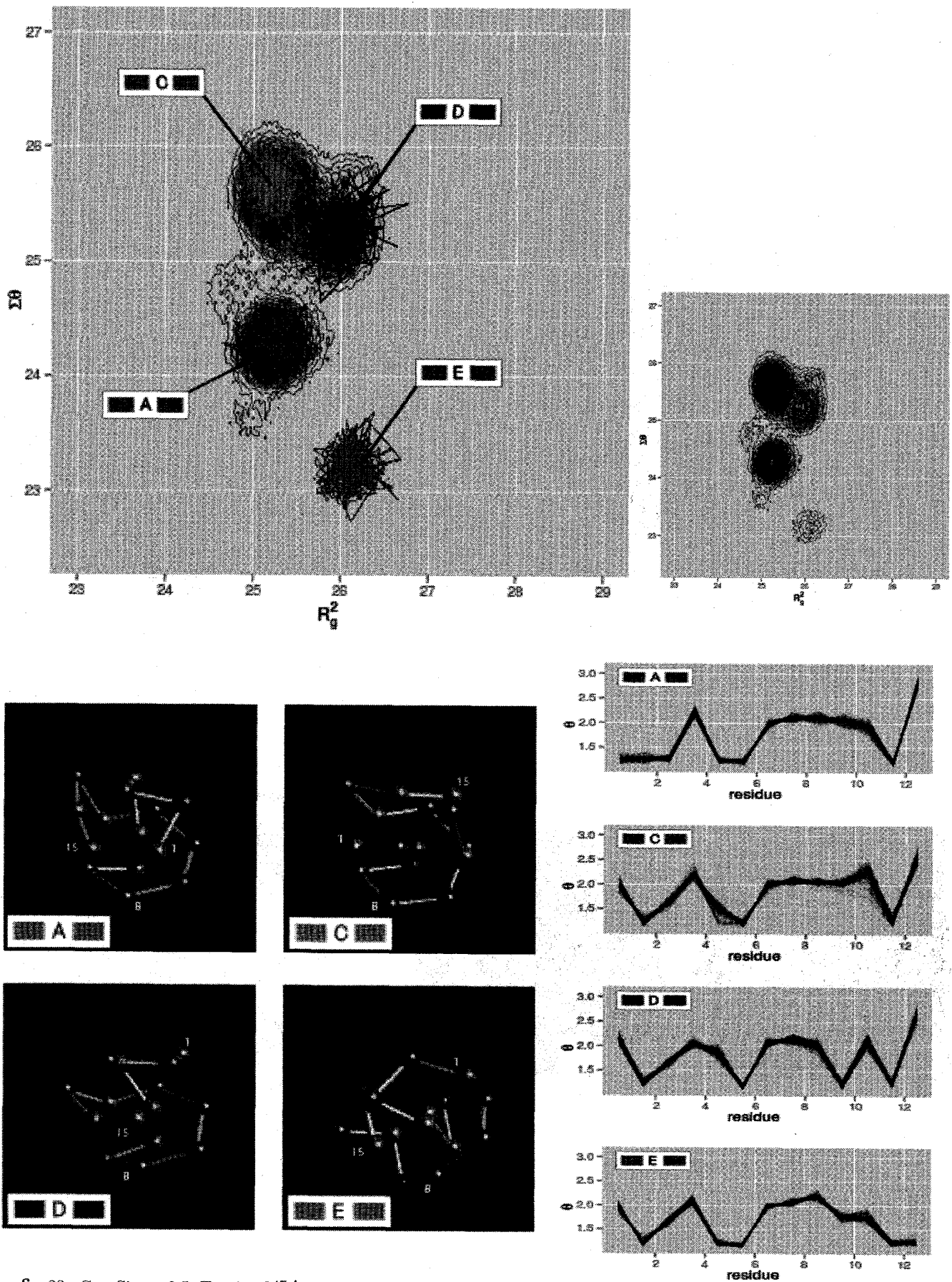


fig.28 :  $\text{CageSize} = 8.5$ ,  $T = 0.3$  の場合

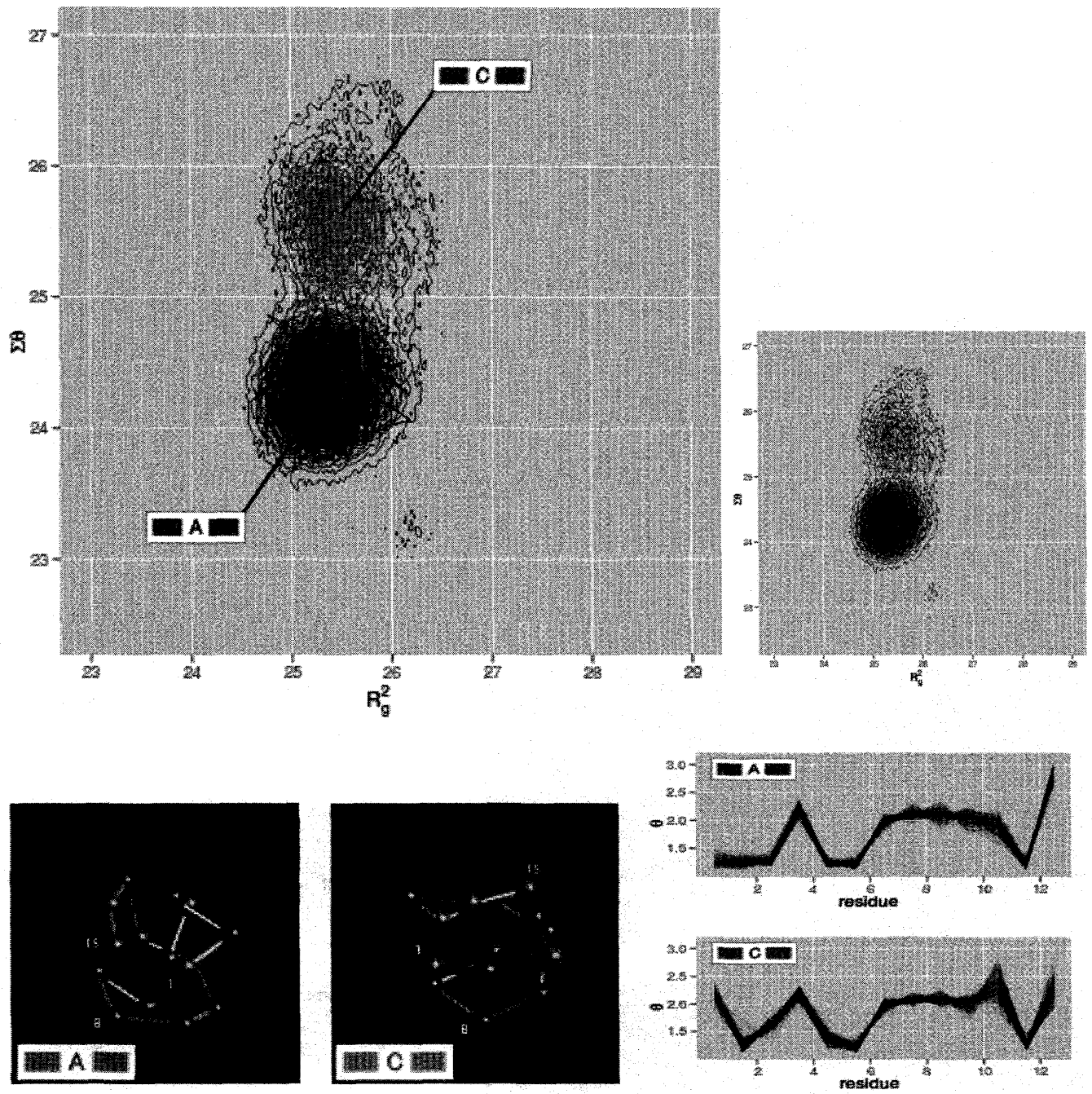


fig.29 : CageSize = 8.5,  $T = 0.6$  の場合

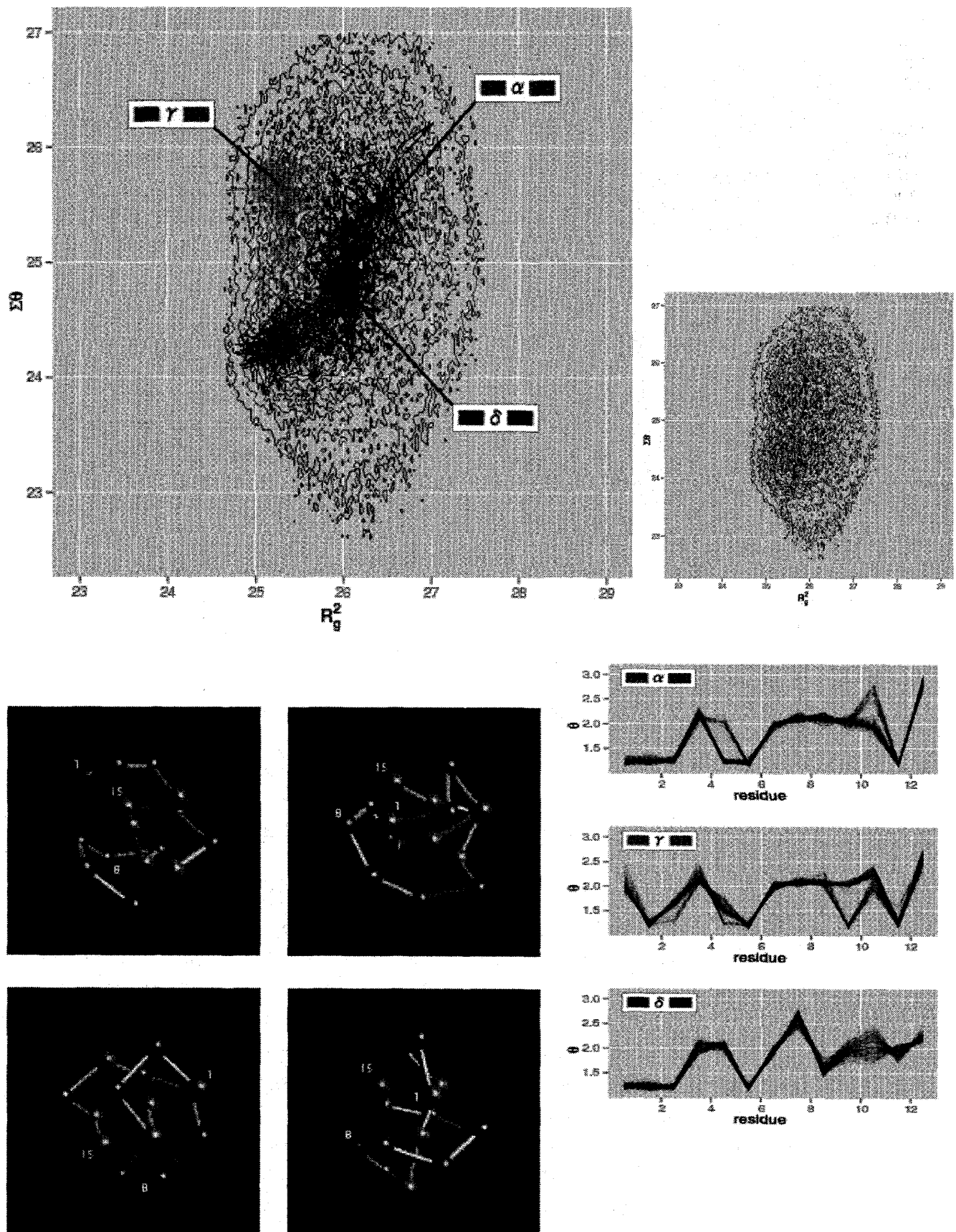


fig.30 :  $\text{CageSize} = 8.5, T = 1.2$  の場合

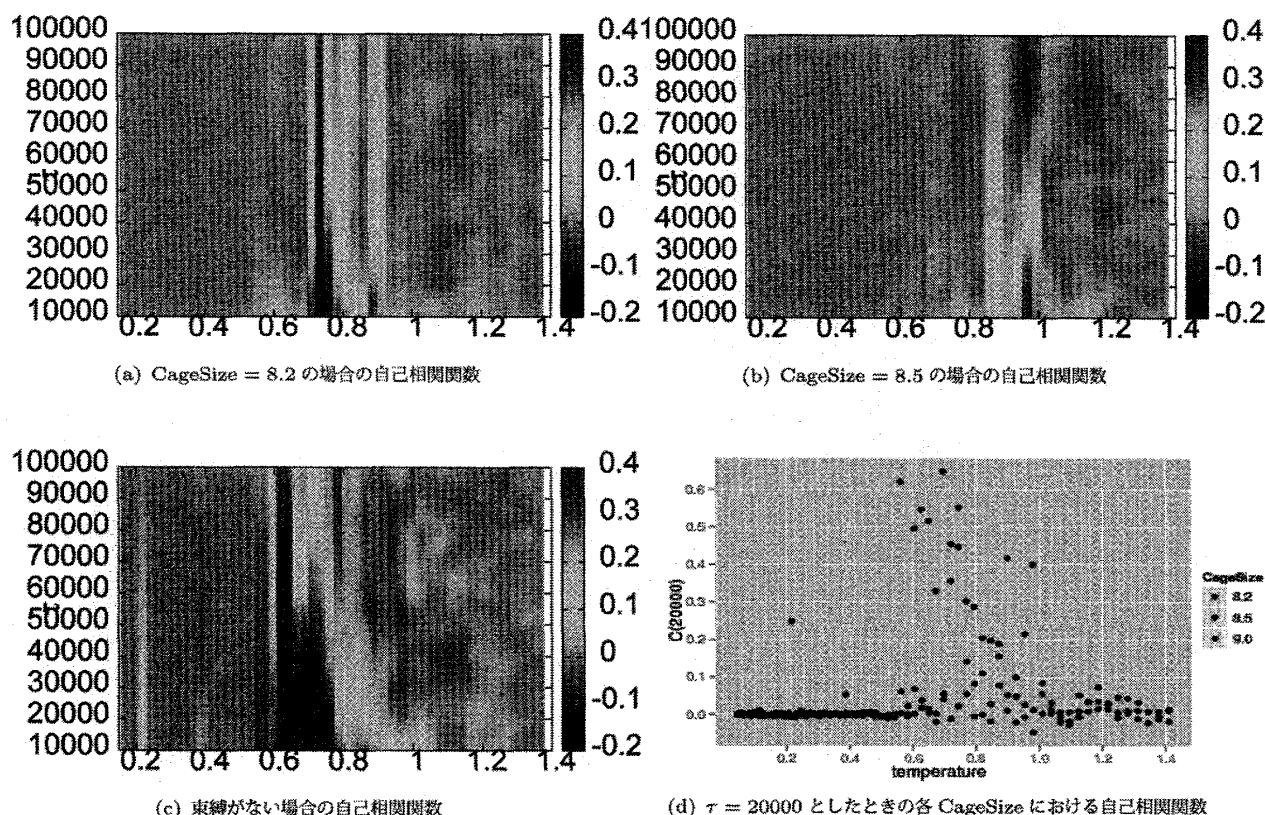


fig.31 : 各 CageSize で測った自己相関関数 (a) CageSize = 8.2 の場合の自己相関関数の様子をカラーマップにより示した. 縦軸に時間の違い  $\tau$  を, 横軸に温度をとり, 自己相関関数の値を敗色で示した. (b) (a) と同様に CageSize = 8.5 の場合の自己相関関数の様子をカラーマップにより示した. (c) (a), (b) と同様に束縛のない場合の自己相関関数の様子をカラーマップにより示した. 図は fig.15(a) と同一である. (d)  $\tau = 20000$  として切ったときの各 CageSize での自己相関関数の様子を示した. 配色が CageSize の違いを表している.

CageSize = 8.2, 8.5、そして束縛のない状態について、 $t = 1.0 \times 10^7$  までの MD シミュレーションを行った。この時系列の長さは 3.1 節で行ったシミュレーションの 2 倍の長さである。最後の  $1.0 \times 10^6 t$  の時系列を用いて各 Shape Parameter の時系列平均を温度ごとに算出しようとしたところ、fig.32 に示すような結果が得られた。特に CageSize = 8.5 の場合に注目すると、 $T = 0.7$  付近において、時系列平均がほとんど揺らなくなる。一方で他の CageSize の場合では揺らぎが大きい。本稿で求める、最安定状態への高速なフォールディングが示されたと期待される。

$T = 0.7$  の場合について、先と同じく各 CageSize において  $t = 1.0 \times 10^7$  までの MD シミュレーションを行い、これを 50 回試行として繰り返した。 $R_g^2$ 、 $\Sigma\theta$  それぞれの平均値を最後の  $1.0 \times 10^6 t$  の時系列を用いて算出し、REMD の結果から得た頻度分布の等高線図の上に二次元プロットしたものを fig.33 に示した。この図を見ると、CageSize = 8.5 の場合では他の CageSize の場合に比べ、ドットの収束が顕著に観察できる。

この  $1.0 \times 10^6 t$  の時系列中では、それぞれの CageSize においてどのような変化が起きているのであろうか。束縛のない状態ではフォールディング転移温度以上となっているため、ガラス的な振舞いが示されると予想される。 $R_g^2$ - $\Sigma\theta$  二次元ヒストグラムを描くことにより様子を調べることにした。

各 CageSize において、この  $1.0 \times 10^6 t$  の  $R_g^2$ 、 $\Sigma\theta$  の時系列プロットを得たものを fig.34 に、また、この時系列から  $R_g^2$ - $\Sigma\theta$  二次元ヒストグラムを得た結果を fig.35 に示した。この図から、CageSize = 8.2、そして束縛のない状態においては広い範囲で状態間遷移がこの時間幅の中で起きていることがわかる。一方で CageSize = 8.5 の場合ではほとんど状態間遷移が起こらない。一見して、フォールド相にあるように見える。

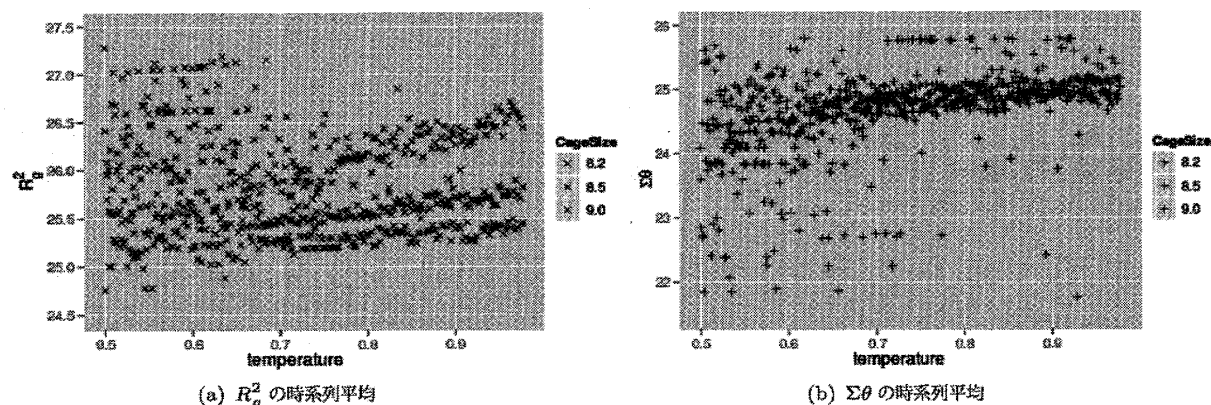


fig.32 : 長時間シミュレーションによる各 Shape Parameter の時系列平均 CageSize = 8.2, 8.5, そして束縛のない状態につき  $t = 1.0 \times 10^7 t$  までの MD シミュレーションを行い, 最後の時間  $1.0 \times 10^6 t$  の時系列を使い  $R_g^2(a)$ ,  $\Sigma\theta(b)$  の時系列平均を算出して温度ごとにプロットした.  $T = 0.7$  付近に注目すると, 特に CageSize = 8.5 のとき, 平均に収束が見られる.

そして注目すべきは、 $T < 0.6$  の低温域の場合と違い、ほとんどの試行で構造 A に再現性よく収束しているということである（フォールド相では別々の状態に落ち着いてしまう）。

本稿の主たるテーマは、天然のタンパク質が示すような最安定状態への速い到達がいかんして実現されているのか、という問いに、それが混雑効果により起こる場合があるということをモデル研究により示すことである。本稿で扱うヘテロポリマー鎖がフォールディングを行う場合、複雑な自由エネルギー地形を持つために立体構造がガラス的となり、実際、通常の MD シミュレーションでは転移温度付近であっても一つの状態へ落ち着くことはなかった。一方で REMD 計算の結果から、その温度では少数個の大きな谷を持つ自由エネルギー地形を持つことがわかっており、これはつまり、タンパク質の高精度シミュレーションで障害となる、準安定状態への滞留により最安定な天然状態へフォールドしないというジレンマを再現したものと言える。しかし束縛ポテンシャルを与えることにより、最安定構造へ短いシミュレーション時間内に到達し、そして状態は揃うということがここに示された。

ここで、なぜこのような振舞いが起こるのか考えてみたい。設定した束縛ポテンシャルへの閉じ込めがモデル鎖にどのような影響を与えるのかについてであるが、この操作は、 $R_g^2$  が大きい準安定状態の自由エネルギーを上げると考えられる。いま注目する系では、 $R_g^2$  が大きい構造は熱平衡状態においてレアイベントとなっており、REMD による解析ではほとんど拾えなかった構造である。しかし、 $T = 0.7$  の束縛のない系でガラス状態となったときに観察される構造はそのようなレアイベントである場合が多く、これらの構造が束縛ポテンシャルへの閉じ込めによりどのように自由エネルギーの変化を示すかが注目される。REMD による解析では FEL 中の特に深い状態をサンプリングするのみで、FEL の浅い領域についてはサンプリングできていない。FEL の浅い領域についてサンプリングする場合には、REMD では長時間のシミュレーションを行わない限り難しい。束縛によりレアイベントとなる構造の自由エネルギーが増すかどうか調べるため、次

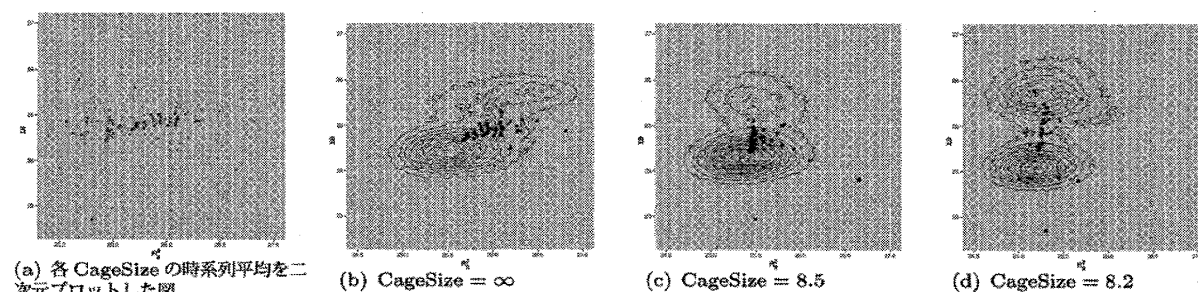


fig.33 : 長時間シミュレーションによる各 Shape Parameter の時系列平均の二次元プロット

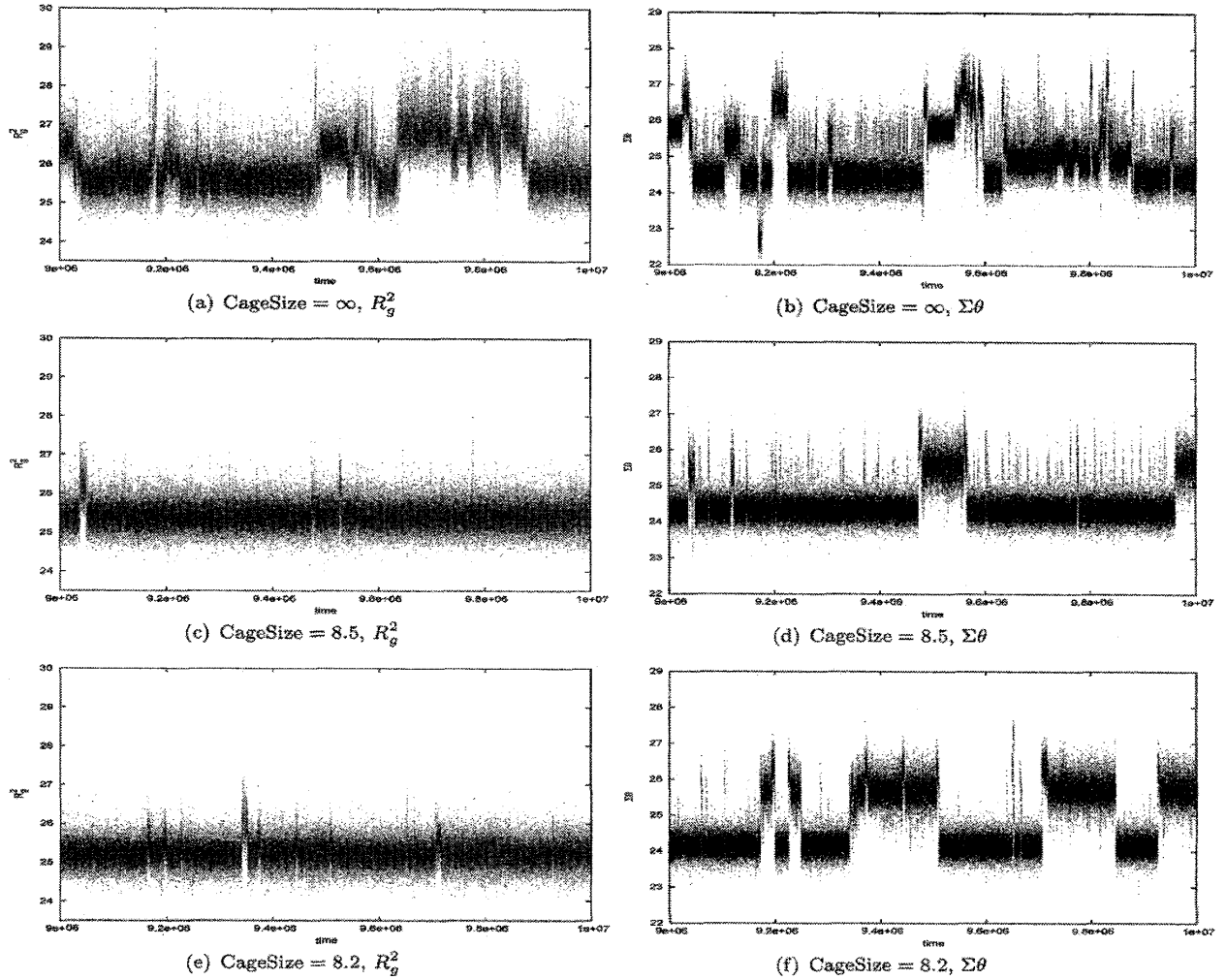


fig.34 : 行った長時間のシミュレーションにおいて得られた時系列

章で述べる解析を行った。

## 6 自由エネルギー地形による議論

5.5 節において、束縛ポテンシャルへの閉じ込めにより高速なフォールディング過程が観察されたことを報告したが、この変化の原因は何であろうか。レイイベントとなる構造の自由エネルギーが増すという FEL の変化が閉じ込めにより起きたと考え、改変したアンブレラサンプリング法により解析を行った。本章ではその結果について報告する。

### 6.1 アンブレラサンプリング法

まずは採用した解析方法について確認を行う。アンブレラサンプリング法は、MD 計算やモンテカルロシミュレーションを行う場合に設定するハミルトニアンに、アンブレラ・ポテンシャルと呼ばれる関数を追加して、そのハミルトニアンを基に仮想的なシミュレーションを行い、得られた時系列データを統計力学に基づいた方法で適切に変換することで、アンブレラ・ポテンシャルを除いた元のポテンシャルを持つ場合の統計平均を算出するというアルゴリズムである。以下でアンブレラサンプリング法のアルゴリズムを確認して

おく [21]。

座標、運動量をそれぞれ  $\mathbf{q}$ 、 $\mathbf{p}$  として、ハミルトニアン  $\mathcal{H}(\mathbf{q}, \mathbf{p})$  を持つ系の分配関数  $Z_{\mathcal{H}}$  は次で表される：

$$Z_{\mathcal{H}} = \int e^{-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \quad (24)$$

これにより、系の中に表れる物理量  $A(\mathbf{q}, \mathbf{p})$  の統計平均は次で表される：

$$\langle A(\mathbf{q}, \mathbf{p}) \rangle_{\mathcal{H}} = \frac{1}{Z_{\mathcal{H}}} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \quad (25)$$

この積分、すなわち全状態のサンプリングであるが、自由度が高い系においては実質上は実行がほぼ不可能である。よって、通常は MD シミュレーションやメトロポリス法などにより、分配関数と同じ確率分布に従い状態のサンプリングを行う手法を通じて統計平均の近似値を算出する。

アンブレラサンプリング法では、ハミルトニアン  $\mathcal{H}(\mathbf{q}, \mathbf{p})$  に更にアンブレラ・ポテンシャル  $V_U(\mathbf{q})$  を加え、シミュレーションを行う。アンブレラ・ポテンシャルによりバイアスを受けた系について物理量  $A(\mathbf{q}, \mathbf{p})$  の統計平均  $\langle A(\mathbf{q}, \mathbf{p}) \rangle_W$  を求めるには次のようにする：

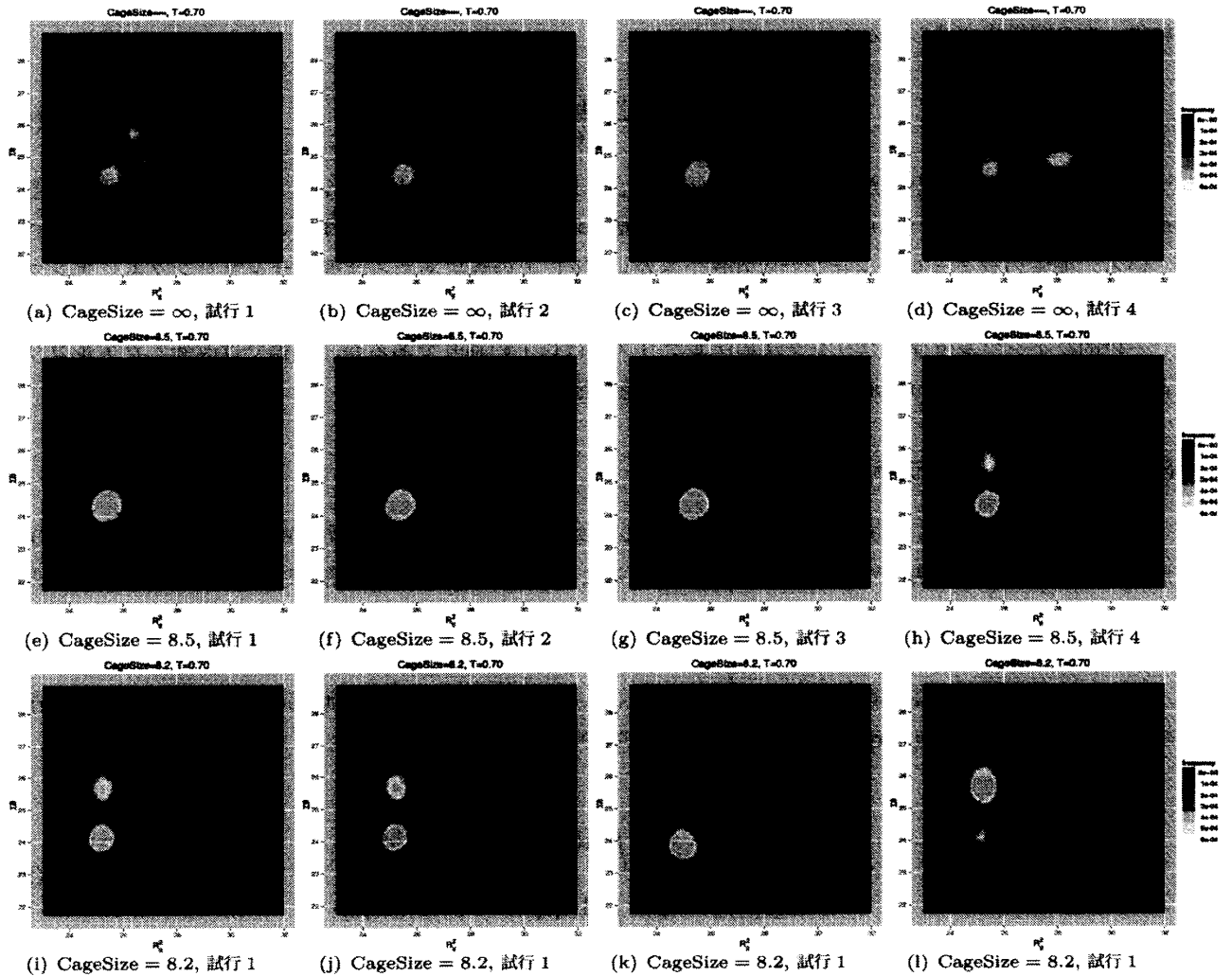


fig.35：長時間シミュレーションによる二次元ヒストグラムの標本

$$\langle A(\mathbf{q}, \mathbf{p}) \rangle_{\mathcal{W}} = \frac{1}{Z_{\mathcal{W}}} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta \mathcal{W}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \quad (26)$$

ここで、 $\mathcal{W}(\mathbf{q}, \mathbf{p}) = \mathcal{H}(\mathbf{q}, \mathbf{p}) + V_u(\mathbf{q})$  とした。 $Z_{\mathcal{W}}$  とはハミルトニアンが  $\mathcal{W}(\mathbf{q}, \mathbf{p})$  である場合の分配関数である。

MD シミュレーションやメトロポリス法などにより物理量  $A(\mathbf{q}, \mathbf{p})$  の統計平均の近似値を求める場合、サンプリングアルゴリズムには元のハミルトニアン  $\mathcal{H}(\mathbf{q}, \mathbf{p})$  にアンブレラ・ポテンシャル  $V_u(\mathbf{q})$  を加えてシミュレーションを行う必要がある。その結果得た統計平均の近似値を  $\langle \rangle_{\text{sim}\mathcal{W}}$  とする。

次のような変形を行う：

$$\begin{aligned} \langle A(\mathbf{q}, \mathbf{p}) \rangle_{\mathcal{H}} &= \frac{1}{Z_{\mathcal{H}}} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \times \frac{Z_{\mathcal{W}}}{Z_{\mathcal{W}}} \\ &= \frac{Z_{\mathcal{W}}}{Z_{\mathcal{H}}} \frac{1}{Z_{\mathcal{W}}} \int A(\mathbf{q}, \mathbf{p}) e^{+\beta V_u(\mathbf{q})} e^{-\beta(\mathcal{H}(\mathbf{q}, \mathbf{p}) + V_u(\mathbf{q}))} d\mathbf{q} d\mathbf{p} \\ &= \frac{Z_{\mathcal{W}}}{Z_{\mathcal{H}}} \langle A(\mathbf{q}, \mathbf{p}) e^{+\beta V_u} \rangle_{\mathcal{W}} \end{aligned} \quad (27)$$

結果、次の関係式が得られる：

$$\langle A(\mathbf{q}, \mathbf{p}) \rangle_{\mathcal{H}} \simeq \frac{\langle A(\mathbf{q}, \mathbf{p}) e^{+\beta V_u(\mathbf{q})} \rangle_{\text{sim}\mathcal{W}}}{\langle e^{+\beta V_u(\mathbf{q})} \rangle_{\text{sim}\mathcal{W}}} \quad (28)$$

すなわち、バイアスをかけたハミルトニアン  $\mathcal{W}(\mathbf{q}, \mathbf{p})$  に振舞いを記述されるシミュレーションを行い、その結果を用いて系が元のハミルトニアン  $\mathcal{H}(\mathbf{q}, \mathbf{p})$  を持つ場合の統計平均を算出することができることが示されている。通常のシミュレーションでは到達できない FEL 上の峰にあたる状態のサンプリングも、アンブレラ・ポテンシャルを適切に設定することで行うことができる仕組みとなっている。

一般的なアンブレラサンプリング法のアルゴリズムは以上の通りであるが、本稿では改変したアルゴリズムを用いる。次節でそのアルゴリズムの導出・確認を行う。

## 6.2 温度アンブレラサンプリング法

前節ではアンブレラ・ポテンシャルを元のハミルトニアンに追加するアルゴリズムを示したが、式 (27) の形は、ハミルトニアン  $\mathcal{H}$  ではなく逆温度  $\beta$  を変えることによってでも同様の計算方法が導かれると考えられる。つまり、高温の状態で広いサンプリングを行わせ、低温の場合の統計平均を間接的に計算するという手法である。高い温度  $T_H$  と低い温度  $T_L$  を考え、それぞれの逆温度を  $\beta_H$ 、 $\beta_L$  とする。 $\Delta\beta = \beta_H - \beta_L$  とする。

温度  $T_H$ 、 $T_L$  における分配関数をそれぞれ  $Z_H$ 、 $Z_L$  とすると、それぞれの温度における物理量  $A(\mathbf{q}, \mathbf{p})$  の統計平均  $\langle A(\mathbf{q}, \mathbf{p}) \rangle_H$ 、 $\langle A(\mathbf{q}, \mathbf{p}) \rangle_L$  は、理想的には次の式により表される：

$$\begin{aligned} \langle A(\mathbf{q}, \mathbf{p}) \rangle_H &= \frac{1}{Z_H} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta_H \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \\ \langle A(\mathbf{q}, \mathbf{p}) \rangle_L &= \frac{1}{Z_L} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta_L \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \end{aligned} \quad (29)$$

式 (27) と同様の式変形を行う：

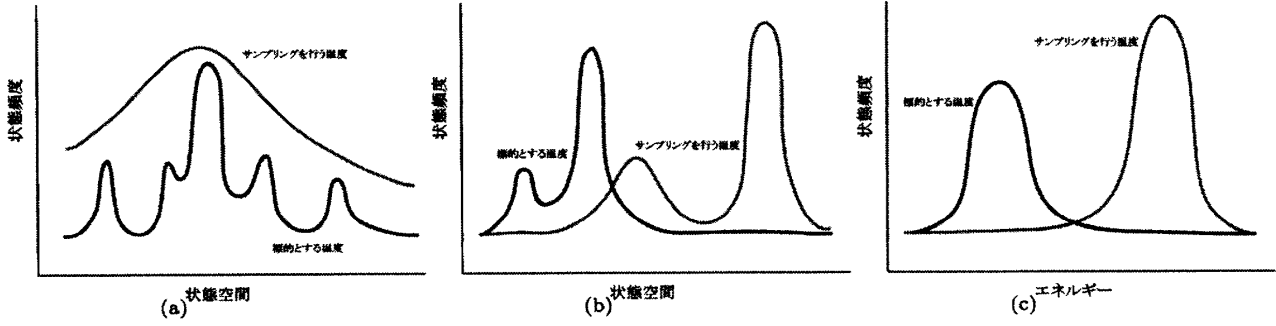


fig.36 : 温度アンブレラサンプリング法が有効に働く場合と働かない場合

$$\begin{aligned}
 \langle A(\mathbf{q}, \mathbf{p}) \rangle_L &= \frac{1}{Z_L} \int A(\mathbf{q}, \mathbf{p}) e^{-\beta_L \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \times \frac{Z_H}{Z_H} \\
 &= \frac{Z_H}{Z_H} \frac{1}{Z_L} \int A(\mathbf{q}, \mathbf{p}) e^{(-\beta_L + \beta_H) \cdot \mathcal{H}(\mathbf{q}, \mathbf{p})} e^{-\beta_H \mathcal{H}(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p} \\
 &= \frac{Z_H}{Z_L} \langle A(\mathbf{q}, \mathbf{p}) e^{+\Delta\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} \rangle_H
 \end{aligned} \tag{30}$$

結果、次の関係式が得られる：

$$\langle A(\mathbf{q}, \mathbf{p}) \rangle_L \simeq \frac{\langle A(\mathbf{q}, \mathbf{p}) e^{+\Delta\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} \rangle_{\text{sim}H}}{\langle e^{+\Delta\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} \rangle_{\text{sim}H}} \tag{31}$$

これにより、サンプリング効率の高い高温域でのシミュレーションにより、エネルギー障壁が多数生じシミュレーションが行えない低温域の統計平均を得る方法が示された。

この方法が有効に働く場合は限られている。

- ・  $\Delta\beta$  が大きすぎるとサンプリングする際の重みが全く異なるようになり、結果が収束しなくなるため、標的となる温度と実際にシミュレーションを行う温度とは近い必要がある
- ・ 高い温度でシミュレーションする場合、シミュレーション時間内に有効にサンプルに掛かる状態が低い温度の場合を包含している必要がある (fig.36(a), (b))
- ・ 粒子数が多い場合は系のそれぞれの状態の実現確率が低分子数の場合と大幅に変わり、サンプリングが困難になるため有効に働かない (fig.36(c))

注目する系は一分子であり、また、シミュレーション時間内に系が取りうる構造が増えるか減じるかという温度領域に注目し、アンフォールド相を利用することでガラス的振舞いを見せる相のサンプリングを行うということなので、上記の条件はクリアしていると考えた。この方法により FEL の浅い領域についてサンプリングした結果を次節にて示す。

### 6.3 シミュレーションの方法

前節で述べた方法により、CageSize = 8.2、8.5、そして束縛のない状態について温度アンブレラサンプリングシミュレーションを行った。実際に系を動かした温度は  $T = 0.85$  である。この温度は、ガラス転移点直下であると考えられ、3.1 節で行ったような短い MD シミュレーションによってもカノニカル平均が安定して得られる温度である。時系列の長さを  $t = 5.0 \times 10^7$  として、それぞれの CageSize について 10 回のサンプリングシミュレーションを行い、結果を平均して最終的に FEL を描く。(31) 式により温度を変換する際に

分配関数を計算するが、いずれの CageSize においても 10 回のサンプリングシミュレーションで得られた分配関数はそれぞれ 5% ほど揺らぐのみであったので、有効なサンプリングが行われているものと考えた。それぞれの構造について自由エネルギーを得たいが、 $R_g^2$  と  $\Sigma\theta$  のみでは構造同士で頻度分布が重なる領域が多いため、 $P_g^2$  も導入し、三次元の Shape Parameter 空間中で FEL を描くことにする ( $P_g^2$  の詳細は 2.5 節参照)。サンプリングされた結果はカノニカル分布によく従っているものと仮定し、三次元の Shape Parameter 空間中の格子それぞれについて次式により自由エネルギーを計算した。

$$G = G_0 + k_B T \ln(p/p_0) \quad (32)$$

ここで、 $G$  は自由エネルギーを、 $G_0$  は基準となる状態の自由エネルギーを、 $p$  は状態頻度を、 $p_0$  は規格化定数を意味する。温度アンブレラサンプリング法により得た Shape Parameter 空間中の状態頻度分布を上式により変換し、次節に示す結果を得た。

#### 6.4 自由エネルギー地形の概形

それぞれの CageSize について FEL を得た。その様子を fig.37~39 に示す。Shape Parameter を三つとったことで分布の重なりを回避できた領域は多く、特に自由エネルギーの低い構造はほとんど図中に示されていると判断する。FEL を描いた図の見方であるが、それぞれの頁で左上に描かれている Scatter Plot が三次元での FEL を示したものであり、配色によりそれぞれの座標での相対的な自由エネルギーを表している。各格子点で状態頻度が 1 以下となった場合は無色とした。Scatter Plot の右には  $R_g^2$ - $P_g^2$  の二次元による FEL、下には  $-\Sigma\theta$ - $P_g^2$  の二次元による FEL ( $\Sigma\theta$  はマイナスをとっていることに注意したい。図の見易さを優先しマイナスの値で FEL を示した)、右下には  $R_g^2$ - $\Sigma\theta$  の二次元による FEL を配置した。

それぞれの図中で示されている構造に、REMD で示された安定構造に加えて新たに名称をつけることにする。それぞれ X1~X6 とし、図中にラベルを加えた。これらレアイベントとなる構造の自由エネルギーは CageSize を狭くしていくに従い、X2 を除いて全くサンプリングされなくなる。すなわち、束縛ポテンシャルの影響により状態が大幅に減ったことがわかる。

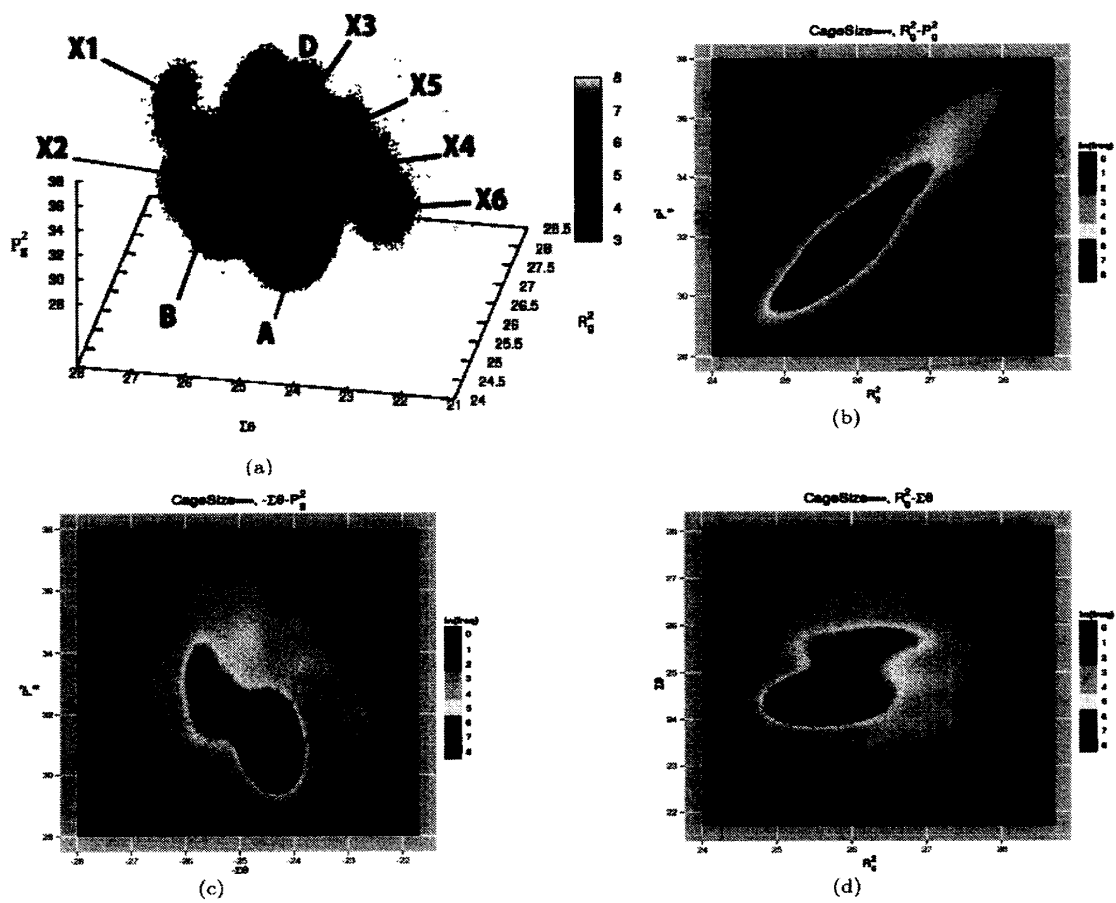


fig.37 : CageSize =  $\infty$

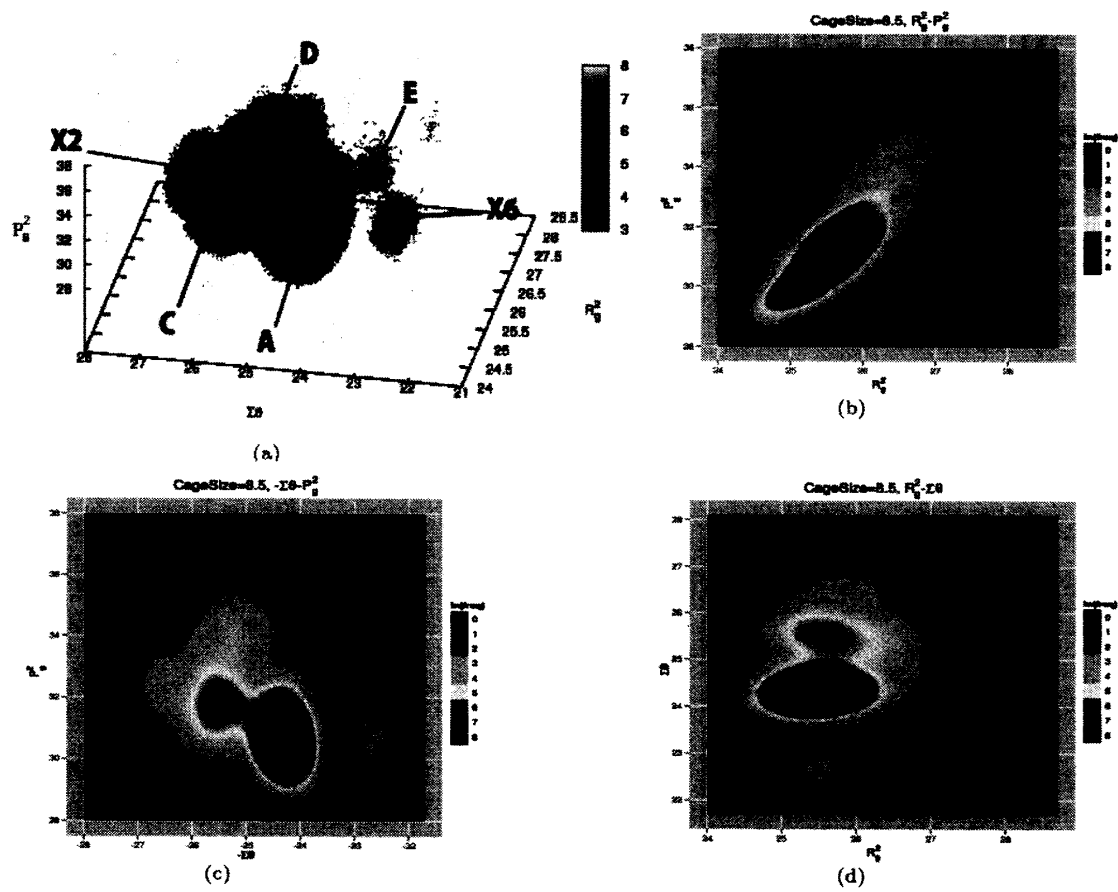


fig.38 :  $\text{CageSize} = 8.5$

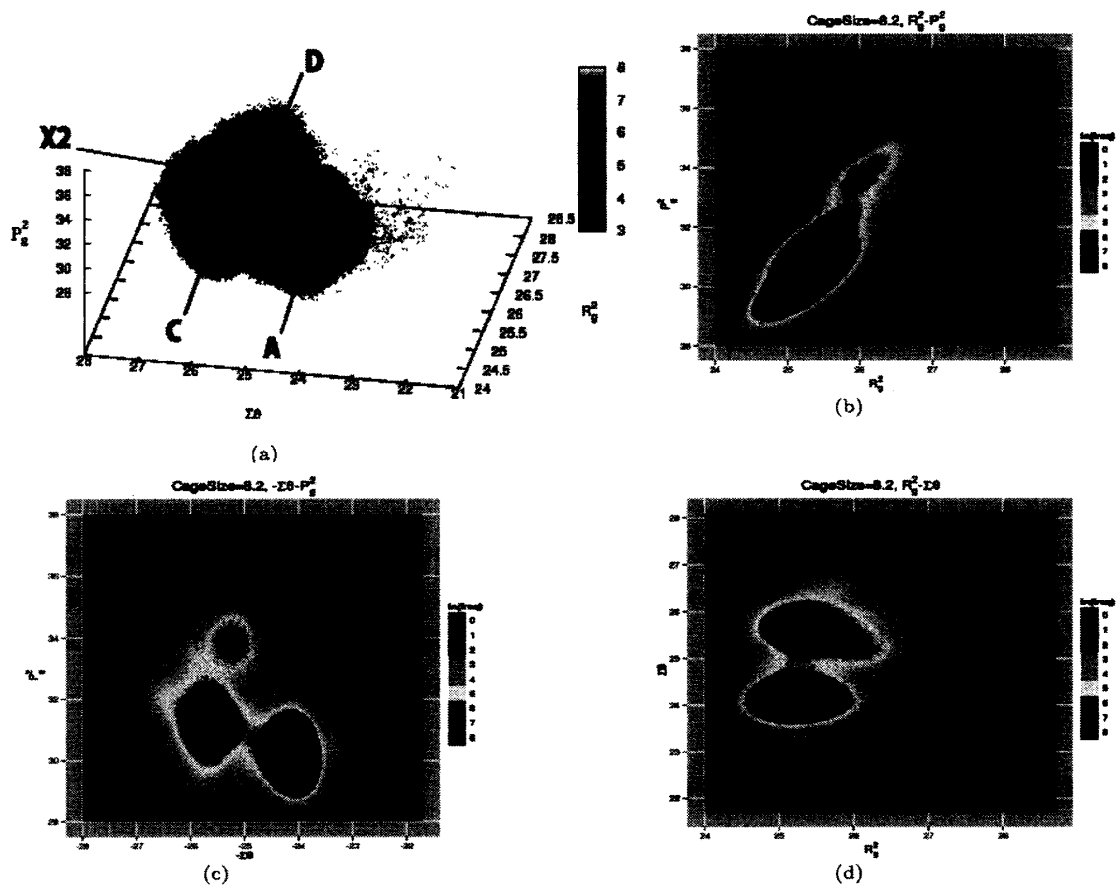


fig.39 : CageSize = 8.2

## 7 Discussion

### 7.1 示された結果のまとめ

本稿では、分子混雑がタンパク質のフォールディングを高速化させているという仮説を立て、モデル研究を通じて答えを模索した。得られた結果をここにまとめておく。

- ・ 3.1 節でフォールディングするモデル鎖を設定し、温度を変えて MD シミュレーションを行ったところ、ある温度以下では Shape Parameter やエネルギー  $E$  の時系列平均が収束しなくなった。これは最安定構造以外の様々な準安定構造に系の状態が落ち着き抜け出せなくなるために起こる。この温度をガラス転移温度  $T_g$  と考え、 $T_g = 0.8 \sim 1.0$  であると考えた。
- ・ REMD シミュレーションを行うと各物理量は収束した (4.4 節)。温度に伴う状態頻度分布の変化も調べることができた (4.6 節)。そして REMD と通常の MD を組み合わせた方法により自己相関関数の温度変化を調べたところ、自己相関関数はある温度で明確に転移を起こすことがわかった (4.7 節)。この温度以下では系は状態遷移を全く起こさなくなると考えられる。この転移点がフォールディング転移点  $T_f$  であると考え、 $T_f = 0.6$  であると判断した。つまり、研究対象とするモデル鎖の転移の様子は fig.40 の緑色部のように模式的に描かれる。 $T_{fg}$  の温度では、状態間遷移は起きるが無秩序ではないという状態になっている。
- ・ 束縛ポテンシャルを導入し、モデル鎖を閉じ込める操作を行った (5.1 節)。これは、細胞内の分子混雑を模擬したものと考えている。その結果、束縛ポテンシャルの半径を意味する CageSize の大きさにより状態頻度分布に変化が現れ (5.2 節)、更に自己相関関数の温度に伴う変化も様相が変わった (5.4 節)。
- ・ CageSize を 8.5 に設定し、 $T = 0.7$  で長時間のシミュレーションを行うと、数個の状態に系が落ち着きやすくなるという現象が観察された (5.5 節)。
- ・ CageSize = 8.5 の場合になぜそのような振舞いが生じるか調べるため、限定的な解析手段である温度アンブレラサンプリング法により FEL 中の浅い領域で起こる束縛に伴う変化について調べた (6.4 節)。結果、レアイベントと考えられる自由エネルギーの比較的高い構造が束縛ポテンシャルへの閉じ込めにより消される (自由エネルギーが高くなる) という振舞いが見られた。

### 7.2 高速なフォールディングはなぜ起こるか

フォールディング転移温度  $T_f$  が束縛ポテンシャルの導入により変化するかどうかは、5.4 節で示した自己相関関数の様子からは判定できなかった。閉じ込めを施された状態にあっても、 $T > 0.6$  の領域では自己相関関数に全く揺らぎがないということではないため、状態遷移は起こっているものと考えられる。ただ、自己相関関数にピークが示される温度が CageSize により変化する。高速なフォールディング過程が示された  $T = 0.7$  の領域では、注目した CageSize においてはいずれの場合もガラス状態にあると判断する。では、なぜ閉じ込めにより数個の状態へフォールディングが起こるのか、という疑問に対する答えを考える。モデル鎖がガラス状態にあるときに状態間遷移が起こる理由は、準安定状態間のエネルギー障壁を頻繁に越えられるほど揺動エネルギーが増したためである。しかしアンフォールド状態となるにはエネルギーが足りないためにフォールド状態の各構造を行き来することが繰り返されている状態にあると考えられる。この状態は、最安定構造の探索をモデル鎖が行う場合には都合がよい。一つの準安定構造に留まり続け、有限時間内に最安

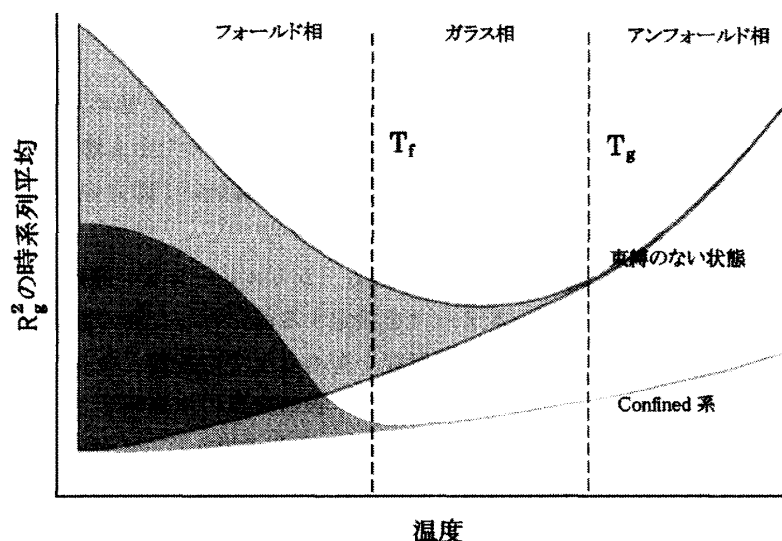


fig.40 : 系の転移の模式図

定となる A の構造に至る確率がほぼゼロとなる可能性が低まるからである。しかし、一方でこの状態間遷移は都合が悪く、なぜなら、最安定構造に至っても再度別の準安定状態へと転じる確率が高くなるためである。

適切な束縛を行っていると、 $R_g^2$  を大きくとる準安定状態は自由エネルギーが高くなり、ガラス相にあってもそこへ到達できなくなる場合があると考えられる。5.5 節で示された結果のように、モデル鎖が束縛を受けた際に見かけ上で系に状態遷移が起こらなくなるという現象は、フォールド相に系が移動しているのではなく、シミュレーション時間内で到達可能な状態が閉じ込めにより大幅に減じられたためであると考えられる。 $R_g^2$  が大きいレアイベントが消されるその様子は 6.4 節において示されている。

### 7.3 本研究から提案されること

REMD により自由エネルギー地形は確かに数個の谷を持つ landscape として描けはするのだが、実際に MD 計算を行うと、自由エネルギー地形から考えてまず起こらないという低確率のイベントがたくさん起き、そこに引っかかってしまう。自由エネルギーの谷には移動せず、レアイベントのみで MD が終わるという事態となる。レアイベントが極めて多くの数あり、そこでは自由エネルギー地形は意味をなさないということだ。ただ、これを揃わせる、自由エネルギー地形に近づける方法があり、その一つが空間的な束縛である。

初期条件をランダムに設定した場合の高精度シミュレーションでしばしば起こる準安定状態への滞留は、実は分子クラウディングを考慮していないために起こることであって、巨大分子同士の衝突という新たなノイズによりフォールディングが高速化しているという可能性を本研究は示唆する。また、この説はファネル型ポテンシャルをなおも支持する。

タンパク質のフォールディングが、補助因子等の作用により特定の構造へ落ち着くよう誘導・フォールディングの高速化がなされていると考えられてきたが、しかし混雑効果・細胞内の混雑によってもフォールディングが高速化されるであろうという予想がなされた。それは束縛ポテンシャルの導入により、レアイベントとして起こる広がった構造の自由エネルギーが高くなるためで、状態同士の競合の結果、安定な小さい状態へと落ち着くのだと考えられる。

タンパク質フォールディングにおける準安定状態への滞留を解消することを考える場合、この準安定状態の安定性をいかに消すかが問題である。

## 謝辞

物理という分野の外にいた自分が、「金子研究室に行きたい」と我儘を言い始めたのは2007年の1月、あれから3年の月日が経ったと思うと、長いとも感じますし、非常に短かったとも感じます。本論文の作成は、達成可能かどうか不安な目標でしたが、お世話になった皆様からの温かいご助力のおかげで形にできました。皆様へ、ここに感謝の意を表します。

金子研究室の石原秀至助教からは、理論面、アルゴリズム、シミュレーション等の各面で、自分がつまづいた際には総合的に助言を頂きました。特に、本テーマを設定する前に行った別の数値実験研究のシミュレーションを彼の指導の下で行わせて頂いたことが、本研究を進める上での基盤となりました。お忙しい中で教育的なテーマを頂きご指導頂いたことに謝意を表します。学振研究員の中島昭彦さんからは、本研究の研究計画、計算結果の解釈について明晰な助言を頂きました。計算負荷の掛かるこの研究は、研究室の計算機サーバーを管理されていた栗川知己さん、嶋岡大輔さんのご支援なしには続けられませんでした。彼らの緻密な管理あつての研究です。ここに謝意を表します。名古屋大学の岡本祐幸教授からは、シンポジウムの合間の時間にレプリカ交換MD法に関する貴重なアドバイスを頂きました。初対面であつたにもかかわらず親身になってご相談を受けて頂いたことに感謝の意を表します。研究室の皆様からは、研究の方向性や結果の解釈の仕方などの視野の大きな議論から、各種計算機ツールの使い方等細かい点まで日々助言を給わって頂きました。修論作成まで漕ぎ着けることができたのは皆様のお陰です、ありがとうございます。そして、テーマ設定の際の議論から本論文作成に至るまで長くご指導いただき、お付き合い頂いた金子邦彦教授に心から感謝の意を表します。

本原稿を提出した日は2011年3月18日、東北関東大震災の発生から7日後のことです。津波・震災・原子力災害により甚大な被害を受けた東北地方をはじめとする日本国の早期復興を心より願いますとともに、大災害後のこの国を担う若い者の一人としてしっかり生きていこうと思います。

## 引用文献

- [1] Han, J. and Herzfeld, J. : Macromolecular diffusion in crowded solutions. *Biophys. J.* 65 1155-1161 (1993)
- [2] Murphy, L.D. and Zimmerman, S.B. : Macromolecular crowding effects on the interaction of DNA with Escherichia coli DNA-binding proteins: a model for bacterial nucleoid stabilization. *Biochim. Biophys. Acta.* 1219 277-284 (1994)
- [3] Spink, C.H. and Chaires, J.B. : Effects of hydration, ion release and excluded volume on the melting of triplex and duplex DNA. *Biochemistry* 38 496-508 (1999)
- [4] Goobes, R. et al. : Metabolic buffering exerted by macromolecular crowding on DNA-DNA interactions: Origin and physiological significance. *Biochemistry* 42 2431-2440 (2003)
- [5] van den Berg, B. et al. : Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *EMBO J.* 19 3870-3875 (2000)
- [6] Sasahara, K. , McPhie, P. and Minton, A.P. : Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J. Mol. Biol.* 222 1227-1237 (2003)
- [7] Tokuriki, N. et al. : Protein folding by the effects of macromolecular crowding. *Protein Sci.* 13 125-133 (2004)

- [8] Tompa,P. *Structure and Function of Intrinsically Disordered Proteins* CRC Press (2010)
- [9] 今井正幸 : ソフトマターの秩序形成 シュプリンガー・ジャパン (2007)
- [10] Levinthal,C. : Are there pathways for protein folding? *J. Chim. Phys.* 65 44-45 (1968)
- [11] Levinthal,C. : How to Fold Graciously, in “Mossbauer spectroscopy in biological Systems”: proceedings of a meeting held at Allerton House, Monticello, Illinois (Debrunnder J.T.P., Munck E., eds.) University of Illinois Press 22-24 (1969)
- [12] Goodsell,D.S. : Inside a living cell. *Trends. Biochem. Sci.* 16 203-206 (1991)
- [13] Huang,K. *Lectures on Statistical Physics and Protein Folding* World Scientific (2005)
- [14] Suenaga,A. *et al.* : Folding dynamics of 10-residue hairpin peptide chignolin. *Chem. Asian J.* 2 591-598 (2007)
- [15] Yomo,T. , Saito,S. and Sasai,M. : Gradual development of protein-like global structures through functional selection. *Nat. Struct. Biol.* 6 743-746 (1999)
- [16] Shimoyama,H. , Yonezawa,Y. and Nakamura,H. : Enhanced free-energy calculation using multiscale simulation. *J. Chem. Phys.* 133 135101 (2010)
- [17] Takagi,F. , Koga,N. and Takada,S. : How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulation. *Proc. Natl. Acad. Sci. USA* 100 11367-11372 (2003)
- [18] Mittal,J. and Best,R.B. : Thermodynamics and kinetics of protein folding under confinement. *Proc. Natl. Acad. Sci. USA* 105 20233-20238 (2008)
- [19] Nakamura,H.K. *et al.* : Scrutinizing the squeezed exponential kinetics observed in the folding simulation of an off-lattice Go-like protein model. *Chem. Phys.* 307 259-267 (2004)
- [20] Sugita,Y. and Okamoto,Y. : *Chem. Phys. Lett.* 314 141-151 (1999)
- [21] 上田顕 : 分子シミュレーション 古典系から量子系手法まで 裳華房 (2003)
- [22] Han,J. and Herzfeld,J. : Macromolecular diffusion in crowded solutions. *Biophys. J.* 65 1155-1161 (1993)
- [23] Onuchic,J.N. *et al.* : Protein folding funnels: the nature of the transition state ensemble. *Folding & Design* 1 441-450 (1996)
- [24] 桑島邦博 : フォールディング機構 *in* シリーズ・ニューバイオフィジックス 1 タンパク質のかたちと物性 (中村春木, 有坂文雄 編) 共立出版 pp.78-97 (1997)
- [25] 高田彰二 : フォールディング反応 *in* タンパク質科学 (後藤祐児, 桑島邦博, 谷澤克行 編) 化学同人 pp.261-273 (2005)
- [26] 神谷成敏, 肥後順一, 福西快文, 中村春木 : タンパク質計算科学 共立出版 (2009)
- [27] Klepeis,J.L. *et al.* : Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19 120-127 (2009)
- [28] Jha,S.K. and Udgaonkar,J.B. : Free energy barriers in protein folding and unfolding reactions. *Curr. Sci.* 99 457-475 (2010)
- [29] Dill,K.A. : Polymer principles and protein folding. *Protein. Sci.* 8 1166-1180 (1999)
- [30] Hart,W.E. and Istrail,S. : Lattice and off-lattice side chain models of protein folding. *J. Comput. Biol.* 4 241-259 (1997)