

修士論文 (1998年度)

計算統計力学的手法による格子タンパク質模型の研究

千見寺浄慈

大阪大学 大学院 理学研究科 物理学専攻

Abstract

タンパク質の折れ畳み問題の解明は現代の科学における最重要課題の1つである。本修士論文はタンパク質の折れ畳み問題を物理学的観点(特に統計力学の立場)から理解しようとするものである。特に本修士論文では平衡論から単純化されたタンパク質模型を議論するために適切な統計力学的な議論をする為の計算物理学的手法(これを計算統計力学と呼ぶことにする)の開発に力を入れた。本修士論文では、はじめにタンパク質の基本的な知識を紹介し、理論的な研究の簡単なレビューを行う。その後、比較的長い格子タンパク質模型に対して熱力学的な議論をする為にはどのような手法を用いたら満足な結果が得られるかを議論する。最後に格子タンパク質模型に対して熱力学的な議論をする為の手法として我々が開発したモンテカルロ法(— Multi-Self-Overlap Ensemble 法 —)を提唱し、その有効性を議論する。またここでは実際にいくつかの模型に対して熱力学量の計算を行なう。

1 Introduction

この章では、本研究で必要になる最小限のタンパク質の知識を簡単に述べる。この章に関連する優れた文献として、文献 [1, 2] がある。

1.1 What are proteins ?

タンパク質は一本の高分子であり、通常 10 ~ 1000 個程度のアミノ酸が 1 次元的につながった高分子である。構成要素であるアミノ酸は 20 種類あり、式 (1) のような構造を持つ。



ここで NH_2 はアミノ基で、 $COOH$ はカルボキシル基である。(ただし環状になっているプロリンは例外である) それぞれのアミノ酸を特徴付けるのは側鎖 R である。(詳しくは [1] 文献を参照) タンパク質を特徴付けているのは、アミノ酸配列であり、この配列のことを“1次構造”と呼ぶ。タンパク質といっても大きくわけて3種類ある。それらは(1) 繊維状タンパク質、(2) 膜タンパク質、(3) 球状タンパク質である。この中で本論文で扱うものは(3)の球状タンパク質である。細胞内ではほとんどの生化学的反応を触媒している酵素は水溶性の球状タンパク質である。球状タンパク質は前にも述べた通り、鎖状の高分子ではあるが、生理条件下では、極めてコンパクトな、あるユニークな構造(ネイティブ構造)をとる。このネイティブ構造は生化学的機能と密接に関係していることが知られている [1]。ネイティブ構造にはしばしば α ヘリックスや、 β シートなどのいわゆる“2次構造”が見られる。そして α ヘリックス、 β シート、その他の2次構造の組み合わせによってできるタンパク質全体の構造のことを“3次構造”と呼ぶ。

1.2 The second genetic code

タンパク質がユニークなネイティブ構造へ折れ畳まれるのは、タンパク質分子内、及び水とタンパク質間の相互作用の結果であるが、その相互作用のバランスはアミノ酸配列の中に書き込まれている。この配列と立体構造の間の関係は“第二の遺伝暗号”と呼ばれている。¹第二の遺伝暗号解読問題、すなわち、「アミノ酸配列の情報からそのタンパク質の立体構造を予測すること」は現代の科学における再重要課題の1つといえよう。

¹第一の遺伝暗号とは DNA 分子の塩基配列とタンパク質のアミノ酸配列の関係で、この関係については既に完全に分かっている

1.3 Anfinsen's dogma and Levinthal's paradox

1960年代、AnfinsenらはリボヌクレアーゼAの巻き戻し実験を行なった[3]。リボヌクレアーゼAとは124残基のアミノ酸からなるタンパク質で、そのネイティブ構造は4本のジスルフィド結合(SS結合)を持っている。(つまり結合の“手”が8本ある)SS結合を還元、切断するとタンパク質の立体構造は壊れ、酵素活性もなくなる。還元されたSH基は適当な条件下におくとSS結合に酸化される。4本のSS結合が全くランダムに形成された場合には105通り($7 \times 5 \times 3 \times 1$)の組み合わせがある。Anfinsenは変性剤が存在しない条件下でSH基を酸化させると、元と同じSS結合をもつ分子が形成されると同時に立体構造や酵素活性も回復することを明らかにした。この結果は「ネイティブ構造はアミノ酸配列の情報のみで決まっており、その構造は自由エネルギー最小の状態である」ことを示している。これは“Anfinsenのドグマ”、もしくは“タンパク質フォールディングの熱力学仮説”と呼ばれており、タンパク質折れ畳みの理論的な研究の基本原理になっている。²

Anfinsenのドグマはタンパク質の折れ畳みの原理は示してくれるが、折れ畳みの形成がどの様にすすんでいくのかは示していない。この点を問題にしたのがLevinthalである[4]。Levinthalは以下の様な考察をした。

例えば、鎖の長さが100残基のタンパク質を考えてみると、1残基あたりのとりうる構造の数は、どんなに少なく見積もっても3通り以上はあるので、鎖全体のとりうる構造の総数は大まかに見積もって $3^{100} \sim 5 \times 10^{47}$ にもなる。従って、ランダムコイルの状態からネイティブ構造へ巻き戻るのに、ランダムサーチ機構で折れ畳むとすると、仮に1つの構造あたりの滞在時間を 10^{-13} 秒と仮定しても、巻き戻りに要する時間は 5×10^{34} 秒 $= 1.6 \times 10^{27}$ 年にもなってしまう。しかし実際のタンパク質は 10^{-3} から 10^3 秒のオーダーで折れ畳まれるのである。それではどうやってタンパク質は素早くネイティブ構造を探することができるのだろうか?というのがLevinthalのパラドクスと呼ばれるもので、タンパク質折れ畳み問題の中心的な問題の1つとなっている。また、このことからLevinthalはタンパク質の折れ畳みにおいてはある決まった経路があるに違いないと述べている。

以上、述べてきたように、タンパク質の折れ畳みに関する中心的な問題は“第二の遺伝暗号問題”と“LevinthalのParadox”である³。我々は、この2つの問題に答えて初めてタンパク質の折れ畳みを理解したと言える。しかしこの問題は今だ未解決である。

2 Simplified protein models

この章ではまずタンパク質を研究する際のアプローチについて概観する。その後、(以下の本修士論文でとる態度である)物理学的手法に関してやや詳しく述べ、対象とする系の問題設定(モデル化)について議論し、いくつかの単純化されたタンパク質モデルを紹介する。

2.1 Approaches

タンパク質立体構造問題の理論的研究としてのアプローチは大きく分けて3つある。(i)[経験的方法.]この方法は現在増加しつつある、タンパク質の立体構造のデータベースを眺め、その知識を基に未知のタンパク質立体構造を予測しようというアプローチである。(日本語の解説として例えば文献[5]がある。)(ii)[分子論的手法.]このアプローチは、なるべく正確に(タンパク質分子のみならず、溶液も)全原子を考慮した系に関して、分子動力学、もしくはモンテカルロ計算を行ない、計算機のなかで実際の折れ畳みを再現しようというものである。(例えば日本語の解説として、文献[6, 7, 8]がある。)この方法は折れ畳み問題を解決する最も直接的な方法ではあるが、対象とする系に含まれる原子の数が膨大であり、その為に例えば分子動力学法では通常 10^{-9} 秒の時間スケールが限界である⁴。(iii)[物理学的手法.]このアプローチは、注目した性質を説明するために、それに関係のないものは切捨て、なるべく簡単なモデルを構築し、出来る限り曖昧さのない計算を行ない、定性的な理解を得る、もしくは折れ畳みの理論的な枠組をつくらうという態度である。(日本語の解説として例えば文献[11, 12, 13]がある。又、若

²ただし全ての蛋白質のネイティブ構造が自由エネルギー最小の状態であるという証明はされていない。ある蛋白質のネイティブ構造は動力学的に辿り着きやすい構造である、という見解もある。

³しかしその他にもまだまだ問題は沢山あるのである。文献[12]の言葉を借りるならば「この領域では興味深い物理的問題に事欠くことは決してないのである！」

⁴ただし最近になって 10^{-6} 秒の時間スケールの分子動力学シミュレーションに成功したという報告がなされた[9]。

干専門的ではあるが大変優れたレビューとして [14, 15, 16] 等がある。) 以下、本論では物理学的手法に基づいたアプローチについて述べる。

2.2 Lattice protein models

物理学者が理論的にタンパク質を理解しようとする、誰しも、あまりにもタンパク質が複雑すぎて呆然とするだろう。しかし、従来、物理学でよく使われてきた発想法が1つのアプローチを示している。それは、文献 [17] において、“牛を球と仮定して議論する話”によく表されている。ことなどを説明している。この話は物理学でよく行なわれている「まず初めに、本筋に関係のないことは全部切捨て、モデル化を行ない、なるべく使えるネタを用いて、なるべく曖昧さのない計算をして議論する」というやり方がよくとらえられている例だと思われる。我々がとるアプローチはこの発想法に基づいているが、我々はタンパク質の折れ畳み問題において「何が本筋に関係ないか?」「何を切捨てて議論すべきか?」という問題は常につきまとう問題である。⁵例えば、アミノ酸を球（もしくは点）だと仮定したらどうだろうか? さらに（使えるネタを使うために）アミノ酸は規則的な格子の上だけにすることにし、とりうるボンドの角度は数種類だけにすると、といった簡単化を行なったらどうだろうか? (例えば Figure 1 に 2 次元正方格子上に丸いアミノ酸がのったタンパク質模型を示してある。) この様な非常に簡単化されたモデルでも、相互作用を適切に選ぶことにより、ユニークな基底状態を持つことがわかっている。また、鎖の構造を格子上の Self-Avoiding-Walk だと近似することにより、Self-Avoiding-Walk の知識を使うことができる。たとえば、Self-Avoiding-Walk の構造の総数は鎖の長さを N とすると、およそ $N^{\gamma-1}\mu^N$ (γ と μ は定数) のような振舞いをする事が分かっている。これは、鎖の長さ N が有限でも比較的大きい数ならば天文学的な数字である。このことから以下のことがいえる。

アミノ酸を“マル”と仮定して、鎖の配置を格子上の Self-Avoiding-Walk だと近似しても、第二の遺伝暗号解読問題になる能力をもっており、さらにそれは Levinthal のパラドクスになっている。

したがって、このような格子モデルにおいて上記の問題（つまり、適当なアミノ酸配列を与えてその情報だけから、基底状態、もしくは自由エネルギー最小の構造を予言することと、その構造を全構造を探索することなく見つけるにはどうしたら良いか? ということ）を完全に解くことが出来れば実際のタンパク質の理解への示唆が与えられるだろう。

2.3 HP model

前節で格子モデルを導入したが、相互作用はどのようにすればよいだろうか? この節では若干の相互作用の考察を行ない、HP model と呼ばれる格子モデルを紹介する。

2.3.1 Local interactions and hydrogen bonding

1936 年、Linus Pauling らのグループは水素結合がタンパク質の折れ畳みにおける支配的な力であるという考えを提唱した [18]。そして、ペプチド結合の幾何学的配置を決定するため、アミノ酸のペプチド鎖模型を作成した。彼らはこの模型を用いて、水素結合がうまく適合する構造を探すことによって α ヘリックス構造と β シート構造を理論的に発見し、それがタンパク質の重要な要素であると予言した [19, 20, 21, 22, 23]。これはまだタンパク質の立体構造が未知の時代になされた予言であるという意味で驚くべき予言である。その後、1960 年になって Kendrew らによる X 線結晶解析でミオグロビンの立体構造が初めて明らかになり、 α ヘリックスが実際の蛋白質中で含まれている事が確かめられ [24]、その後も実験的研究により β シート構造等も実際に存在することが確かめられた。こうして Pauling らの予言が正しかったことが証明され、水素結合が支配的な力であるという考えが受け入れられた。

2.3.2 Nonlocal forces: hydrophobic interactions are important

しかし、1950 年代に Walter Kauzmann は異なる見解を持っていた。すなわち、水素結合は球状タンパク質の形を決める主要な力ではない、という見解である。これは「折れ畳まれていないタンパク質

⁵たとえ不要な情報であっても捨てたくないというのが人情だろう

と、それを囲む水の水素結合の強さは、近似的に天然状態のタンパク質間に働く水素結合の大きさと等しい」という実験事実に基づいている。そして、彼は非極性分子が水を避ける傾向（疎水性相互作用）がタンパク質折れ畳みの最も主要な駆動力であると主張した [25]。Kauzmann の研究以後、非常に多くの疎水性相互作用に関する理論的及び実験的研究が行なわれたが疎水性相互作用の分子論的なメカニズムは必ずしも明らかとはなっていない。1つの分子論的解釈として以下のものがある。

一般に溶質を水に溶解させるときのことを考えると、まず水分子間の水素結合を切断して溶質分子が入れるだけの穴をあけなければならない。しかし、溶質を溶かした後、系全体のエネルギーは出来るだけ復元して安定しなければならない。溶質分子が水と水素結合を作りうる基（例えば OH 基）を持っているときには、断ち切られた水分子内の水素結合は溶質分子と水素結合をつくることによってエネルギーが復元できる。しかし、溶質分子が非極性の場合には、溶質分子と水分子との間には水素結合が出来ないので、溶質分子のまわりの水分子間に水素結合をできるだけつくってエネルギーを修復しようとする。このために水の構造がより規則的になってエントロピーが現象する。（文献 [26] より抜粋）

一見したところ疎水性相互作用は、分子をコンパクトにし、疎水性残基は内側に押し込めようとするので、タンパク質天然立体構造のユニークな構造を決めるには、あまりに非特異的な力であるように思われる。実際、1975 年までは、疎水性相互作用は主にタンパク質をコンパクトにしておくのに重要な因子で、水素結合などの力がユニークな立体構造を決める、といった見解が受け入れられてきた [27]。それ以降、現在に至るまで上記の見解は広く受け入れられ、「折れ畳みに寄与しているのは 1 種類の力だけではない」と多くの研究者が考えている。

しかし、近年 Ken.A.Dill によって「疎水性相互作用はタンパク質をコンパクトにするばかりでなく、特異的でユニークな構造をも、さらには折れ畳み経路さえをも担っている」という見解が発表された [28, 29]。実際、最近の実験では（側鎖のタイトパッキング等の細かい因子を忘れて）疎水性相互作用のみを考慮すればユニークな天然構造が得られるという結果も現れている [30]。

2.4 HP model

前小節の議論に基づき、Lau と Dill は疎水性相互作用のみを考慮した格子モデル - HP model - を提案した [31]。このモデルは鎖の構造を格子上の Self-Avoiding Walk で近似し、さらにアミノ酸は疎水性 (H: hydrophobic) のものと親水性 (P: polar) の 2 種類からなるとし、タンパク質のアミノ酸配列は H と P の配列として表現されるものである。そして、「H モノマー同士が最近接格子にきた場合、自由エネルギーが得をする」というモデルである。これは整理条件下では、疎水性モノマーが集まったほうが自由エネルギーが得であるという考えに基づいている。Figure 1 に 2 次元正方格子上の長さ 14 の配列 $HPHPHP^2HPHP^2H^2$ の HP モデルの例を示した。この様に、H-H コンタクトが増えれば増えるほど自由エネルギーは低くなり熱力学的に安定になるモデルである。この模型のネイティブ構造は H-H コンタクトが最大のもので定義する。また、H-H コンタクトが最大の構造が 1 つしかないとき（縮退していないとき）、それはユニークなネイティブ構造をもつといい、実際のタンパク質に対応していると考えられる。文献 [31] において、長さ 18 までの 2 次元 HP モデルに対して、全配列空間を全構造空間探索で調べられ、調べた範囲ではユニークなネイティブ構造を持つ配列は全配列空間のうち、約 2.1 ~ 2.4% であることが報告されている。

上記の定義では H-H コンタクトが増えれば自由エネルギーが下がるというものであったが、それは生理条件下での話であって、例えば熱力学的振舞い等を議論するときは上記のモデルの拡張が必要である。この拡張として、H-H コンタクトがある場合、自由エネルギーではなくてエネルギーが -1 だけ下がるという拡張がなされた。これは式で書くと以下のようなモデルとなる。

$$E(r) = \sum_{i < j+1} u(S_i, S_j) \Delta(r_i - r_j) \quad (2)$$

ここで r は鎖の構造を表し、 r_i は i 番目のモノマーの位置を表している。 $\Delta(r_i - r_j)$ はもし r_i と r_j が最近接格子を表しているならば -1 で、その他は 0 をとる記号である。又、記号 u は、 $u(H, H) = -1$, $u(H, P) = u(P, H) = u(P, P) = 0$ で定義される。これは $0^\circ \sim 100^\circ\text{C}$ の間では、疎水性モノマーが集まる傾向にあり、自由エネルギーの温度依存性はこの温度範囲においては小さいという実験事実 [32] を基にしている。

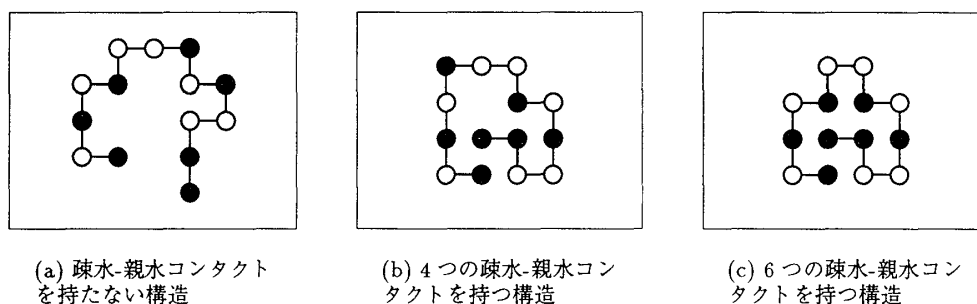


図 1: 配列 $HPHPHP^2HPHP^2H^2$ の HP モデル：白丸を P、黒丸を H で示した。

この様に H-H コンタクトをエネルギーと解釈する場合に議論できない点は“低温変性”である。この様に拡張した HP モデルでは、ネイティブ構造とは式 (2) のエネルギーにおける基底状態で定義される。

2.5 Other lattice models

格子蛋白質模型をはじめとする、簡単化されたモデルは HP model の他にも様々なものが提案されている。モデルの選択は議論したい内容によって適切なものを選ぶべきだろう。ここでは、紙面の都合上文献のみを羅列するに留めるが、実に様々なモデルが使われている [33, 34, 35, 36, 37, 38, 39, 40, 41, 42]。

3 Tools

この章では格子タンパク質において、構造空間を探索する道具を紹介する。構造空間を探索する優れた方法を開発することには 2 つの意味がある。1 つ目はアミノ酸配列だけから全構造探索せずに基底状態を探すことができればそれは Levinthal のパラドクスを解決する方法かもしれない、ということ。2 つ目は格子タンパク質模型の熱力学を議論する際に、きちんと重要な情報を得るためである。

3.1 exact enumeration of all conformation

格子上の短い高分子鎖についての厳密な計算を最初に行なったのは 1974 年の W.J.C. Orr である [43]。W.J.Orr は 10 残基より短い高分子鎖で格子上の全ての自己回避歩行の構造を計算した。その後、計算機の技術向上に伴い、全構造探索が可能な残基数も増加しつつある。現在までに公表されている最も長い鎖の全構造数え上げは 2 次元正方格子上で長さ 51 である [44]。ただしこの計算は 10GB のメモリーを積んだ計算機でなければ不可能である。通常よく行なわれているのは長さ 18 までの数え上げがほとんどである。この様に厳密数え上げ法は、確かに曖昧さが全くない為に計算結果は強いことを主張できるが、長さが非常に制限されているという難点がある。

3.2 exact enumeration of maximally compact conformation

コンパクトな構造の数は全構造数に比べると、非常に少ない。例えば格子上の self-avoiding walk を考えた時、最近接格子の数を z とすると、長さ N の鎖のとり得る全構造数は z^N のオーダーである。一方、完全にコンパクトな構造の数は $(z/e)^N$ となり、全構造を探索するよりも遥かに少ない。ここで $e > 1$ の定数である。

3.3 cubic lattice of the size $3 \times 3 \times 3$

格子タンパク質の研究で良く用いられているテクニックが、長さ 27 の模型に対して maximally compact 構造 ($3 \times 3 \times 3$ の立方格子に収まる構造) を数え上げる方法である。これは Shakhnovich らによっ

て初めて用いられた方法である [45]。3 × 3 × 3 の maximally compact 構造は回転対象などの対象性をすべて除いて 103346 個ある。この方法は完全自動化は不可能で、対象性を“手”で抜く必要がある。

3.4 related topics

その他、maximally compact ではないが、compact な構造（ただし人間が指定した形）に対して数え上げを行なう方法が Mai らによって提唱されたが [46]、本質的には、3 × 3 × 3 の maximally compact 構造の方法と同様である。また、比較的大きい maximally compact 構造の数え上げとして、Pande らが 4 × 4 × 3 の maximally compact 構造の数え上げを報告している [47]。しかし通常の計算機ではこれは不可能である。

又、少し洗練された方法として、転送行列を用いた maximally compact 構造の発生法が Kloczkowski らによって提唱されている [48, 49]。この方法の利点は 2 次元では $m \times n$ のサイズの maximally compact 構造が、3 次元では $m \times n \times l$ のサイズの maximally compact 構造が行列のかけ算だけで得られるということと、例えば $3 \times n$ や $4 \times n$ の 2 次元正方格子の maximally compact 構造を発生させるのに必要な転送行列を一旦求めてしまえば、 n はいくらでも大きくできるという点である。しかし一般にこの転送行列を求めることは難しい。

以上、述べてきた maximally compact 構造の数え上げ法や、転送行列による構造発生法などはくまなく maximally compact 構造を探索でき、基底状態探索などに有効であるかのように見えるが、（例えば HP モデル等では）maximally compact 構造ではない構造がネイティブ構造になることが良く知られている [91]。したがって、人工的にコンパクト化するような相互作用を入れたモデルに対しては上記の方法は有効であるが、一般にはこの方法で基底状態が見つかるという保証はない。

3.5 Monte Carlo scheme for thermodynamics

1992 年、O'Toole と Panagiotopoulos は HP モデルの熱力学を議論するためにはどのようなアルゴリズムが適切かを議論した [50]。まず、O'Toole と Panagiotopoulos は長さ 48 の 3 次元 HP モデルに対して、「手で」配列をデザインし、その配列に関して 2 つの異なるモンテカルロ法のアルゴリズムを用いて、どちらが低エネルギー構造をサンプルする能力を持っているかを調べた。2 つのアルゴリズムとは、通常のメトロポリス法（マルコフ連鎖モンテカルロ）と、Rosenbluth & Rosenbluth 法（以下、RR 法と略す）[51] の変型モンテカルロ法⁶である。基底状態探索の結果、メトロポリス法では低エネルギー構造をサンプルすることは事実上不可能で、RR 法の変型版の方がより低いエネルギー構造をサンプルできることがわかった。しかし、変型 RR 法でも必ずしも精度の良い結果を与えてくれるものではないということも分かった。この研究では、いかに HP モデル（より一般にヘテロポリマー）のモンテカルロ法が、鎖の長さが少し長くなった場合、難しいかを示している。

1997 年、Peter Grassberger はさらに RR 法を発展させたモンテカルロ法 —Pruned-enriched Rosenbluth method（以後、PERM と略す）— を開発した [52]。この方法は基本的には RR 法と同じであるが、異なる点は「複数の系を同時に用意する」点である。そして、「複数の鎖を同時に伸ばしていき、ある程度まで伸ばしたところで、重みの小さい者は捨ててしまい、その代わり、重みの大きいものをコピーして増やす」というものである。文献 [52] ではホモポリマーに対して用いられたが、その後、若干の修正を行ない格子タンパク質模型に対して応用された [53]。PERM を様々な相互作用の格子タンパク質模型の基底状態探索に用いて（この時点では）最も効率の良いアルゴリズムであることを示した。⁷

3.6 optimization scheme - finding a needle in a haystack -

格子モデルの基底状態探索方法の開発は、単に最終的に得られた構造に興味があるから開発する意味がある、というだけでなく、「実際のタンパク質は、なぜ構造空間の完全探索を行わずに、素早く最適化された構造へ折れ畳まれるのか？」という問いに対する答えを提供してくれるかもしれないという意味で重要である。ただし、後者の様な意味を持つためには明らかに非物理的な過程を踏んでいるスキームであってはならない。

⁶通常 RR 法は 1 つずつ鎖にモノマーを足していく方法であるが、O'Toole らが用いたのは一度に 3 つのモノマーを足すが、3 つ足す際にありとあらゆる伸ばし方を数え上げでチェックし、最もウエイトが大きいものを選ぶという方法である。

⁷しかし PERM を用いて有限温度の計算をした場合、バイアスなく構造サンプリングができるという保証はない

Constraint Hydrophobic Core Construction(CHCC)

配列が与えられたときの HP モデル (長さ 88 まで) の基底状態のエネルギーと、その構造、縮退度が厳密に計算できるスキームが Yue と Dill によって提出された [54]。この方法は Constraint Hydrophobic Core Construction(CHCC と略す) と呼ばれる方法で、「HP モデルの基底状態の構造 (H-H コンタクト数最大) は、疎水性コアの表面積が最小の構造である」という事実を用いて、数え上げ的方法ではあるが (したがって非物理的な過程を踏んでいる)、厳密に計算ができるという意味で非常に優れている。実際に報告された鎖の長さは 88 までであるが、この範囲内で計算時間 T の鎖の長さ L に対する依存性は大雑把に

$$T \sim 0.001 D^{0.9} 1.125^L \quad (3)$$

という振舞いをすることも報告された [55]。つまり、数え上げ的方法なので指数関数的に長さ L とともに計算時間はかかるが、それでも 1.125 という小さい値の冪乗である。ここで D は与えられた HP 配列の基底状態の縮退度である。

また、HP モデルにおいて、配列が与えられたときにネイティブ構造を厳密に予言できるという意味で、第二の遺伝暗号解読問題が (HP モデルに限定されてはいるが) 完全に解けたと言える。しかし残念な事にこの方法は HP モデルに特化した方法であるので実際のタンパク質には応用は不可能である。また、CHCC は非物理的な過程によって計算する方法なので、Levinthal のパラドクスの解決にはなり得ない。

Hydrophobic zipper

格子タンパク質において、ランダムな構造からネイティブ構造へたどり着くにはどのような物理的道筋を通れば良いだろうか？この問の答えに関して興味深い提案が Chan, Fiebig, Dill らによってなされた。紙面の都合状、提案されたアルゴリズムの名前と文献を挙げるに留めるが、彼らは広がった構造から、ネイティブ構造へ全構造探索せずに折り畳むには、あたかもジッパーが閉じていくように、なるべくローカルな 2 次構造をつくるよう、折り畳んでいくとうまく行くだろうという Hydrophobic zipping 説を提唱した [58]。関連文献は [56, 57, 59, 91] である。これは必ずしも蛋白質の天然構造予測に成功したとは言えないが、折り畳みの初期段階を非常にうまく説明できる [60]。

4 Native structure of lattice protein models

この章では格子タンパク質模型のネイティブ構造を調べることによって実際のタンパク質に対して何が示唆できたか？を簡単に紹介する。

4.1 Symmetry and native structure in lattice proteins

実際のタンパク質のネイティブ構造を眺めると、しばしば非常に幾何学的に対称性の良い構造が見受けられる。例えば、フォーヘリックスバンドルや、 α - β バレルなどが挙げられる [1]。なぜこのような規則的な構造がネイティブ構造に選ばれるのだろうか？3 次元 HP モデルを用いてこの問いに対する回答の提案がなされた [55]。Yue と Dill は CHCC を用いて、比較的長い ($N \sim 67$) 3 次元 HP モデルに対して、基底状態の縮退度 (g_N) の少ない ($g_N < 100$) 配列を探した [55]。その結果、縮退度が小さいものは、その基底状態は非常に幾何学的対象性が良いという観察結果を得た。例えば、顕著な物として、 α - β バレルに似た基底状態 ($g_N = 3$)、4 ヘリックスバンドルに似た基底状態 ($g_N = 67$)、 β ヘリックスに似た基底状態 ($g_N = 4$) 等があげられた。

以上の観察から Yue と Dill は「実際のタンパク質のネイティブ構造が幾何学的に対象性の良いのは (縮退度が小さいという意味で) 良い疎水性コアを作った結果である」と主張した。

4.2 Designability

近年 X 線結晶解析と NMR 分光法によって指数関数的にタンパク質の立体構造が明らかになりつつある。それにも関わらず新しい構造トポロジーの数は少しずつしか増えていない。いったいなぜ構造トポロジーの数はそんなに増えないのであろうか？1992 年に Chothia は以下の事実から、タンパク質の折れ畳みのトポロジーの数は有限であり、約 1000 種類であると推定した [61]。

現在、立体構造既知のタンパク質を立体構造分類法に基づいて分類すると、異なった折れ畳み構造をもつ球状ドメインの種類は約 120 程度といわれる。遺伝子の解析で新たに同定されるタンパク質のアミノ酸配列の約 1/3 は既知のタンパク質と相同である。そして、既知のタンパク質の 1/4 は、この 120 種類の折れ畳み構造をもつ球状ドメインのうちのどれかとアミノ酸の一致度が 25% 以上の相同性を持つ。このことは、きちの 120 の折れ畳み構造全体の 1/12 であることを意味する。すなわち、全体の折れ畳み構造は $120 \times 12 =$ 約 1000 種類となる。

上記の見積もり方には疑問が残るものの、タンパク質のネイティブ構造には約 1000 種類の構造トポロジーしかないということは多くの研究者が支持している。もし上記の命題が正しいとするならば、なぜそんなに構造トポロジーは少ないのだろうか？

Hao Li らは HP モデルを用いた上記の問題に対する回答にもなりうる研究結果を発表した [40]。Hao Li らは（広い意味での）HP モデルを用いて、全配列空間 (2^N) に対して全 maximally compact 構造空間⁸で基底状態探索を行ない、ユニークな基底状態を持つ配列とその基底状態の構造のみを記録した。鎖の長さは 2 次元正方格子では 25 と 30 と 36、3 次元立方格子では 27 の物が用いられた。その結果まず、3 次元の解析では全配列 (2^{27}) のうち、4.75% の配列が縮退しない基底状態を持つことが明らかとなった。さらに、それらの配列と基底状態の構造との関係を調べたところ、沢山の配列が、同じ構造を縮退しない基底状態として持ち、そのような構造は少ないことが分かった。例えば、最も顕著な例として、ある構造を“縮退しない基底状態”として持つ配列は 3794 個ある。Li らは、与えられた構造を縮退しない基底状態として持つ“配列の数”を“Designability(N_s)”と呼び、 N_s の大きい構造のことを“Highly designable structure”と呼んだ。もし、今後の研究で実際の蛋白質の天然構造の特徴と、格子モデルの“Highly designable structure”の特徴に関連付けができたならば、Chotiha の提案を裏付けることができるかもしれない。無論、どういう構造が“良い疎水性コア”を作りうるのか？ということを一般的に明らかにすることは非常に重要な問題で、今後の課題として残されている。その後も Designability に関する研究は続いており、関連する論文として例えば文献 [62, 63, 64, 65, 66] がある。又、日本語の解説としては文献 [67] がある。

5 Monte Carlo sampling method

前章では格子タンパク質モデルのネイティブ構造の研究により、いくつかの見地が得られたことを概観した。しかし、いままでの議論からもわかるとうり、Levinthal のパラドクスは格子モデルの枠組の中でも依然として解決されていない。どうすればこの問題の理解に一步でも近づくことができるだろうか？

近年、蛋白質の折り畳みは、(超多次元空間である) いわゆるエネルギーランドスケープの形状がネイティブ構造に近ければ近いほどエネルギーが下がるようにデザインされていれば、速い折り畳みや、熱力学的安定性など、様々な実験結果を説明できる、という提案がなされた [68, 69]。これは、エネルギーランドスケープの形状が漏斗 (funnel) 状になる、ということから、しばしば funnel 理論などと呼ばれている。funnel 理論は多くの研究者にインパクトを与えたが、より深い理解のためには以下の 2 つの課題が残されている。(i) エネルギーランドスケープは、非常に膨大な自由度を持っているために人間がそれを表現しようとしたときに、どうやったら（気持ちでわかるのではなく）定量的に表現できるか？ どのような量がどのような振る舞いをすれば漏斗状であると定量的に結論付けられるか？ など、超多次元空間を人間が理解できる範囲内で表現する方法がまだ確立されていない。(ii) もし、上記の問題が解決されても、リアルなモデルでも格子モデルでも、全構造探索が出来ない場合にどうやって重要な情報を持った構造をピックアップすることができるか？ といった問題が難しい問題として残っている。⁹

本修士論文では、(i) の問題には（とりあえず）触れないとして、主に (ii) の問題に取り組んだ。以後、(ii) の問題を克服するための手法を議論する。我々が採用した方法はモンテカルロ法であるが、そのための準備として、この章では格子タンパク質のモンテカルロを行なう為の最小限必要な知識を極簡単にまとめる。まず、ポリマーの動かし方として Self-Avoiding Walk で良く用いられているものを紹介

⁸文献 [40] では maximally compact 構造に限って計算が行なわれたが、HP model では必ずしも maximally compact 構造が基底状態になるということは保証されていない。しかし、定性的な議論としてはこれでも良いようである。

⁹解析的に蛋白質を扱うことができれば良いだろうが。

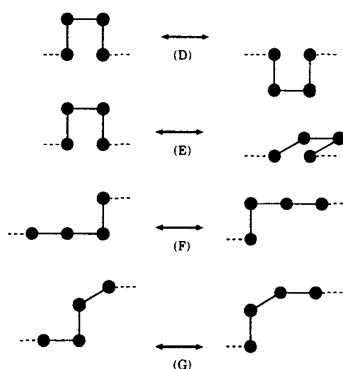


図 2: 全ての two-bead flip:(D)180° crankshaft. (E)90° crankshaft. (3次元のみ). (F)Two-bead L-flip. (G)Cube permutation(3次元のみ)

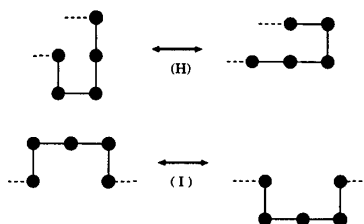


図 3: three-bead flip の 2 つの例:(H)Three-bead J-flip (I)Three-bead 180° crankshaft

し、次いで統計力学でよく用いられるモンテカルロの手法を復讐する。ここで重要なポイントは素過程の Ergodicity である。

5.1 Classification of moves of Self-Avoiding Walk

格子上に載ったポリマーの Monte Carlo においてはポリマーをどう動かすかが基本的である。ここでは Self-Avoiding Walk(SAW) の Monte Carlo でよく用いられている非常に一般的な素過程を紹介する。ここでのムーブの名前等は文献 [70] に従うことにする。

5.1.1 Local N-conserving moves

ローカルムーブとは、SAW の内の数個（通常 1~4 個）のつながったサイト（“ビーズ” もしくは “たま”）のみを動かし、他のサイトは動かさない素過程である。以下で具体的なローカルムーブをやや羅列的であるが紹介する。

Figure ??に全ての可能な “one-bead moves”（1つのサイトを動かすムーブ）を示す。ムーブ (A) は “one-bead flip” と呼ばれる素過程である。（“kink-jump” と呼ばれることもある）鎖の内部の “one-bead moves” はこれだけが可能である。ムーブ (B) (C) は “end-bond rotation” である。ムーブ (D) は “180° crankshaft” と呼ばれる素過程である。ムーブ (E) は “90° crankshaft” と呼ばれる素過程で、これは 3 次元以上で可能なムーブである。ムーブ (F) は “Two-bead L-flip” である。ムーブ (G) は “Cube permutation” で、これも 3 次元以上で可能である。ローカルムーブとして、他にも鎖の端を 2 つ動かす “two-bead end-group moves” や、鎖の中や端を 3 つ動かす “three-bead moves” があるが、非常に多数あるので、興味ある “three-bead moves” の例を 2 つ紹介するに留める。(Figure 3)

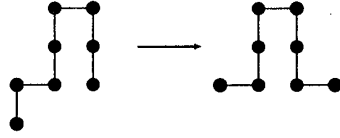


図 4: slithering-snake(reptation)

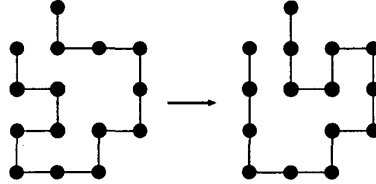


図 5: kink-transport

5.1.2 Bilocal N-conserving moves

Bilocal move とは元の鎖の一部を消去し、その消去した部分を鎖の他の部分にくっつけるアルゴリズムである。以下にその例をいくつか紹介する。

- slithering-snake move (もしくは reptation と呼ばれる)。これは一方の端のボンドを消去し、新しいボンドを多方の端に付け加えるものである。(Figure 4)
- kink transport move。これは鎖のいわゆる“こぶ”の部分が合った場合、そのみを消去し、他の部分にくっつけるムーブである。(Figure 5)
- kink-end reptation move。これは kink transport move と同様にまず鎖上の“こぶ”を消去するが、その後 2 つのボンドを (いずれかの) 端にくっつけるものである。(Figure 6の→の方向)
- end-kink reptation move。これは kink-end reptation move の逆で、端の 2 つのボンドを消去し、それを鎖の内部の“こぶ”としてくっつける方法である。(Figure 6の←の方向)

5.1.3 Non-local N-conserving moves

ノンローカルムーブは多数のサイトを動かす素過程であるので、ヴァリエーションは無限にある。しかし、Monte Carlo のアルゴリズムにおいて最も便利な素過程を見つけることは非常に難しい。というのも、(1) ノンローカルムーブを一回行なうと鎖の形状は劇的に変わってしまうので、試しに作ってみた構造は排除体積条件を満足していない可能性が非常に大きい。それゆえに鎖の長さが大きくなってもアクセプタンスが 0 にならないようなノンローカルムーブを開発するという事は簡単な仕事ではない。(2) ノンローカルムーブは通常非常に時間のかかるものである。(通常 CPU time で N のオーダー。

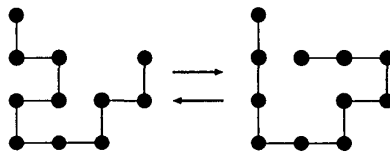


図 6: kink-end reptation(←の方向) と end-kink reptation(→) の方向

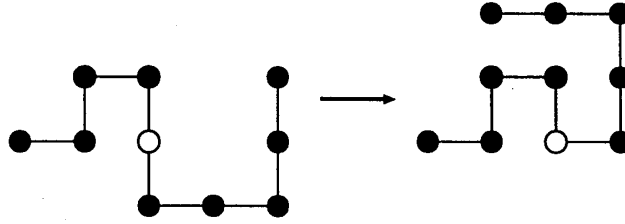


図 7: pivot move: (ここでは例として 90° rotation を示してある。) 回転軸を表す点 (pivot point) を白丸で表してある。

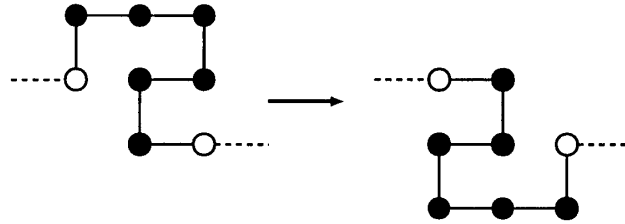


図 8: Cut-and-paste move。白丸から白丸の間を一度切り離し、回転させてもう一度くっつける

それに対してローカールムーブは 1 のオーダーである。) 従って、時間がかかってもやる価値があるかどうか? という問題がある。

以上の様な理由により、以下ではノンローカールムーブの羅列的な紹介はしない。ただ、現在までのところ SAW の Monte Carlo で非常に有効であると認められている幾つかのムーブを紹介するに留める。

SAW で広く用いられているノンローカールムーブとして pivot move と cut-and paste move がある。

pivot move では、まず鎖の中のある一点 (k と記す) を選び、その点に関して格子における対象な演算 (例えば回転、反転等) を k 以上の全てのサイトに関して行なう。(Figure 7) すなわち、新しい構造 w' は

$$w'_i = \begin{cases} w_i & \text{for } 0 \leq i \leq k \\ w_k + g(w_i - w_k) & \text{for } k+1 \leq i \leq N \end{cases} \quad (4)$$

ここで、 w_i は $i-1$ 番目のサイトからのボンドの角度を、 N は鎖の長さ、 g は回転などの演算を表す。

cut-and-paste move では、2 もしくはそれ以上のサイトを選び (k と l と記しておく)、その間を一旦取り除き、取り除いた部分だけに対して逆点、回転、反転、もしくは置換といった演算を施し、その後再び取り付けるといったものである。(Figure 8) 例えば逆点の演算では、新しい構造 w' は

$$w'_i = \begin{cases} w_k + w_l - w_{k+l-i} & \text{for } k \leq i \leq l \\ w_i & \text{for otherwise} \end{cases} \quad (5)$$

経験的に、pivot move と cut-and-paste move は鎖の長さを N としたときのアクセプタンスが大体 $\sim N^{-p}$ の様に振舞うことが知られている。ここで p は $0.1 \sim 0.2$ である。特に、pivot move の方は 2 次元で $p = 0.19$ 、3 次元で $p = 0.11$ となることが数値実験で示されている [71]。したがって大体 N^p 回のトライアルを行なうことによって 1 回は新しい構造がアクセプトされることを意味する。また、pivot move は劇的に構造が変化するために、一回アクセプトされるとそれは本質的に全く新しい構造である。このことから、非常に大まかには自己相関関数が N^p のような振舞いをするだろうと予想がつく。一方、local move と bilocal move を用いた場合の自己相関関数はそれぞれ、 $N^{3.2}$ と N^2 という振舞いをする事がわかっており、[71] pivot move が SAW に対しては最も効率良いアルゴリズムであることがわかる。

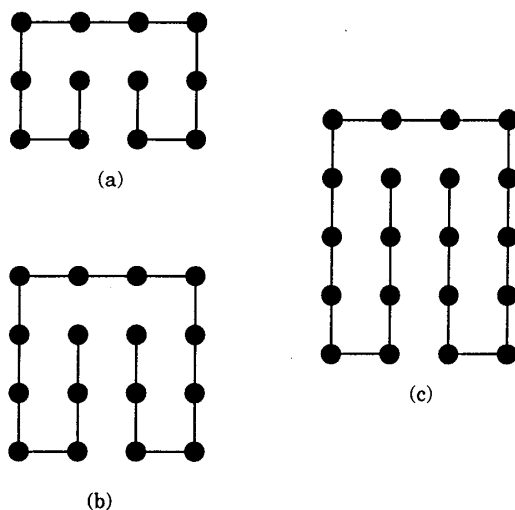


図 9: double cul-de-sac 構造：この構造から他の構造へ、ローカルムーブでは抜け出すことが出来ない。

5.1.4 Ergodicity

格子タンパク質模型でもそうなのだが、SAW の Monte Carlo でも素過程が Ergodic であることは必須の条件である。ここではどのような素過程を選べば Ergodic であるかを紹介する。

ローカルムーブは極少数のサイトを動かすだけなのでプログラミングが簡単であるし、CPU time もかからないので広く使われてきた。しかし致命的な欠点がある。それは Ergodic ではないということである。例えば 2 次元正方格子の SAW を考えよう。Figure 9 の構造（この様な構造は “double cul-de-sac” と呼ばれる）を初期状態に選んだとするとローカルムーブではもはやこの構造から抜け出せないことがわかるだろう。3 次元でも同様のことが知られており、ローカルムーブは Ergodic ではない。

一方、bilocal move に関しては以下の組み合わせにより、Ergodic になることが知られている。

- reptation + kink-end and end-kin reptation
- kink-end and end-kin reptation + 90° end-bond rotation
- reptation + kink transport + on-bead flip （ただし 2 次元で）
- kink transport + on-bead flip + 90° and 180° end-bond rotation （ただし 2 次元で）

さらに、pivot move に関しては Sokal らによって厳密に Ergodic であることが証明されている [70]。

以上の様に、（エネルギーが入っていない）最も簡単な高分子の模型の SAW でさえ、排除体積効果によって緩和が遅くなり、効率良く鎖の構造を発生させる事が難しい。さらには単純なムーブでは Ergodic でさえなくなってしまうということを見た。このことから、いかに格子タンパク質模型のモンテカルロシミュレーションが難しいかがわかるであろう。

5.2 Metropolis Algorithm

後で必要になるので、統計力学でよく用いられているマルコフ連鎖モンテカルロ法（特にメトロポリス法）の原理を簡単に復讐しておく。ここでは興味ある分布はカノニカル分布なので、逆温度 β での分布を得るためのアルゴリズムについて述べる。詳しくは [72, 73, 74, 75] 等を参照されたい。

格子 G 上に載ったヘテロポリマー（例えば HP model）を考えよう。鎖の構造を Γ と表記した時のヘテロポリマーのエネルギーを $E(\Gamma)$ で表すことにする。 t ステップでの状態分布関数 $P(\Gamma, t)$ に対し、マルコフ過程

$$P(\Gamma, t+1) = \sum_{\Gamma'} w(\Gamma' \rightarrow \Gamma) P(\Gamma', t) \quad (6)$$

を定義する。ここで $w(\Gamma' \rightarrow \Gamma)$ は状態 Γ' から状態 Γ への遷移確率である。遷移確率に詳細つりあいの条件

$$w(\Gamma' \rightarrow \Gamma)e^{-\beta E(\Gamma')} = w(\Gamma \rightarrow \Gamma')e^{-\beta E(\Gamma)} \quad (7)$$

を課せば、この過程の定常分布がカノニカルアンサンブルになる。この条件を満たす遷移確率として、よくメトロポリス法

$$w(\Gamma' \rightarrow \Gamma) = \min[1, e^{-\beta(E(\Gamma) - E(\Gamma'))}] \quad (8)$$

が用いられている。

ここで注目すべきことは、なぜ定常分布がカノニカル分布になるか？ということである。ということである。それは明らかに式 (8) でカノニカル分布になることを要請しているからである。したがって、任意の関数 $f(E)$ を持ってきて、定常分布が

$$P(E) = e^{-f(E)} \quad (9)$$

となるようにすることは容易であり、(メトロポリス法的には) 遷移確率を

$$w(\Gamma' \rightarrow \Gamma) = \min[1, e^{-\beta(f(\Gamma) - f(\Gamma'))}] \quad (10)$$

とすれば良いだけである。これは至極当然のことであるが、カノニカルアンサンブルだけでなく、任意のアンサンブルが構成できるという意味において重要であるということを強調しておく。

上記の様な確率過程で作られた時系列は原理的には現実のダイナミクスとは全く関係ない。しかし、現実にもあり得そうな「素過程」を選んでシミュレーションに適用した場合、それは実際に熱浴に接した実際の系の模倣をシミュレートしているという解釈が広く認められている。¹⁰ 実際、格子タンパク質模型のダイナミクスの研究として、one-bead flip と crank-shaft を素過程に選んでメトロポリス・モンテカルロを行ない、それを実際の折れ畳み過程だと解釈して議論していることが多い [65, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88]。

5.3 Histogram reweighting technique

Histogram reweighting technique とは、ある温度で実際にシミュレーションを行なったデータを用いて、違う温度の熱力学量も計算してしまおうというものである [89, 90]。Histogram reweighting technique の原理は非常に簡単で、以下の通りである。

ある温度 $T_0 = 1/\beta_0$ の物理量 O の期待値 $\langle O(\beta_0) \rangle$ は

$$\langle O(\beta_0) \rangle = \frac{\sum O(E)e^{-\beta_0 E}}{\sum e^{-\beta_0 E}} \quad (11)$$

である。ここでの $\sum$ は全配位空間に関する和を意味する。これはモンテカルロ法を用いると、以下の様に書ける。

$$\langle O(\beta_0) \rangle = \frac{\sum_{MC(\beta_0)} O(E)}{\sum_{MC(\beta_0)} 1} \quad (12)$$

ここで $\sum_{MC(\beta_0)}$ とは、各状態を確率 $P(E) \propto e^{-\beta_0 E}$ でサンプリングしたものの和を意味する。このとき、

物理量 O の T_0 とは異なる温度 $T = 1/\beta$ での平均値は

$$\langle O(\beta) \rangle = \frac{\sum_{MC(\beta_0)} O(E)e^{-(\beta - \beta_0)E}}{\sum_{MC(\beta_0)} e^{-(\beta - \beta_0)E}} \quad (13)$$

¹⁰ 無論、それは解釈であって何らかの原理から導かれて正しいと証明されたものではない。このような解釈が正しいかどうかの議論も必要だろう。

と計算できる。又、物理量 $\langle O(\beta) \rangle$ の温度微分の平均値 $\langle \partial O(\beta) / \partial T \rangle$ も

$$\begin{aligned} \left\langle \frac{\partial O(\beta)}{\partial T} \right\rangle &= -\beta^2 \left\langle \frac{\partial O(\beta)}{\partial \beta} \right\rangle \\ &= -\beta^2 \left[\langle (O \cdot E)(\beta) \rangle - \langle O(\beta) \rangle \cdot \langle E(\beta) \rangle \right] \end{aligned} \quad (14)$$

と計算できる。上の計算は正しいが、実際にこれを適用する際には注意が必要である。実際のシミュレーションでは配位空間の極一部しかサンプリングしていない為に、あまり温度を大きくシフトすると、ヒストグラム法で大きな重みに付け変えられる状態を全くサンプリングしていないという事が起る。(この問題点の詳しい解説は文献 [75] を参照)

ここでの教訓としては任意の重み関数 $f(E)$ を用いて確率 $e^{-f(E)}$ でモンテカルロを行なった場合でも有限温度の物理量の計算はヒストグラム法を用いて

$$\langle O(\beta) \rangle = \frac{\sum_{f(E)} O(E) e^{-\beta E + f(E)}}{\sum_{f(E)} e^{-\beta E + f(E)}} \quad (15)$$

と計算できてしまうという点である。

6 Extended Ensemble Monte Carlo

前の章で述べたメトロポリス法などのマルコフ連鎖モンテカルロ法は、数値計算の方法ではあるが、「実際の物理過程の模倣」という側面も持っている。しかし、「自然の真似」をしているだけで真に興味ある物理量をきちんと求めることができるだろうか？というのも実際の自然でも非常に緩和時間の長い系はいくらでも存在し、往々にして興味ある現象はそういうものが多いのであり、そのような系を対象として「自然の真似」をしてシミュレーションしたのでは実用上（計算時間など）限界がある。例えば格子タンパク質模型では比較的長い鎖に対して「自然の真似」をしているだけではほとんど基底状態をサンプルすることが不可能であることが知られている [91]。そこで、「真似」することを諦め、うまく目的（通常カノニカル平均の物理量を求める事）を遂げるような“探索者”としてのシステムを人為的に設計し、それを自立的に働かせるで、目的を達成させるということが考えられる。

このような考え方をを用いたモンテカルロ法として、「拡張アンサンブル法」とよばれる一連の方法がある。これはシミュレートしたい系をなんらかのかたちで「拡張」もしくは「結合」し、困難（例えば遅い緩和の原因等）を除去しようとするものである。例えば、カノニカルモンテカルロでは重み

$$w_{\text{Boltzman}}(E) = e^{-\beta E} \quad (16)$$

でシミュレーションを行なうのに対し、かつてな関数 $f(E)$ を用いた重み

$$w_{\text{extended}}(E) = e^{-f(E)} \quad (17)$$

でシミュレーションすることが考えられる。この「重み」を用いることによって緩和の促進や、（自明でない）基底状態を探しやすくなる等の理点があるならば積極的にこの様な重みを使った方が良いだろう。というのも、どんな変な重みを使ってアンサンブルを構成しようとも、結局 Histogram reweighting してしまえば、カノニカルアンサンブルが再構成できてしまうからである。もともとの Histogram reweighting 自体はそれほど意義があるとは思えないが¹¹、「どんなアンサンブルからでもカノニカルアンサンブルが reweight することにより、再構成できるのである」という事を示したという点で、特に拡張アンサンブル法にもたらした意義があるのではないかと思われる。

以下では、特に格子タンパク質模型への応用を念頭に置いて、いくつかの拡張アンサンブルの例を紹介する。

¹¹少なくとも著者には

6.1 Multicanonical Ensemble

マルチカノニカル法 [92, 93] とは、「状態の出現確率を、エネルギー状態密度 $\Omega(E)$ の逆数にとる」方法である。すなわち、ある関数 $f(E)$ を $f(E) = \ln \Omega(E)$ ととり、重み $w(E)$ を $w(E) = e^{-f(E)} = e^{-\ln \Omega(E)}$ と選んでシミュレーションすることにより、得られる分布 $H(E)$ が

$$\begin{aligned} H(E) &= \Omega(E) e^{-f(E)} \\ &= \Omega(E) e^{-\ln \Omega(E)} \\ &= \text{const} \end{aligned} \quad (18)$$

となり、全てのエネルギー状態が等しい確率で出現する。もともとこの方法は 1 次相転移の転移点のように状態間の自由エネルギー障壁が高く平衡分布が得にくい問題に対して開発された。ヒストグラムを平らにすることは直ちに自由エネルギー障壁を低くすることに対応するので多状態ポッツモデルの様に 1 次相転移を起こす系には極めて効果的である [92, 93]。

さらに、マルチカノニカル法を用いる理点のもう一つの側面として、エネルギーのローカルミニマムに長い時間トラップされないという点が挙げられる。(これに関しては和文の解説では、参考文献 [6, 7, 8] で強調されている。) というのも、ヒストグラムが平らになるということは、シミュレーションを行なうと、系はエネルギー空間を (疑似的にはあるが) ランダムウォークすることに対応するからである。この利点に着目して、スピングラス [94] やタンパク質 [95]、格子上のヘテロポリマー [96, 97] 等の様に配位空間がいわゆる “Rugged Energy Landscape” になっている系に対しても適用され、従来のメトロポリス法等よりも遥かに有効であることが示されている。

6.1.1 Multicanonical Recursion

Multicanonical 法とは、状態の出現確率を $e^{-\ln \Omega(E)}$ と取るということを前に述べた。しかし、そもそも $\Omega(E)$ を知るということは一般に困難であり、もし $\Omega(E)$ が厳密に分かっていればシミュレーションを行なう必要もない。それではどうやって $\Omega(E)$ を求めるのであろうか？ここでは Berg & Neuhaus [92, 93] の方法に従って関数 $f(E)$ を近似的に求める方法について述べる。

状態密度 $\Omega(E)$ を求める基本的な考え方は「厳密に $\Omega(E)$ が求まらなくても、必要なエネルギー範囲で、おおよそ $f(E) = 1/\Omega(E)$ となれば良い」ということと、「短いランを数回行なうことにより、逐次ヒストグラムが平らになるように、重み関数を改良していく」ということの 2 点である。なお、ここでの手続きの様に重みを求める事を “Preliminary Run”、もしくは “学習” と呼ぶ。

まず、便宜上エネルギーを間隔 ϵ を用いて、

$$[E^0, E^0 + \epsilon], [E^0 + \epsilon, E^0 + 2\epsilon], \dots, [E^0 + (m-1)\epsilon, E^0 + m\epsilon]$$

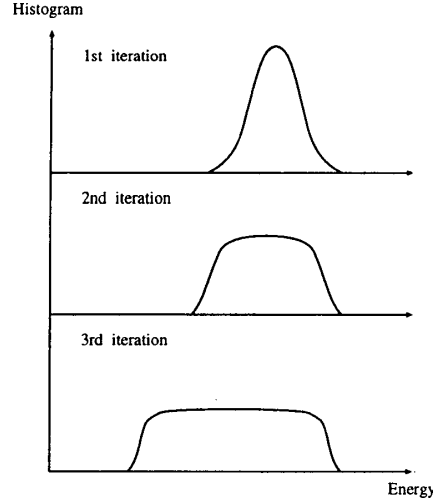
の様に m 個の区間に分割する。(以下で E_i と書いたら $[E^0 + (i-1)\epsilon, E^0 + i\epsilon]$ に属するエネルギーのことを指す。) そして、関数 $f(E)$ を形式的に以下の様に書いておく

$$f(E) = \alpha(E) + \beta(E)E \quad (19)$$

ここで、 $\alpha(E)$ と $\beta(E)$ は “Multicanonical parametrization” と呼ばれる量で、エネルギー E の関数である。理想的な $f(E)$ は、 $\ln \Omega(E)$ の近似式であるので、式 (19) は $\ln \Omega(E)$ を折れ線で近似することにはならない。

上の準備のもと、以下の手順に従って逐次的に $\alpha(E)$ と $\beta(E)$ を改良していく。以下で、 $\alpha^n(E), \beta^n(E)$ と $H^n(E)$ は、それぞれ領域 D_i に対する n 回目のシミュレーションで用いる係数、傾き、ヒストグラムを表している。

1. 全ての D_i に対して $\alpha^1(E) = 0$ 、 $\beta^1(E) = 0$ として、短いシミュレーションを 1 回行なう。その結果、エネルギーヒストグラム $H^1(E)$ が得られるとする。又、得られた最低エネルギーが属するエネルギー領域を E_{min} と気録しておき、この重みで得られたエネルギー平均値 (最も良く出現したエネルギー) が属するエネルギー領域を E_{max} としておく。以後、 E_{min} は更新するが、 E_{max} は固定しておく。

図 10: マルチカノニカルの重み関数 $f(E)$ の学習の様子

2. n 回目以降は ($n \geq 2$)、 $n-1$ 回目に得られたヒストグラム H^{n-1} を用いて $\alpha^n(E)$ と $\beta^n(E)$ を以下の式に従って改良する。

$$\beta^n(E) = \begin{cases} \beta^{n-1}(E) & \text{for } E > E_{max} \\ \beta^{n-1}(E) + \frac{1}{\epsilon} \ln \frac{H^{n-1}(E+\epsilon)}{H^{n-1}(E)} & \text{for } E_{min} \leq E \leq E_{max} \\ \beta^{n-1}(E) & \text{for } E \leq E_{min} \end{cases} \quad (20)$$

全ての $\beta^n(E)$ を決めた後、以下の式で、 $\alpha^n(E)$ を決める。

$$\alpha^n(E) = \begin{cases} 0 & \text{for } E > E_{max} \\ \alpha^n(E+\epsilon) + [\beta^n(E) - \beta^n(E+\epsilon)] & \text{for } E < E_{max} \end{cases} \quad (21)$$

3. 新しい重み因子 $f(E) = \alpha^n(E) - \beta^n(E)E$ を用いて重み

$$w_{mu} = e^{-\alpha^n(E) - \beta^n(E)E} \quad (22)$$

でシミュレーションを行ない、ヒストグラム $H^n(E)$ を得る。

4. 必要なエネルギー領域のヒストグラム $H(E)$ が実用上問題ない程度に平らになるまで上記の 2 と 3 を繰り返す。

この様にして重み因子 $f(E)$ を改良していくと、その都度得られるヒストグラムは図 (10) の様になり、だんだんヒストグラムが平らになっていくはずである。以上の方法からわかるように、Berg & Neuhaus の方法は、「隣の領域のヒストグラムを見て、少なかったら重みを重くし、多かったら重みを軽くする」というものである。また、全ての領域のヒストグラムが等しくなったら重み関数を更新しても変わらないことは明らかであろう。

6.1.2 calculation of thermodynamics quantity

十分にヒストグラムが平らになる重み因子 $f(E)$ が求まったら、その重みを用いて一回長い Run を走らせる。(このことを Measurement Run と呼ぶ) そのデータを用いてヒストグラム法を用いて

$$\langle O(\beta) \rangle = \frac{\sum_{f(E)} O(E) e^{-\beta E + f(E)}}{\sum_{f(E)} e^{-\beta E + f(E)}} \quad (23)$$

とすることにより、熱力学量が計算できるわけである。ここで、Multicanonical Ensemble では興味あるエネルギー領域すべてに関してほぼ等しい出現頻度で状態が出現するので、ヒストグラム法のところで述べた様な問題点はなく、一回のシミュレーションで任意の温度の物理量が計算可能である。

6.2 Entropic sampling method

Entropic sampling 法 [98] とは、後で見るように本質的には Multicanonical と全く同じであるが、重み因子を求める手順が異なる。以下では、Lee[98] に従って Entropic sampling 法による重み関数の求め方を述べる。

Multicanonical のときと同様にエネルギーを区間に分割する。便宜上重み関数を $S(E)$ と書くことにする。具体的な手順は以下の通りである。

1. すべてのエネルギー領域 D_i に対して $S(E) = 0$ として短いシミュレーションを行なう。
2. 上でヒストグラム $H(E)$ が得られる。一回目のシミュレーションで最も良く出現したエネルギーを E_{max} と定義しておく。(以下で、 E_{max} は固定しておく) 又、上のシミュレーションで得られた最低エネルギーを E_{min} としておく。

$$S^n(E) = \begin{cases} S^{n-1}(E) & \text{for } H^{n-1}(E) = 0 \\ S^{n-1}(E) + \ln H^{n-1}(E) & \text{otherwise} \end{cases} \quad (24)$$

以上の手続きを、必要なエネルギー範囲のヒストグラムが実用上平らになるまで繰り返す。Multicanonical 法では「隣のヒストグラムと見比べて重みを調節した」のに対し、Entropic sampling 法ではそのようなことをしないで済む。すなわち、隣の状態が出ようが出まいが重み因子は求まる、という点で実際にプログラムをコードする際、Entropic sampling 法の方がより簡単である。熱力学量の計算は式 23 と同様である。

6.3 Multicanonical and Entropic sampling are essentially identical

この節では、Multicanonical 法と Entropics Sampling 法が本質的に同等であることを示す。ここでの論法は Berg & Hansmann & Okamoto[99] に従うことにする。まず、Entropic Sampling 法での重み $S^n(E)$ を以下の様に表現したとする。

$$S^n(E) = \alpha^n(E) + \beta^n(E)E \quad (25)$$

ただし、ここで関数 $S^n(E)$ は連続であると仮定すると

$$\alpha^n(E) = \alpha^n(E + \epsilon) + [\beta^n(E) - \beta^n(E + \epsilon)]\epsilon \quad (26)$$

が成り立つ。 $S^n(E + \epsilon) - S^n(E)$ と、式 26 を用いて $\alpha^n(E)$ を消去すると、

$$S^n(E + \epsilon) - S^n(E) = \beta^n(E)\epsilon \quad (27)$$

を得る。ところで、Entropic Sampling 法での重み $S^n(E)$ のアップデートの仕方は式 (24) であったから式 (24) を式 (26) に代入すると、

$$\beta^{n+1}(E) = \beta^n(E) + \frac{1}{\epsilon} \ln \frac{H^{n-1}(E + \epsilon)}{H^{n-1}(E)} \quad (28)$$

となる。これは Multicanonical 法のアップデートと同じ式である。このことから、Entropic Sampling 法と Multicanonical 法は本質的に同じだといえる。

以上で見たように、Multicanonical 法と Entropics Sampling 法は原理的には同じである。ではいったい何が違うのであろうか？ 答えは「Multicanonical 法は、状態密度を“折れ線”で近似する」のに対し、「Entropics Sampling 法は“棒グラフ”で近似する」という点で事なる。従って、エネルギースペクトルが極少数しかない場合（例えば HP model）に、エネルギーの分割をエネルギー対分割の領域を 1:1 にとれば両者は完全に一致する。又、両者に於いて、エネルギー区間を分割する際、十分に区間の長さ $\Delta E = E_i - E_{i-1}$ を小さくとれば、エネルギースペクトルが連続的になっている場合でも両者の差異はなくなる。

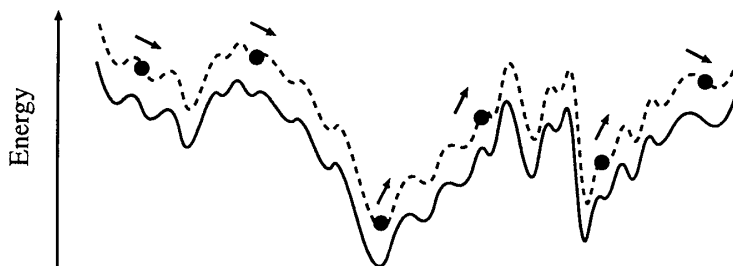


図 11: エネルギー極小に留まらないアルゴリズムを用いた場合の配位空間上の配位の移り変わりの様子

6.4 Improvement of Multicanonical recursion

前々節、前節の方法で重みを求められれば問題ないが、通常以下の様な問題がある。

- 例えば Multicanonical 法で $f^0(E)$ を厳密に $\log \Omega(E)$ にとったとしよう。そうすると、Multicanonical 法の重みのアップデートを行なうと、厳密な値からずれていってしまう。
- 初期の重みは通常高温極限にしてある ($f^0(E) = 0$) これは高いエネルギー状態では理想値に近いかもしれないが、低エネルギー状態の重みとしては理想値にほど遠い為に preliminary run を非常に沢山行なわなければ低エネルギー状態の重みが精度良く求まらない。

このような問題を解決するために様々な工夫がなされている。ここでは具体的に触れずに文献だけを紹介するに留める。[100, 101, 102, 103]

7 Multi-Self-Overlap Ensemble

Multicanonical をはじめとする拡張アンサンブル法は確かに従来のモンテカルロ法よりも遥かに有効である。だからといって今、我々が調べようとしている格子蛋白模型に対して満足な結果を与えてくれる手法だろうか？ Multicanonical や、交換法など、「エネルギー極小に留まらないアルゴリズム」を用いた場合、Figure (11) の様に、ローカルミニマムに引っかからずに、スルスルと排位空間の曲面に沿って構造は移り変わっていく。しかし、だからといって直ちにこれが基底状態をサンプルできることを保証してくれているわけではない。十分に長い時間（しかし遥かに従来のものよりは短いのであるが）をかければ、基底状態までいきつくだろうが、「エネルギー極小に留まらない」だけでは、いわば「逃げの姿勢」であり、消極的である。

一次転移を引き起こす系に関して言えば、「ヒストグラムを平らにする」ということは直ちに基底状態等、秩序相への“パス”を作ったことに対応するが、いわゆる “Rugged energy landscape” になっている系に対してはそうっていない。格子蛋白質模型も一般には “Rugged energy landscape” であるが、Figure (12) の様に “非物理的過程” を使ってでも基底状態をはじめとする “重要な情報を担った排位” をすばやくサンプルするにはどうしたら良いだろうか？

この章では以上のことを動機に、「どうすれば格子蛋白質模型に対して満足な結果を与えてくれる手法になるだろうか？」ということを議論する。その問いに対する 1 つの答えとして、我々が提唱した排除体積条件をシステムティックに緩めるモンテカルロ法 “Multi-Self-Overlap Ensemble” [104] を紹介する。

7.1 weakening self-avoiding condition

格子タンパク質模型をはじめとするヘテロポリマーの熱力学量をマルコフ連鎖モンテカルロを用いて計算する際の困難となっている原因はいったいなんだろうか？ いままで、配位空間が凸凹になっていて、ローカルなエネルギー極小値に配位がつかまってしまうからだということは何度も述べてきた。

しかしそれだけが問題を難しくしているのではない。例えば、(エネルギーの入っていない) 単純な Self-Avoiding-Walk でさえ効率良く発生させることは難しいのである。これは深刻な問題である。とい

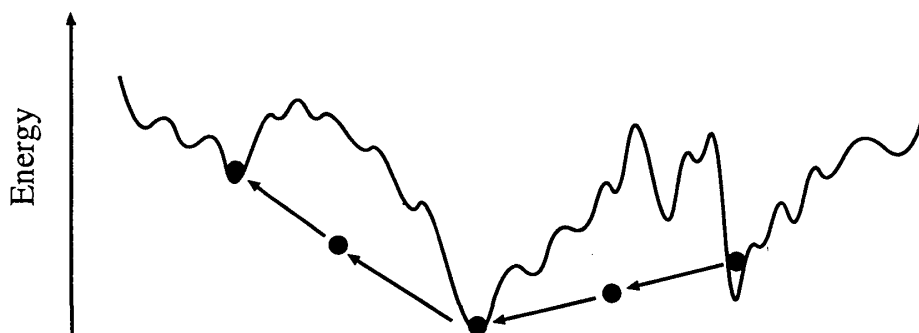


図 12: 状態間に新しい“パス”を作ることにより重要な情報を持った配位をすばやく見つける方法の概念図

うのも、排除体積条件はエネルギーに全く関係ないので、いくら従来の拡張アンサンブル法を用いて配位の凍結を防いでも、排除体積条件に伴う緩和の遅さは残ってしまう。

従って排除体積条件を緩めることにより、上記の困難を解消できるのでは？というアイデアが浮かぶ。実際、文献 [81, 87] で、以下の様な試みがなされた。

3次元立方格子におかれた長さ27の鎖を考え、系のエネルギー E を

$$E = B_0 \sum_{\langle ij \rangle} \Delta(r_i - r_j) + \sum_{\langle ij \rangle} B_{ij} \Delta(r_i - r_j) + D_2 \sum_{\langle ij \rangle} \delta(r_i - r_j) + D_3 \sum_{\langle ijk \rangle} \delta(r_i - r_j) \cdot \delta(r_i - r_k) \quad (29)$$

としたモデルを考えた。ここで、 $\Delta(r_i - r_j)$ は、 $r_i - r_j = 1$ のとき値1をとり、その他のときは0をとる記号である。このモデルを簡単に説明しよう。第一項は $B_0 < 0$ ととり、“コンパクトになるとエネルギーが下がる”効果を表している。第二項は“アミノ酸間の相互作用”を表している。つまり、 i 番目のアミノ酸と j 番目のアミノ酸が最近接格子に来た場合、 B_{ij} の相互作用をするというものである。第三項は i 番目のアミノ酸と j 番目のアミノ酸が同じサイトにきたら、つまり排除体積条件を侵したらエネルギーが $D_2 (D_2 > 0)$ だけ上がる、というものである。第四項は、同様に同じサイトに3つのモノマー i, j, k が来たらエネルギー $D_3 (D_3 > 0)$ だけ上がる、というものである。ここで文献 [81] では $D_2 = 10$, $D_3 = 14$ とした。

このモデルと、式 (29) の第二、三項を入れずに排除体積条件を完全に課したモデルとで、ランダム構造から、基底状態への折れ畳みの速さを比較したところ、圧倒的に式 (29) のモデルの方が速かった。

この様に、同じ格子点上に2つ、もしくは3つまでの多重占有をゆるして、その変わり多重占有があった場合エネルギーが上がる、といったモデルでも緩和は随分速くなる様である。文献 [81, 87] では、熱力学量を求める手法として排除体積条件を緩めたわけではなく、速く折れ畳ませる為の手法ととらえている。又、パラメーター D_2 , D_3 は、折れ畳みが速くなるようにトライ&エラーで選んでいる。従ってこのままではどうすれば適切なパラメーターを選べるかはわからないのである。

しかしここで教訓的なことは「排除体積条件を少し緩めるだけで緩和が速くなる」ということである。

7.2 How to controll self-avoiding condition

前節で、排除体積条件を少し緩めただけで緩和が速くなるという事実を紹介した。従って排除体積条件を緩めてトポロジカルな障壁（通常これは無限大の高さの障壁である！）を乗り越えられるようにすることは有用であろうと推測される。後で示す様に、実際、多重占有を許した構造というのは、基底状態など、興味ある低エネルギー状態への“橋渡し”の役目をする構造になりうるのである。しか

し、前節のような制御の仕方では自己回避歩行の構造がどの位の割合でシミュレーションで出現するかが分からず、悲惨な場合、ほとんど全ての構造が非自己回避歩行になってしまって熱力学量が計算できないといった状況も考えられる。と、いうのも明らかに自己回避歩行のエントロピーの方が、多重占有を許した構造のエントロピーよりも遥かに小さいからである。それではいったいどうやって排除体積条件を制御すれば最も効率が良くよくなるだろうか？その問いに対する我々の提案は「多重占有を表す変数を導入し、それを Multicanonical 的な制御をする」ことである。そうすれば、マルコフ連鎖の中で鎖は自己回避歩行と、非自己回避歩行の間を行ったり来たりする。これは排除体積条件を強めたり弱めたりすることに相当する。このようにすることによって望みの「自己回避歩行の出現する割合」を人間が制御できるわけである。以下で、より具体的な方法について述べる。

まず、“多重占有の割合”を表現するための変数を導入する。鎖の構造（これは多重占有がある場合にも拡張した構造）を Γ と表そう。そのときに“多重占有の割合”もしくは“重なり具合”を表す変数 $V(\Gamma)$ を以下の式で定義する。

$$V(\Gamma) = \sum_{i \in G'} [n_i(\Gamma) - 1]^2 \quad (30)$$

ここで、 $n_i(\Gamma)$ とは、格子点 i 上に載っているモノマーの数、 G' とは、1 つ以上のモノマーが存在する格子点の集合を表す。以下、この変数 V を“罰金”と呼ぶことにする。この変数は式 (30) から明らかな様に、鎖の構造 Γ が自己回避歩行の構造の場合、 $V(\Gamma) = 0$ で、その逆の極限、すなわち 2 つの点に“かたまってしまった場合”の構造 Γ のときは $V(\Gamma) \sim N^2$ となる。

上で導入した“罰金”を用いて「エネルギー - 罰金の 2 変数空間でヒストグラムが平らになる様なアンサンブル」を構成することを考える。これが実現できれば、系は局所的なエネルギー障壁に引っかからないばかりか排除体積条件に伴うトポロジカルな障壁も乗り越えられ、興味ある情報をもつ構造を効率良くサンプルできるであろう。このようなアンサンブルを“Multi-Self-Overlap Ensemble”と呼ぶことにする。

“Multi-Self-Overlap Ensemble”を実現する方法は容易である。まず、エネルギー - 罰金の 2 変数空間でヒストグラムが平らになる様な重み関数 $f_{MSOE}(E, V)$ を求める。この重み関数を求める方法は Multicanonical 法や Entropic sampling 法と全く同じ方法で求めることができる。異なる点はヒストグラムと重み関数が 2 変数であるという点だけである。2 変数のマルチカノニカルは、例えば Ising model のエネルギーと磁化の 2 変数で [105]、タンパク質模型のエネルギーと慣性半径で [106] で既に行なわれているが、“非物理的方向”へ拡張した例は“Multi-Self-Overlap Ensemble”が初めての試みである。

重複をいとわず、具体的な重み関数を求める方法を述べる。

まず、便宜上エネルギーを間隔 ϵ を用いて、

$$[E_{min}, E_{min} + \epsilon], [E_{min} + \epsilon, E_{min} + 2\epsilon], \dots, [E_{min} + (m-1)\epsilon, E_{min} + m\epsilon]$$

の様に m 個の区間に分割する。（以下で E_i と書いたら $[E_{min} + (i-1)\epsilon, E_{min} + i\epsilon]$ に属するエネルギーのことを指す。）ここでの E_{min} とは、排除体積条件を考慮したとき基底状態のエネルギーが厳密に分かる場合はその値を、分からない場合は明らかにそれより低いエネルギーは存在しないであろうと思われる値にする。罰金は整数の値しか取り得ないので間隔 1 で分割する。（場合に依っては罰金の分割も幅を 2 や 3 などにするのも良いだろう）また、あまりに多重占有が多過ぎる構造は良い“橋渡し”の役目をしてくれるとは期待できないし、第一排除体積条件を緩め過ぎると自己回避歩行の構造のサンプル数が非常に少なくなってしまう。そこで、罰金に上限 V_{max} を設け、それ以上の多重占有がある構造は禁止することにする。 V_{max} の値は鎖の長さを N としたとき、 $N/10 \sim N/5$ 位に設定するのが経験的に良い。こうすると、とり得る（エネルギー、罰金）の状態は通常 Figure 13 の様になる。

上の準備のもと、以下の手順に従って逐次的に 2 変数の重み関数 $f(E, V)$ を改良していく。ここでは Lee による Entropic sampling [98] の方法に基づいた方法を述べるが、Berg らの方法 [92, 93] でも同様に求めることが可能である。

以下で、 $f^n(E)$ と $H^n(E)$ は、それぞれ領域 D_i に対する n 回目のシミュレーションで用いる重み、ヒストグラムを表している。

1. 全ての E, V に対して $f^1(E, V) = 0$ として、短いシミュレーションを 1 回行なう。その結果、エネルギーヒストグラム $H^1(E, V)$ が得られるとする。

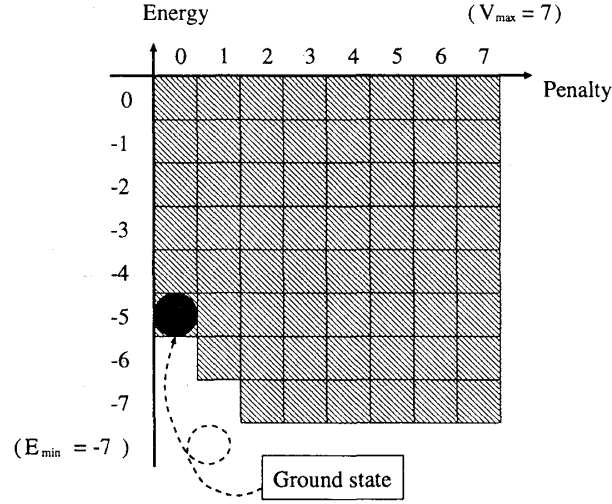


図 13: 例として $V_{max} = 7$, $E_{min} = -7$ としたときのとりうる状態: この例では、基底状態 ($E = -5$) を知らないとして、 $E_{min} = -7$ と置いたときの様子を示している。又、この例ではエネルギーの分割の幅を 1 にしている。

2. n 回目以降は ($n \geq 2$)、 $n-1$ 回目に得られたヒストグラム H^{n-1} を用いて $f^n(E, V)$ を以下の式に従って改良する。

$$f^n(E, V) = \begin{cases} f^{n-1}(E, V) & \text{for } H^{n-1}(E, V) = 0 \\ f^{n-1}(E, v) + \ln H^{n-1}(E, V) & \text{otherwise} \end{cases} \quad (31)$$

3. 新しい重み因子 $f^n(E, V)$ を用いて重み

$$w_{MSOE} = e^{-f^n(E, V)} \quad (32)$$

でシミュレーションを行ない、ヒストグラム $H^n(E)$ を得る。

4. 必要なエネルギー領域のヒストグラム $H(E, V)$ が実用上問題ない程度に平らになるまで上記の 2 と 3 を繰り返す。

この様にして重み関数を改良していくと、その都度得られる 2 次元ヒストグラムはだんだん平らになっていくはずである。

熱力学量の計算は、排除体積条件を制限する度合いを λ ($\lambda \rightarrow \infty$ で排除体積条件を満たしたもののしか許さなくなる) と書くと¹²

$$\begin{aligned} \langle O \rangle_\beta &= \lim_{\lambda \rightarrow \infty} \frac{\sum_i O_i P_f^{-1}(E_i, V_i) \exp(-\beta E_i - \lambda V_i)}{\sum_i P_f^{-1}(E_i, V_i) \exp(-\beta E_i - \lambda V_i)} \\ &= \frac{\sum_i' O_i P_f^{-1}(E_i, 0) \exp(-\beta E_i)}{\sum_i' P_f^{-1}(E_i, 0) \exp(-\beta E_i)} \end{aligned} \quad (33)$$

と求められる。ここで $\sum_i'$ は排除体積条件を満たしたデータだけで和をとることを意味し、 P_f は重み $f(E, V)$ を使った確率であることを、 i は i 番目のモンテカルロステップであることを意味している。式 (33) からわかる様に Multi-Self-Overlap を用いたときは熱力学量の計算は排除体積条件を満たした構造のデータだけを使ってあとのデータは捨ててしまっても良いことがわかる。

¹² E と β の関係は V と λ の関係に似ている

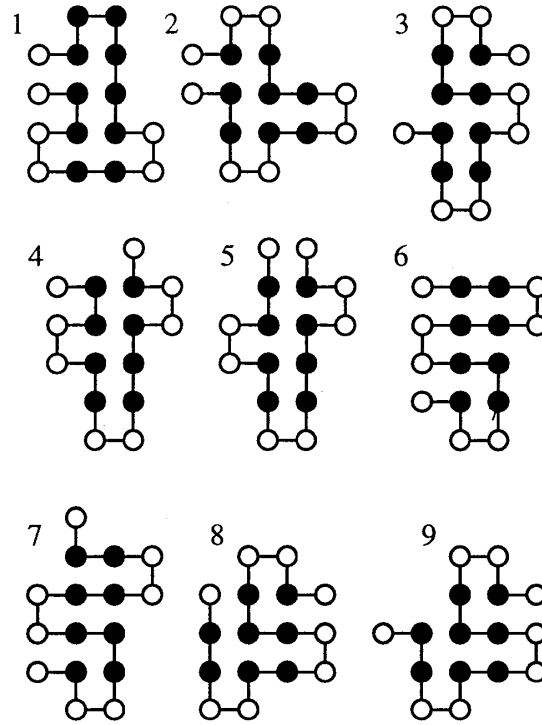


図 14: このモデルの 9 つの基底状態：黒丸が H モノマー、白丸が P モノマーを表している

7.3 Numerical test

この節では前節で導入した “Multi-Self-Overlap Ensemble” 法が他のアルゴリズムに比べてどのように改善されたか等、効率比較を行なう。

7.3.1 Does Multi-Self-Overlap Ensemble give accurate result for short chain ?

まず、Multi-Self-Overlap Ensemble が正しい答えを与えてくれるかどうかをチェックする為に、全構造探索が可能な、長さ 16 の HP model を用いて厳密計算と Multi-Self-Overlap Ensemble で計算した結果とを比較する。ここで用いる配列は $PH^2P^2H^2P^2H^3P^2$ で、基底状態は Figure 14 に示した様に 9 つある事が知られている。使ったムーブは (1)one-bead flip, (2)90° end-bond rotation, (3)180° end-bond rotation, (4)±90° rotation, (5)180° rotation, (6) rotation and reflection の 6 つと、さらに Figure 15 に示した新しい 2 つのものを使った。Figure 15 は効率良く Self-Avoiding Walk とそうでないものをいったりきたりさせる為に必要なムーブである。また、使ったパラメーターは $V_{max} = 4, E_{min} = -8$ を用いた。厳密数え上げの結果と比較する量は非常に微妙な量である “ローカルフィールド” と呼ばれる量を計算した。ローカルフィールド f_i は以下の式で定義される。

$$f_i = \sum_j u(S_i, S_j) \Delta(r_i - r_j) \quad (34)$$

ここで $u(S_i, S_j)$ は i 番目のモノマーと j 番目のモノマーが共に H モノマーだったら -1 をとりその他は 0 をとる記号で、 $\Delta(r_i - r_j)$ は r_i と r_j が最近接格子に来たとき 1 でその他は 0 をとる記号である。Figure 16 にローカルフィールド 1, 6, 11, 15 を Multi-Self-Overlap を使って計算した結果と厳密数え上げで計算した結果を示す。Figure 16 から厳密数え上げと非常に良い一致を示していることがわかり、Multi-Self-Overlap が正しいアルゴリズムであることが示された。

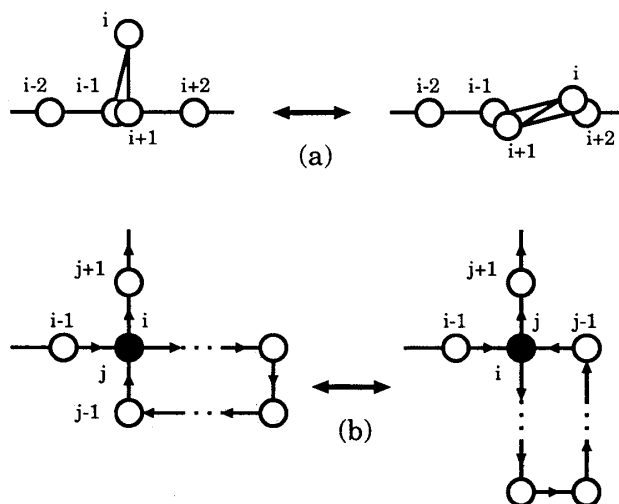


図 15: 新たに開発したムーブ:(a)“Jackknif”:モノマー $i-1$ と $i+1$ が同一サイトにあった場合、モノマー i をモノマー $i \pm 1$ の最近接格子のいずれかにおき直す。(b)“loop overturn”: i 番目と j 番目のモノマーが同じサイトにあってループを形成していた場合、そのループをひっくり返す。

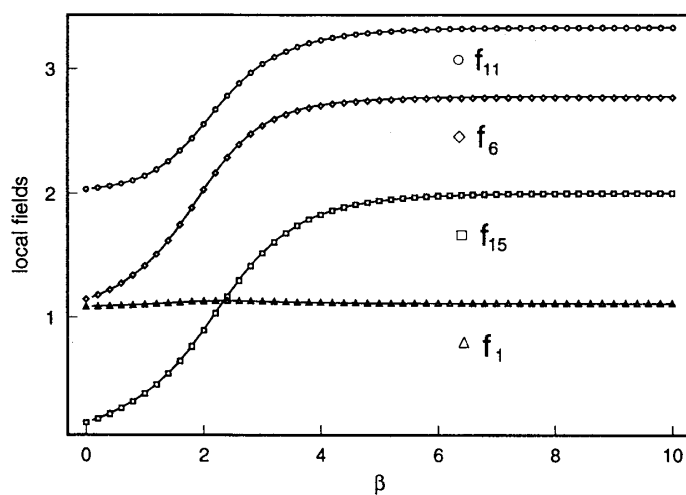


図 16: Multi-Self-Overlap Ensemble による計算と厳密数え上げによる計算の比較

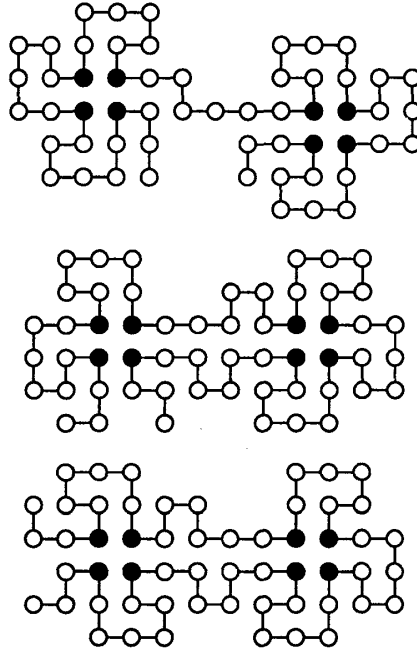


図 17: このモデルの 3 つの基底状態のタイプ: H モノマーを黒丸で、P モノマーを白丸で表している。基底状態のグループとして、上から順にグループ 1、グループ 2、グループ 3 と分類した。

7.3.2 Comparison to Multicanonical on 8-sticker HP model

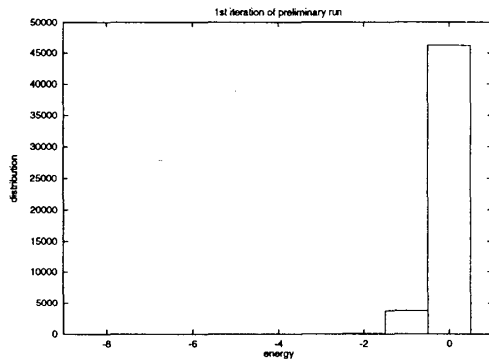
この小節では重み関数の学習の性能に関して Multicanonical と Multi-Self-Overlap Ensemble の効率比較を行なう。ここで用いるモデルは 2 次元 HP model で、配列は $P^3(HP^6)^7HP^3$ というものである。このモデルの基底状態のエネルギーは $E = -8$ で、2 つの疎水性コアを作る。基底状態は、2 つの疎水性コアを作る P モノマーの組み合わせ方によって、Figure 17 の様に 3 つのグループに分類することができる。グループ 1 は、(4,11,18,25) と (18,25,32,39) 番目の H モノマーからなる 2 つの疎水性コアをつくり、グループ 2 は、(4,11,18,53) と (18,25,32,39)、もしくは (4,39,46,53) と (11,18,25,32) からなり、グループ 3 は、(4,11,46,53) と (18,25,32,39) からなる。これらの基底状態の構造は大きく異なる為にそれぞれの基底状態グループ間での移り変わりは、一旦鎖が開いてからでないと実現しない。¹³このように、この配列は基底状態等が自明であるにもかかわらず、シミュレーションが難しいという側面も持っており、効率比較に敵しているといえよう。Multi-Self-Overlap のムーブとしては前小節で使った 8 つのムーブを用いた。パラメーターは $V_{max} = 4, E_{min} = -8$ を使った。一方、Multicanonical の方は前小節で使った (1)~(6) の他に、(7) 180° crankshaft と (8) three-bead J-flip を用いた。又、1 モンテカルロステップ (MCS) はいずれも「すべてのモノマーに対して全てのムーブのトライアルする」ことによって 1MCS を定義した。

7.3.3 Comparison of Preliminary run

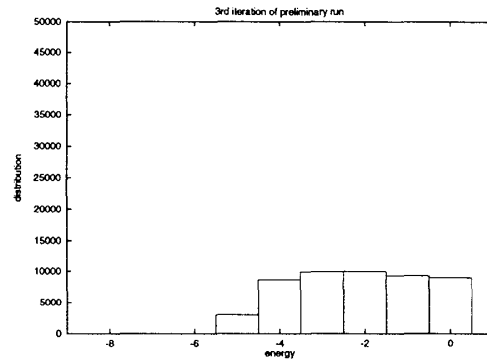
まず、重み関数を求める “Preliminary run” の比較を行なう。ここでは Multicanonical と、Multi-Self-Overlap 共に一回の学習につき 10000MCS を捨て、50000MCS を使った。Figure. 18 に Multicanonical での重み関数の学習によってヒストグラムが平らになっていく様子を示した。Figure.18 から分かるように、始めから 5 回目までは順当にヒストグラムが平らになっていくが、¹⁴その後はヒストグラムが“振

¹³ただしこれは今用いているムーブを用いた場合そうになっている、ということである。しかし、エネルギーが上がらずにグループ間を移り買われるムーブは非常に複雑であり、この配列には有用かもしれないが、一般性がない。ここではアルゴリズムのチェックが目的なので特変したムーブは使わない方が良さだろう。

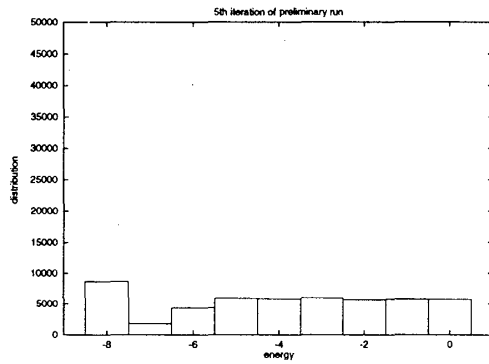
¹⁴一見、5 回目でヒストグラムが平らになったかのように見えるが、実はこの重みで長い Run を走らせると、ほとんど基底状態を拾わない、もしくは、基底状態に“溜ってしまう”。すなわち、5 回目でヒストグラムが平らになったのは完全に偶然



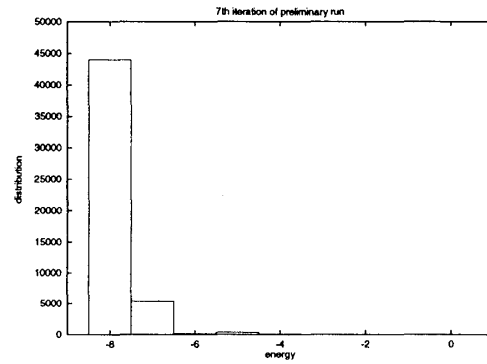
(a) 1 回目



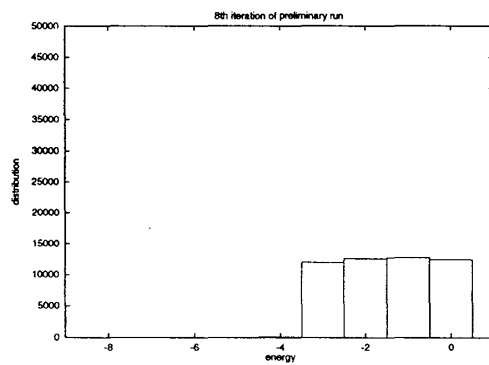
(b) 3 回目



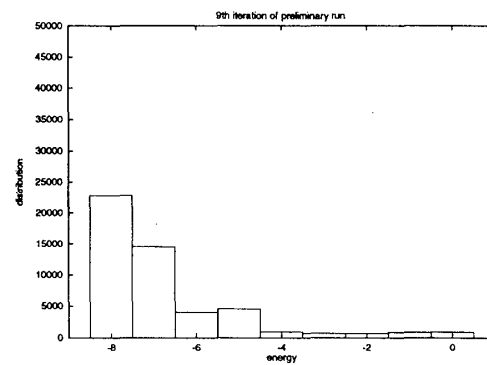
(c) 5 回目



(d) 7 回目

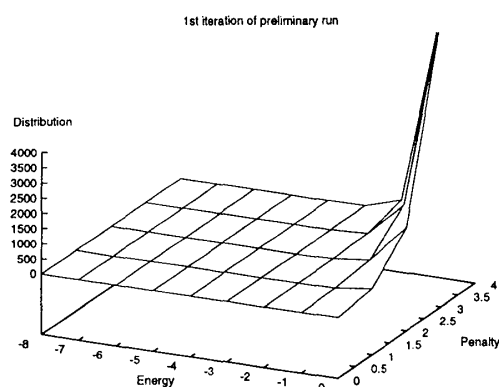


(e) 8 回目

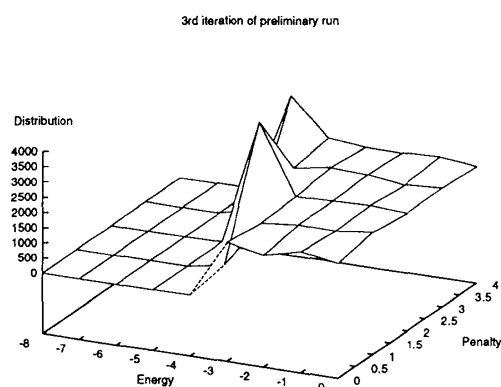


(f) 9 回目

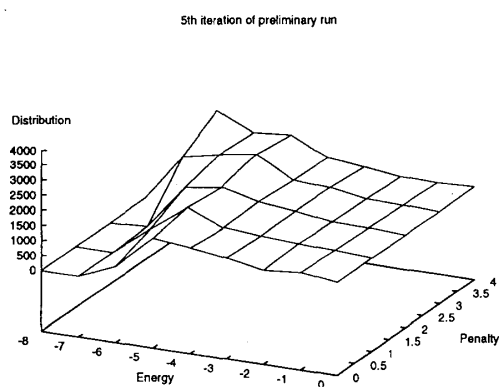
図 18: 重み関数の学習によって2次元ヒストグラムが平らになって行く様子



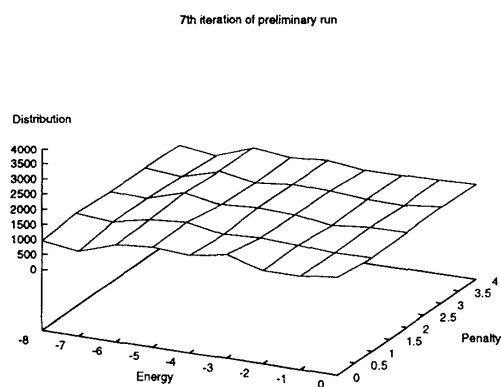
(a) 1回目



(b) 3回目



(c) 5回目



(d) 7回目

図 19: 重み関数の学習によって2次元ヒストグラムが平らになって行く様子

動”してしまう。このことから分かることは、「50000MCSでは学習のモンテカルロステップとしては足りない」ということである。実際、十分ヒストグラムが平らになる様な重み関数を求める為には 10^6 ものモンテカルロステップを使った学習が必要であった。

一方、Multi-Self-Overlap の方を見てみよう。(Figure. 19) 排除体積条件を緩めた為に、状態が増えてしまい多くのモンテカルロステップが必要になると思われるかもしれないが、Figure.19に示したように、50000MCS の Run を7回行なったところで2次元ヒストグラムは十分平らになり、その後、学習を続けてもヒストグラムは安定している。また、7回目に使った重み関数を用いて長い Run を行なってもヒストグラムは50000MCSのものとはほとんど同じである。Figure.20に、この重み関数を用いて 2.5×10^6 MCS の Run を行なって得られたヒストグラムを示してある。

結局、ここで明らかになったことは、「ヒストグラムが平らになるような、重み関数を求める際、Multi-Self-Overlapの方が圧倒的に少ないモンテカルロ数で、求めることが出来る」ということである。

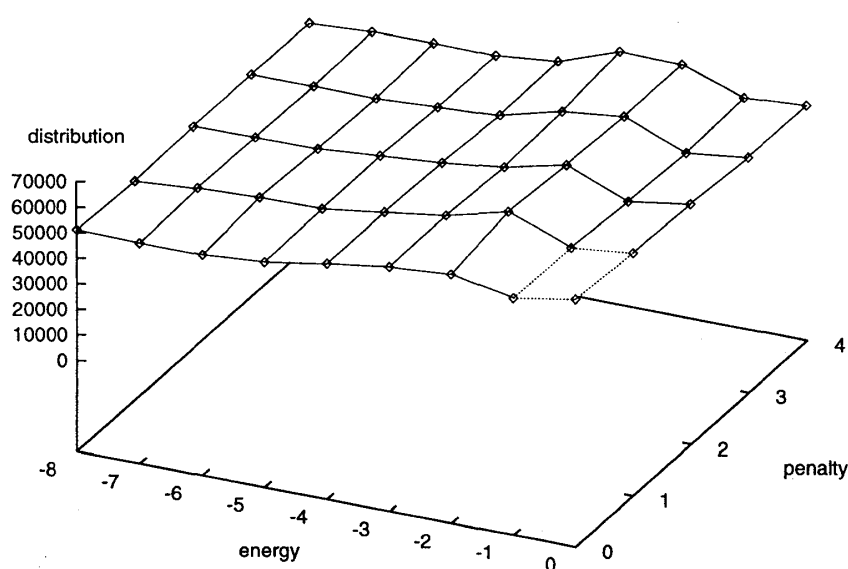


図 20: 最終的に得られた2次元ヒストグラム

7.3.4 Comparison of measurement run

次に、十分にヒストグラムが平らになる重み関数が求まったとして、その重みを使ってシミュレーションをしたときの効率を比較する。ここでは主に“緩和の速さ”に着目する。両者ともに 10000 MCS を捨てた後の 1×10^6 MCS の Run を行なった結果について述べる。まず、エネルギーの時系列を Figure.21(a) に示す。

明らかに Multi-Self-Overlap の方が上下に激しく振動していることがわかる。このことから Multi-Self-Overlap の方が遥かに緩和が速いことがわかる。ここで注目すべきことは Multi-Self-Overlap の方は“排除体積条件を満たした時だけプロットしてある”ということである。¹⁵

次に基底状態のグループの移り変わりを見てみよう。Figure.21(b) に「エネルギーが $E = -8$ になったときだけその基底状態がどの基底状態のグループに属しているか」をプロットしたものである。この図から、圧倒的に Multi-Self-Overlap の方が頻繁に基底状態のグループ間を移り変わっていることがわかる。

結局、ここで明らかになったことは「Multi-Self-Overlap の方が遥かに短い時間で重要な情報を得ることが出来る」ということである。

¹⁵ 今の場合、 $V_{max} = 4$ にしてあるので、プロットしてある点の数は、Multicanonical の約 1/5 しかない。

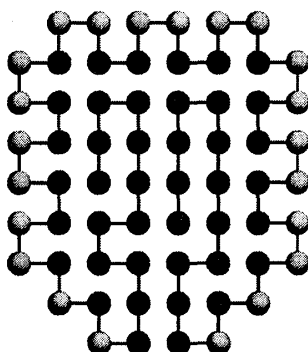


図 22: 2 次元 HP モデルの基底状態の構造の例

である。重み関数の学習には Dec Alpha 500MHz のマシンを用いて約 50 時間を要した。これは決して速いとはいえないが、重み関数が求まってしまうえば任意の温度の熱力学量が計算できることを考慮すれば、これはリーズナブルな時間であるといえる。

この配列を排除体積条件を緩めない Multicanonical を用いたが、基底状態をサンプルすることは 1 週間ほどの時間では不可能であった。

以上のことから強調したいことは「この配列に関して、任意の温度の熱力学量を計算できるアルゴリズムは Multi-Self-Overlap Ensemble しかない」ということである。後の章でこの配列の熱力学的振舞いを議論する。

7.3.7 2d HP model with chain length 100 No.1

2 次元正方格子上的 HP モデルに関して、以下の長さ 100 の配列が文献 [112] で提出された。

$$P_6 H P H_2 P_5 H_3 P H_5 P H_2 P_2 (P_2 H_2)_2 P H_5 P H_{10} P H_2 P H_7 P_{11} H_7 P_2 H P H_3 P_6 H P H_2 \quad (36)$$

表 1: 基底状態探索の結果

<i>Author</i>	<i>Method</i>	<i>Thermal</i>	<i>Energy</i>	<i>Reference</i>
Unger & Moult	Genetic Algorithm	×	-37	[107]
Toma & Toma	Contact Interaction Method	×	-40	[108]
Beutler & Dill	Core-directed chain Growth	×	-42	[109]
Bastolla <i>et al</i>	Pruned-Enriched Rosenbluth Method	○	-40	[110]
Chickenji & Kikuchi & Iba	Multi-Self-Overlap Ensemble	○	-42	[111]

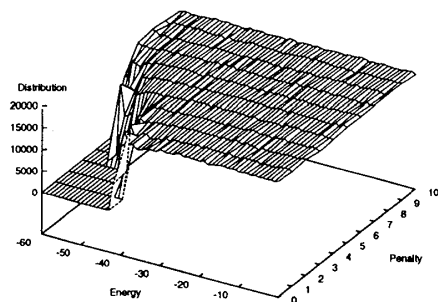


図 23: Multi-Self-Overlap Ensemble によって得られた 2 次元ヒストグラム

まず、文献[112]では“breaking and patching move”と呼ばれる maximally compact 構造のみをサンプリングするモンテカルロ法を用いて基底状態探索（ここではシミュレーテッドアニーリング）を行ない、Figure.24 の様なエネルギー -44 の構造を得、この構造がこの配列の基底状態であると主張した。ところが、後に Grassberger らの開発した“Pruned-Enriched Rosenbluth Method”(通称 PERM)を用いて Bastolla らがこの配列の基底状態探索を行なったところ、Figure.25(a) の様に maximally compact 構造であり、エネルギーが $E = -46$ のものや、maximally compact 構造でないエネルギーが $E = -47$ のものが発見された [110]。

我々も Multi-Self-Overlap を用いてこの配列の基底状態探索を行なった [111]。パラメーターは $V_{max} = 10$ と $E_{min} = -70$ にセットした。結果として、 $E = -47$ の状態が発見でき、それ以上低いエネルギーの低い構造は確認されなかった。Figure.26 に Multi-Self-Overlap を用いて発見した $E = -47$ の構造の例が示してある。 $E = -47$ の状態は非常に多数、しかも構造が大きく異なるものが確認された。つまり、この配列の基底状態は非常に縮退しており、良い配列ではないと言える。また、約 40 時間ほどで ($E = -47$ を含めて) 2 次元ヒストグラムは十分平らになる重みを求めることに成功した。

以上、この配列の Folding Game のまとめを Table2 に示す。

表 2: 基底状態探索の結果

Author	Method	Thermal	Energy	Reference
Ramakrishnan et al	breaking and patching move	×	-44	[112]
Bastolla et al	Pruned-Enriched Rosenbluth Method	○	-47	[110]
Chikenji & Kikuchi & Iba	Multi-Self-Overlap Ensemble	○	-47	[111]

7.3.8 2d HP model with chain length 100 No.2

全述の配列と同じく、2 次元、長さ 100 の配列が文献 [112] で提出された。配列は以下の通り。

$$P_3H_2P_2H_4P_2H_3(PH_2)_3H_2P_8H_6P_2H_6P_9HPH_2PH_{11}P_2H_3PH_2PHP_2HPH_3P_6H_3 \quad (37)$$

前述の配列と同様に、Ramakrishnan ら [112]、Bastolla ら [110]、Chikenji & Kikuchi & Iba [111] によって基底状態探索が行なわれた。Ramakrishnan らは“breaking and patching move”を用いたアニーリングで、Figure.27 に示した様にエネルギー $E = -46$ の構造を発見した。Bastolla らは PERM を用いて Figure.28(a) に示したエネルギー $E = -46$ の構造や、 $E = -47$ の構造を発見した。さらに Bastolla らは「H-P contact を禁止する」といった付加的な方法を加えた PERM を用いることによって Figure.28(b) に示したエネルギー $E = -49$ の構造を発見した。我々も Multi-Self-Overlap を用いて、パラメーター

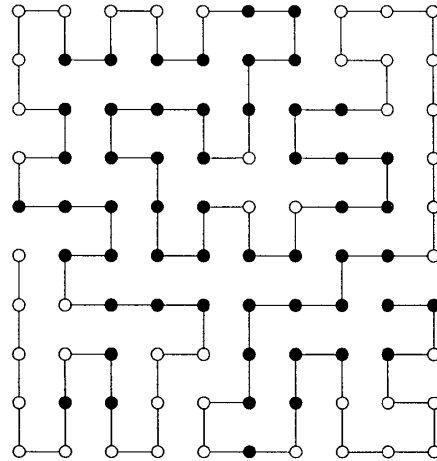
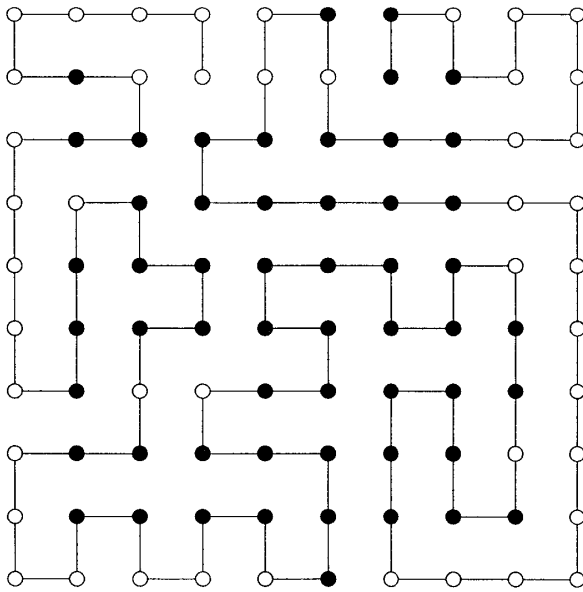
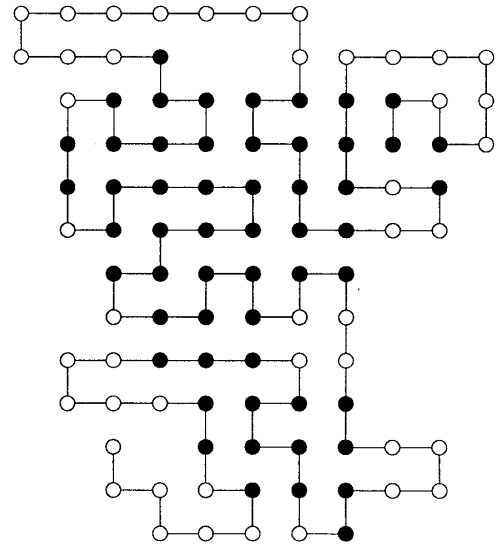


図 24: Ramakrishnan らが主張した基底状態の構造 ($E = -44$)



(a) $E = -46$



(b) $E = -47$

図 25: Grassberger グループが発見した低エネルギー構造の例：(a) $E = -46$ ：(b) $E = -47$

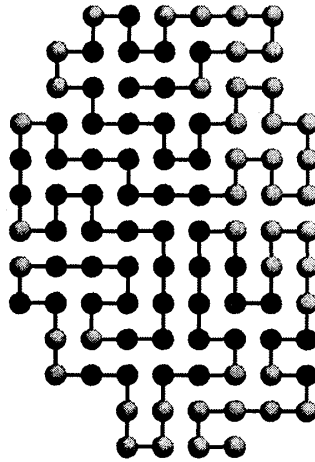


図 26: Multi-Self-Overlap Ensemble を用いて発見した低エネルギー構造 ($E = -47$) の一例

を $V_{max} = 10$ 、 $E_{min} = -65$ にして基底状態を行なった。その結果エネルギー $E = -50$ の構造を発見した。これを Figure.29に示した。Multi-Self-Overlap を用いてこの構造を発見するのに約 48 時間かったが、この状態を含めてヒストグラムを平らにすることは約 1 週間かけたが不可能であった。したがってこの配列に関しては信頼をもって熱力学量を計算できないが、基底状態探索としてはいままでのアルゴリズムでは発見できなかった低いエネルギー構造を発見できたという意味で従来のものと比較しても最も優れていると言って良いだろう。Table3に結果のまとめを記した。

表 3: 基底状態探索の結果

<i>Author</i>	<i>Method</i>	<i>Thermal</i>	<i>Energy</i>	<i>Reference</i>
Ramakrishnan et al	breaking and patching move	×	-47	[112]
Bastolla et al	Pruned-Enriched Rosenbluth Method	○	-49	[110]
Chikenji & Kikuchi & Iba	Multi-Self-Overlap Ensemble	○	-50	[111]

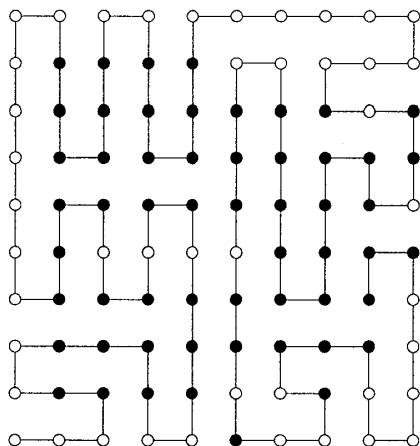
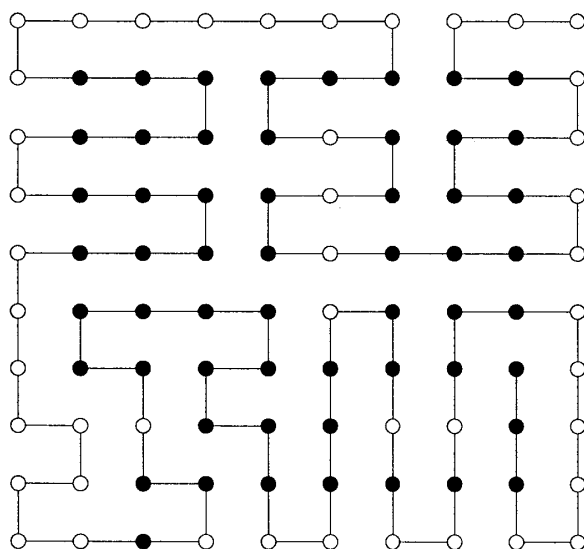
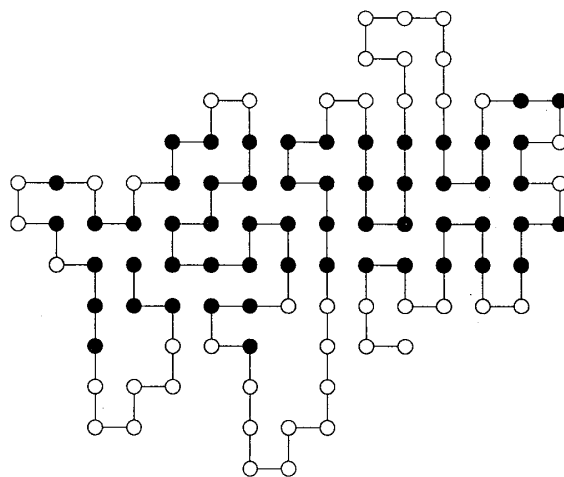


図 27: Ramakrishnan らが主張した基底状態の構造 ($E = -46$)



(a) $E = -46$



(b) $E = -49$

図 28: Grassberger グループが発見した低エネルギー構造: (a) $E = -46$: (b) $E = -49$ の構造。H-P contact を禁止するといった付加的な PERM を用いた。

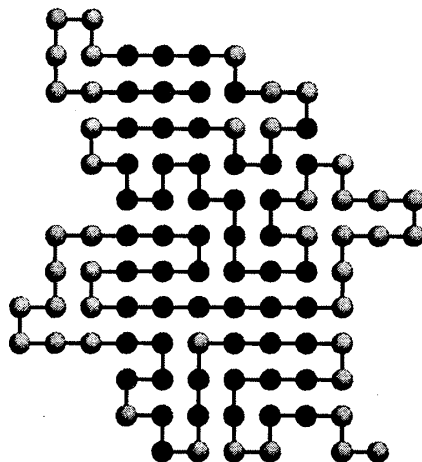


図 29: Multi-Self-Overlap Ensemble を用いて発見した低エネルギー構造 ($E = -50$)

7.4 Thermodynamic Properties of HP model

前々節、前節とで Multi-Self-Overlap Ensemble が「緩和が速いこと」と「基底状態探索としても優れている」ことを示してきた。この2点から、このアルゴリズムが格子タンパク質模型の熱力学量を計算することに対して現在のところ最もすぐれた方法であるといえるだろう。この節では今までの方法では計算不可能であった鎖に対して実際に Multi-Self-Overlap Ensemble を適用し、モデルの熱力学的振舞いを調べる。

7.4.1 2d HP model with chain length 64

まず、前節の一番はじめに基底状態探索を行なった長さ 64 の 2 次元 HP model に関して述べる。配列は前にも記した通りだが、もう一度示しておく。

$$H_{12}P H P H P_2 H_2 P_2 H_2 P_2 H P_2 H_2 P_2 H_2 P_2 H P_2 H_2 P_2 H_2 P_2 H P H P H_{12} \quad (38)$$

まず、比熱と慣性半径の温度依存性を Figure.30 に示す。Figure.30(a) からわかるように、比熱にピークが二つあることがわかる。低温側のピークは約 $T = 0.4$ で、高温側のピークは約 $T = 0.55$ である。又、慣性半径の温度微分にも 2 つのピーク (約 $T = 0.4$ と約 $T = 0.68$) があるが、高温側のピーク的位置が比熱のものと異なる。今までの格子タンパク質模型の研究では、ホモポリマーと同様に「比熱のピークがコイル⇄グロビュール転移」であると考えられてきた [113]。しかし今回の計算ではコイル⇄グロビュール転移の転移点 ($d < R_g > / dT$ のピーク) T_θ と比熱のピーク T_c には有意なズレが認められた。実際、Figure.30 (b) の T_c での慣性半径の値を見ると、すでに鎖はかなりコンパクトな形状をしていることがわかる。文献 [113] を始めとする従来までの研究は鎖の長さが短いものに限られており、そのようなものに関しては $T_c = T_\theta$ であった。しかし、これは鎖が短いが為に T_c と T_θ のズレが“つぶれている”という可能性があり、実際、今回の様な比較的長い ($N = 64$) ものを計算することによって始めてズレが見えてきたと考えられる。

次に、それぞれの転移点で何が起きているかを見るために、その温度の近傍でのエネルギー又は慣性半径の分布を見てみよう。まず、 $T = 0.4$ 近傍でのエネルギーと慣性半径の分布の様子を Figure.31 に示す。いずれも分布のピークが 2 つあり、 $T = 0.41$ で 1 次転移的な転移が起きていることがわかる。

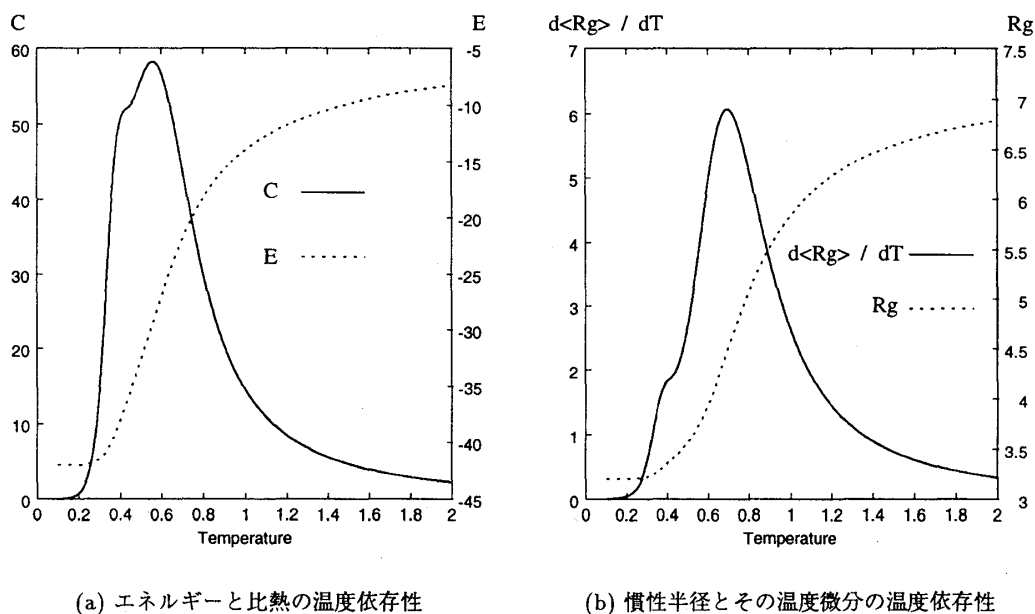


図 30: 物理量の温度依存性：(a) エネルギーと比熱の温度依存性：点線でエネルギー E を、実線で比熱 C を表示してある。(b) 慣性半径とその温度微分の温度依存性点線で慣性半径 R_g を、実線でその温度微分 $d\langle R_g \rangle / dT$ を表示してある。

又、 $T = 0.41$ を境に、それより低温になると急激に基底状態 ($E = -42$) の出現する確率が増加していることから、 $T = 0.41$ はネイティブ (基底状態) への転移が起こる温度である。次に高温側の比熱、もしくは慣性半径の温度微分のピークの近傍の様子を見ることにする。Figure.32に $T = 0.6$ の近傍のエネルギーと慣性半径の分布を示す。Figure.32(a) より、比熱のピーク $T = 0.55$ は2次転移的な振舞いであることがわかる。又、Figure.32(b) を見ても不連続的な転移は見られず、コイル-グロブユール転移は2次転移的もしくは連続的であり、転移点は約 $T = 0.7$ である。より詳しい議論をするためにはさらなる計算が必要だが、以上のことからこのモデルの相図が描け、熱力学的振る舞いの概略が理解できた。

7.4.2 3d HP model with chain length 42 (β -helix sequence)

基底状態の縮退度が4しかない、長さ42の3次元 HP model の配列が文献で [55] 提出された。この配列は Constrained Hydrophobic Core Construction (CHCC)[54] と呼ばれる HP model の基底状態とその縮退度を厳密に計算できる方法を用いて、配列空間の中で縮退度になるべく少ないものを選ぶ様に“最適化”したところ得られた配列である。この配列の基底状態は Figure 33の様に非常に規則的であり β シートが支配的である。また、このような構造は実験的にも観測されており [114]、この構造トポロジーは“ β ヘリックス”と呼ばれている。

この様にこの配列は実際のものとネイティブ構造が非常に似ているという点と、縮退度が非常に少ないという点で、タンパク質の良いモデルになっているといえる。しかし、この配列を CHCC 以外でサンプルできるアルゴリズムは現在までのところ報告されておらず、無論熱力学的振舞いも明らかとなっていない。我々は Multi-Self-Overlap Ensemble をこの配列に適用したところ、基底状態 ($E = -34$) を含めてヒストグラムが平らになる重み関数を求めることに成功した [111]。(パラメーターは $V_{max} = 8$ 、 $E_{min} = -44$ を用いた。)

この小節ではその結果について報告する。まず、大まかな熱力学的振舞いを掴む為に種々の物理量の温度依存性を Figure.34に示す。Figure.34(a) のエントロピーの計算は

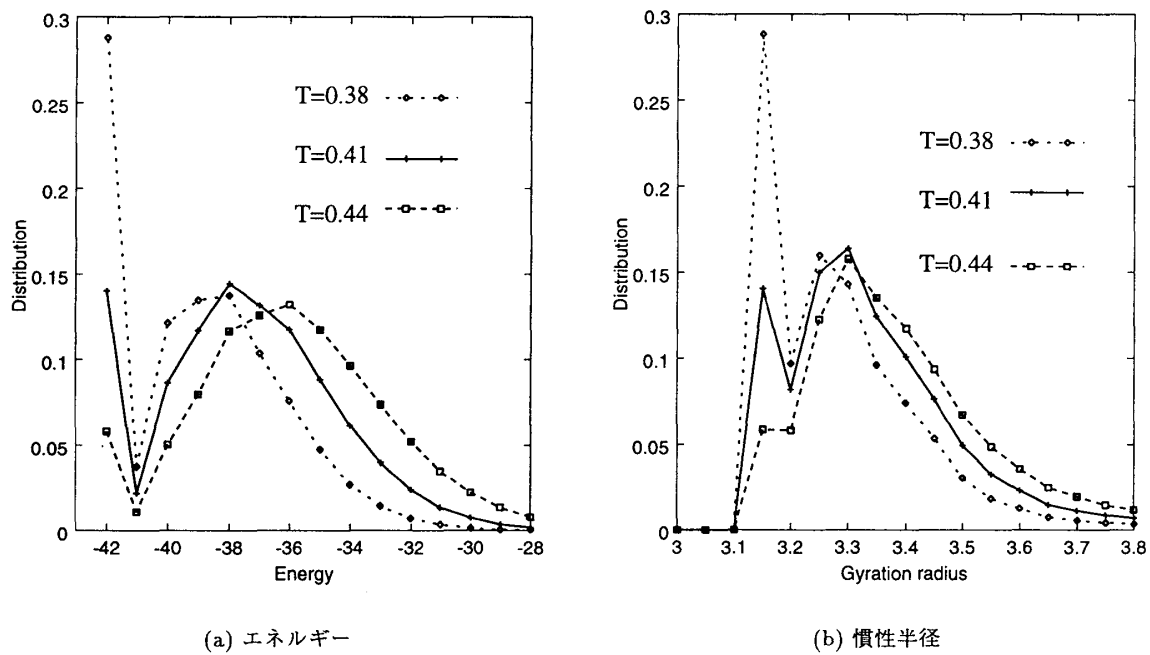


図 31: 転移点（低温側）近傍での分布の振舞い

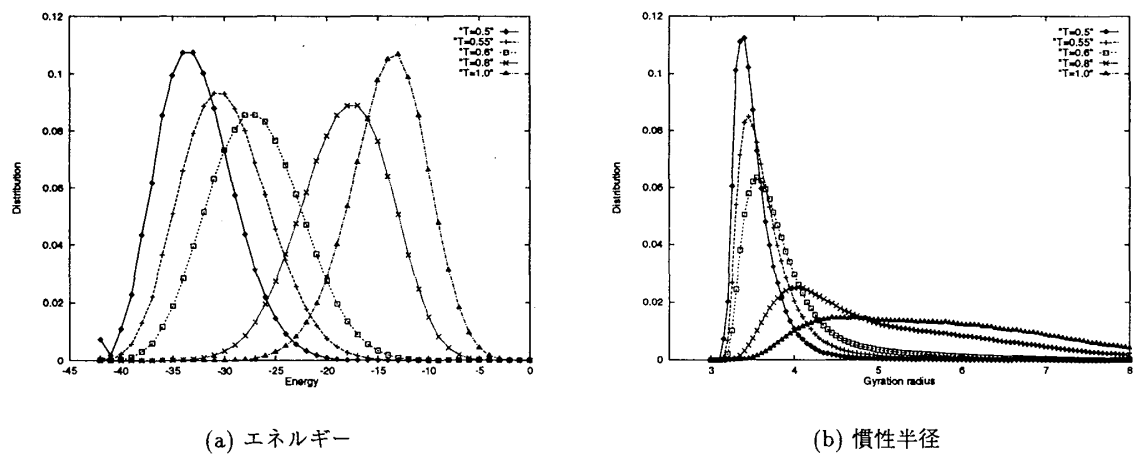


図 32: 転移点（高温側）近傍での分布の振舞い

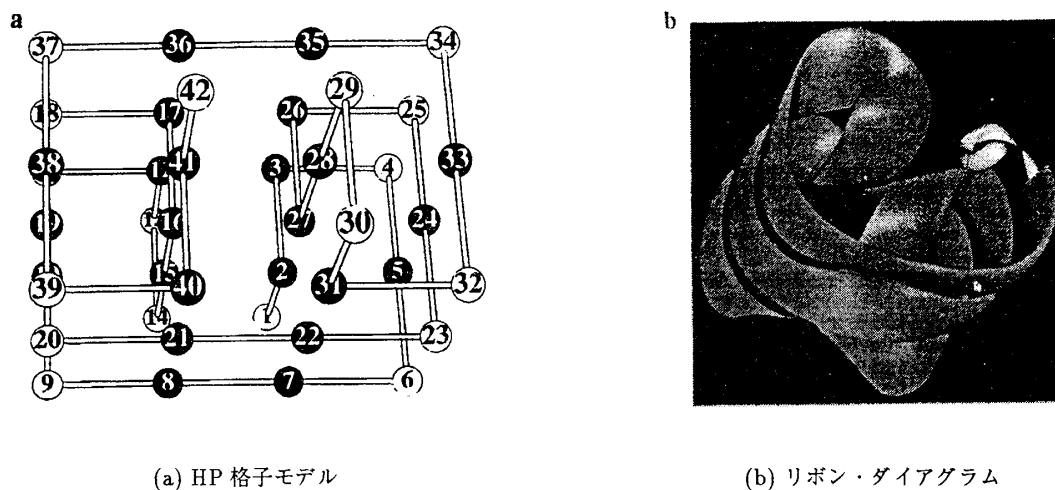


図 33: β ヘリックス構造：(a) HP 格子モデルの β ヘリックス構造（縮退度 4）(b) β ヘリックス構造のリボン・ダイアグラム

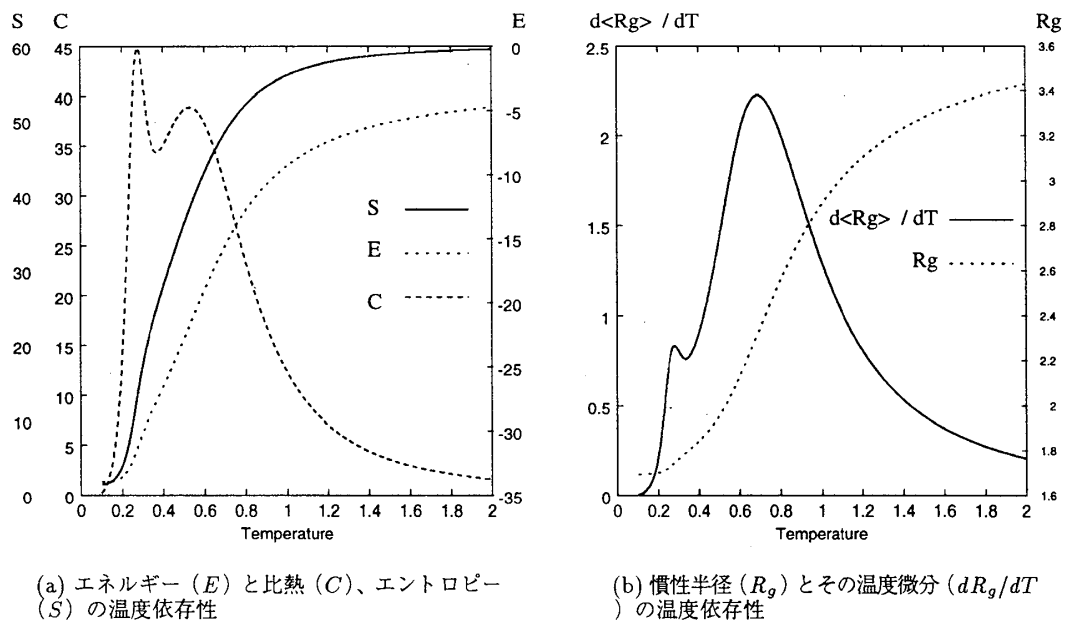
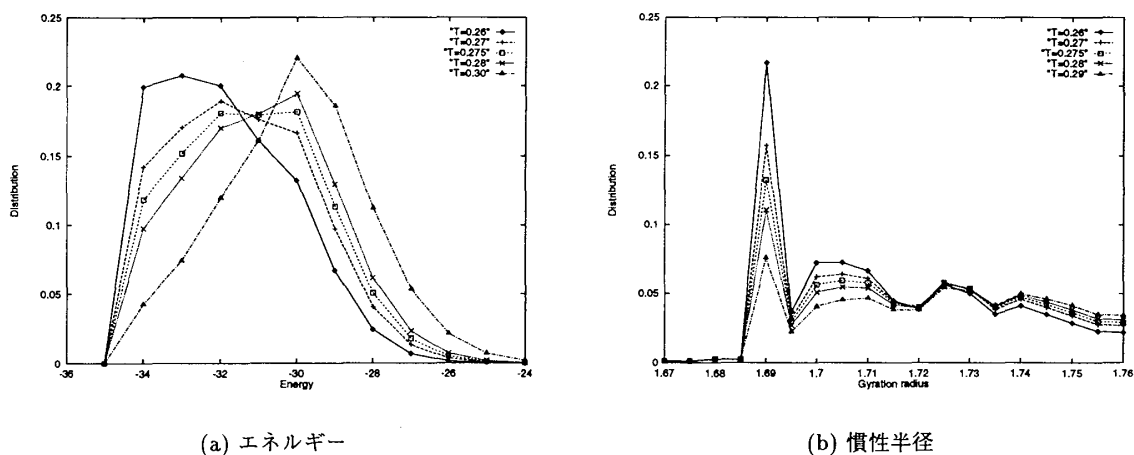


図 34: β -helix の物理量の温度依存性：(a) エネルギーと比熱、エントロピーの温度依存性：点線でエネルギー E を、破線で比熱 C を、又実線でエントロピー S を表示してある。(b) 慣性半径とその温度微分の温度依存性点線で慣性半径 R_g を、実線でその温度微分 $d\langle R_g \rangle / dT$ を表示してある。

図 35: 転移点 ($T=0.28$) 近傍での分布の振舞い

$$dS = \frac{1}{T} C_v dT \quad (39)$$

を用いて、比熱を数値積分することにより求めた。ここで、積分領域の下限を $T_0 = 0.1$ とし、この温度では基底状態のみしか出現しないと近似してエントロピーが $S(T_0) = \log(4)$ だとみなして計算した。¹⁷Figure 34(a) の比熱のプロットを見ると $T = 0.28$ の近傍と $T = 0.52$ の近傍でピークを持っているのがわかる。又、Figure 34(b) の慣性半径の温度微分のプロットからもやはり $T = 0.28$ と $T = 0.69$ の近傍でピークを持っている。やはり前小節で扱った 2 次元のモデルと同様、低温側のピークは両者ともに同じ温度だが、高温側のピークは有意にズレている。¹⁸

より深い理解のためには、例えばコンタクトマップ等、さらなる計算が必要であるが、Multi-Self-Overlap Ensemble を用いることによって初めてこのような計算ができるのであるという事を強調したい。

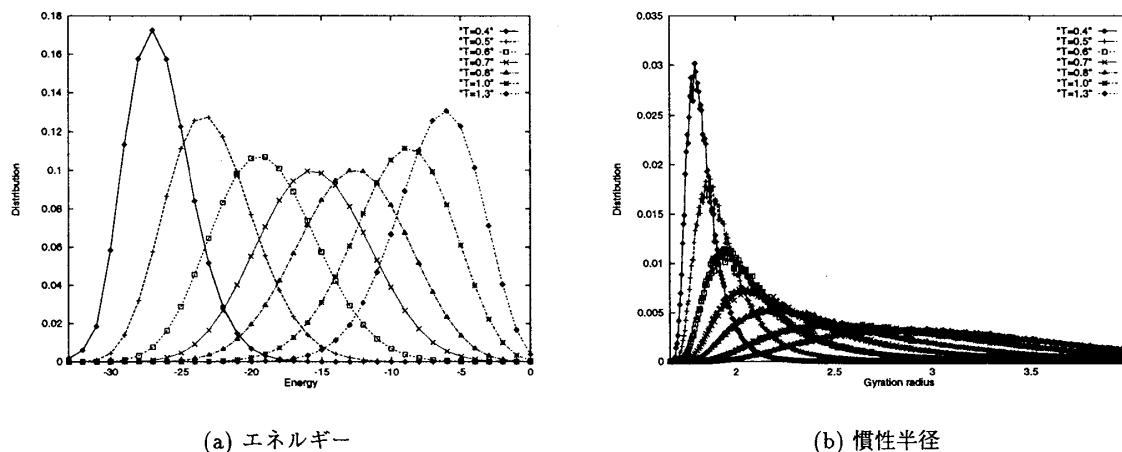
8 Summary

タンパク質の折れ畳み問題を理解するための 1 つのアプローチとしての格子タンパク質模型の研究と概観し、これから必要な研究として、なるべく正確に比較的長い格子タンパク質模型の熱力学を議論する必要があることを指摘した。そのための手法としては何が適切かを議論し、我々は「排除体積条件をシステムティックに緩めるモンテカルロ法」— Multi-Self-Overlap Ensemble — を開発し、その有効性の検証を行なった。現在までに存在する熱力学量を求めるアルゴリズムと我々のアルゴリズムを比較し、格子タンパク質模型の熱力学を議論するのに最も敵した方法が Multi-Self-Overlap Ensemble であることを示した。したがって、格子タンパク質模型の熱力学的な振舞いを精度の良い計算で議論できる準備ができたといえる。具体的な熱力学的振舞いの研究は現在も進行中であるが、実際にいくつかの比較的長い模型に対して精度良い計算が出来た。

これからの課題としてはモルテングロビュール状態などの熱力学的、動力学的役割を解明し、折れ畳み機構の理解を一步でも進めることが重要であると考ええる。

¹⁷繰り返すが、この配列の基底状態は 4 つに縮退している

¹⁸ここでは予備的な計算であった為 1 回しか計算を行なっており、統計誤差を計算していないが、十分長い Run であったので信頼置ける結果である。また、確認のためにもう一回違う乱数でシミュレーションを行なったが、ほとんど同じ結果が得られたことも記しておく

図 36: $T = 0.4$ より高温側での分布の振舞い

9 Acknowledgments

本研究を行なうにあたり、指導教官である菊池誠助教授、共同研究者である統計数理研究所の伊庭幸人助手には本質的な議論から、下らない質問にも答えて頂き、さらには私の大量の計算ミス（バグ）にもめげずに研究を温かく見守っていただきました。また、阿久津泰弘教授、時田恵一郎助手、宮下研M2の有森君にはタンパク質の論文を100本読むセミナー（プロセミ）に、忙しいなか参加していただき、論文紹介や、貴重なコメントをして頂きました。このセミナーは本研究において非常に重要で、有意義なものでした。さらに、阿久津研究室の永尾太郎助手、奥西さん、日永田さん、前島君には普段から様々な面でサポートしていただきました。共同研究者であり、日頃からタンパク質の議論に火花を散らせていただいた巽さんとの議論も大いに役立ちました。各氏に深く感謝します。

参考文献

- [1] C. Branden and J. Tooze, *Introduction to Protein Structure* Garland (1991).
- [2] 中村春木・有坂文雄 [編集], タンパク質のかたちと物性, 共立出版 (1997).
- [3] A.C. Anfinsen, *Science* **181**, 223-230 (1973).
- [4] C. Levinthal, *J. Chim. Phys.* **65**, 44-45 (1968).
- [5] 西川 建, 蛋白質 — この絶妙な設計物 — 吉岡書店, 79-102 (1994).
- [6] 岡本裕幸, 数理科学 **401** 11月号, 66-73 (1996).
- [7] 岡本裕幸, 日本物理学会誌 **51-4** 4月号, 279-287 (1997).
- [8] 岡本裕幸, 物性研究 **70-6** 9月号, 719-742 (1998).
- [9] Y. Duan and P.A. Kollman, *Science* **282**, 740-744 (1998).
- [10] Y. Okamoto, T. Kikuchi, T. Nakazawa, and H. Kawai, *Int. J. Peptide. Prot. Res* **42**, 300-303 (1993).
- [11] 笹井理生, 日本物理学会誌 **52-12** 12月号, 901-908 (1997).
- [12] H.S. Chan and K.A. Dill (美宅成樹 [訳]), *パリティー* **8-12** 12月号, 24-35 (1993).
- [13] H. Frauenfelder, and P.G. Wolynes (倭 剛久 [訳]), *パリティー* **9-11** 11月号, 4-13 (1994).
- [14] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, and H.S. Chan, *Protein Science* **4**, 561-602 (1995).
- [15] J.N. Onuchic, Z. Luthey-Schulten, and P.G. Wolynes, *Annu. Rev. Chem.* **48**, 545-600 (1997).
- [16] V.S. Pande, A.Y. Grosberg, and T. Tanaka, *Rev. Mod. Phys.* (in press).
- [17] ローレンス・クラウス [著], 青木薫 [訳], 物理の超発想, 講談社 (1996).
- [18] A.E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. USA* **22**, 439-447 (1936).
- [19] L. Pauling, R.B. Corey, H.R. Branson, *Proc. Natl. Acad. Sci. USA* **37**, 205-211 (1951).
- [20] L. Pauling and R.B. Corey, *Proc. Natl. Acad. Sci. USA* **37**, 235-240 (1951).
- [21] L. Pauling and R.B. Corey, *Proc. Natl. Acad. Sci. USA* **37** 251-256 (1951).

- [22] L. Pauling and R.B. Corey, *Proc. Natl. Acad. Sci. USA* **37**, 272-281 (1951).
- [23] L. Pauling and R.B. Corey, *Proc. Natl. Acad. Sci. USA* **37**, 729-740 (1951).
- [24] J.C. Kendrew, R.E. Dickerson, B.E. Strandberg, R.G. Hart, D.R. Dabies, *Nature* **185**, 422-427 (1960).
- [25] W. Kauzmann, *Adv. Prot. Chem.* **14**, 1-63 (1959).
- [26] 浜口浩三, 改訂 蛋白質機能の分子論, 学会出版センター (1990).
- [27] C.B. Anfinsen and H.A. Scheraga, *Adv. Prot. Chem.* **29**, 205-300 (1975).
- [28] K.A. Dill, *Biochemistry* **24**, 1501-1509 (1985).
- [29] K.A. Dill, *Biochemistry* **29**, 7133-7155 (1990).
- [30] S. Kamtekar, J.M. Schiffer, H. Xiong, J.M. Badik, and M.H. Hecht, *Science* **262**, 1680-1685 (1993).
- [31] K.F. Lau, K.A. Dill, *Macromolecules* **22**, 3986-3997 (1989).
- [32] P.L. Privalov and S.J. Gill, *Adv. Prot. Chem.* **39**, 181-234 (1988).
- [33] H. Taketomi, Y. Ueda, and N. Gō (1975) *Int. J. Pept. Protein Res.* **7**, 445-459 (1975).
- [34] J.S. Bryngelson and P.G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84**, 7524-7528 (1987).
- [35] J.S. Bryngelson and P.G. Wolynes, *Biopolymer* **30**, 177-188 (1990).
- [36] S. Miyazawa and R.L. Jernigan, *Macromolecules* **18**, 534-552 (1985).
- [37] S. Miyazawa and R.L. Jernigan, *J. Mol. Biol.* **256**, 623-644 (1996).
- [38] 宮澤三造, 日本物理学会誌 **52-5** 5月号, 312-319 (1997).
- [39] H. Li, C. Tang, and N.S. Wingreen, *Phys. Rev. Lett.* **79**, 765-768 (1997).
- [40] H. Li, R. Helling, C. Tang, and N.S. Wingreen, *Science* **273**, 666-669 (1996).
- [41] C. Micheletti, J.R. Banavar, A. Maritan, and F. Seno, *Phys. Rev. Lett.* **80**, 5683-5686 (1998).
- [42] K. Iguchi, *Mod. Phys. Lett. B* **12**, 499-506 (1998).
- [43] W.J.C. Orr, *Trans. Faraday. Soc.* **43**, 12-27 (1947).
- [44] A.R. Conway and A.J. Guttmann, *Phys. Rev. Lett.* **77**, 5284-5287 (1996).
- [45] E. Shakhnovich and A. Gutin, *J. Chem. Phys.* **93**, 5967-5971 (1990).
- [46] J. Mai, I.M. Sokolov, and A. Blumen, *J. Chem. Phys.* **106**, 7829-7833 (1997).
- [47] V.S. Pande, C. Joerg, A.Y. Grosberg, and T. Tanaka, *J. Phys. A: Math. Gen.* **27**, 6231-6236 (1994).
- [48] A. Kloczkowski and R.L. Jernigan, *J. Chem. Phys.* **109**, 5134-5146 (1998).
- [49] A. Kloczkowski and R.L. Jernigan, *J. Chem. Phys.* **109**, 5147-5159 (1998).
- [50] E.M. O'Toole and A.Z. Panagiotopoulos, *J. Chem. Phys.* **97**, 8644-8652 (1992).
- [51] M.N. Rosenbluth and A.W. Rosenbluth, *J. Chem. Phys.* **23**, 356 (1955).
- [52] P. Grassberger, *Phys. Rev. E* **56**, 3682-3693 (1997).
- [53] H. Frauenkron, U. Bastolla, E. Gerstner, P. Grassberger, and W. Nadler, *Phys. Rev. Lett.* **80**, 3149-3152 (1998).
- [54] K. Yue, K.A. Dill, *Phys. Rev. E* **48**, 2267-2278 (1993).
- [55] K. Yue, K.A. Dill, *Proc. Natl. Acad. Sci. USA* **92**, 325-329 (1995).
- [56] H.S. Chan and K.A. Dill, *J. Chem. Phys.* **90**, 492-509 (1989).
- [57] H.S. Chan and K.A. Dill, *J. Chem. Phys.* **92**, 3118-3135 (1990).
- [58] K.M. Fiebig and K.A. Dill, *J. Chem. Phys.* **98**, 3475-3487 (1993).
- [59] K.A. Dill, K.M. Fiebig, and H.S. Chan, *Proc. Natl. Acad. Sci. USA* **90**, 1942-1946 (1993).
- [60] E.E. Lattman, K.M. Fiebig, and K.A. Dill, *Biochemistry* **33**, 6158-6166 (1994).
- [61] C. Chothia, *Nature* **357**, 543-544 (1992).
- [62] H. Li, C. Tang, and N.S. Wingreen, *Proc. Natl. Acad. Sci. USA* **95**, 4987-4990 (1998).
- [63] A. Irbäck and E. Sandelin, *J. Chem. Phys.* **108**, 2245-2250 (1998).
- [64] M.R. Ejtehadi, N. Hamedami, H. Seyed-Allaek, M. Shahrezaei, and M. Yahyanejad, *J. Phys. A: Math. Gen.* **31**, 6141-6155 (1998).
- [65] R. Melin, H. Li, N.S. Wingreen, C. Tang, *J. Chem. Phys.* **110**, 1252-1262 (1999).
- [66] R. Tatsumi and G. Chikenji, *Phys. Rev. E* **60**, 4696-4700 (1999).
- [67] 巽 理恵, 大阪大学 大学院 修士論文 (1998).
- [68] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes, *Proteins: Struct. Funct. Genet.* **21**, 167-195 (1995).
- [69] K.A. Dill and H.S. Chan, *Nature Struct. Biol.* **4**, 10-19 (1997).
- [70] A.D. Sokal, Monte Carlo methods for the Self-Avoiding Walk, in *Monte Carlo and Molecular Dynamics Simulation in Polymer Science*, ed. K. Binder (Oxford University Press, New York, Oxford, 1995) 47.
- [71] N. Madras and A.D. Sokal, *J. Stat. Phys.* **50**, 109-186 (1988).
- [72] K. Binder and D.W. Heermann, *Monte Carlo Simulation in Statistical Physics - An Introduction*, 2nd. edition, (Springer, 1992).
- [73] 宮下精二, 熱・統計力学, (倍風館 1993).
- [74] 伊庭幸人, *ISM Research Memorandum* **635** (1997). (<http://www.ism.ac.jp/iba/jdoc.html> より入手可能)
- [75] 菊池誠, 物性研究 **11** (1994).

- [76] A. Sali, E. Shakhnovich, and M. Karplus, *Nature* **369**, 248-251 (1994).
- [77] P.E. Leopold, M. Montal, and J.N. Onuchic, *Proc. Natl. Acad. Sci. USA* **89**, 8721-8725 (1992).
- [78] D. K. Klimov and D. Thirumalai, *Phys. Rev. Lett.* **76**, 4070-4073 (1996).
- [79] C.J. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **90**, 6369-6372 (1993).
- [80] A.M. Gutin, V.I. Abkevich, and E.I. Shakhnovich, *Phys. Rev. Lett* **77**, 5433-5436 (1996).
- [81] E. Shakhnovich, G. Farztdinov, A.M. Gutin, and M. Karplus, *Phys. Rev. Lett.* **67**, 1665-1668 (1991).
- [82] N.D. Socci and J.N. Onuchic, *J. Chem. Phys.* **103**, 4732-4744 (1995).
- [83] N.D. Socci and J.N. Onuchic, P.G. Wolynes, *J. Chem. Phys.* **104**, 5760-5868 (1996).
- [84] M. Cieplak, S. Vishveshwara and J.R. Banavar, *Phys. Rev. Lett.* **77**, 3681-3684 (1996).
- [85] R. Miller, C.A. Danko, M.J. Fasolka, A.C. Balazs, H.S. Chan and K.A. Dill, *J. Chem. Phys.* **96**, 768-780 (1992).
- [86] L.A. Mirny, V.I. Abkevich, and E.I. Shakhnovich, *Proc. Natl. Acad. Sci. USA* **95**, 4976-4981 (1998).
- [87] A. Sali, E. Shakhnovich and M. Karplus, *J. Mol. Biol.* **235**, 1614-1636 (1994).
- [88] M. Cieplak, M. Henkel, J. Karbowski, and J.R. Banavar, *Phys. Rev. Lett.* **80**, 3654-3657 (1998).
- [89] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61**, 2635-2638 (1988).
- [90] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **63**, 1195-1198 (1989).
- [91] K.Yue, K.M.Fiebig, P.D.Thomas, H.S.Chan, E.I.Shakhnovich, and K.A.Dill, *Proc. Natl. Acad. Sci. USA* **92**, 325-329 (1995).
- [92] B.A. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249-253 (1991).
- [93] B.A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9-12 (1992).
- [94] B.A. Berg and T. Celik, *Phys. Rev. Lett.* **69**, 2292-2295 (1992).
- [95] U.H.E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333-1338 (1993).
- [96] N. Urakami and M. Takasu, *J. Phys. Soc. Jpn* **65**, 2694-2699 (1996).
- [97] N. Urakami and M. Takasu, *Molecular Simulation* **19**, 63-73 (1997).
- [98] J. Lee, *Phys. Rev. Lett.* **71**, 211-214 (1993).
- [99] B.A. Berg, U.H.E. Hansmann, and Y. Okamoto, *J. Phys. Chem.* **99**, 2236-2237 (1995).
- [100] B.A. Berg, *J. Stat. Phys.* **82**, 323-342 (1996).
- [101] G.R. Smith and A.D. Bruce, *J. Phys. A : Math. Gen.* **28**, 6623-6643 (1995).
- [102] U.H.E. Hansmann and Y. Okamoto, *Phys. Rev. E* **56**, 2228-2233 (1997).
- [103] U.H.E. Hansmann, *Phys. Rev. E* **56**, 6200-6203 (1997).
- [104] Y. Iba, G. Chikenji, and M. Kikuchi, *J. Phys. Soc. Jpn.* **67**, 3327-3330 (1998).
- [105] I. Shteto, J. Linares, and F. Varret, *Phys. Rev. E* **56**, 5128-5137 (1997).
- [106] J. Higo, N. Nakajima, H. Shiraki, A. Kidera, H. Nakamura, *J. Comp. Chem.* **18**, 2086-2092 (1997).
- [107] R. Unger and J. Moulton, *J. Mol. Biol.* **231**, 75-81 (1993).
- [108] L. Toma and S. Toma, *Protein Science* **5**, 147-153 (1996).
- [109] T.C. Beutler and K.A. Dill, *Protein Science* **5**, 2037-2043 (1996).
- [110] U. Bastolla, H. Frauenkron, E. Gerstner, P. Grassberger and W. Nadler, *Proteins: Struct.Funct.Genet.* **32**, 52-66 (1998).
- [111] G. Chikenji, M. Kikuchi, and Y. Iba,, *Phys. Rev. Lett.* **83**, 1886-1889 (1999).
- [112] R. Ramakrishnan, B. Ramachandran, and J.F. Pekny, *J. Chem. Phys.* **106**, 2418-2425 (1997).
- [113] D.K. Klimov and D. Thirumalai, *J. Chem. Phys.* **109**, 4119-4125 (1998).
- [114] M.D. Yonder, N.T. Keen, and F. Jurnak, *Science* **260**, 1503-1507 (1993).