

学位論文

**The Study of Generalized-Ensemble Algorithms
for Determining the Density of States:
Applications to the Ising Model, Ice I_h Model,
and Biomolecular System**

状態密度推定のための
拡張アンサンブル法の研究：
イジング模型、氷 I_h 系および
生体高分子系への応用

名古屋大学 大学院理学研究科 物質理学専攻(物理系)
理論生物化学物理(TB)研究室

林 卓弥

2021年3月25日

目次

第1章	序論	1
第2章	計算機シミュレーション手法	3
2.1	カノニカル分布	3
2.2	モンテカルロ法	5
2.3	分子動力学法	8
2.4	拡張アンサンブル法	9
2.4.1	マルチカノニカル法	11
2.4.2	ワン-ランダウ法	14
2.4.3	レプリカ交換法	16
2.4.4	マルチカノニカルレプリカ交換法	19
2.4.5	レプリカ交換ワン-ランダウ法	23
2.5	REWL-MUCAREM プロトコル	25
第3章	Ising model	29
3.1	導入	29
3.2	計算モデルと初期条件	31
3.2.1	2次元イジング模型	31
3.2.2	初期条件	31
3.3	結果と考察	31
3.3.1	比熱の比較	31
3.3.2	状態密度の比較	34
3.4	結論	36
第4章	The Residual Entropy of Ice I_h	41
4.1	導入	41
4.2	手法	42
4.2.1	Ice rule と残余エントロピー	42
4.2.2	2-State Model	44
4.3	計算条件	45
4.4	結果と考察	47
4.4.1	REWL-MUCAREM の動作確認	47
4.4.2	残余エントロピー	53
4.4.3	擬似乱数発生器の評価	55
4.5	結論	57

第 5 章	REWL-MUCAREM with Genetic Algorithm for Protein Folding Problem and Microcanonical Inflection-Point Analysis	61
5.1	導入	61
5.2	Models	62
5.3	シミュレーション手法	63
5.3.1	遺伝的アルゴリズム	63
5.3.2	REWL-MUCAREM with Genetic Algorithm	65
5.3.3	Microcanonical Inflection-Point Analysis Method	68
5.4	初期条件	69
5.5	結果と考察	69
5.6	結論	72
第 6 章	Modified Wang-Landau Algorithm with Lorentzian Fitting Algorithm	75
6.1	導入	75
6.2	Multicanonical method with Lorentzian Fitting Algorithm	75
6.2.1	逐次近似法	75
6.2.2	Lorentz 型の逐次近似法	77
6.3	Lorentzian 型のフィッティングを考慮したワン-ランダウ法	77
6.4	結論	79
第 7 章	本論文の結論	83
第 8 章	謝辞	85

第1章 序論

スピニングラス系や生体高分子系のような多自由度複雑系におけるエネルギーランドスケープではエネルギーの極小値が無数に存在する。ボルツマン因子で特徴付けられた、温度一定のカノニカルアンサンブルを再現するような、従来のモンテカルロ (Monte Carlo :MC) 法や分子動力学法 (Molecular Dynamics :MD) によるシミュレーションを実行すると、エネルギーランドスケープの極小値にシミュレーション状態がトラップされるという問題がある。従って、多自由度複雑系における通常のシミュレーションでは、十分な状態空間を探索することができず、シミュレーションの信頼度が低下する。

このような問題に対処するため、拡張アンサンブル法 (Generalized Ensemble Method) と総称される手法が提案されている。拡張アンサンブル法では、非ボルツマン因子によって特徴付けられる人工的な統計力学集団に従ったシミュレーションが実行される。これにより、ポテンシャルエネルギー空間や他のパラメータ空間上のランダムウォークを実現することで、エネルギー極小値にトラップされるという問題を回避する。結果として、従来の手法に比べて広い状態空間を探索することが可能となる。また、拡張アンサンブル法によって得られたサンプルデータを再重法と呼ばれる手法によって解析することにより、様々な物理量の任意の温度における期待値を得ることができる。代表的な拡張アンサンブル法として、マルチカノニカル法 (Multicanonical algorithm:MUCA) やレプリカ交換法 (Replica-Exchange method:REM) などが広く用いられる。これらの代表的な拡張アンサンブル法の開発後、シミュレーションの目的に合わせて様々な改良アルゴリズムが提案されてきた。

拡張アンサンブル法によって得られる重要な物理量として、状態密度 (Density of States :DOS) がある。これは、DOS を推定することで任意の物理量の期待値や自由エネルギーを統計力学的に推定できるためである。しかし拡張アンサンブル法であっても DOS を高精度で推定することが困難な場合が多い。本研究では、DOS を高精度で推定する計算プロトコル (REWL-MUCAREM) を提案した (第2章)。REWL-MUCAREM では、2つの拡張アンサンブル法、すなわちレプリカ交換ワン-ランダウ (Replica-Exchange Wang-Landau: REWL) 法とマルチカノニカルレプリカ交換 (Multicanonical Replica-Exchange: MUCAREM) 法を適切な順序で組み合わせ、DOS を推定する。

第3章では、REWL-MUCAREM の有用性を検証するため、REWL、MUCAREM、REWL-MUCAREM の3つの方法で正方格子イジングモデルの MC シミュレーションを実行し、得られた DOS を比較した。その結果、REWL-MUCAREM によって得られた DOS は、REWL および MUCAREM によって単独で推定された DOS よりも正確であることが示された。

第4章では、REWL-MUCAREM によって氷 I_h 系の残余エントロピーを推定した。氷 I_h 系は、絶対零度で0ではないエントロピー (残余エントロピー) が実験的に観測された系である。その残余エントロピーの起源は、氷中の水分子の配向の規則 (Ice rule) に由来するものとして説明された。氷 I_h 系の残余エントロピーは、理論的にも実験的にも未だに推

定することが難しい物理量の1つであるが、近年の計算科学の発展に伴って、コンピューター・シミュレーションによる残余エントロピーの推定値は、従来のアプローチで得られる推定値と同等の見積りを与えるようになってきた。現在では、氷 I_h の残余エントロピーは計算アルゴリズムを検証するベンチマークのひとつになりつつある。しかし、シミュレーションによる推定値の間には不一致が残っている。すなわち氷 I_h の正確な残余エントロピーはいまだ得られていない。本研究では、REWL-MUCAREM MC プロトコルによって氷 I_h の残余エントロピーの推定した。その結果は、本研究グループの先行研究の推定値を改善し、他のシミュレーション手法を使用したグループと良く一致した。また第4章では、乱数発生器の生成する乱数の均一性の重要性についても議論した。これによりマルチカノニカル法を大規模複雑系に適用する場合、均一性の高い乱数を用いなければ信頼性の低い結果につながるということがわかった。そして、マルチカノニカル法を使用した本研究グループのひとつの先行研究の結果が他の研究結果と一致していなかった原因が、乱数発生プログラムに起因することを見出した。

第5章では、タンパク質系のフォールディング問題について、REWL-MUCAREM プロトコルを適用した。計算対象は、ヘリックス-コイル構造転移を示すアラニン重合体とした。ここでは、タンパク質系における REWL-MUCAREM MC プロトコルのサンプリング効率を上げるために、遺伝的アルゴリズムを導入した REWL-MUCAREM を提案した。結果として、タンパク質のような複雑多体系においても DOS を推定することができ、REWL-MUCAREM の実用性が示された。またヘリックス-コイル構造転移の解析では、ミクロカノニカルエントロピーを使用した構造転移分類法である、Microcanonical Inflection-Point Analysis Method を使用した。この手法で定義されるミクロカノニカルエントロピーは DOS と密接な関係があるため、DOS を高精度で推定する REWL-MUCAREM は Microcanonical Inflection-Point Analysis Method と相性が良い。結果として、アラニン重合体のヘリックス-コイル構造転移は従来の分類法とは異なる解析手法によって、水の影響を考慮しなかった場合に1次転移、陰溶媒モデルによって水の影響を考慮した場合に2次転移のような振る舞いをすることを示した。

第6章では、REWL-MUCAREM 法の分子動力学法への拡張を考えた。通常の Wang-Landau 分子動力学法アルゴリズムでは、シミュレーションの初期段階にアルゴリズムとして非物理的な状態が生じやすく、シミュレーションが破綻することが報告されている。本章では、マルチカノニカル MD 法で提案された Lorentzian Fitting Algorithm を Wang-Landau 法に組み込むアルゴリズムを提案した。

第2章 計算機シミュレーション手法

本章では、本研究で提案した状態密度 (DOS) を高精度で推定するための計算プロトコル、REWL-MUCAREM[1] を導入する。はじめに統計力学の基礎としてカノニカルアンサンブルを復習した後、基本的なシミュレーション手法であるモンテカルロ法および分子動力学法について解説する。次に、代表的な拡張アンサンブル法であるマルチカノニカル法 [2, 3, 4]、ワン-ランダウ法 [5, 6] およびレプリカ交換法 [7, 8, 9] について解説する。そして、これら3つの拡張アンサンブル法の応用として、マルチカノニカルレプリカ交換法 [11, 12, 13] およびレプリカ交換ワン-ランダウ法 [14, 15] について解説する。以上のシミュレーション手法の詳細な解説は、例えば文献 [16]–[20] などが詳しい。最後に、本研究で提案した REWL-MUCAREM の計算プロトコルを解説する。

2.1 カノニカル分布

統計力学で扱う系は物質 (原子) の集団である。この集団がどのような環境 (外的条件)のもとにあるかによって、系の振る舞いが規定される。個々の統計力学的集団 (統計力学的アンサンブル) とは、それらの環境を指定した物質の集団を表している。計算機シミュレーションの目的のひとつは、統計力学的集団に従う系を計算機上に発生させ、その微視的状態の具体的なサンプルを得て、興味のある物理現象を解析することである。ここで統計力学的集団を計算機上に発生させる手法がモンテカルロ法と分子動力学法であるが、これらの手法を解説する前に、代表的な統計力学的集団であるカノニカルアンサンブルについて復習する。

カノニカルアンサンブル (Canonical Ensemble: CA) とは、温度一定の熱浴中の部分系を表す統計力学的集団である。粒子数 N , 体積 V , 温度 T が一定であるカノニカルアンサンブルの例として、質量が $m_k (k = 1, 2, \dots, N)$ の3次元の粒子系を考える。各粒子の座標と運動量は $q \equiv \{q_1, q_2, \dots, q_N\}$ および $p \equiv \{p_1, p_2, \dots, p_N\}$ とする。ここで、 q_k および p_k は k 番目の粒子の座標と運動量を表す。従って全自由度は $6N$ となる。これらの座標と運動量で張られる位相空間上の一点を微視的状态と呼び、 $x \equiv \{p, q\}$ と定義する。またこのとき、系のハミルトニアン $H(q, p)$ は、運動エネルギー $K(p)$ とポテンシャルエネルギー $E(q)$ を用いて、

$$H(q, p) = K(p) + E(q). \quad (2.1)$$

と表される。

温度一定のカノニカルアンサンブルでは、各微視的状态 x は次のボルツマン因子 $W_{\text{CA}}(x)$ に従った重みで実現される。

$$W_{\text{CA}}(x) = e^{-\beta H(q, p)}. \quad (2.2)$$

ここで β はボルツマン定数 k_B と絶対温度 T から、

$$\beta = \frac{1}{k_B T}. \quad (2.3)$$

である。このとき、系の分配関数 $Z(N; V; T)$ は以下で与えられる (前にかける定数係数を 1 とした)。

$$\begin{aligned} Z(N; V; T) &= \int dx W_{CA}(x) \\ &= \int dqdp e^{-\beta H(q,p)}. \end{aligned} \quad (2.4)$$

カノニカルアンサンブルでは粒子数 N , 体積 V , 温度 T が一定であり、これらの値ごとに分配関数が決まることを明示するために、 $Z(N; V; T)$ とした。しかし以下では簡単のために、粒子数 N と体積 V への依存性は省略し、温度依存性のみをあらわに書いて、 $Z(T)$ と書くことにする。系の分配関数 $Z(T)$ が求められれば、温度 T のカノニカルアンサンブルにおける任意の統計力学的物理量が $Z(T)$ から求められるので、この系は厳密に解けたことになる。

分配関数 $Z(T)$ によって得られる統計力学的物理量の具体例として、ヘルムホルツ自由エネルギー $F(T)$ がある。これは、

$$Z(T) = \exp[-\beta F(T)]. \quad (2.5)$$

で定義されるため、

$$F(T) = -k_B T \ln Z(T). \quad (2.6)$$

である。また、系の内部エネルギー $U(T)$ は、

$$\begin{aligned} U(T) &= \langle H \rangle_T \\ &= \frac{\int dqdp H(q,p) e^{-\beta H(q,p)}}{Z} \\ &= -\frac{\partial \ln Z}{\partial \beta} \\ &= -\frac{\partial Z}{\partial \beta} \cdot \frac{1}{Z}. \end{aligned} \quad (2.7)$$

で与えられる。ここで、 $\langle A \rangle_T$ は、物理量 A の温度 T における平均値 (期待値)

$$\langle A \rangle_T = \frac{\int dqdp A(q,p) e^{-\beta H(q,p)}}{Z}. \quad (2.8)$$

を表す。このように、分配関数 $Z(T)$ を得ることで様々な物理量が得られる。

また分配関数 (2.4) は状態密度 (Density of States: DOS) $g(\varepsilon)$ を用いることで、系のエネルギーの積分として、

$$Z(T) = \int d\varepsilon g(\varepsilon) e^{-\beta \varepsilon}. \quad (2.9)$$

と表すことができる。ここで ε は系のエネルギーであり、 $g(\varepsilon)\Delta\varepsilon$ は ε と $\varepsilon + \Delta\varepsilon$ の間に存在する状態の数を表す。そして物理量 A が系のエネルギーに依存する場合に式 (2.8) は、

$$\langle A \rangle_T = \frac{\int d\varepsilon A(\varepsilon)g(\varepsilon)e^{-\beta\varepsilon}}{Z}. \quad (2.10)$$

となる。従って $g(\varepsilon)$ を得ることで、この系における分配関数 (式 (2.4)) や統計力学的物理量 (式 (2.8)) が計算できるようになるため。 $g(\varepsilon)$ は統計力学的に重要な物理量であるといえる。

式 (2.1) のように座標 q と運動量 p が分離されている場合、分配関数 $Z(T)$ は座標分配関数 $Z_q(T)$ と運動量分配関数 $Z_p(T)$ の積で表すことができる。従って、

$$Z(T) = Z_q(T)Z_p(T). \quad (2.11)$$

である。ここで、 $Z_q(T)$ と $Z_p(T)$ は、

$$Z_q(T) = \int dq e^{-\beta E(q)}, \quad (2.12)$$

$$Z_p(T) = \int dp e^{-\beta K(p)}. \quad (2.13)$$

と定義される。特に物理量 A が q のみの関数、または p のみの関数の場合、式 (2.8) は、共通部分が分子と分母で打ち消しあって、それぞれ以下のように書くことができる。

$$\langle A(q) \rangle_T = \frac{\int dq A(q)e^{-\beta E(q)}}{Z_q}, \quad (2.14)$$

$$\langle A(p) \rangle_T = \frac{\int dp A(p)e^{-\beta K(p)}}{Z_p}, \quad (2.15)$$

2.2 モンテカルロ法

前節のカノニカルアンサンブル系を計算機上で再現する手法として、乱数 (random number) を使用して確率的に系の状態を発生させていくモンテカルロ (MC) 法が広く使われる。モンテカルロ法では通常、運動量 p を考慮せず座標やスピンを表す q のみを変化させていく。従って物理量 A も q のみの関数を扱い、その分配関数と A の期待値は式 (2.12) と式 (2.14) で与えられる。式 (2.12) と式 (2.14) は座標に関する積分をポテンシャルエネルギー $E(q)$ の積分に置き換えて、次のように書き直せる。

$$\begin{aligned} Z_q &= \int dq e^{-\beta E(q)} \\ &= \int dE P_{CA}(E). \end{aligned} \quad (2.16)$$

$$\begin{aligned} \langle A \rangle_T &= \frac{\int dq A(q)e^{-\beta E(q)}}{Z_q} \\ &= \frac{\int dE A(E)P_{CA}(E)}{\int dE P_{CA}(E)}. \end{aligned} \quad (2.17)$$

このとき、ポテンシャルエネルギー E のカノニカル分布 $P_{\text{CA}}(E)$ は系の状態密度 $g(E)$ とボルツマン因子 $W_{\text{CA}}(E)$ の積として、次のように定義される。

$$P_{\text{CA}}(E) = g(E)W_{\text{CA}}(E). \quad (2.18)$$

$$W_{\text{CA}}(E) = e^{-\beta E}. \quad (2.19)$$

$P_{\text{CA}}(\varepsilon)$ は、急激に増加する状態密度 $g(E)$ と指数関数的に減少するボルツマン因子 $W_{\text{CA}}(E)$ のつり合いで、ベル型の関数となっている。

また (2.2) より、系のある状態 x ($x \equiv \{q\}$) は、ポテンシャルエネルギー $E(x)$ に比例した形で、

$$P_{\text{CA}}(x) \propto e^{-\beta E(x)} \quad (2.20)$$

に従って出現する。そのため MC 法では、確率分布 $P_{\text{CA}}(x)$ に従った状態 x を、乱数を用いて計算機上に発生させることになる。以下では、MC 法で満たすべき条件について解説する。この条件を満足するように状態 x を発生させることで、最終的な x のサンプルの集合がカノニカル分布 $P_{\text{CA}}(E)$ を形成しているとみなせる。本研究では、第 6 章を除いたすべてのシミュレーションで MC 法を使用した。

原理

MC 法では、シミュレーション中の 1 STEP ごとに系の状態 x を発生させていく。

$$x^1 \rightarrow x^2 \rightarrow x^3 \dots \rightarrow x^\nu \rightarrow \dots \quad (2.21)$$

この新しい状態 x を発生させる動作を MC STEP と呼ぶ。MC 法では確率に従って x^ν の状態から $x^{\nu+1}$ の状態を生成する。シミュレーションが 1 MC STEP 進み、状態が $x^\nu (= x)$ から $x^{\nu+1} (= x')$ に遷移する確率を遷移確率 $w(x \rightarrow x')$ と呼ぶ。以下の議論の重要な点は、 x のサンプルの集合がカノニカル分布 $P_{\text{CA}}(E)$ を形成しているとみなせるように、遷移確率 w を決定することである。本節では簡潔に遷移確率を導出するが、詳細な議論は例えば [16, 21] が詳しい。

いま、MC シミュレーションが ν 番目の MC STEP まで進み、状態 x_j にあったとする ($x^\nu = x_j$)。これが $\nu + 1$ 番目の MC STEP で状態 x_k に遷移すると考える ($x^{\nu+1} = x_k$)。本来 $x_j \rightarrow x_k$ への遷移確率 w は全ての過去の状態 ($x^\nu : \nu = 1, 2, \dots, \nu$) に依存するが、直前の状態 x_j のみに依存するという仮定 (マルコフ過程) を課すことで、

$$w = w(x_j \rightarrow x_k). \quad (2.22)$$

と単純な形で表すことができる。

ν 番目の MC STEP において系がある状態 x を取る確率分布を $P^\nu(x)$ とすると、確率論の定義より、 $\nu + 1$ 番目に状態 x_k をとる確率は、

$$P^{\nu+1}(x_k) = \sum_j P^\nu(x_j)w(x_j \rightarrow x_k). \quad (2.23)$$

である。MC STEP を繰り返す ($\nu \rightarrow \infty$) ことにより、 $P^\nu(x)$ が平衡分布 $P_{\text{eq}}(x)$ に収束すると仮定すると、

$$P_{\text{eq}}(x_k) = \sum_j P_{\text{eq}}(x_j) w(x_j \rightarrow x_k). \quad (2.24)$$

が成立する。この式は、詳細釣り合いの条件

$$P_{\text{eq}}(x_j) w(x_j \rightarrow x_k) = P_{\text{eq}}(x_k) w(x_k \rightarrow x_j). \quad (2.25)$$

を満たすことで満足する。従って、式 (2.25) の詳細釣り合いの条件を満たしながら、新しい $\nu+1$ 番目の状態を生成できれば、シミュレーションは平衡分布 $P_{\text{eq}}(x)$ を実現していることになる。

詳細釣り合いの条件式 (2.25) を満足する具体的な遷移確率 $w(x_j \rightarrow x_k)$ の決め方は、メトロポリス法が有名である。この方法では遷移確率を、

$$w(x_j \rightarrow x_k) = \begin{cases} 1 & , \text{if } \frac{P_{\text{eq}}(x_k)}{P_{\text{eq}}(x_j)} \geq 1, \\ \frac{P_{\text{eq}}(x_k)}{P_{\text{eq}}(x_j)} & , \text{if } \frac{P_{\text{eq}}(x_k)}{P_{\text{eq}}(x_j)} < 1. \end{cases} \quad (2.26)$$

とすることで、詳細釣り合いの条件 (2.25) を満たす。

カノニカル分布を実現したい場合、 $P_{\text{eq}}(x) = P_{\text{CA}}(x)$ とすれば良いので式 (2.20) より、遷移確率 (2.26) は、

$$w(x_j \rightarrow x_k) = \begin{cases} 1 & , \text{if } \Delta E \leq 0, \\ e^{-\beta \Delta E} & , \text{if } \Delta E > 0. \end{cases} \quad (2.27)$$

となる。 ΔE は、

$$\Delta E \equiv E(x_k) - E(x_j) \quad (2.28)$$

で定義される。また、式 (2.27) は簡単に

$$w(x_j \rightarrow x_k) = \min \left(1, \frac{W_{\text{CA}}(E(x_k))}{W_{\text{CA}}(E(x_j))} \right) = \min(1, \exp[-\beta \Delta E]). \quad (2.29)$$

と表記できる。

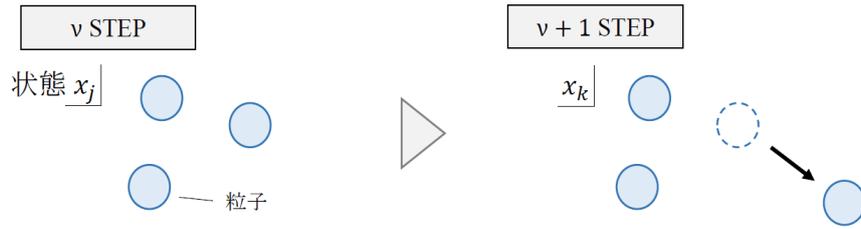
アルゴリズム

実際の計算アルゴリズムは以下のような手順となる。モンテカルロシミュレーションが ν 番目の MC STEP まで進んだと仮定して、その状態を x_j とする ($x^\nu = x_j$)。

1. $\nu+1$ 番目の状態候補 $x_{\text{cand}}^{\nu+1}$ を用意する。この状態を x_k とする。
2. $\Delta E = E(x_k) - E(x_j)$ を計算する。
3. $\Delta E \leq 0$ ならば、 $x_{\text{cand}}^{\nu+1}$ ($= x_k$) を $x^{\nu+1}$ として採択する。 $\Delta E > 0$ ならば、 $[0, 1]$ までの擬似乱数 r を発生させ、 $r \leq \exp(-\beta \Delta E)$ を満足していたら、 $x_{\text{cand}}^{\nu+1}$ ($= x_k$) を $x^{\nu+1}$ として採択する。そうでなければ、 x^ν ($= x_j$) を $x^{\nu+1}$ として採択する。

図 2.1 はメトロポリスモンテカルロ法のフローチャートを示したものである。メトロポリス法では、この操作を十分な回数繰り返すことによってカノニカルアンサンブルを計算機上に発生させサンプルを得る。

1. 状態の変化



2. メトロポリス判定

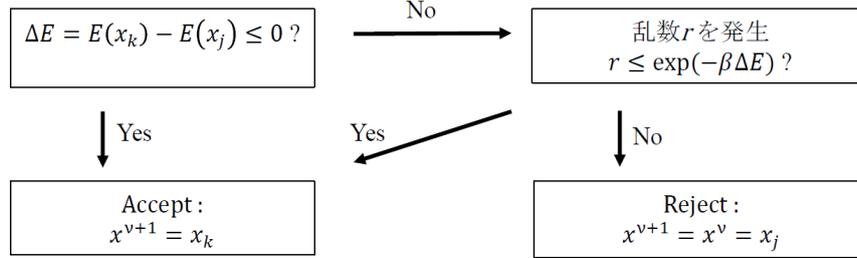


Fig. 2.1: メトロポリスモンテカルロ法のフローチャート

2.3 分子動力学法

ニュートンの運動方程式を差分方程式として表現し、これを数値的に解くことで、新しい状態 x を発生させる手法を分子動力学法 (Molecular Dynamics:MD) という。MD 法では MC 法とは異なり運動量 p の変化も考慮に入れる。ニュートンの運動方程式は差分方程式として、

$$\dot{\mathbf{q}}_k = \frac{\Delta \mathbf{q}_k}{\Delta t} = \frac{\mathbf{p}_k}{m_k}, \quad (2.30)$$

$$\dot{\mathbf{p}}_k = \frac{\Delta \mathbf{p}_k}{\Delta t} = -\frac{\partial E}{\partial \mathbf{q}_k} = \mathbf{F}_k. \quad (2.31)$$

と表される。 $\Delta \mathbf{q}_k$ および $\Delta \mathbf{p}_k$ は、時間 Δt の間の k 番目の粒子の座標変化と運動量変化を表し、 \mathbf{F}_k は k 番目 ($k = 1, \dots, N$) の粒子に働く力である。しかし、上式をそのまま解いた場合、時間変化に対して全エネルギーが不変となるため、ミクロカノニカル分布が実現される。そのため温度一定のカノニカルアンサンブルを実現するために、運動方程式に工夫をする必要がある。

MD 法においてカノニカルアンサンブルを実現する手法はいくつか提案されているが、ここでは能勢の方法を考える。その他の手法や能勢の方法の詳細な解説は、例えば参考文献 [19, 20] などが詳しい。結論として、以下の微分方程式を差分方程式として解けばよい。

$$\dot{\mathbf{q}}_k = \frac{\mathbf{p}_k}{m_k}, \quad (2.32)$$

$$\dot{\mathbf{p}}_k = -\frac{\partial E}{\partial \mathbf{q}_k} - \frac{\dot{s}}{s} \mathbf{p}_k = \mathbf{F}_k - \frac{\dot{s}}{s} \mathbf{p}_k, \quad (2.33)$$

$$\dot{s} = s \frac{P_s}{Q}, \quad (2.34)$$

$$\dot{P}_s = \sum_{k=1}^N \frac{\mathbf{p}_k^2}{m_k} - 3Nk_B T = 3Nk_B(T(t) - T). \quad (2.35)$$

ここでは、 s は温度を調節するため導入された仮想粒子の一般化座標、 P_s は s の共役運動量、 Q は仮想粒子の質量を表す。また T はカノニカルアンサンブルで実現したい温度であり、 $T(t)$ は、

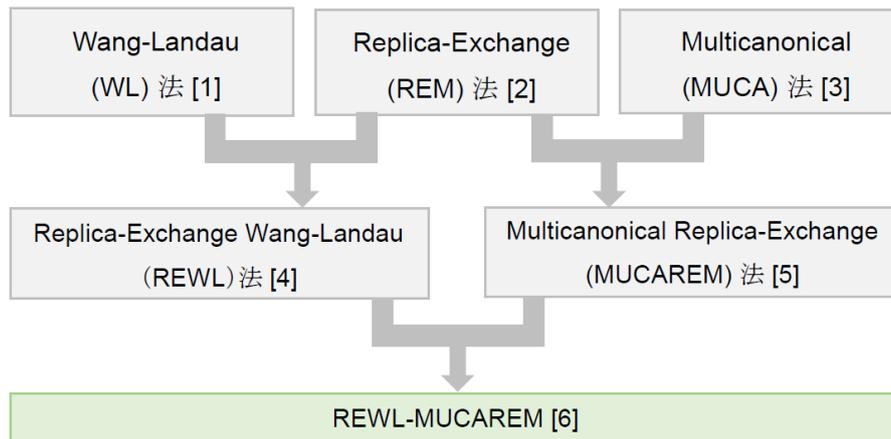
$$T(t) = \frac{1}{3Nk_B} \sum_{k=1}^N \frac{\mathbf{p}_k^2(t)}{m_k}. \quad (2.36)$$

として定義された時刻 t における系の瞬間温度である。能勢の方法では、系の粒子に加えて温度を調節するための仮想粒子を導入することによって、系全体の瞬間温度 $T(t)$ が設定温度 T となるようにコントロールしている。

2.4 拡張アンサンブル法

スピングラス系、生体高分子系などの多自由度複雑系では、エネルギー極小値が無数に存在する。前述したカノニカルアンサンブルを再現するような MC 法や MD 法では、そのようなエネルギー極小状態にシミュレーションが留まり、構造空間を十分に探索できない。有効なサンプルデータが少なくなり、誤ったシミュレーション結果を与える原因となる。このような困難を乗り越えるために開発された手法が、拡張アンサンブル法である。拡張アンサンブル法では、非ボルツマン因子に基づく重み因子によって再現される人工的な統計力学集団によってシミュレーションが実行される。この人工的に作られた確率分布は、シミュレーションがエネルギー極小値にトラップされても、抜け出せるように設計されている。従って、従来のカノニカルアンサンブルよりも幅広い構造空間を探索できる。

本章では、代表的な拡張アンサンブル法であるマルチカノニカル法、ワン-ランダウ法、レプリカ交換法を説明し、次にこれらを組み合わせた手法であるマルチカノニカルレプリカ交換法とレプリカ交換ワン-ランダウ法を説明する。最後に、DOS を高精度で推定するため本研究で提案した、マルチカノニカルレプリカ交換法とレプリカ交換ワン-ランダウ法を組み合わせた計算プロトコルである REWL-MUCAREM を説明する。各拡張アンサンブル法の関係は図 2.2 の通りである。



- [1] F. Wang and D. P. Landau, Phys. Rev. Lett 86 2050 (2001).
- [2] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. 65, 1604 (1996).
- [3] B. A. Berg and T. Neuhaus, Phys. Lett. B 267, 249 (1991).
- [4] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. 110,210603 (2013).
- [5] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. 329, 261 (2000).
- [6] T. Hayashi and Y. Okamoto, Phys. Rev. E 100, 043304 (2019).

Fig. 2.2: 拡張アンサンブル法の関係性。本研究では、状態密度を高精度で推定するために REWL-MUCAREM を提案した。

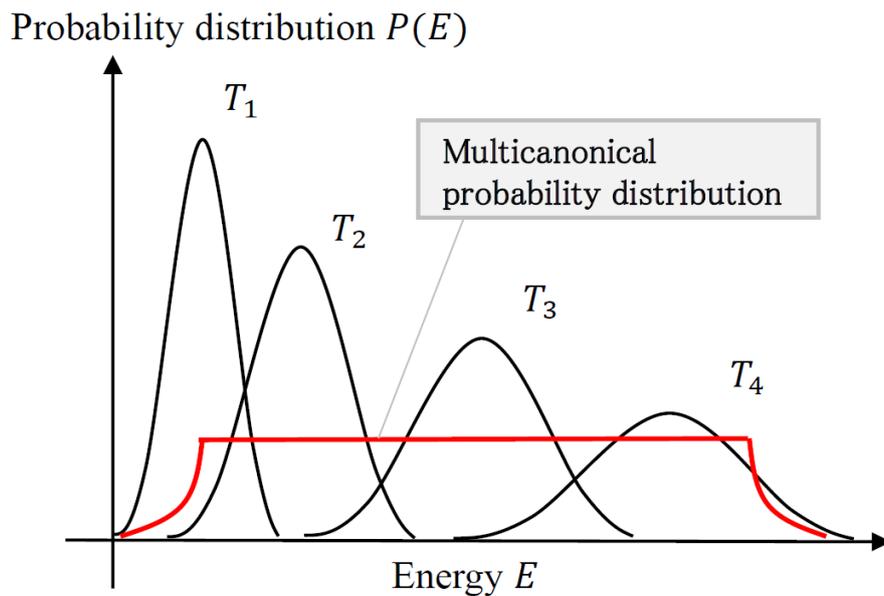


Fig. 2.3: マルチカノニカル分布とカノニカル分布を比較した概念図。カノニカル分布は温度 T に依存したベル型の確率分布であるが、マルチカノニカル分布は広いエネルギー領域で $P_{\text{MUCA}}(E) \propto \text{const.}$ となった確率分布である。

2.4.1 マルチカノニカル法

原理

はじめに、マルチカノニカル法 (Multicanonical algorithm: MUCA) について説明する。MUCA では、ポテンシャルエネルギー空間でのランダムウォークを実現することによって、シミュレーションがエネルギー極小値に留まることを避ける。ポテンシャルエネルギー空間でのランダムウォークを実現する確率分布として、

$$P_{\text{MUCA}}(E) \propto g(E)W_{\text{MUCA}}(E) \equiv \text{const}, \quad (2.37)$$

が考えられる。このポテンシャルエネルギー空間上で一様な確率分布を実現する人工的な確率分布をマルチカノニカル確率分布 $P_{\text{MUCA}}(E)$ と呼ぶ。図 2.3 はカノニカル分布 $P_{\text{CA}}(E)$ とマルチカノニカル分布 $P_{\text{MUCA}}(E)$ を比較した図である。カノニカル分布が温度 T に依存したベル型の確率分布をとるのに対し、マルチカノニカル分布は幅広いエネルギー領域で一定の値をとる。

式 (2.37) で、 $W_{\text{MUCA}}(E)$ をマルチカノニカル重み因子と呼び、MUCA 法を特徴づける人工的な非ボルツマン重み因子である。 $W_{\text{MUCA}}(E)$ は、

$$W_{\text{MUCA}}(E) = \frac{1}{g(E)}. \quad (2.38)$$

とすることで、式 (2.37) を満足する。ここで比例係数は省略した。つまり、MUCA 法では、重み因子 $W_{\text{MUCA}}(E)$ として系の DOS の逆数を用いることで、ポテンシャルエネルギー空間上での一様な確率分布 (マルチカノニカル分布 $P_{\text{MUCA}}(E)$) を実現できる。

しかし、一般的に $g(E)$ はあらかじめわかっていないので、MUCA 重み因子 $W_{\text{MUCA}}(E)$ ($= 1/g(E)$) を、シミュレーションによって事前に推定する必要がある。実際にはシミュレーションに $g(E)$ の仮の推定値 ($\tilde{g}(E)$) を用いて、 $W_{\text{MUCA}}(E) = 1/\tilde{g}(E)$ としてシミュレーションを実行する。この準備段階のシミュレーションでは、 $P_{\text{MUCA}}(E)$ は図 2.3 ほど平坦な確率分布をならない。 $\tilde{g}(E)$ を調節して短いシミュレーションの繰り返すことで、試行錯誤しながらマルチカノニカル分布を実現する $\tilde{g}(E)$ を得る手法 (逐次近似法: 第 6 章) などが提案されているが、ワン-ランダウ法は特に有効な手法であるため次節で解説する。

アルゴリズム

モンテカルロ法における MUCA シミュレーションは以下のように実行される。まず MUCA MC では、メトロポリス遷移確率 $w(x \rightarrow x')$ は、式 (2.24) と式 (2.25) において $P_{\text{eq}}(E) \rightarrow P_{\text{MUCA}}(E)$ とすることより、次のように定義される。

$$w(x \rightarrow x') = \min \left[1, \frac{W_{\text{MUCA}}(E(x'))}{W_{\text{MUCA}}(E(x))} \right] = \min \left[1, \frac{g(E(x))}{g(E(x'))} \right]. \quad (2.39)$$

$E(x)$ は元の配位 x のポテンシャルエネルギーであり、 $E(x')$ は新たに提案された配位 x' のポテンシャルエネルギーである。従って、MUCA MC 法は系の DOS 変化によるモンテカルロ法といえる。

議論を簡潔にするため、参照温度 $T_0 = (1/k_B\beta_0)$ と、マルチカノニカルポテンシャルエネルギー $E_{\text{MUCA}}(E; T_0)$ を導入する。ここで T_0 は任意の温度であり、 β_0 は逆温度である。 $E_{\text{MUCA}}(E; T_0)$ は、

$$E_{\text{MUCA}}(E; T_0) = k_B T_0 \ln g(E) = T_0 S(E), \quad (2.40)$$

と定義される。 $S(E)$ は(ミクロカノニカルアンサンブルにおける)エントロピーに対応する。これにより、式(2.39)は、

$$w(x \rightarrow x') = \min[1, \exp(-\beta_0 \Delta E_{\text{MUCA}})] . \quad (2.41)$$

と変形できる。ここで、 $\Delta E_{\text{MUCA}} = E_{\text{MUCA}}(E(x'; T_0)) - E_{\text{MUCA}}(E(x; T_0))$ である。このように $E_{\text{MUCA}}(E; T_0)$ を導入することで、カノニカルアンサンブルでの遷移確率(式(2.29))と比較できるようになる。従って実際の MUCA MC シミュレーションは次の手順で実行される。MC シミュレーションが ν 番目の MC STEP まで進んだと仮定して、その状態を x_j とする。

1. $\nu + 1$ 番目の状態候補 $x_{\text{cand}}^{\nu+1}$ を用意する。この状態を x_k とする。
2. $\Delta E_{\text{MUCA}} = E_{\text{MUCA}}(E(x_k; T_0)) - E_{\text{MUCA}}(E(x_j; T_0))$ を計算する。
3. $\Delta E_{\text{MUCA}} \leq 0$ ならば、 $x_{\text{cand}}^{\nu+1} (= x_k)$ を $x^{\nu+1}$ として採択する。 $\Delta E_{\text{MUCA}} > 0$ ならば、 $[0, 1]$ までの乱数 r を発生させ、 $r \leq \exp(-\beta \Delta E_{\text{MUCA}})$ を満足すれば、 $x_{\text{cand}}^{\nu+1} (= x_k)$ を $x^{\nu+1}$ として採択する。そうでなければ、 $x^\nu (= x_j)$ を $x^{\nu+1}$ として採択する。
4. 手順1へ戻る

これにより、マルチカノニカルアンサンブルを実現する MC シミュレーションが実行できる。原理的には、もし厳密な $g(E)$ によって MUCA シミュレーションを実行できたならば、図2.3のような完全に一様な MUCA 分布 $P_{\text{MUCA}}(E)$ が得られる。

マルチカノニカルアンサンブルは MD シミュレーションにおいても、カノニカルアンサンブルを再現する能勢の運動方程式(2.32)–(2.35)のポテンシャルエネルギー E を、 E_{MUCA} に置き換えることによって実現できる。従って、

$$\dot{\mathbf{q}}_k = \frac{\mathbf{p}_k}{m_k}, \quad (2.42)$$

$$\dot{\mathbf{p}}_k = -\frac{\partial E_{\text{MUCA}}(E; T_0)}{\partial \mathbf{q}_k} - \frac{\dot{s}}{s} \dot{\mathbf{p}}_k = \frac{\partial E_{\text{MUCA}}(E; T_0)}{\partial E} \mathbf{F}_k - \frac{\dot{s}}{s} \dot{\mathbf{p}}_k, \quad (2.43)$$

$$\dot{s} = s \frac{P_s}{Q}, \quad (2.44)$$

$$\dot{P}_s = \sum_{k=1}^N \frac{\mathbf{p}_k^2}{m_k} - 3Nk_B T_0 = 3Nk_B (T(t) - T_0). \quad (2.45)$$

となる。このとき、熱浴の温度として適切な T_0 (例えば、シミュレーションしたい物理現象が生じる温度)をあらかじめ選択しておく必要がある。

実践的な MUCA 法では、マルチカノニカル分布を実現したいエネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ をあらかじめ指定する必要がある。ここで $[E_{\text{min}}, E_{\text{max}}]$ は T_0 と同様に興味のある物理現

象が生じるエネルギー領域とする。\$[E_{\min}, E_{\max}]\$ を指定しなかった場合、興味のないエネルギー領域をシミュレーションすることになり効率が悪化する。\$[E_{\min}, E_{\max}]\$ の領域を中心にサンプリングを行うためには、\$E_{\text{MUCA}}(E)\$ を \$\varepsilon_{\text{MUCA}}(E)\$ として次のように定義し直すことが提案されている [11, 12]。まず \$E_{\min}\$ と \$E_{\max}\$ がある温度 \$T_{\min}, T_{\max}\$ でのエネルギーの期待値 \$\langle E \rangle_T\$ に対応するとして、

$$\begin{cases} E_{\min} = \langle E \rangle_{T_{\min}} \\ E_{\max} = \langle E \rangle_{T_{\max}} \end{cases} \quad (2.46)$$

とする。このとき \$\varepsilon_{\text{MUCA}}(E)\$ を、

$$\varepsilon_{\text{MUCA}}(E) = \begin{cases} \left. \frac{\partial E_{\text{MUCA}}(E; T_0)}{\partial E} \right|_{E=E_{\min}} (E - E_{\min}) + E_{\text{MUCA}}(E_{\min}; T_0), & \text{for } E < E_{\min} \\ E_{\text{MUCA}}(E; T_0), & \text{for } E_{\min} \leq E \leq E_{\max} \\ \left. \frac{\partial E_{\text{MUCA}}(E; T_0)}{\partial E} \right|_{E=E_{\max}} (E - E_{\max}) + E_{\text{MUCA}}(E_{\max}; T_0), & \text{for } E_{\max} < E \end{cases} \quad (2.47)$$

として定義し直す。モンテカルロ法の場合は式 (2.40) より、

$$W_{\text{MUCA}}(E) = \exp[-\beta_0 \varepsilon_{\text{MUCA}}(E)] \begin{cases} \exp[-\beta_{\min} E], & \text{for } E < E_{\min} \\ \frac{1}{g(E)}, & \text{for } E_{\min} \leq E \leq E_{\max} \\ \exp[-\beta_{\max} E], & \text{for } E_{\max} < E \end{cases} \quad (2.48)$$

と定義し直すことによって、\$[E_{\min}, E_{\max}]\$ で一様なポテンシャルエネルギー分布 (マルチカノニカル分布) を実現し、\$E < E_{\min}\$ では温度 \$T_{\min}\$ の、\$E_{\max} < E\$ では温度 \$T_{\max}\$ のカノニカル分布が実現される。

MUCA 法では、重み因子 \$W_{\text{MUCA}}(E)\$ として状態密度の逆数 \$1/g(E)\$ が重要な役割を果たす。もし仮に、系の \$g(E)\$ がわかっていたら、物理量 \$A(E)\$ の任意の温度 \$T\$ における期待値 \$\langle A \rangle_T\$ は、

$$\begin{aligned} \langle A \rangle_T &= \frac{\sum_E A(E) P_{\text{CA}}(E; T)}{\sum_E P_{\text{CA}}(E; T)} \\ &= \frac{\sum_E A(E) g(E) e^{-\beta E}}{\sum_E g(E) e^{-\beta E}}, \end{aligned} \quad (2.49)$$

より、シミュレーションを行うことなく厳密に計算できる(式(2.17))。ポテンシャルエネルギーの関数で表せられない物理量の場合は、MUCAシミュレーションを実行し、シミュレーション中に一定間隔で保存した物理量 $A(x_k)$ のサンプリングデータを使用して、

$$\langle A \rangle_T = \frac{\sum_{k=1}^{n_s} A(x_k) W_{\text{MUCA}}^{-1}(E(x_k)) \exp[-\beta E(x_k)]}{\sum_{k=1}^{n_s} W_{\text{MUCA}}^{-1}(E(x_k)) \exp[-\beta E(x_k)]}. \quad (2.50)$$

で期待値を計算できる。ここで、 x_k は k 番目のサンプル、 n_s はその総サンプル数である。 $W_{\text{MUCA}}(E)$ は実際にシミュレーションで用いた重み因子である。

また、シミュレーションに使用した重み因子 $W_{\text{MUCA}}(E)$ と、シミュレーション中に一定の時間 STEP 間隔で保存したエネルギーのヒストグラム $H_{\text{MUCA}}(E)$ から、状態密度の最適解 $g(E)$ は

$$g(E) = \frac{H_{\text{MUCA}}(E)}{W_{\text{MUCA}}(E)}. \quad (2.51)$$

として計算することが可能である。式(2.51)を単ヒストグラム再重法 (single-histogram reweighting techniques) という。

2.4.2 ワン-ランダウ法

原理

ワン-ランダウ法 (Wang-Landau Algorithm: WL) は、マルチカノニカル分布を利用して状態密度 $\tilde{g}(E)$ を推定するシミュレーション手法である。WL法も、重み因子 $W_{\text{WL}}(E)$ として $1/g(E)$ を使用する。従って

$$W_{\text{WL}}(E) = \frac{1}{g(E)}. \quad (2.52)$$

となる。これは、式(2.40)と同様に参照温度 $T_0 = (1/k_B\beta_0)$ を使うことで、ワン-ランダウのポテンシャルエネルギー $E_{\text{WL}}(E; T_0)$ を、

$$E_{\text{WL}}(E; T_0) = k_B T_0 \ln g(E) = T_0 S(E), \quad (2.53)$$

と定義できる。よって WL MC は式(2.39)の $W_{\text{MUCA}}(E(x))$ を $W_{\text{WL}}(E(x))$ に置き換えた式を、WL MD は(2.42)–(2.45)の $E_{\text{MUCA}}(E; T_0)$ を $E_{\text{WL}}(E; T_0)$ に置き換えた式を使用することで実行できる。

WL法は、シミュレーション中1STEPごとに重み因子 $W_{\text{WL}}(E) (= 1/\tilde{g}(E))$ を修正することで、確率分布をマルチカノニカル分布に収束させるアルゴリズムである。WLでははじめに $\tilde{g}(E)$ に1あるいは $\exp[-\beta E]$ を初期値として与え、系を時間発展させる。シミュレーション中、各時間STEPごとに系が取るポテンシャルエネルギー値 $E(x^\nu)$ に対応した $\tilde{g}(E(x^\nu))$ の値が、

$$\tilde{g}(E(x^\nu)) \rightarrow f \times \tilde{g}(E(x^\nu)), \quad (2.54)$$

として更新される。 $E_{\text{WL}}(E; T_0)$ を使用した場合は、

$$\tilde{E}_{\text{WL}}(E^*; T_0) \rightarrow \tilde{E}_{\text{WL}}(E^*; T_0) + \ln f, \quad (2.55)$$

となる。ここで f は重みの修正係数である。 f の値は計算アルゴリズムの組み立て方に依存するが、 $f = \exp[1]$ を初期値とすることが多い。また各時間 STEP でのエネルギーの値 $E(x^\nu)$ をヒストグラム $H(E)$ として保存しておく。

式(2.54)による $W_{\text{WL}}(E)$ の更新は、 $H(E)$ が DOS を求めたい全エネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ で平坦になる (マルチカノニカル分布を形成する) まで続けられる。一定の時間 STEP 間隔で $H(E)$ の平坦度を確認し、 $H(E)$ が十分に平坦な場合にシミュレーションを停止する。結果として、修正因子 f の大きさに応じた精度の $\tilde{g}(E)$ が得られることになる。ここで、 $H(E)$ の平坦度の評価は様々な方法で行うことができる。例えば、 $[E_{\text{min}}, E_{\text{max}}]$ のヒストグラムの平均値 H_{ave} を基準に、すべての E の値で、

$$H(E) > 0.8 \times H_{\text{ave}}, \quad (2.56)$$

を満足する場合に、ヒストグラムは十分に平坦である”が提案されている。

シミュレーションの停止後、得られた $W_{\text{WL}}(E)$ ($= 1/\tilde{g}(E)$) が十分な精度ではなかった場合、 $H(E)$ をリセット ($H(E) = 0$) し、修正係数 f を減らした後 (通常は $f \rightarrow \sqrt{f}$)、前回の WL シミュレーションで得られた $\tilde{g}(E)$ を初期重み因子 $W_{\text{WL}}(E)$ ($= 1/\tilde{g}(E)$) として次の WL シミュレーションを開始する。修正係数 f を小さくしたことで、より高精度な $\tilde{g}(E)$ が得られると期待される。

この手順を、満足の行く精度の $\tilde{g}(E)$ が得られるまで繰り返す。十分高精度な $\tilde{g}(E)$ を得る修正係数 f の値 f_{final} として、 $\exp[10^{-8}] \simeq 1.000\,000\,01$ が用いられることが多い。注意点として、初めから小さい重み修正因子 f を設定すると、収束までの時間 STEP が極めて長くなることが指摘されている。

アルゴリズム

WL MC 法のアルゴリズムは、次のようなものとなる。はじめに重み因子 $W_{\text{WL}}(E)$ ($= 1/\tilde{g}(E)$) に初期値 ($\tilde{g}(E) = 1$ あるいは $\tilde{g}(E) = \exp[-\beta E]$) を与える。WL MC シミュレーションが ν 番目の MC STEP まで進んだと仮定して、その状態を x_j とする。

1. $\nu + 1$ 番目の状態候補 $x_{\text{cand}}^{\nu+1}$ を用意する。この状態を x_k とする。
2. $\Delta \tilde{E}_{\text{WL}} = \tilde{E}_{\text{WL}}(E(x_k; T_0) - \tilde{E}_{\text{WL}}(E(x_j; T_0)))$ を計算する。
3. $\Delta \tilde{E}_{\text{WL}} \leq 0$ ならば、 $x_{\text{cand}}^{\nu+1}$ ($= x_k$) を $x^{\nu+1}$ として採択する。 $\Delta \tilde{E}_{\text{WL}} > 0$ ならば、 $[0, 1]$ までの乱数 r を発生させ、 $r \leq \exp(-\beta \Delta \tilde{E}_{\text{WL}})$ を満足していたら、 $x_{\text{cand}}^{\nu+1}$ ($= x_k$) を $x^{\nu+1}$ として採択する。そうでなければ、 x^ν ($= x_j$) を $x^{\nu+1}$ として採択する。
4. $E(x^{\nu+1})$ に対応した重み因子 $W_{\text{WL}}(E)$ ($= 1/\tilde{g}(E)$) を $\tilde{g}(E) \rightarrow f \times \tilde{g}(E)$ として更新する。これは、 $\tilde{E}_{\text{WL}}(E; T_0) \rightarrow \tilde{E}_{\text{WL}}(E; T_0) + k_{\text{B}} T_0 \ln f$ に対応する。
5. 一定の時間 STEP 間隔で、 $\tilde{g}(E)$ を求めたい全エネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ のヒストグラム $H(E)$ の平坦さを確認する。 $H(E)$ が十分に平坦であれば、 $f \rightarrow \sqrt{f}$, $H(E) = 0$ とし、推定した $W_{\text{WL}}(E)$ ($= 1/\tilde{g}(E)$) を初期重みとした WL シミュレーションを再度実行する。

6. $f < f_{\text{final}}$ であれば、シミュレーションを終える。

7. 手順1へ戻る。

この繰り返しによって $\tilde{g}(E)$ を推定する。

WL MC 法では MUCA MC 法と異なり、シミュレーション中に重み因子を更新するため、厳密には詳細釣り合いの条件を満たしていない。従って得られたデータに統計力学的なアンサンブルからサンプルされたという保障はないが、WL 法によって推定された $\tilde{g}(E)$ は、 f_{final} によって設定された精度の範囲内で系の真の $g(E)$ に収束する傾向があるとされている。ただし WL 法にはアルゴリズムに由来する消すことのできない系統誤差が、 $\tilde{g}(E)$ に存在することが報告されている [24]。

WL MD シミュレーションの場合も、MUCA MD と同様に式 (2.42)–(2.45) に従って、 $E_{\text{WL}}(E; T_0)$ を更新しながら、系を時間発展させればよい。しかし MD シミュレーションの場合は、 $E_{\text{WL}}(E; T_0)$ のエネルギー微分を計算する必要がある。この微分はシミュレーションの初期段階で変動が大きく、粒子に人工的な大きい力を加えてしまう。従って非物理的な状態が生じやすく、シミュレーションの破綻につながることが多い。そのため、Statistical Temperature Molecular Dynamics (STMD) [22] など、MD のための WL アルゴリズムが開発されている。本研究においても、Lorentzian fitting algorithm [23] を導入した Wang-Landau MD 法を提案した (第7章)。

2.4.3 レプリカ交換法

原理

レプリカ交換法 (Replica-Exchange Method: REM) では、 M 個の異なる温度 $T_m (m = 1, 2, \dots, M)$ での系のコピー (レプリカ) を用意し、従来のカノニカルアンサンブルに基づく MC または MD を、各レプリカで独立して同時に実行する。レプリカにはひとつの温度 T_m が対応し、あらかじめ定めたシミュレーション時間間隔 (STEP 数) でレプリカの温度を交換 (レプリカ交換) することで、温度空間でのランダムウォークが実現し、ポテンシャルエネルギー空間でのランダムウォークが誘発される。これによりエネルギー極小値からの脱出を図る。レプリカ交換法の概念図を示したものが図 2.4 である。

通常、カノニカルアンサンブルを実現するシミュレーション中に温度を切り替えることは、厳密には詳細釣り合いの条件を破るため、カノニカルアンサンブルからの正しいサンプリングを行えている保証はない。しかし、メトロポリス法によって定められた遷移確率に従ってレプリカ交換を行い温度を切り替えることで、カノニカルアンサンブルからの統計的なサンプリングが保証される。以下では、レプリカ交換に関する遷移確率について解説する。

はじめに、温度の異なる複数のレプリカから構成される系の、重み因子 $W_{\text{REM}}(X)$ を導出する。ここで X は全レプリカの状態を表し、各レプリカの状態の集合である。まずレプリカを表すラベルを $i (i = 1, 2, \dots, M)$ 、温度を表すラベルを $m (m = 1, 2, \dots, M)$ とする。レプリカと温度は一対一対応するため、

$$i = i(m), \quad (2.57)$$

$$m = m(i). \quad (2.58)$$

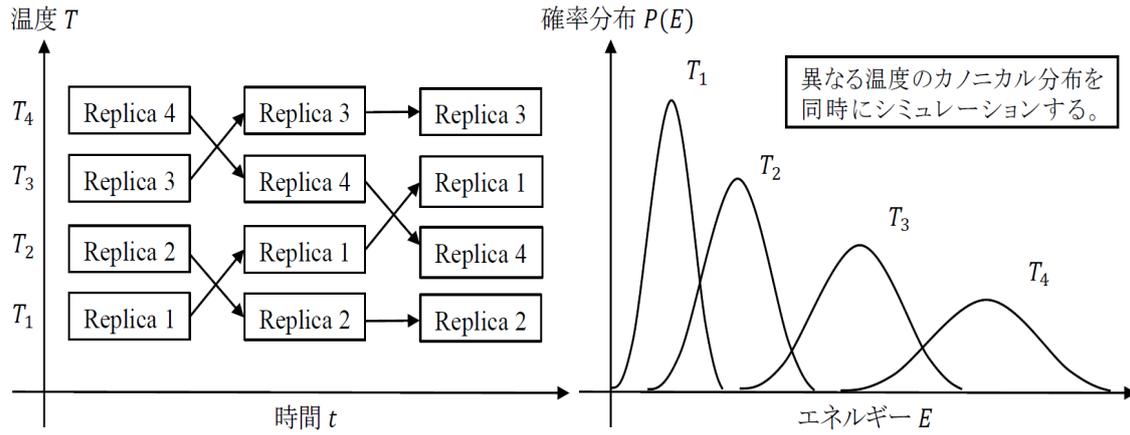


Fig. 2.4: レプリカ交換法概念図。左図はレプリカ交換法における各レプリカの時間発展を示す。レプリカ交換が一定の時間STEPで実行され、交換が受け入れられると、各レプリカの温度が更新される。右図は重み因子 $W_{\text{REM}}(X)$ によって得られる確率分布を示す。

である。それぞれの関数は、温度番号 m (あるいはレプリカ番号 i) を入れるとレプリカ番号 i (あるいは温度番号 m) を返す関数である。各レプリカの系の状態を、

$$x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m. \quad (2.59)$$

と表す。ここで上付きの i はレプリカの状態、下付きの m は温度を表す。従って全状態 X は、

$$X = \{x_1^{[i(1)]}, x_2^{[i(2)]}, \dots, x_M^{[i(M)]}\} = \{x_{m(1)}^{[1]}, x_{m(2)}^{[2]}, \dots, x_{m(M)}^{[M]}\} \quad (2.60)$$

となる。各レプリカは独立してMCあるいはMDシミュレーションを行う。レプリカ間での相互作用は存在しないため、全レプリカの重み因子は、各レプリカのボルツマンの重み因子の積として、

$$\begin{aligned} W_{\text{REM}}(X) &= \prod_{i=1}^M \exp \{-\beta_{m(i)} H(q^{[i]}, p^{[i]})\}, \\ &= \prod_{m=1}^M \exp \{-\beta_m H(q^{[i(m)]}, p^{[i(m)]})\}, \\ &= \exp \left[-\sum_{i=1}^M \beta_{m(i)} H(q^{[i]}, p^{[i]}) \right], \\ &= \exp \left[-\sum_{i=1}^M \beta_m H(q^{[i(m)]}, p^{[i(m)]}) \right]. \end{aligned} \quad (2.61)$$

と定義される。

レプリカ交換は、MCでもMDでも、メトロポリス判定に基づいて実行される。そのためメトロポリス判定に従ったレプリカ交換の遷移確率を考える。いま温度 T_m と T_n に対応するレプリカ i, j を交換するとして、温度交換前の系の状態を X 、後の状態を X' とすると、

$$X = \{\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots\} \rightarrow X' = \{\dots, x_m^{[j]'}, \dots, x_n^{[i]'}, \dots\}. \quad (2.62)$$

となる。このレプリカ交換は、次のように書くことができる。

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_n^{[j]'} \equiv (q^{[j]}, p^{[j]'})_m, \\ x_m^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_m^{[i]'} \equiv (q^{[i]}, p^{[i]'})_n, \end{cases} \quad (2.63)$$

ここでの $p^{[i]}'$ と $p^{[j]}'$ は後述する。この操作はレプリカがそれぞれ i と j に対応する温度 T_m と T_n を交換することに等しいので、

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_n^{[i]'} \equiv (q^{[i]}, p^{[i]'})_m, \\ x_m^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_m^{[j]'} \equiv (q^{[j]}, p^{[j]'})_n, \end{cases} \quad (2.64)$$

と書いてもよい。

従って、式 (2.61) の重みに基づくメトロポリス判定の式は、

$$\begin{aligned} w(X \rightarrow X') &= \min \left(1, \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)} \right) \\ &= \min(1, \exp(-\Delta)). \end{aligned} \quad (2.65)$$

となる。ここで、 Δ は、

$$\Delta = (\beta_m - \beta_n) (E(q^{[j]}) - E(q^{[i]})) \quad (2.66)$$

である。

Δ は、MC シミュレーションでは運動量 p^i を考慮する必要がないので式 (2.66) で良いが、MD シミュレーションでは運動量を考慮する必要がある。これは、レプリカ交換の後に

$$\begin{cases} p^{[i]'} = \sqrt{\frac{T_n}{T_m}} p^{[i]}, \\ p^{[j]'} = \sqrt{\frac{T_m}{T_n}} p^{[j]}. \end{cases} \quad (2.67)$$

のように運動量をスケールングすることで、MD シミュレーションにおいても式 2.65 による交換判定を使用できる。この証明は参考文献 [16] に詳しい。

アルゴリズム

REM におけるアルゴリズムは次のようなものとなる。

1. 各レプリカ i ($i = 1, 2, \dots, M$) において、温度 T_m ($m = 1, 2, \dots, M$) のカノニカルアンサンブルを再現する MC あるいは MD シミュレーションを決められた STEP 数実行する。
2. レプリカ対 (i, j) を作り、遷移確率 (2.65) に従ってレプリカ交換を行う。レプリカ交換が受託された場合、各レプリカの温度を更新する。
3. 手順 1 へ戻る。

REMによるシミュレーションを実行した後、シミュレーション中に得られた状態 X のサンプリングデータから、系の状態密度 $g(E)$ と任意の温度 T における物理量の期待値 $\langle A \rangle_T$ を計算できる。これには、十分に長いシミュレーション（プロダクションランという）の後、それぞれのレプリカでサンプリングされたポテンシャルエネルギーのヒストグラムから、Multi-Histogram Reweighting Technique（Weighted Histogram Analysis Method (WHAM) とも呼ばれる）を使用し、次の WHAM 方程式を自己無撞着に解くことで最適解が得られる。

$$\begin{cases} g(E) = \frac{\sum_{m=1}^M H_m(E)}{\sum_{m=1}^M n_m e^{f_m - \beta_m E}}, \\ e^{-f_m} = \sum_E g(E) e^{-\beta_m E}, \quad (m = 1, 2, \dots, M). \end{cases} \quad (2.68)$$

ここで、 f_m は無次元化されたヘルムホルツの自由エネルギーであり、 $H_m(E)$ は温度 T_m で得られたエネルギーのヒストグラム、 n_m はその総サンプル数である。DOS の最適解が求まれば、物理量の期待値は式 (2.49) によって求められる。ポテンシャルエネルギーの関数で表せられない物理量であっても、シミュレーション中に一定間隔で保存した物理量 $A(x_m(k))$ のデータを使用して、

$$\langle A \rangle_T = \frac{\sum_{m=1}^M \sum_{k=1}^{n_m} A(x_m(k)) \frac{1}{\sum_{l=1}^M n_l \exp[f_l - \beta_l E(x_m(k))]} \exp[-\beta E(x_m(k))]}{\sum_{m=1}^M \sum_{k=1}^{n_m} \frac{1}{\sum_{l=1}^M n_l \exp[f_l - \beta_l E(x_m(k))]} \exp[-\beta E(x_m(k))]} \quad (2.69)$$

で求めることができる。 β は任意の温度 T における逆温度 $\beta = 1/k_B T$ である。ここで $(x_m(k))$ は温度 T_m で得られた k 番目のサンプルデータである。

これまでの節で導入した3つの拡張アンサンブル法 (MUCA, WL, REM) は、組み合わせることによって計算効率を高めることができる。以降の節では、REM と MUCA を組み合わせた手法 (Multicanonical Replica-Exchange Method : MUCAREM) と、REM と WL を組み合わせた手法 (Replica-Exchange Wang-Landau Algorithm : REWL) を紹介する。

2.4.4 マルチカノニカルレプリカ交換法

原理

マルチカノニカルレプリカ交換法 (Multicanonical Replica-Exchange Method: MUCAREM) では、マルチカノニカル重み因子が与えられた後、少ないレプリカによるマルチカノニカルアンサンブル上のレプリカ交換シミュレーションを実行する。このとき、各レプリカは異なったエネルギー領域をカバーするマルチカノニカルシミュレーションを実行する。

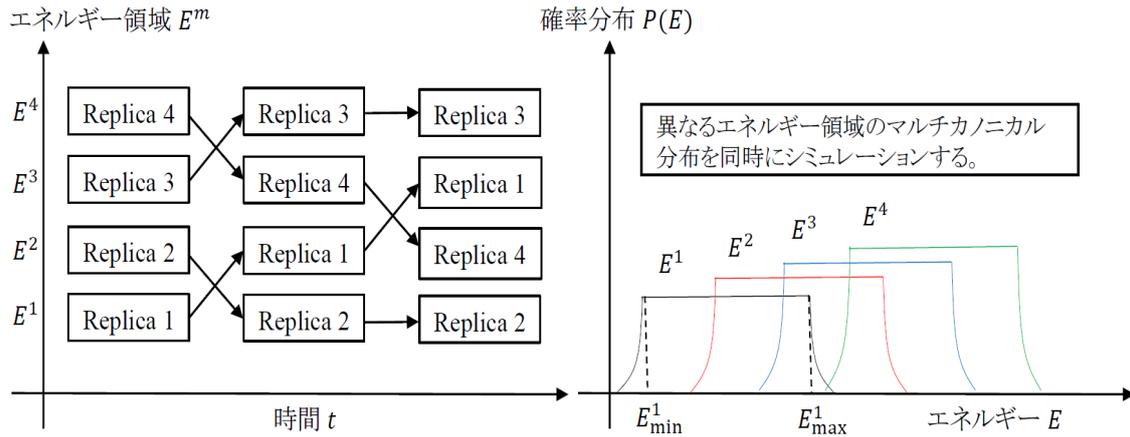


Fig. 2.5: マルチカノニカルレプリカ交換法の概念図。左図はマルチカノニカルレプリカ交換法における各レプリカの時間発展を示す。レプリカ交換が受託されると、各レプリカは MUCAREM シミュレーションを行うエネルギー領域を切り替える。右図は重み因子 $W_{\text{MUCAREM}}(X)$ によって得られる確率分布を示す。

図 2.5 はマルチカノニカルレプリカ交換法の概念図を示す。マルチカノニカルシミュレーションは、従来のカノニカルシミュレーションよりも幅広いエネルギー領域を一度のシミュレーションでカバーできるため、MUCAREM に必要なレプリカ数は普通の REM よりも少なく、計算コストを抑えることができる。

はじめにシミュレーションしたい全エネルギー領域 $[E_{\min}, E_{\max}]$ を、 M 個の sub region $[E_L^{\{m\}}, E_H^{\{m\}}]$ ($m = 1, 2, \dots, M$) に分割する。MUCAREM においても REM のように、元の系の M 個のコピー (レプリカ) を用意する。それぞれのレプリカにはひとつのエネルギー領域 $[E_L^{\{m\}}, E_H^{\{m\}}]$ が割り当てられる。このとき、MUCAREM の拡張アンサンブルの重み因子は REM と同様に各レプリカの重み因子の積として、

$$W_{\text{MUCAREM}}(X) = \prod_{i=1}^M W_{\text{MUCA}}^{\{m(i)\}}(E(x_m^{[i]})) \quad (2.70)$$

$$= \prod_{m=1}^M W_{\text{MUCA}}^{\{m\}}(E(x_m^{[i(m)]})) . \quad (2.71)$$

で定義される。ここでエネルギー領域についての MUCA 重み因子 $W_{\text{MUCA}}(E)$ は、各エネルギー領域 m ごとに準備する必要があり、次のように定義される [11, 12, 13]。

$$W_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]})) = \exp \left[-\beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]})) \right] \quad (2.72)$$

$$\equiv \frac{1}{g^{\{m\}}(E(x_m^{[i]}))} . \quad (2.73)$$

ここで、 M 個の任意の参照温度 $T_m (= 1/k_B \beta_m)$ を準備したが、最終的な結果はこれらの温度には依存しない。

各 MUCA 重み因子 $W_{\text{MUCA}}^{\{m\}}(E)$ または状態密度 $g^{\{m\}}(E)$ は以下で定義される。それぞれの m ($m = 1, 2, \dots, M$) の値について、温度 $T_L^{\{m\}}, T_H^{\{m\}}$ を用意する。ここでは、 $T_L^{\{m\}} < T_L^{\{m\}}$ を仮定し、温度分布が隣り合う領域について十分なオーバーラップを持つようにする。一

般性を失うことなく、 $T_L^{\{1\}} < T_L^{\{2\}} < \dots < T_L^{\{M\}}$ および $T_H^{\{1\}} < T_H^{\{2\}} < \dots < T_H^{\{M\}}$ と仮定することができる。そして $[E_L^{\{m\}}, E_H^{\{m\}}]$ を次のように定義する。

$$\begin{cases} E_L^{\{m\}} = \langle E \rangle_{T_L^{\{m\}}}, \\ E_H^{\{m\}} = \langle E \rangle_{T_H^{\{m\}}}, (m = 1, 2, \dots, M). \end{cases} \quad (2.74)$$

ここで、 $E_L^{\{a\}}$ は E_{\min} に、 $E_H^{\{M\}}$ は E_{\max} に対応する。

いま、対象としている全エネルギー領域 $[E_{\min}, E_{\max}]$ において、マルチカノニカル重み因子 $W_{\text{MUCA}}(E)$ 、あるいは式(2.40)のマルチカノニカルポテンシャルエネルギー $E_{\text{MUCA}}(E; T_0)$ が得られているとする。このとき、それぞれの $(m = 1, 2, \dots, M)$ について、以下のマルチカノニカルポテンシャルエネルギーが定義できる。

$$\varepsilon_{\text{MUCA}}^{\{m\}}(E) = \begin{cases} \frac{\partial E_{\text{MUCA}}(E_L^{\{m\}}; T_m)}{\partial E} (E - E_L^{\{m\}}) + E_{\text{MUCA}}(E_L^{\{m\}}; T_m), & \text{for } E < E_L^{\{m\}}, \\ E_{\text{MUCA}}(E; T_m) & \text{for } E_L^{\{m\}} \leq E \leq E_H^{\{m\}}, \\ \frac{\partial E_{\text{MUCA}}(E_H^{\{m\}}; T_m)}{\partial E} (E - E_H^{\{m\}}) + E_{\text{MUCA}}(E_H^{\{m\}}; T_m), & \text{for } E > E_H^{\{m\}}. \end{cases} \quad (2.75)$$

モンテカルロ法の場合、MUCAREM の重み因子は式(5.10)を式(5.12)に代入することで、以下ようになる。

$$W_{\text{MUCA}}^{\{m\}}(E(x)) = \begin{cases} e^{-\beta_L^{\{m\}} E}, & \text{for } E < E_L^{\{m\}}, \\ \frac{1}{g_m(E)}, & \text{for } E_L^{\{m\}} \leq E \leq E_H^{\{m\}}, \\ e^{-\beta_H^{\{m\}} E}, & \text{for } E > E_H^{\{m\}}, \end{cases} \quad (2.76)$$

ここで、 $\beta_L^{\{m\}} = 1/k_B T_L^{\{m\}}$ 、 $\beta_H^{\{m\}} = 1/k_B T_H^{\{m\}}$ となる。

最後に MUCAREM におけるメトロポリス判定に従ったレプリカ交換の遷移確率を考える。 m 番目と $m+1$ 番目の隣接したエネルギー領域のマルチカノニカルアンサンブルに対応するレプリカ対 i と j を交換する。すなわち、REM の式(2.60)と同様に $X = \{\dots, x_m^{[i]}, \dots, x_{m+1}^{[j]}, \dots\} \rightarrow X' = \{\dots, x_m^{[j]}, \dots, x_{m+1}^{[i]}, \dots\}$ を考える。このレプリカ交換の遷移確率 $w_{\text{MUCAREM}}(X \rightarrow X')$ は、

$$\begin{aligned} w_{\text{MUCAREM}}(X \rightarrow X') &= \min \left[1, \frac{W_{\text{MUCAREM}}(X')}{W_{\text{MUCAREM}}(X)} \right], \\ &= \min [1, \exp[-\Delta]]. \end{aligned} \quad (2.77)$$

ここで Δ は、

$$\begin{aligned} \Delta &= \beta_m \left\{ \varepsilon_{\text{MUCA}}^{\{m\}} (E(q^{[j]})) - \varepsilon_{\text{MUCA}}^{\{m\}} (E(q^{[i]})) \right\} \\ &\quad - \beta_{m+1} \left\{ \varepsilon_{\text{MUCA}}^{\{m+1\}} (E(q^{[j]})) - \varepsilon_{\text{MUCA}}^{\{m+1\}} (E(q^{[i]})) \right\}. \end{aligned} \quad (2.78)$$

となる。ここで、 $(E(q^{[i]}))$ と $(E(q^{[j]}))$ はそれぞれ i 番目と j 番目のレプリカのポテンシャルエネルギーである。

アルゴリズム

実際の MUCAREM シミュレーションは次の STEP を繰り返すことによって実行される。

1. 各レプリカ i ($i = 1, 2, \dots, M$) において、それぞれ、一定のマルチカノニカル重み因子 $W_{\text{MUCA}}^{\{m\}}(E)$ ($m = 1, 2, \dots, M$) に基づく MUCA MC あるいは MUCA MD シミュレーションを、決められた STEP 数同時に独立して実行する。
2. 遷移確率 (2.77) に従ってレプリカ交換を行う。
3. 手順 1 へ戻る。

この MUCAREM においては、 m 番目のマルチカノニカルアンサンブルは、 $E < E_L^{\{m\}}$ の領域では温度 $T = T_L^{\{m\}}$ のカノニカルシミュレーションを、 $E_L^{\{m\}} < E < E_H^{\{m\}}$ の領域では温度マルチカノニカルシミュレーションを、 $E_H^{\{m\}} < E$ の領域では温度 $T = T_H^{\{m\}}$ のカノニカルシミュレーションを実行することになる。

任意の中間温度 T におけるカノニカル分布を得るために、次のように修正された WHAM 方程式が使用される。MUCAREM シミュレーション後に、マルチカノニカル重み因子 $W_{\text{MUCA}}^{\{m\}}(E)$ によって得られたポテンシャルエネルギー分布のヒストグラムと全サンプル数を、それぞれ $H_m(E)$ と N_m とする。物理量 A の期待値 $\langle A \rangle$ は式 (2.49) によって得られるが、その状態密度 $g(E)$ の最適解は次の WHAM 方程式を自己無撞着的に解くことで得られる [19]。

$$\left\{ \begin{aligned} g(E) &= \frac{\sum_{m=1}^M H_m(E)}{\sum_{m=1}^M N_m \exp[f_m] W_{\text{MUCA}}^{\{m\}}(E)}, \\ &= \frac{\sum_{m=1}^M H_m(E)}{\sum_{m=1}^M N_m \exp\left[f_m - \beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E)\right]}, \\ \exp[-f_m] &= \sum_E g(E) W_{\text{MUCA}}^{\{m\}}(E), \\ &= \sum_E g(E) \exp\left[-\beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E)\right]. \end{aligned} \right. \quad (2.79)$$

この修正された WHAM 方程式では、REM における WHAM 方程式 (式 (2.68)) のボルツマン因子 $\exp[-\beta_m E]$ の代わりに、マルチカノニカル重み因子 $W_{\text{MUCA}}^{\{m\}}$ が使用されている。

これらの MUCAREM サンプルングおよび WHAM のプロセスは、原理的には、より正確な DOS を取得するために繰り返すことができる [20]。この場合、DOS の最初の推定値を取得するために REM がよく使用され、REM シミュレーション後に通常の WHAM を解くことで、MUCAREM における初めの DOS の推定値を準備する。

従来のマルチカノニカル法においてレプリカ交換を導入する理由は、通常のモンテカルロ法や分子動力学法では、配位の局所的な変更 (local update) しか実行しないのに対して、レプリカ交換の操作は大局的な変更 (global update) を可能にするためである。これにより、配位空間上での探索効率を大幅にあげることが期待される。

最後に、(ポテンシャルエネルギーの関数で表せないものを含む) 一般の物理量 A の任意の温度 $T (= 1/k_B T)$ における期待値は、シミュレーションの途中の状態 (全粒子の座標) をある間隔で保存することによって、以下のように計算することができる。まず、式 (2.79) を自己無撞着的に解くことによって、 $f_m (m = 1, 2, \dots, M)$ を得る。すると以下の式で任意の温度における期待値が求められる。

$$\langle A \rangle_T = \frac{\sum_{m=1}^M \sum_{k=1}^{n_m} A(x_m(k)) \frac{1}{\sum_{l=1}^M n_l \exp[(f_l)] W_{\text{MUCA}}^{\{l\}}(E(x_m(k)))} \exp[-\beta E(x_m(k))]}{\sum_{m=1}^M \sum_{k=1}^{n_m} \frac{1}{\sum_{l=1}^M n_l \exp[(f_l)] W_{\text{MUCA}}^{\{l\}}(E(x_m(k)))} \exp[-\beta E(x_m(k))]} \quad (2.80)$$

ここで、 $x_m(k)$ ($k = 1, \dots, n_m$) は、それぞれのマルチカノニカル重み因子 $W_{\text{MUCA}}^{\{m\}}(E)$ ($m = 1, \dots, M$) に基づくマルチカノニカルシミュレーションによって別々に得られた k 番目の MC (または MD) の STEP における状態 (全粒子の座標) であり、 n_m はその総数である。

2.4.5 レプリカ交換ワン-ランダウ法

原理

MUCAREM 法と同じように、従来のワン-ランダウ法にレプリカ交換を実装することによって、シミュレーション中に大局的な変更 (global update) を与えることができ、これによりサンプルング効率の向上が期待できる。レプリカ交換法を導入したワン-ランダウ法をレプリカ交換ワン-ランダウ法 (Replica-Exchange Wang-Landau Algorithm: REWL) という。

REWL 法においても、シミュレーションを実行したい全エネルギー領域 $[E_{\min}, E_{\max}]$ を、 M 個の sub region $[E_L^{\{m\}}, E_H^{\{m\}}]$ ($m = 1, 2, \dots, M$) に分割する。そして、各エネルギー領域 m に対応したレプリカでは WL シミュレーションが実行される。従って、REWL 法における拡張アンサンブルの重み因子 $W_{\text{REWL}}(X)$ は、MUCAREM の場合 (式 2.70) と同様に

各レプリカの重み因子の積として、

$$W_{\text{REWL}}(X) = \prod_{i=1}^M W_{\text{WL}}^{\{m(i)\}}(E(x_{m(i)}^{[i]})) \quad (2.81)$$

$$= \prod_{m=1}^M W_{\text{WL}}^{\{m\}}(E(x_m^{[i(m)]})) . \quad (2.82)$$

で定義される。ここで各エネルギー領域についてのワン-ランダウ重み因子 $W_{\text{WL}}^{\{m\}}(E(x_m^{[i(m)]}))$ は、各エネルギー領域 m ごとに準備する必要があり、

$$W_{\text{WL}}(E) = \exp \left[-\beta_m \varepsilon_{\text{WL}}^{\{m\}} \left(E \left(x_m^{[i]} \right) \right) \right], \quad (2.83)$$

$$= \frac{1}{g^{\{m\}} \left(E \left(x_m^{[i]} \right) \right)} . \quad (2.84)$$

と定義される。ここでも M 個の参照温度 $T_m (= 1/k_B \beta_m)$ を準備した。

MUCAREM と同様の議論によって、REWL の重み因子 $W_{\text{REWL}}(X)$ (式 (2.81)) に基づき、

$$\varepsilon_{\text{WL}}^{\{m\}}(E) = \begin{cases} \frac{\partial E_{\text{WL}}(E_L^{\{m\}}; T_m)}{\partial E} (E - E_L^{\{m\}}) + E_{\text{WL}}(E_L^{\{m\}}; T_m), & \text{for } E < E_L^{\{m\}}, \\ E_{\text{WL}}(E; T_m) & \text{for } E_L^{\{m\}} \leq E \leq E_H^{\{m\}}, \\ \frac{\partial E_{\text{WL}}(E_H^{\{m\}}; T_m)}{\partial E} (E - E_H^{\{m\}}) + E_{\text{WL}}(E_H^{\{m\}}; T_m), & \text{for } E < E_H^{\{m\}}. \end{cases} \quad (2.85)$$

となる。モンテカルロ法の場合、各エネルギー領域の WL の重み因子は式 ((2.83)) を式 ((2.85)) に代入することで、以下ようになる。

$$W_{\text{WL}}^{\{m\}}(E(x)) = \begin{cases} e^{-\beta_L^{\{m\}} E}, & \text{for } E < E_L^{\{m\}}, \\ \frac{1}{g_m(E)}, & \text{for } E_L^{\{m\}} \leq E \leq E_H^{\{m\}}, \\ e^{-\beta_H^{\{m\}} E}, & \text{for } E_H^{\{m\}} < E, \end{cases} \quad (2.86)$$

ここで、 $\beta_L^{\{m\}} = 1/k_B T_L^{\{m\}}$, $\beta_H^{\{m\}} = 1/k_B T_H^{\{m\}}$ となる。従って、メトロポリス法によるレプリカ交換の遷移確率 $w_{\text{REWL}}(X \rightarrow X')$ も、式 (2.77) および (2.78) の $W_{\text{MUCAREM}}(X)$ を、 $W_{\text{REWL}}(X)$ に置き換えたものを使用すれば良い。

MUCAREM と REWL の相違点は、REWL の重み因子 $W_{\text{REWL}}(X)$ がシミュレーション中に更新されることである。REWL において各レプリカでは独立して WL シミュレーションを実行し、エネルギー領域 m に対応したワン-ランダウ重み因子 $W_{\text{WL}}^{\{m\}}(E)$ が、式 (2.54)

あるいは式 (2.55) によって各時間 STEP で修正される。このとき、重みの修正係数 f は各エネルギー領域 m に応じた修正係数 f_m となる。またエネルギーのヒストグラム $H(E)$ も、各エネルギー領域 m に応じたヒストグラム $H_m(E)$ として保存する。 $W_{\text{WL}}^{\{m\}}(E)$ の修正は、それぞれのエネルギー領域 $[E_L^{\{m\}}, E_H^{\{m\}}]$ ($m = 1, 2, \dots, M$) で $H_m(E)$ が十分に平坦になるまで続けられる。そして WL 法と同様に $H_m(E)$ の平坦度の評価 (式 (2.56)) を一定の時間 STEP 間隔で行いながら、レプリカ交換を実行し、シミュレーションを時間発展させる。この過程を繰り返し、すべてのエネルギー領域 m ($m = 1, 2, \dots, M$) の修正係数 f_m があらかじめ決められた最小値 f_{final} に達したときに終了する。

アルゴリズム

実際の REWL シミュレーションは次の STEP を繰り返すことによって実行される。

1. 各レプリカ i ($i = 1, 2, \dots, M$) において、それぞれのエネルギー領域 $[E_L^{\{m\}}, E_H^{\{m\}}]$ ($m = 1, 2, \dots, M$) で、ワン-ランダウの重み因子 $W_{\text{MUCA}}^{\{m\}}(E)$ ($m = 1, 2, \dots, M$) に基づく WL MC あるいは WL MD シミュレーションを独立して同時に決められた STEP 数実行する。
2. 遷移確率 (2.77) に従ってレプリカ交換を行う。
3. 手順 1 へ戻る。
4. すべてのエネルギー領域 m ($m = 1, 2, \dots, M$) の修正係数 f_m が $f_m < f_{\text{final}}$ となっていれば、シミュレーションを終了する。

REWL のシミュレーションを終了した後、 M 個に分割されたエネルギー領域 m ごとに $\tilde{g}^{\{m\}}(E)$ が得られる。 $\tilde{g}^{\{m\}}(E)$ から、全エネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ にわたる最終的な $\tilde{g}(E)$ は次のように決定される。まず隣り合うエネルギー領域 $(m, m+1)$ のエネルギーがオーバーラップする領域で、 $\tilde{g}^{\{m\}}(E)$ の対数のエネルギー微分 ($\tilde{\beta}^{\{m\}}(E) = \partial \ln[\tilde{g}^{\{m\}}(E)] / \partial E$) を計算する。この $\tilde{\beta}^{\{m\}}(E)$ と $\tilde{\beta}^{\{m+1\}}(E)$ が最もよく一致するエネルギーの値が接続点 E_{join} として選択される。従って、 $E \leq E_{\text{join}}$ では $\tilde{g}^{\{m\}}(E)$ が、 $E_{\text{join}} \leq E$ では $\tilde{g}^{\{m+1\}}(E)$ が、最終的な DOS の推定値 $\tilde{g}(E)$ となる。この接続過程による誤差を減らすために、

$$g_{\text{rescaled}}^{\{m+1\}}(E) = g^{\{m+1\}}(E) \frac{g^{\{m+1\}}(E_{\text{join}})}{g^{\{m\}}(E_{\text{join}})} \quad (2.87)$$

というリスケーリングなどが提案されている [28, 29]。この過程をすべての m について行うことで、全エネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ にわたる $\tilde{g}(E)$ が得られる。

2.5 REWL-MUCAREM プロトコル

拡張アンサンブル法では、各手法を適切な順序で組み合わせることによってより高い効果を発揮することが報告されている [16]。本研究では、高精度の DOS を得るために、レプリカ交換ワン-ランダウ法とマルチカノニカルレプリカ交換法を組み合わせた計算プロトコル (REWL-MUCAREM) を提案した。この計算プロトコルでは、シミュレーション

の前半で REWL を使用し、MUCAREM の重み因子 $W_{\text{MUCAREM}}(X)$ を見積もり、続けて MUCAREM を実行することで、WHAM 法によって最も高い精度を持った DOS を計算できる。この手順でシミュレーションを実行することによって、WL 法 (REWL 法) で報告されている計算アルゴリズムに伴った系統誤差 [24] を抑えることができ、REWL における DOS の接続も行う必要はない。また MUCAREM シミュレーションでは、あらかじめ重み因子 $W_{\text{MUCAREM}}(X)$ を準備する必要があり、これには REM 法が推奨されてきたが、REWL 法を用いることで、REM 法で困難とされる一次相転移を持つ系であっても、重み因子 $W_{\text{MUCAREM}}(X)$ を得ることができる。このようにして得られた DOS は、同じ計算コストであっても、それぞれの手法で独立して推定された精度よりも、高い精度を持つとわかった (第 3 章)。本研究では、REWL-MUCAREM プロトコルを 3 種類の異なる性質を持った系 (Ising 系、氷 I_h 系、生体高分子系) に適用し、その有効性を示した。

References

- [1] T. Hayashi and Y. Okamoto, Phys. Rev. E **100**, 043304 (2019).
- [2] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
- [3] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
- [4] W. Janke, Physica A **254**, 164 (1998).
- [5] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
- [6] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
- [7] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
- [8] C. J. Geyer, in Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, E. M. Keramidas (ed.), (Interface Foundation, Fairfax Station, 1991) pp. 156-163.
- [9] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140, (1997).
- [10] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
- [11] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
- [12] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6664 (2003).
- [13] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6676 (2003).
- [14] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
- [15] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. E **90**, 023302 (2014).
- [16] 岡本祐幸. 拡張アンサンブル法. 「計算科学講座」第9巻「超多自由度への最適化」第2章. 古橋武, 笹井理生 (編), 共立出版, 110 (2013).
- [17] Bernd A. Berg, Markov Chain Monte Carlo Simulation and Their Statistical Analysis, World Scientific, Singapore, (2004).
- [18] David P. Landau and Kurt Binder, A guide to Monte Carlo simulations in statistical physics. Cambridge university press, (2014).
- [19] 岡崎進; 吉井範行. 「コンピュータ・シミュレーションの基礎: 分子のミクロな性質を解明するために」. 化学同人, (2011).

- [20] Michael P. Allen and Dominic J. Tildesley, *Computer simulation of liquids*. Oxford university press, (2017).
- [21] 伊庭幸人; 種村正美; 大森裕浩; 和合肇; 佐藤整尚; 高橋明彦. 「計算統計Ⅱ マルコフ連鎖モンテカルロ法とその周辺」. 岩波書店, (2005).
- [22] J. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.* **97**, 050601 (2006).
- [23] H. Shimizu, K. Uehara, K. Yamamoto and Y. Hiwatari, *Mol. Simul.*, 22, 285 (1999)

第3章 Ising model

T. Hayashi and Y. Okamoto, Physical Review E 100, 043304 (2019).

3.1 導入

状態密度 (DOS) が与えられれば、物理量の統計力学的な期待値を計算できる (第2章)。しかし多くの場合、DOS は事前にわかっておらず、また理論的にも実験的にも推定することが難しい。そのため、モンテカルロ (MC) 法や分子動力学 (MD) といった計算機シミュレーションによって、DOS を決定するため手法が、過去数十年で開発されてきた [1]–[31]。最も初期のそのような方法の1つは Umbrella Sampling 法 [1] であり、続いて Muticanonical Algorithm [2]–[4]、Simulated Tempering [5]–[7]、Replica-Exchange Method [8]–[11]、Wang-Landau Method [12, 13]、および Meta-dynamics [14]–[16] が開発された。さらに、これらの手法を一般化したものも提案されている [17]–[31]。これらの手法には密接な関係性がある。たとえば、適切な条件下で Statistical Temperature Molecular Dynamics [22] は Meta-dynamics と同等であることが示されている [30]。また、Meta-dynamics は、エネルギー空間ではなく反応座標空間における Wang-Landau 法と見なすことができる [32]。従って、各手法で提案された改良アルゴリズムは、別の手法に応用できる可能性がある。これらのシミュレーション手法は、スピングラス、液晶、ポリマー、タンパク質など、物性物理学や統計物理学の様々な問題に適用されてきたが、それでもなお、シミュレーション対象となる系が大きく複雑になるにつれて、DOS を正確に計算できないという問題が残っている。

本研究では、Replica-Exchange Wang-Landau (REWL) method [28, 29] と Multicanonical Replica-Exchange Method (MUCAREM) [18, 19, 20] を組み合わせることにより、最も正確な DOS を得るための効率的なシミュレーションプロトコル REWL-MUCAREM を提案した (第2章)。本章では、REWL-MUCAREM の有用性を検証するため、REWL、MUCAREM、REWL-MUCAREM の3つの方法で正方格子イジングモデルの MC シミュレーションを実行し、得られた DOS を比較した。その結果、REWL-MUCAREM によって得られた DOS は、REWL および MUCAREM によって単独で推定された DOS よりも正確であることを示した。

Table 3.1: Conditions of the present simulations.

Methods	Number of spins N	Number of replicas M	Frequency (in MC sweeps) of flatness evaluation	Total MC sweeps per replica
MUCAREM	64	4	NA	200 000
	256	8		200 000
	1024	16		200 000
	4096	32		500 000
	16384	64		3 000 000
REWL	64	4	1000	200 000
	256	8		200 000
	1024	16		200 000
	4096	32		500 000
	16384	64		3 000 000
REWL–MUCAREM	64	4	1000–NA	100 000–100 000
	256	8		100 000–100 000
	1024	16		100 000–100 000
	4096	32		250 000–250 000
	16384	64		1 500 000–1 500 000

3.2 計算モデルと初期条件

3.2.1 2次元イジング模型

周期境界条件を課した2次元正方格子イジング模型のシミュレーションを実行した。この模型では、系の全ポテンシャルエネルギー E は次の式で定義される。

$$E = -J \sum_{\langle i,j \rangle} S_i S_j, \quad (3.1)$$

ここで、 i と j は格子点の番号を表す。 J は隣接するスピン間の相互作用の大きさである。簡単のため、本章では J とボルツマン因子 k_B を 1 とした。 $\sum_{\langle i,j \rangle}$ は最隣接スピンの全ての対に対して和をとった。 S_i は、格子点 i 上のスピンの状態であり、 ± 1 の値を取る。有限系の2次元イジング模型のDOSの厳密解は参考文献 [34, 35] で既に得られている。

3.2.2 初期条件

Table 3.1 に、今回のシミュレーションの初期条件を示した。スピンの総数 N は L^2 個であり、ここで L は正方格子の一辺の長さを表す。シミュレーションを実行した系のスピン総数は $N = 64, 256, 1024, 4096, 16384$ とした。またメトロポリス判定 N 回を 1 MC sweep として定義し、各手法での計算コスト (MC sweep の総数) は等しくなるように設定した。注意点として、通常のREWLアルゴリズムは、重み修正係数 f が f_{final} に収束したときに終了するが、本シミュレーションでは一定回数の MC sweep 後に終了した。

擬似乱数として、Marsaglia 擬似乱数発生器 [36] を採用した。これにはオープンソースのプログラムコードを使用した [37]。レプリカ数は系の大きさに合わせて $L/2$ 個とした。各レプリカは分割されたエネルギー領域内で、MUCAREM では MUCA シミュレーションを、REWL では WL シミュレーションを実行した。隣接するエネルギー領域間には 80% のオーバーラップを設け、レプリカ交換は 100 MC sweep ごとに試行された。REWL および REWL-MUCAREM シミュレーションの場合、WL のヒストグラム平坦度判定は 1000 MC sweep ごとにテストされた。このとき、エネルギーのヒストグラムが十分平坦である場合、WL の重み修正係数を $f \rightarrow \sqrt{f}$ としてを更新した (式 (2.54) あるいは式 (2.55))。Table 3.1 の条件に従い、異なる乱数の種を使用して 25 回 ($n = 25$) の独立したプロダクションランを行った。本研究では簡単のため、MUCAREM と WHAM を繰り返さなかった。従って今回の MUCAREM では、全 MC sweep の前半は REM を実行し、残りは REM シミュレーションから取得した DOS を使用した MUCAREM を行った。付録 A では、この MUCAREM と WHAM プロセスの反復の有効性を評価した。MUCAREM における REM では、逆温度 β は $\beta_{\text{min}} = 0.01$ と $\beta_{\text{max}} = 1$ の間でレプリカ数に合わせて均等に分布させた。

3.3 結果と考察

3.3.1 比熱の比較

図 3.1(A) の 4 つの図は、各手法で推定した DOS から、次の式によって計算された比熱 $C(T)$ を示す。これにより、各手法による $g(E)$ の推定値の測定精度が、物理量の計算に与

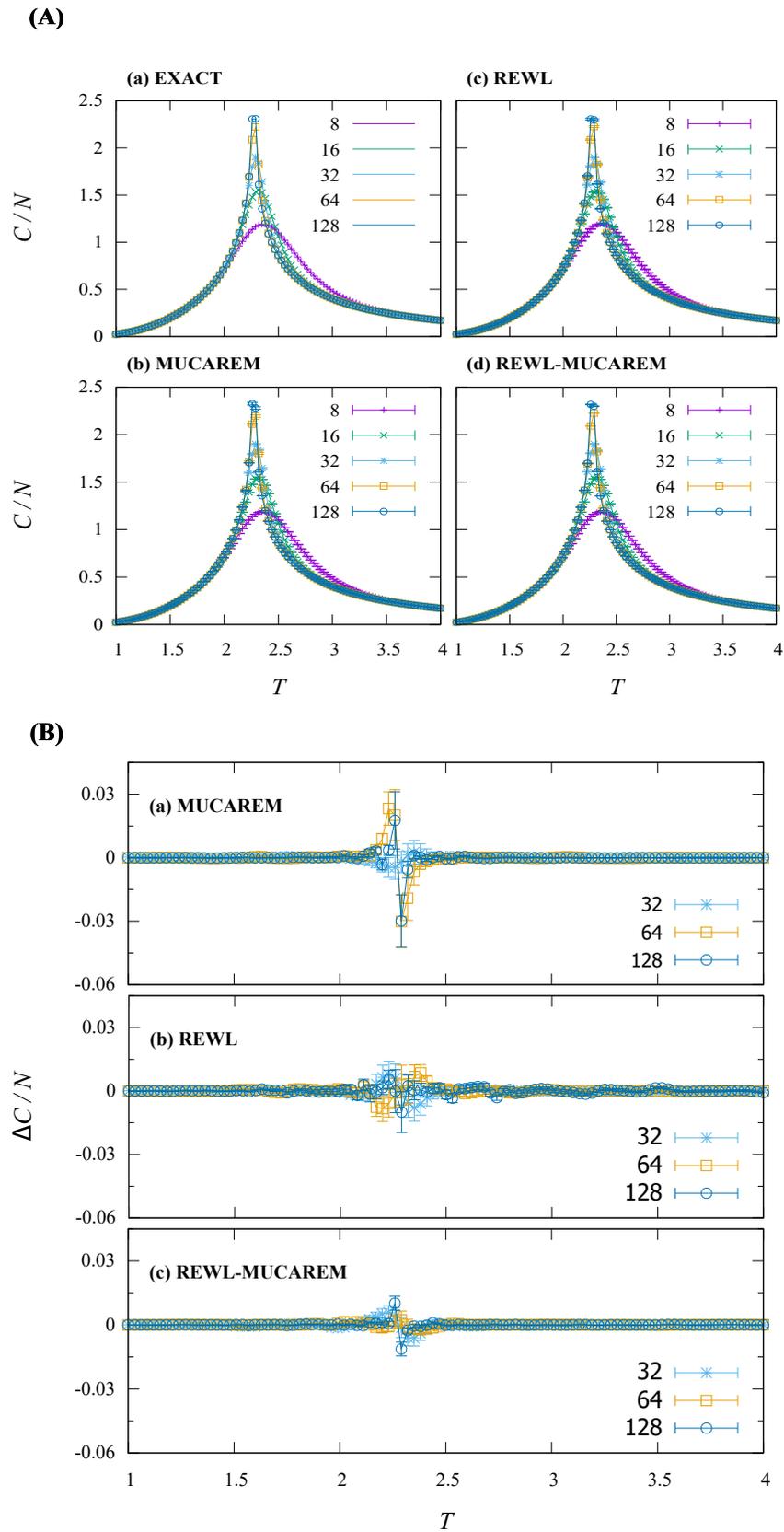


Fig. 3.1: 比熱. (A) の (a) は状態密度の厳密解 [34, 35] によって計算された比熱であり、(b), (c) (d) はそれぞれ MUCAREM、REWL、REWL-MUCAREM によって計算された比熱の値である。(B) は厳密解とシミュレーションで得られた比熱の差分 $\Delta C(T) = C_{\text{sim}}(T) - C_{\text{EXACT}}(T)$ をプロットした図である.

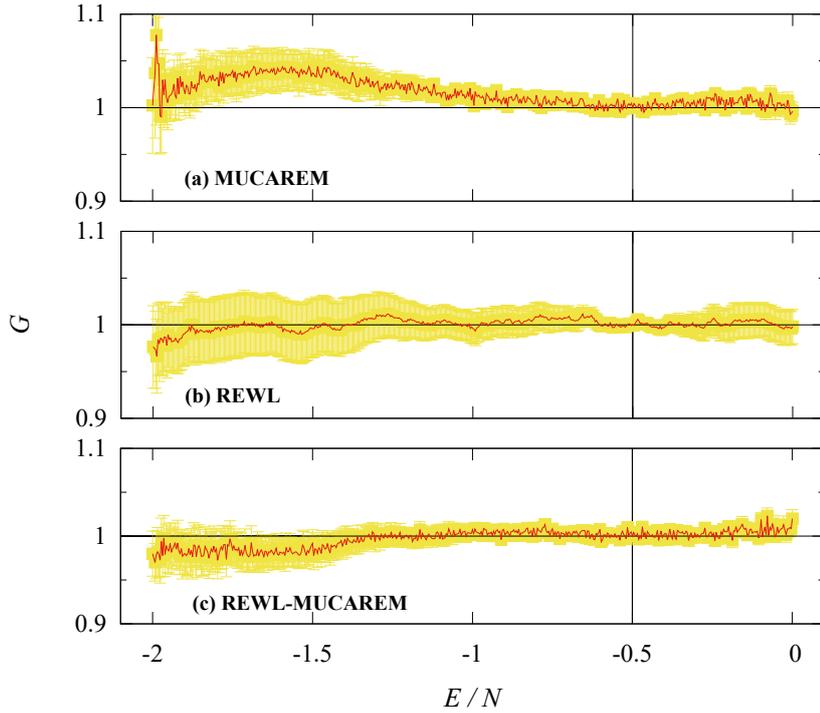


Fig. 3.2: Eq. (3.5) によって定義された Mean local flatness $G(E)$ の平均値。全スピンの数は $N = 32 \times 32$ であった。 (a) (b) (c) はそれぞれ MUCAREM、REWL、REWL-MUCAREM に対応する。シミュレーションによって得られた DOS が厳密解に近いほど $G(E)$ は 1 に近い値を取る。

える影響の大きさを確認した。

$$C(T) = \frac{\langle E^2 \rangle_T - \langle E \rangle_T^2}{T^2}, \quad (3.2)$$

ここで

$$\langle A \rangle_T = \frac{\sum_E A(E) g(E) e^{-\beta E}}{\sum_E g(E) e^{-\beta E}}, \quad (3.3)$$

であり、 $A(E)$ はポテンシャルエネルギー E に依存した物理量を表す。誤差は次の式、

$$\varepsilon_A = \sqrt{\frac{\sum_{i=1}^n (A^{\{i\}} - \bar{A})^2}{n(n-1)}}, \quad \bar{A} = \frac{\sum_{i=1}^n A^{\{i\}}}{n}. \quad (3.4)$$

で求めた。 $A^{\{i\}}$ は i 回目のシミュレーションから得られた値である ($i = 1, 2, \dots, n$)。有限系の比熱の厳密解は Ferdinand と Fisher によって得られているが [38]、今回は DOS の厳密解 $g_{\text{EXACT}}(E)$ [34, 35] から、(3.2) と (3.3) 式を使用して、図 3.1 (A) (a) の比熱を得た。 $g_{\text{EXACT}}(E)$ の計算には、[35] に公開されている Mathematica の計算プログラムを使用した。REWL, MUCAREM, REWL-MUCAREM による比熱の推定値は、厳密解の値を良く再現した。

厳密解とシミュレーション結果の差分 $\Delta C(T)$ を図 3.1 (B) に示す。 $\Delta C(T)$ はどの手法においても相転移温度 $T_c = 2 / \ln(1 + \sqrt{2}) \simeq 2.269$ 付近で最大値を取った。すなわ

ち、DOSのシミュレーションによる推定値と厳密解との差が、 T_c 周辺における比熱の推定値に最も大きな影響を与えたことを示す。 $\Delta C(T)$ は、REWL-MUCAREM、REWL、MUCAREMの順で厳密解に近い値を与えていた。これは、REWL-MUCAREMが他の2つの方法と比較して、より正確なDOSを取得できたことを意味する。

3.3.2 状態密度の比較

3つの手法で得られたDOSを直接比較した。ここでは、DOSの厳密解($g_{\text{EXACT}}(E)$)とシミュレーションで推定されたDOS($g_{\text{sim}}(E)$)を用いて、正確さを表すための指標としてLocal Flatness $G(E)$ を次のように定義した。

$$\begin{cases} G(E) &= \frac{1}{n} \sum_{i=1}^n G^{\{i\}}(E), \\ G^{\{i\}}(E) &= \frac{g_{\text{sim}}^{\{i\}}(E)}{g_{\text{EXACT}}(E)}. \end{cases} \quad (3.5)$$

ここで、 $g_{\text{sim}}^{\{i\}}$ は*i*番目のシミュレーションから推定されたDOSである($i = 1, 2, \dots, n$)。 $g_{\text{sim}}(E)$ が $g_{\text{EXACT}}(E)$ と完全に等しい場合、 $G(E)$ はシミュレーションを行った全エネルギー領域 $[E_{\text{min}}, E_{\text{max}}]$ で1を取る。MUCA法によるDOSの推定値 $g_{\text{sim}}(E)$ と厳密解 $g_{\text{EXACT}}(E)$ との間には、アルゴリズムに由来する定数倍の不確実性が存在する。これを除くため、 $E/N = -0.5$ で、 $\ln g_{\text{EXACT}}(E)$ と $\ln g_{\text{sim}}(E)$ の値を一致させた。

$N = 32 \times 32$ の系のLocal Flatness $G(E)$ の平均値を図3.2に示す。誤差は式(3.4)によって見積もった。全体的な傾向として、赤の曲線と黒の直線($G(E) = 1$)との差は、低エネルギー領域で大きくなった。相転移がより強くなる大規模な系ではこの傾向がより強くなった。どの系においても低エネルギー領域でのREWL-MUCAREMの誤差は3つの方法の中で最も小さく、これはREWLおよびMUCAREMよりも正確なDOSを取得できたことを示唆する。

DOSの精度をより詳細に調べるため、次の式でGlobal Flatness F を定義した。

$$F \equiv \frac{G_{\text{min}}}{G_{\text{max}}}, \quad (3.6)$$

ここで、 G_{min} は全エネルギー領域にわたる $G(E)$ の最小値であり、 G_{max} は最大値である。従って F は0から1までの値を取る。シミュレーションによって得られた $g_{\text{sim}}(E)$ が全エネルギー領域に渡って $g_{\text{EXACT}}(E)$ に近いほど、 F は1に近い値を取る。図3.3は、各大きさの系で測定されたGlobal Flatness F を示す。このとき F の誤差は、 G の最小値と最大値を用いて、

$$\begin{cases} F_{\text{min}} &= \frac{G_{\text{min}} - \frac{1}{2}\varepsilon(E_{\text{glmin}})}{G_{\text{max}} + \frac{1}{2}\varepsilon(E_{\text{glmax}})}, \\ F_{\text{max}} &= \frac{G_{\text{min}} + \frac{1}{2}\varepsilon(E_{\text{glmin}})}{G_{\text{max}} - \frac{1}{2}\varepsilon(E_{\text{glmax}})}. \end{cases} \quad (3.7)$$

と定義した。ここで、 $\varepsilon(E)$ は $G(E)$ の標準誤差であり、 E_{glmin} と E_{glmax} は、それぞれ G_{min} と G_{max} が得られるエネルギー値である。系が大きくなると、 F が小さくなることが見て

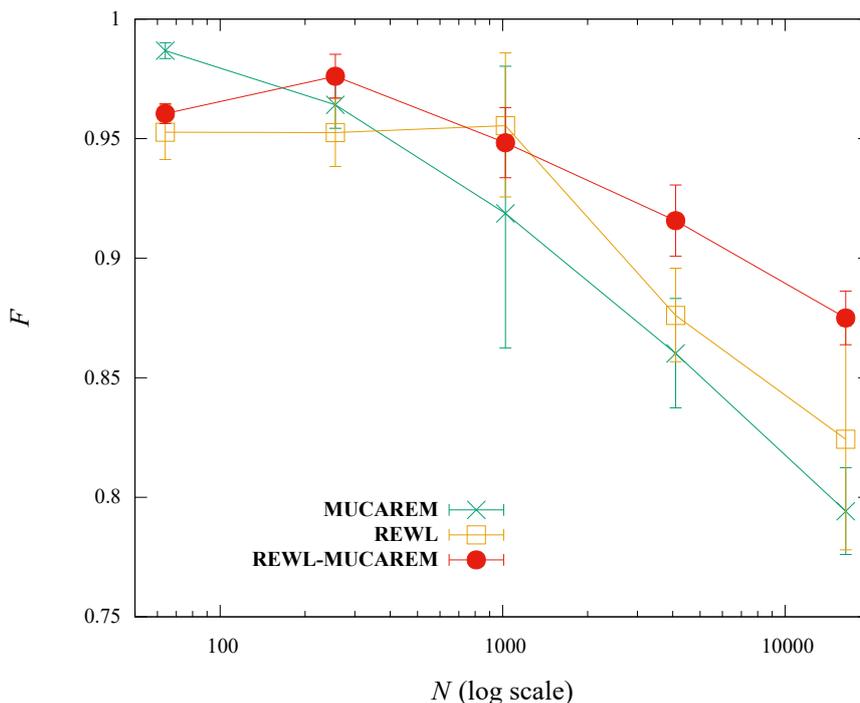


Fig. 3.3: Eq. (3.6) で定義された Global flatness F 。全エネルギー領域で $g_{\text{sim}}^{\{i\}}(E)$ が $g_{\text{EXACT}}(E)$ に近いほど、 F は 1 に近い値をとる。

とれる (図 3.3)。これは、系の自由度が大きくなるにつれて DOS の推定が困難になったことを意味する。 F の測定値に関して、 $N = 16 \times 16$ まで 3 つの方法の間に大きな違いは見られなかった。また MUCAREM では $N = 32 \times 32$ には、1 つの外れ値が見つかり、他の方法よりも誤差が大きくなっていった。 $N = 32 \times 32$ より大きな系では、REWL-MUCAREM は 3 つの方法の中で最も大きい F を与えた。これは REWL-MUCAREM によって測定された DOS が REWL, MUCAREM それぞれの手法で別々に推定された DOS よりも正確であったことを示す。

本研究の MUCAREM シミュレーションでは、2 回の多ヒストグラム再重法 (WHAM) を実行した。1 つは計算前半の REM シミュレーションの後、もう 1 つは後半の MUCAREM シミュレーションの後である。2 回目の WHAM は、1 回目の WHAM で推定された DOS をもとに自己無撞着的に再計算するため比較的収束が早いですが、はじめの WHAM では収束が遅くなる可能性がある。すなわち大規模な系では、WHAM 収束のための計算コストが無視できなくなる。REWL-MUCAREM シミュレーションでは、2 回目のみ WHAM を使用し、1 回目の WHAM を行う必要はない。これは REWL シミュレーションによって DOS の適切な推定値がすでに準備されているためであり、少ない計算コストで 2 回目の WHAM を実行できる。また、REWL-MUCAREM では、WHAM によって全エネルギー領域で DOS を自動的に求めるため、REWL で必要な、レプリカに対応したエネルギー領域ごとの DOS を接続する過程は必要ない。従って、はじめに REWL を行い、次に MUCAREM を実行することが、REWL および MUCAREM を最も効率よく扱える順序であるといえる。

3.4 結論

本研究では、状態密度を高精度で推定するためのシミュレーションプロトコルを調査し、REWLとMUCAREMの利点を組み合わせたREWL-MUCAREMを提案した。この手法では、REWLを最初に実行し、次にMUCAREMが実行される。このプロトコルを正方格子イジングモデルを使用して、REWLとMUCAREMで独立に得られた状態密度と比較した。その結果、REWL-MUCAREMが最も正確な状態密度を与えることを示した。REWL-MUCAREMは他の系でも効果的であると考えられる。

Appendix A: MUCAREMにおける最適な初期条件

REWL-MUCAREMでより高精度なDOSを得るため、MUCAREMでの最適な計算条件について考察した。そのため、初期条件の異なる2つのMUCAREMシミュレーションを追加で実行した。Table 3.2に、追加のシミュレーション（MUCAREM2およびMUCAREM3）と、本文中でのMUCAREM（MUCAREM1）の初期条件を示す。MUCAREM1と、MUCAREM2,3の主な違いは、REMとMUCAREMのMC sweepの数、REMに使用されるレプリカの数、およびMUCAREMとWHAMの反復回数である。MUCAREM2およびMUCAREM3では、MC sweep全体のはじめ10%でREMを実行し、残りの90%を、REMシミュレーションから取得したDOSを重み因子として使用したMUCAREMを実行した。このときのREMでは、十分な回数のレプリカ交換を行うため、10 MC sweepごとにレプリカ交換を実行した。一方で続けて実行されるMUCAREMシミュレーションでは、100 MC sweepごとにレプリカ交換を行った。一般的にMUCAREMとWHAMを繰り返すことで、DOSがより高精度になると言われている [20]。従ってMUCAREM3では、MUCAREMシミュレーションを1回繰り返した。

また25回の測定のうち明らかに誤ったDOSを確認したため、MUCAREM2の $N = 32 \times 32$ および $N = 128 \times 128$ では1回余分に追加計算を実行し、誤ったDOSは平均値の計算に用いなかった。（本文中のREWLおよびREWL-MUCAREMシミュレーションでは、明らかに誤った結果は得なかった。）これは、REMで広いエネルギー範囲にわたって均一にサンプリングすることが難しいことに起因すると考えられる。すなわち最初のREMから得られたDOSの精度が悪く、この不正確さが続けて実行されたMUCAREMのサンプリングに悪影響を及ぼしたと考えられる。間違った推定値を与えることは、MUCAREMがREWLおよびREWL-MUCAREMと比較して不安定であることを示唆する。

Table 3.2: Conditions of MUCAREM simulations.

Methods	Number of spins N	MUCAREM iterations	Number of replicas		Number of MC sweeps per replica		Total MC sweeps
			REM	MUCAREM	REM	MUCAREM	
MUCAREM1	64	1	4	4	100 000	100 000	$200\,000 \times 4$
	256		8	8	100 000	100 000	$200\,000 \times 8$
	1024		16	16	100 000	100 000	$200\,000 \times 16$
	4096		32	32	250 000	250 000	$500\,000 \times 32$
	16384		64	64	1 500 000	1 500 000	$3\,000\,000 \times 64$
MUCAREM2	64	1	8	4	10 000	180 000	$200\,000 \times 4$
	256		16	8	10 000	180 000	$200\,000 \times 8$
	1024		32	16	10 000	180 000	$200\,000 \times 16$
	4096		64	32	25 000	450 000	$500\,000 \times 32$
	16384		128	64	150 000	2 700 000	$3\,000\,000 \times 64$
MUCAREM3	64	2	8	4	10 000	90 000	$200\,000 \times 4$
	256		16	8	10 000	90 000	$200\,000 \times 8$
	1024		32	16	10 000	90 000	$200\,000 \times 16$
	4096		64	32	25 000	225 000	$500\,000 \times 32$
	16384		128	64	150 000	1 350 000	$3\,000\,000 \times 64$

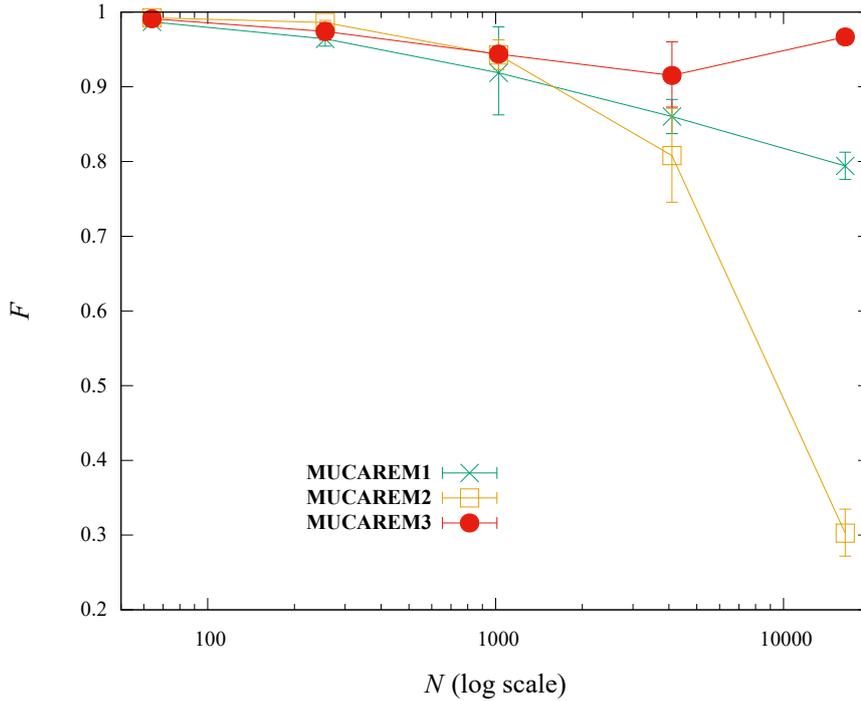


Fig. 3.4: Eq. (3.6) で定義された Global flatness F 。全エネルギー領域で $g_{\text{sim}}^{\{i\}}(E)$ が $g_{\text{EXACT}}(E)$ に近いほど、 F は 1 に近い値をとる。 $N = 64 \times 64$ の系を除き、MUCAREM3 の誤差はデータ点よりも十分に小さかった。

図 3.4 は、式 (3.6) で定義される Global Flatness F を示す。 $N = 64 \times 64$ までの F には、3 つの MUCAREM シミュレーション間でほとんど違いはないが、 $N = 128 \times 128$ に大きな差異が見て取れた。MUCAREM3 は、MUCAREM1 および MUCAREM2 と比較して、真の DOS に近い DOS の推定値を得た。これは、長時間の MUCAREM から得た多くのサンプルを使用した計算条件 (MUCAREM2) よりも、MUCAREM シミュレーションを繰り返す計算条件 (MUCAREM3) が、より高精度な DOS が得られることを意味する。

また MUCAREM3 の $N = 64 \times 64$ の誤差は、他の大きさの系と比較した場合に大きかった。これは、25 回のシミュレーションのうち 1 回で DOS の推定値に外れ値があり、これにより誤差が大きくなっていったとわかった。これもまた、MUCAREM シミュレーションが REWL および REWL-MUCAREM シミュレーションと比較して不安定であることを示唆した。MUCAREM は不安定であるように見えるが、イテレーションを行った MUCAREM3 の $N = 128 \times 128$ の DOS は、図 3.3 の REWL-MUCAREM の推定よりも、より 1 に近い値の F を与えた。従って、REWL-MUCAREM は、MUCAREM の不安定性を取り除くことができ、REWL-MUCAREM での MUCAREM シミュレーションを繰り返すことで、より良い DOS が得られると期待される。すなわち REWL-MUCAREM では、REWL によってシミュレーションの初期段階で安定して DOS を取得でき、MUCAREM と WHAM を繰り返すことで DOS がより正確になる。最適な計算条件はシミュレーション対象となる系によって異なるが、REWL と MUCAREM を組み合わせることで、不良データを気にすることなく正確な DOS を推定できることが期待される。

References

- [1] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- [2] B. A. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249 (1991).
- [3] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [4] W. Janke, *Physica A* **254**, 164 (1998).
- [5] A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).
- [6] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- [7] A. Irbäck and F. Potthast, *J. Chem. Phys.* **103**, 10298 (1995).
- [8] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [9] C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, E. M. Keramidas (ed.), (Interface Foundation, Fairfax Station, 1991) pp. 156-163.
- [10] U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140, (1997).
- [11] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [12] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [13] F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- [14] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002).
- [15] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Comp. Aid. Mol. Des.* **8**, 695 (1994).
- [16] H. Grübmueller, *Phys. Rev. E* **52**, 2893 (1995).
- [17] Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
- [18] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **329**, 261 (2000).
- [19] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6664 (2003).
- [20] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6676 (2003).
- [21] K. Hukushima and Y. Iba, *AIP Conf. Proc.* **690**, 200 (2003).

- [22] J. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.* **97**, 050601 (2006).
- [23] H. Okumura and Y. Okamoto, *J. Comput. Chem.* **27**, 379 (2006).
- [24] R. E. Belardinelli, S. Manzi and V. D. Pereyra, *Phys. Rev. E* **75**, 067701 (2008).
- [25] A. Mitsutake and Y. Okamoto, *J. Chem. Phys.* **130**, 214105 (2009).
- [26] Y. Mori and Y. Okamoto, *J. Phys. Soc. Jpn.* **79**, 074003 (2010).
- [27] J. Kim, J. E. Straub, and T. Keyes, *J. Phys. Chem. B* **116**, 8646 (2012).
- [28] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, *Phys. Rev. Lett.* **110**, 210603 (2013).
- [29] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, *Phys. Rev. E* **90**, 023302 (2014).
- [30] C. Junghans, D. Perez, and T. Vogel, *J. Chem. Theory Comput.* **10**, 1843 (2014).
- [31] F. Yaşar, N. A. Bernhardt, and U. H. E. Hansmann, *J. Chem. Phys.* **143**, 224102 (2015).
- [32] Y. Okamoto, *Mol. Sim.* **38**, 1282 (2012).
- [33] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [34] P. D. Beale, *Phys. Rev. Lett.* **76**, 78 (1996).
- [35] <http://spot.colorado.edu/~beale/>
- [36] G. Marsaglia, A. Zaman, and W. W. Tsang, *Stat. Prob. Lett.* **8**, 35 (1990).
- [37] B. A. Berg, *Markov Chain Monte Carlo Simulation and Their Statistical Analysis* (World Scientific, Singapore, 2004).
- [38] A. E. Ferdinand and M. E. Fisher, *Phys. Rev.* **185**, 832 (1969).
- [39] T. Nagasima, A. R. Kinjo, T. Mitsui, and K. Nishikawa, *Phys. Rev. E* **75**, 066706 (2007).

第4章 The Residual Entropy of Ice I_h

T. Hayashi, C. Muguruma, and Y. Okamoto, The Journal of Chemical Physics 154, 044503 (2021).

4.1 導入

氷 I_h 系は、絶対零度で 0 ではない有限のエントロピー (残余エントロピー) が実験的に観測された [1] 系である。後にこの残余エントロピーの起源についての理論的説明は、氷中の水分子の配向の規則 (Ice rule[2, 3]) から生じる結果として与えられた。文献 [3] によれば、残余エントロピーは水分子の配向の自由度に由来しており、1 つの水分子あたりの残余エントロピー S_0 は、1 つの水分子の Ice rule を満足する配向の自由度の数 W_0 の自然対数に比例する。

$$S_0 = k_B \ln W_0. \quad (4.1)$$

ここで、 k_B はボルツマン定数である。Pauling によれば、1 水分子あたりの配向の自由度は $W_0^{\text{Pauling}} = 1.5$ であり、その残余エントロピーの値は $S^{\text{Pauling}} = R \ln W_0^{\text{Pauling}} \simeq 0.806$ であった [3]。ここでは、気体定数 R として $R = 8.314\,462\,618\dots$ [J/(mol K)] という値と、1[cal]=4.184[J] という関係を用いた。従って $R = 1.987\,42$ [cal/(mol K)] である。 S^{Pauling} は実験値 $S_0^{\text{Experiment}} = 0.82$ (5) と良く一致していた [1] (本章での誤差は括弧で示す) が、後に Onsager と Dupuis によって Pauling の推定値 S^{Pauling} は理論的な残余エントロピーの下限であることが示され [4]、より発展的な理論的近似値が Nagle によって与えられた [5]。

計算機シミュレーションに関しては、基底状態で Ice rule を満足する 2 つのシミュレーションモデル (2-state model と 6-state model) [6, 7, 8, 9] が提案され、そのモデルをもとにマルチカノニカルモンテカルロ法 [10, 11, 12, 13] によって、残余エントロピーが計算された。これらのモデルが提案された後、多くの研究グループが過去 10 年間で様々な計算アプローチによって残余エントロピーを推定した [14, 15, 16, 17, 18]。現在では、コンピュータシミュレーションによる測定値は、Nagle による理論的近似値と同等以上の精度を持つようになった。そのため氷の残余エントロピーはシミュレーションアルゴリズムの計算精度を確認するための良いベンチマークのひとつになりつつあるが、それぞれの手法による推定値には誤差よりも大きい不一致が残る。すなわち、氷 I_h の正確な残余エントロピーはいまだに得られていない。

本研究では、2-state model[6] と REWL-MUCAREM [19] によって、氷 I_h の残余エントロピーを計算した。第 3 章で示したように、REWL-MUCAREM は適切な計算条件下で、高精度な状態密度 (DOS) とエントロピーを与える。本研究による最新の結果は、他のシミュレーション手法を使用したいくつかの研究グループによる推定とよく一致した。さらに、残余エントロピーの高精度計算のためには、擬似乱数生成アルゴリズムの発生させる乱数の均一性が重要であったことも報告する。

4.2 手法

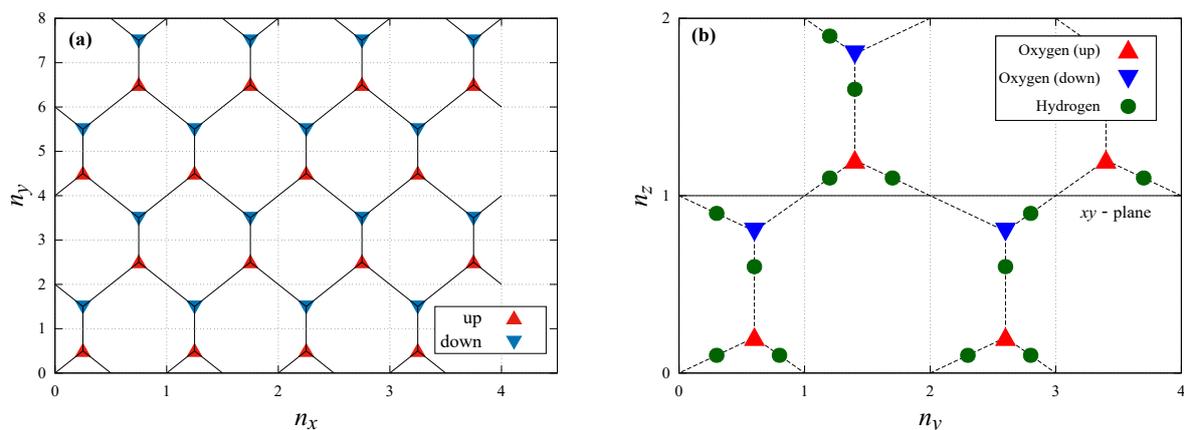


Fig. 4.1: 氷 I_h 系の 2 次元平面への射影。(a) は xy 平面への、(b) は yz 平面への射影を表す。簡単のため、縮尺は実際の氷 I_h 系とは異なる。 n_x, n_y, n_z は x, y, z 軸に沿った格子番号であり、全水分子数 N は $n_x \times n_y \times n_z$ で与えられる。(a) の赤い三角形の記号は格子点が xy 平面より上側に位置することを表し、青い三角形の記号は格子点が xy 平面より下側に位置することを表す。酸素原子は各格子点上に配置されており、(b) における三角形は酸素原子を表す。酸素原子間の点線は水素結合を表し、緑の円は結合上に配置された水素原子を表す。結合上の水素原子は 2 つの Ice rule に従って配置される。

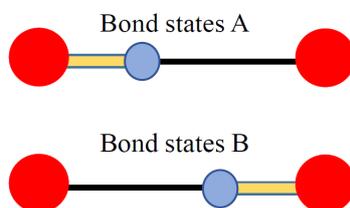


Fig. 4.2: 各結合で考えられる 2 つの状態。水素原子は、結合を形成する 2 つの酸素原子のどちらか一方の近傍に位置する。

4.2.1 Ice rule と残余エントロピー

図 4.1 は、氷 I_h の六方晶構造を 2 次元に射影したものである。図 4.1(a) と (b) は、それぞれ xy 平面と yz 平面への射影に対応する。水中においても水分子は H_2O 分子の形を保ったまま存在し、水素原子は各結合上の 2 つの場所 (ボンドを形成する 2 つの酸素原子のうち、どちらか一方の近傍の位置) の 1 つを占めることができると仮定する (図 4.2)。このとき、Ice rule は次の (1)(2) として表現される [2, 3]。

1. 各結合上に 1 つの水素原子のみ存在する (水素結合の条件)。
2. 各酸素原子の近くに 2 つのみ水素原子が存在し、イオン状態は存在しない (水分子の条件)。

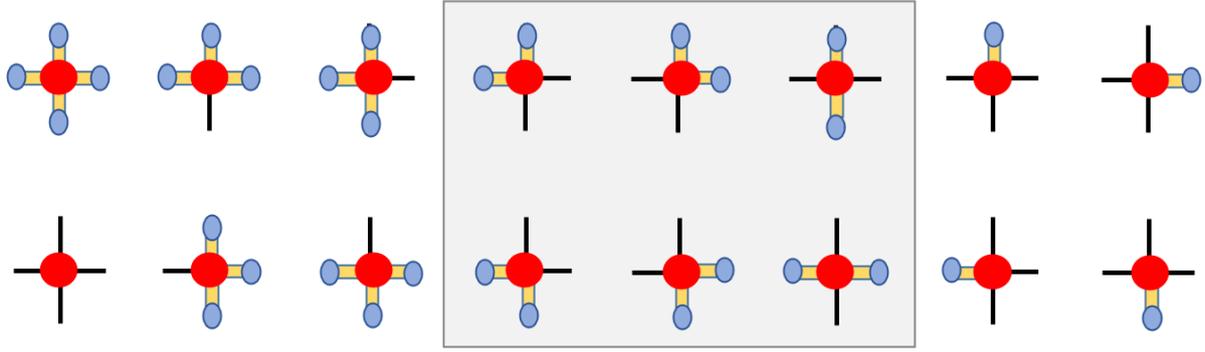


Fig. 4.3: Ice rule 1 を無視した場合に考えられる酸素原子周辺の水素原子配置の 16 通りの組み合わせ。このうち、ハイライトされた 6 通りのみが ice rule 1 を満足する。

氷中の水分子は上記の Ice rule を満足するように配向し、Ice rule を満足する配向の組み合わせの数の自由度が、残余エントロピーを生じさせる。

N 個の水分子があると仮定すると、水素原子の数は $2N$ 個となる。このとき、水分子 1 つあたりの理論上の残余エントロピー S_0 は次のように定義される。

$$S_0 = \frac{k_B \ln W}{N} = k_B \ln W_0, \quad (4.2)$$

ここで

$$W = (W_0)^N. \quad (4.3)$$

という関係を用いた。 W は 2 つの Ice rule を満たす全水分子の配位の総数である。 W_0 を 1 つの水分子あたりの配位数として定義することで、Pauling はその値を

$$W_0^{\text{Pauling}} = 1.5. \quad (4.4)$$

と推定した [3]。Pauling の考えは次のようなものである。2 番目の Ice rule (水分子条件) を無視すると、各水素原子には、結合上に 2 つの位置の選択肢が与えられる (図 4.2)。このとき考えられる全水素原子の配位数は、 2^{2N} となる。次に、ある 1 つの酸素原子を考える。1 つの酸素原子の周りの水素結合上には、4 つの水素原子による 16 通りの配置があり、そのうちの 6 つの配置だけが 2 番目の Ice rule を満たすことができる (図 4.3)。従って、Ice rule(1) と Ice rule(2) を同時に満たす配位の総数 W は次のように見積もることができる。

$$W = (W_0^{\text{Pauling}})^N = 2^{2N} \times \left(\frac{6}{16}\right)^N = \left(\frac{3}{2}\right)^N = (1.5)^N. \quad (4.5)$$

従って Pauling の推定した残余エントロピーとして、

$$\begin{aligned} S_0^{\text{Pauling}} &= k_B \ln (W_0^{\text{Pauling}}) \\ &= 0.80574 \cdots [\text{cal}/(\text{mol K})]. \end{aligned} \quad (4.6)$$

を得る。

Onsager と Dupuis は、Pauling の議論が水素原子の配置が形成する長距離ネットワーク (closed loop) に由来する自由度への影響を省略しており、 $W_0^{\text{Pauling}} = 1.5$ が実際には下限値

であったことを示した [4]。そして Nagle は Series Expansion Method によって closed loop の影響を考慮し、より高精度な推定値を得た [5]。短い closed loop からの寄与は、loop の図形を直接カウントすることで取り入れ、長い closed loop の影響は、短いループの結果に基づく外挿によって近似した。その近似値は、

$$W_0^{\text{Nagle}} = 1.50685(15), \quad (4.7)$$

であり、

$$S_0^{\text{Nagle}} = 0.81480(20) \cdots [\text{cal}/(\text{mol K})]. \quad (4.8)$$

であった。ここでの誤差は統計的なものではなく、Series Expansion Method における展開の高次項を無視したことに起因する。また Series Expansion Method と異なる理論的近似計算法として、Numerical Linked Cluster (NLC) 展開を使用する級数展開法が提案されている [20]。そして現在では、計算科学の発展に伴い、多くの研究グループが様々な計算アプローチ（熱力学的積分法、Wang-Landau アルゴリズム、PEPS アルゴリズムなど）によって残余エントロピーを推定する試みが行われている [14, 16, 17, 15, 18]。

4.2.2 2-State Model

本研究では 2-state model[6] を使用した。このモデルでは、水分子の明確な配向を考慮しない (Ice rule(2) を無視する) が、2つの酸素原子間の結合上の水素原子に、2つの位置のうち1つの位置をとることを許可する (従って Ice rule(1) は常に満たされる)。この系の全ポテンシャルエネルギー E を、次の式で与える。

$$E = - \sum_i f(i, b_i^1, b_i^2, b_i^3, b_i^4), \quad (4.9)$$

ここで、 i は酸素原子の位置番号を表し、和はすべての i についてとる。 $b_i^1, b_i^2, b_i^3, b_i^4$ は酸素原子 i 周りの水素結合の状態を表す。水素原子が酸素原子 i の近くに配置されれば 1, そうでなければ 0 をとる。そして関数 f は次のように定義される。

$$f(i, b_i^1, b_i^2, b_i^3, b_i^4) = \begin{cases} 2 & \text{for two hydrogen nuclei close to } i, \\ 1 & \text{for one or three hydrogen nuclei close to } i, \\ 0 & \text{for zero or four hydrogen nuclei close to } i. \end{cases} \quad (4.10)$$

このとき、基底エネルギー状態 E_{ground} でのエネルギーは $-2N$ となる。

このようなポテンシャルエネルギーで定義された 2-state model の基底状態は Ice rule(1)(2) を完全に満足する。 $n(E)$ をあるエネルギー E での状態の数として、基底状態の状態数 $n(E_{\text{ground}})$ は、水分子数 N 個の氷 I_h 系における水分子の配向の自由度 W に対応する。従って、2 state model における $n(E_{\text{ground}})$ を計算することができれば、式 (4.2) より、残余エントロピーを計算できる。そして $n(E)$ は MUCA 法における DOS として推定できる。

第2章で紹介した MUCA 法による DOS の推定値 (本章では $\tilde{n}(E)$) は、真の DOS ($\tilde{n}(E)$) との間に定数倍の不確実性が存在する。この不確実性は、2-state model の全状態数 (配位

の総数 $\sum_E n(E) = 2^{2N}$ を用いて、 $\tilde{n}(E)$ と $\sum_E \tilde{n}(E)$ の比率を計算することにより、取り除くことができる [21]。従って基底状態の状態数は、

$$W = (W_0)^N = n(E_{\text{ground}}) = 2^{2N} \times \frac{\tilde{n}(E_{\text{ground}})}{\sum_E \tilde{n}(E)}. \quad (4.11)$$

として計算できる。ここで、 $\tilde{n}(E)$ は、MUCA シミュレーションから得られた規格化されていない状態数を表す。

4.3 計算条件

水分子の総数 N は、 $n_x \times n_y \times n_z$ で与えられる。ここで、 n_x, n_y, n_z は、 x, y, z 軸に沿った格子点の数である (図 4.1 参照)。格子点の総数 (すなわち酸素原子の総数) は N であり、水素原子の総数は $2N$ となる。周期境界条件を使用したため、 n_x, n_y, n_z は、 $n_x = 1, 2, 3, \dots : n_y = 4, 8, 12, \dots : n_z = 2, 4, 6, \dots$ に制限される。シミュレーションに使用した水分子の総数は $N = 128, 288, 360, 576, 896, 1600, 2880, 4704$ とし、各 N での W_0 を求め、その結果に外挿を行うことで、 $N \rightarrow \infty$ における W_0 および残余エントロピーを計算した。主な計算条件を Table 4.1 に示す。本 MC シミュレーションでは、1 MC STEP ごとに水素原子の位置が更新される。各 MC STEP 後に水分子の配位に対応したエネルギーの値を保存し、ヒストグラムを作成した。1 回の MC sweep は、メトロポリス判定 $2N$ 回 (全水素原子の数) として定義した。高精度で $n(E)$ を推定するために、REWL-MUCAREM プロトコルを使用した。

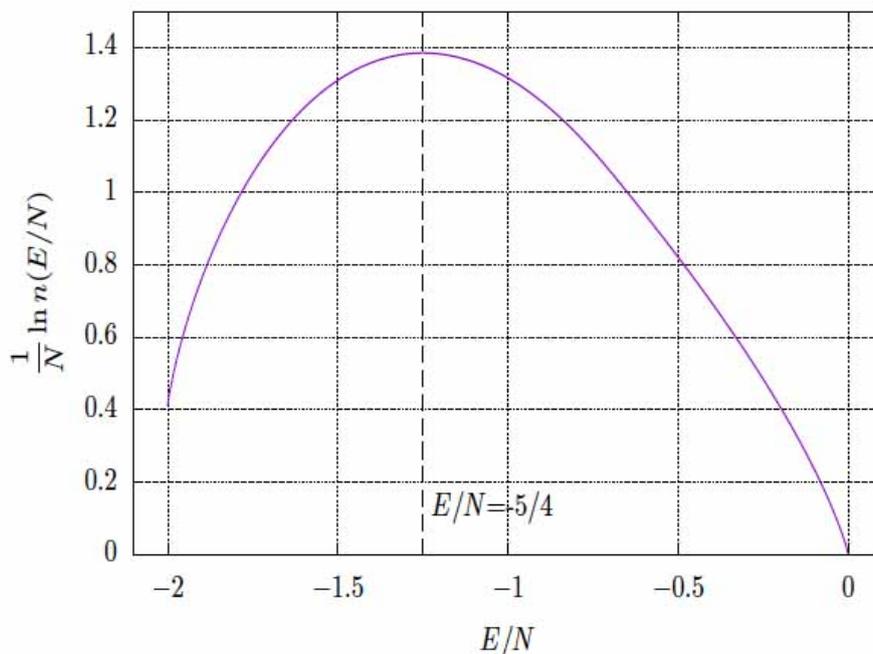


Fig. 4.4: 氷 I_h の 2-state model における、1 水分子あたりの無次元のエントロピー ($\ln n(E/N)$)。水分子数 $N = 2880$ での REWL-MUCAREM シミュレーションによって得られた結果を表す。 $E/N = 0$ での考えられる $n(0)$ が 2 であるため、 $\ln n(0)$ は $\ln(2)$ とした。 $\ln n(E/N)$ は $E/N = -5/4$ で最大値を取る。

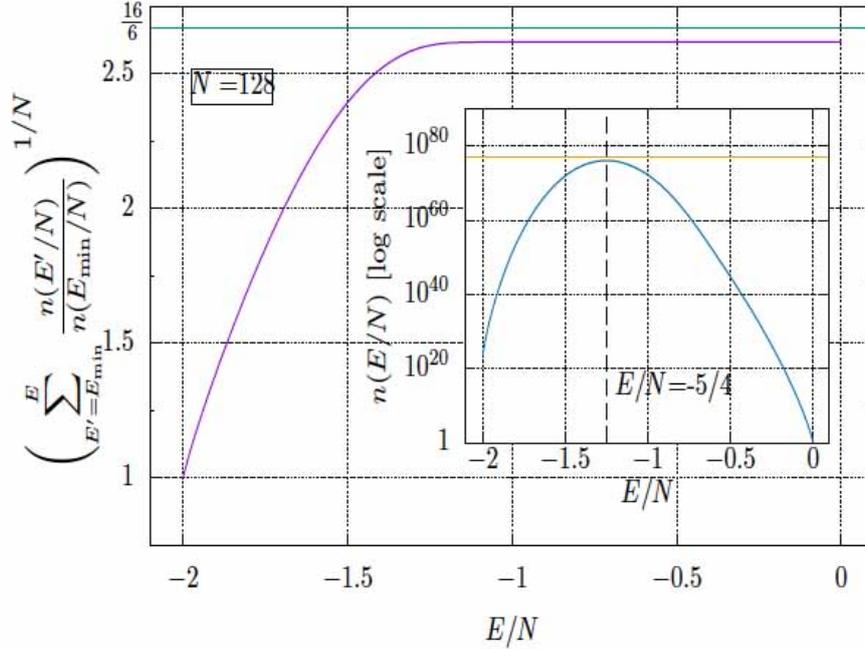


Fig. 4.5: 水分子数 $N = 128$ における、 E_{\min} から E までの $n(E/N)$ の総和を表す。この値は式 (4.5) における $6/16$ の逆数に対応するように規格化されている。緑の直線は Pauling によって得られた推定値 ($16/6$) を表す。 $n(E/N)$ の総和は $E/N = -5/4$ あたりで飽和する。挿入図は、 $n(0)$ が 2 で規格化された $n(E/N)$ の値を表す。挿入図におけるオレンジ色の直線は、全水分子数 128 個の 2-state model における全状態数 ($\sum_E n(E/N) = 2^{2N}$) を表す。 $n(E/N)$ は $E/N = -5/4$ あたりで最大値を取り、これは全状態数の大部分が $E/N = -5/4$ に近いエネルギー値を与えるとわかる。

MUCA および WL シミュレーションでは、シミュレーションを開始する前に、DOS を求める全エネルギー領域を $[E_{\min}, E_{\max}]$ をあらかじめ決めておく必要がある。本研究では、 $[E_{\min}, E_{\max}] = [-2N, -5N/4]$ とした。ここで、 E_{\min} は基底状態のエネルギーに対応し、 E_{\max} はエントロピーが最大値をとるエネルギーの値に対応する (図 4.4 を参照)。4.4 は、氷 I_h の 1 水分子あたりの典型的な無次元のエントロピー ($\ln n(E)$) を表す。これは、 $[E_{\min}, E_{\max}] = [-2N, 0]$ とした事前計算によって得られた図である。

無次元エントロピーは、 $E/N = -5/4$ で最大値を取る。従って、逆温度 β は E_{\max} で値 0 ($T \sim \infty$) となる。本研究での使用したエネルギー領域 $[E_{\min}, E_{\max}] = [-2N, -5N/4]$ の条件下では、MUCA での平坦な確率分布は $[-2N, -5N/4]$ で実現され、 $T = \infty$ のカノニカル分布が $[E > -5N/4]$ で実現する。規格化 (式 4.11) での配位の総数を推定するために、 $n(E)$ は、シミュレーション中に得られた最大エネルギー値まで和を取った (式 (4.11))。本研究での規格化によって W_0 を高精度で推定する場合は、 E_{\max} を 0 とすることが望ましいが、全状態数 ($\sum_E n(E) = 2^{2N}$) のほとんどは $E = -5N/4$ とその周辺のエネルギー値に分布している (図 4.5 の内挿図) ため、 E_{\max} は $-5N/4$ で十分であるとした。すなわち、 $E = -5N/4$ よりも十分高いポテンシャルエネルギーの状態数は無視できる (図 4.5)。図 4.5 は、 $N = 128$ の系の、 E_{\min} で正規化された 1 水分子あたりの $n(E)$ の、 E_{\min} からあるエネルギー E まで和を示す。この和は、 $E/N = -5/4$ よりも少し大きいエネルギーで飽和する。この漸近値と、Pauling の推定値 $6/16$ の逆数である値 $16/6$ の差 (式 (5) と (11) を比較) は、参考文献 [4] で指摘されている closed loop の影響によるものである。図 4.5 の挿入図は、 $n(E)$ を直接表したものである。全状態数のほとんどは、 $E/N = -5/4$ の周

りに分布していることがわかる。

これらの結果は、 $E/N = -5/4$ よりも十分に高いエネルギーの状態数の和は、今回の測定方法における残余エントロピーの推定に影響を与えないほど十分に小さいことを意味する。実際、 $N = 2880$ まで、 $[-2N, -5N/4]$ の条件で得た W_0 の推定値を、 $[-2N, 0]$ の条件で得た W_0 の推定値の差を比較したが、その差は誤差の範囲内であり、十分に小さかった。以上より、本シミュレーションによって DOS を求めるエネルギー領域は $[E_{\min}, E_{\max}] = [-2N, -5N/4]$ とした。またこの条件を採用した結果、残余エントロピーの推定のための MUCA シミュレーションで、最も重要な状態である基底エネルギーの状態をより多くのサンプルできた。

REWL-MUCAREM によって DOS を推定した後、残余エントロピーの平均値と誤差を評価するため、得られた DOS を MUCA 重み因子 ($1/n(E)$) として使用し、MUCA プロダクションランを、32 回独立して実行した ($M = 32$)。平均値と誤差は、次の式によって見積もった。

$$\bar{n}(E_{\min}) = \frac{\sum_{i=1}^M n(E_{\min})^{\{i\}}}{M},$$

$$\varepsilon_n = \sqrt{\frac{\sum_{i=1}^M (n(E_{\min})^{\{i\}} - \bar{n}(E_{\min}))^2}{M(M-1)}}. \quad (4.12)$$

ここで、 $n(E_{\min})$ は、 i 番目の MUCA プロダクションラン ($i = 1, 2, \dots, M$) から得た測定値である。測定用の MC sweep の総数は、MUCA プロダクションランごとに 6.4×10^8 MC sweep であった。単ヒストグラム再重法 (式 (2.51)) を W_0 の最終的な推定値を得るために使用した。

擬似乱数発生器は MC 法のメトロポリス判定に大きな影響を与える。本研究では、メルセンヌツイスター乱数発生器を使用した [22]。これにはオープンソースのプログラムコードを使用した [23]。

4.4 結果と考察

4.4.1 REWL-MUCAREM の動作確認

はじめにレプリカ交換法が正常に動作していることを、レプリカのエネルギー遷移の時系列データによって確認する。図 4.6 は、全水分子数 4704 の MUCAREM シミュレーションにおける、レプリカ (レプリカ 1) のエネルギー領域遷移の時系列データを示す。この系では 32 個のレプリカを使用したため、全エネルギー領域 $[E_{\min}, E_{\max}]$ を、32 個の sub energy region ($m = 1, 2, \dots, 32$) に分割した。 E_{\min} は -9408 で、 E_{\max} は -5880 であった。図 4.6 から、レプリカ 1 は $m = 1$ から $m = 32$ に遷移し、再び $m = 1$ へ遷移したことがわかる。これは、レプリカ交換のアルゴリズムが正しく機能し、ひとつのレプリカがすべてのエネルギー領域に渡ってサンプリングできたことを示す。

図 4.7 は、図 4.6 と同じシミュレーションで、1 つのレプリカ (レプリカ 1) のポテンシャルエネルギーの時系列データを示す。レプリカ 1 は $[E_{\min}, E_{\max}] = [-9408, -5880]$ までの

Table 4.1: Initial conditions in REWL-MUCAREM simulations.

N	n_x	n_y	n_z	No. of replicas	Replica exchange ^a	WL criteria ^b	Total MC sweeps for REWL ^c	Total MC sweeps for MUCAREM ^d
128	4	8	4	4	250	500	8.150×10^4	$2.0 \times 10^7 \times 5$
288	4	12	6	8	250	500	2.795×10^5	$2.0 \times 10^7 \times 5$
360	5	12	6	8	250	500	2.850×10^5	$2.0 \times 10^7 \times 5$
576	6	12	8	16	500	1000	4.410×10^5	$2.0 \times 10^7 \times 5$
896	7	16	8	16	500	1000	1.334×10^6	$2.0 \times 10^7 \times 5$
1600	8	20	10	32	2500	5000	2.530×10^6	$2.0 \times 10^7 \times 5$
2880	10	24	12	32	2500	5000	5.110×10^6	$2.0 \times 10^7 \times 5$
4704	12	28	14	32	5000	10000	1.637×10^7	$2.0 \times 10^7 \times 5$

^a The interval of replica exchange trial (MC sweeps) in REWL and MUCAREM.

^b The interval of WL criteria check (MC sweeps) in REWL.

^c Total MC sweeps per each replica that is required for all WL weight factors f to converge to f_{final} in REWL.

^d Total MC sweeps per each replica in MUCAREM. MUCAREM simulations were iterated 5 times.

エネルギー空間をランダムウォークした。図 4.6 のエネルギー遷移と図 4.6 のポテンシャルエネルギーの間には強い正の相関関係があるとわかる。図 4.8 の 4 つの図は、 $N = 4704$ の系の MUCAREM シミュレーションによって得られたポテンシャルエネルギーのヒストグラムを示す。ここでは、 $m = 1, 2, 3, 4$ に対応するエネルギー領域のみを図 4.8 に示した。各ヒストグラムは平坦な分布を実現した。

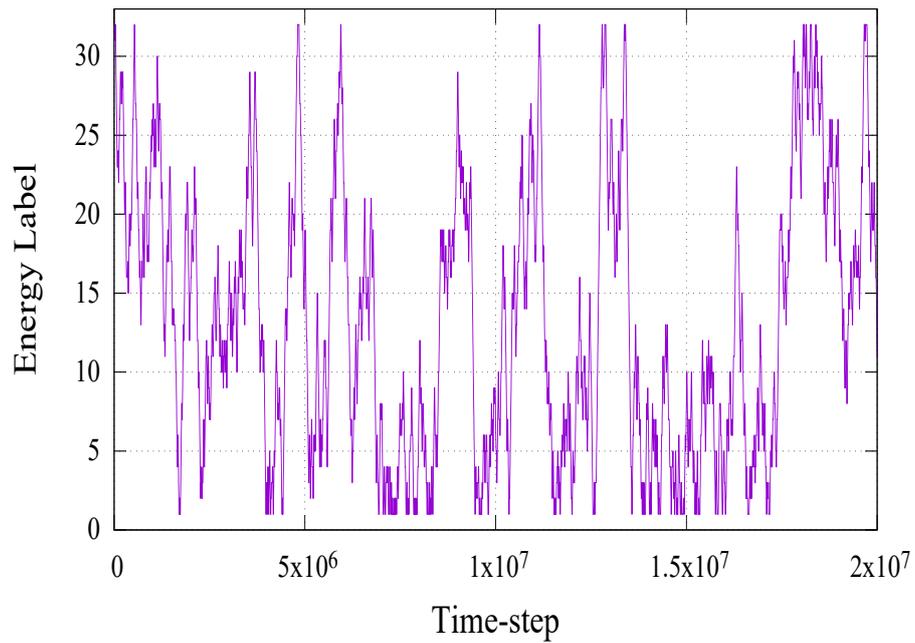


Fig. 4.6: 全水分子数 4704 個における MUCAREM シミュレーション中の一つのレプリカ (レプリカ 1) のエネルギー領域 m (energy label) の時間発展。

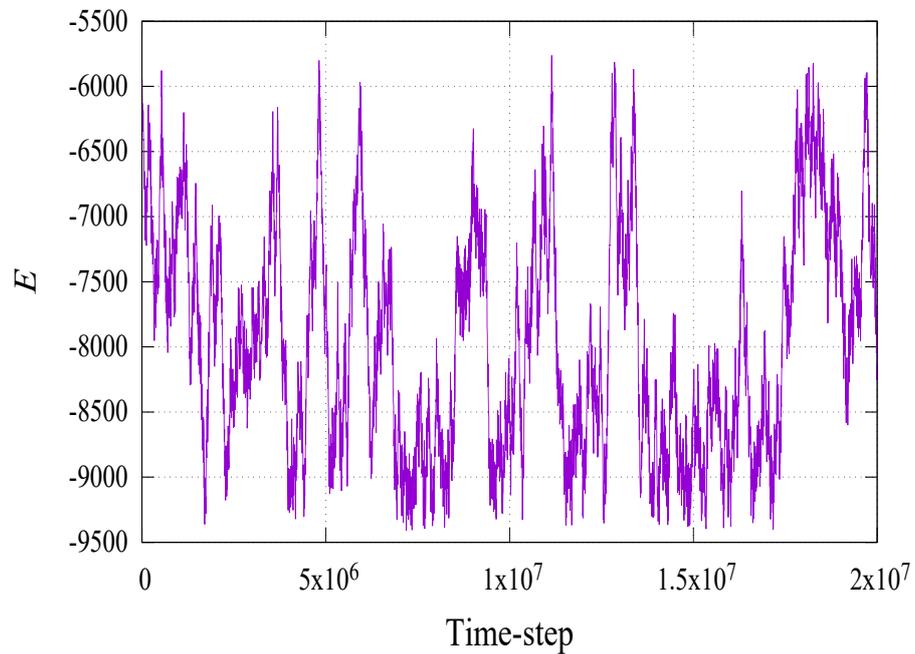


Fig. 4.7: 全水分子数 4704 個における MUCAREM シミュレーション中の一つのレプリカ (レプリカ 1) のエネルギー E の時間発展。

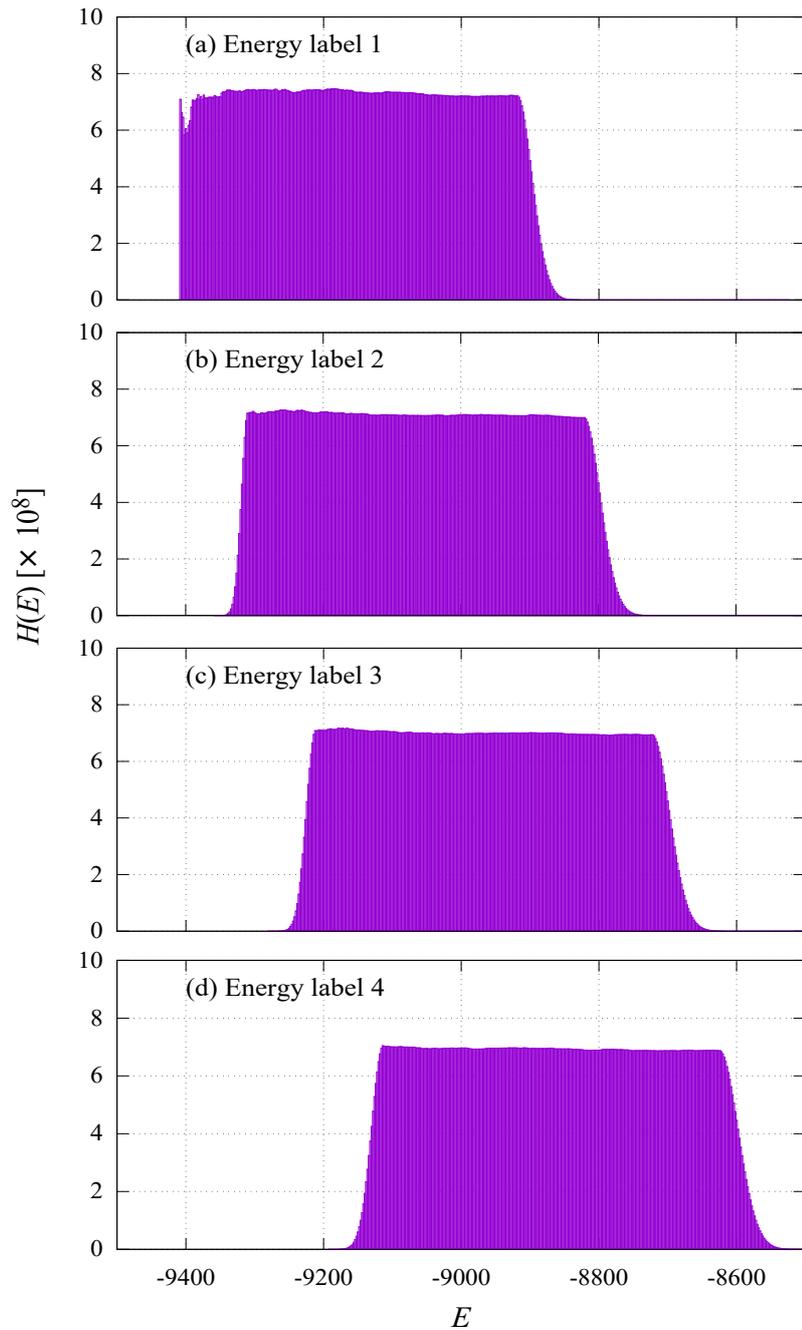


Fig. 4.8: 全水分子数 4704 個の系において、MUCAREM シミュレーションから得られたエネルギーのヒストグラム。全エネルギー領域は 32 個に分割 ($m = 1, 2, \dots, 32$) されており、 $m = 1, 2, 3, 4$ に対応したヒストグラムを示す。隣り合うエネルギー領域は約 80% の重なりを持つ。各ヒストグラムは十分に平坦であった。

Table 4.2: Estimated residual entropy of Ice I_h.

N	n_x	n_y	n_z	Tunneling ^a	Flatness ^b	W_0^*	S_0^*
128	4	8	4	7637621	0.99068	1.5286237(338)	0.8433037(404)
288	4	12	6	1603919	0.97233	1.5175992(309)	0.8289200(405)
360	5	12	6	1019315	0.94745	1.5155995(314)	0.8262998(412)
576	6	12	8	412944	0.95416	1.5127845(264)	0.8226054(347)
896	7	16	8	180083	0.94447	1.5109637(237)	0.8202121(312)
1600	8	20	10	57543	0.93800	1.5095176(236)	0.8183093(311)
2880	10	24	12	18577	0.91332	1.5086491(313)	0.8171657(413)
4704	12	28	14	6591	0.83009	1.5082029(360)	0.8165779(475)
∞	fitting					1.5074723(474) [†]	0.8156150(625) [†]

^a The total counts of observed tunneling events during 32 MUCA production runs.

^b The value of flatness (H_{\max}/H_{\min}) calculated by accumulating the histogram after 32 MUCA production runs.

* The values in parentheses represent the errors obtained by 32 MUCA production runs and fitting, using Eq. (4.12). The error bars for S_0 , ΔS_0 were calculated by $\Delta S_0 = k_B \Delta W_0 / W_0$.

[†] The data point for the smallest lattice size ($N = 128$) was not included in the fitting in Eq. (4.13). The errors of each estimate of W_0 were considered in the fitting.

図 4.9 は、 $N = 4704$ の REWL-MUCAREM によって得られた $n(E)$ を自然対数でプロットしたもの (従って無次元のエントロピー) である。また、図 4.10 は、図 4.9 の DOS を MUCA 重み因子とした MUCA シミュレーションによって得られたエネルギーヒストグラムを示す。理想的な MUCA 重み因子は、完全に平坦なヒストグラムを実現する。MUCA プロダクションランによって得られたヒストグラムの平坦度 (H_{\min}/H_{\max}) を Table 4.2 に示した。平坦度の値は、すべての系で 0.8 より大きい値を得た。この平坦度の定義において、通常 DOS が十分な精度で得られていると判断する平坦度の値が 0.5 とされること考慮すると、本研究における REWL-MUCAREM によって得られた DOS の値が十分に正確であったことを意味する。

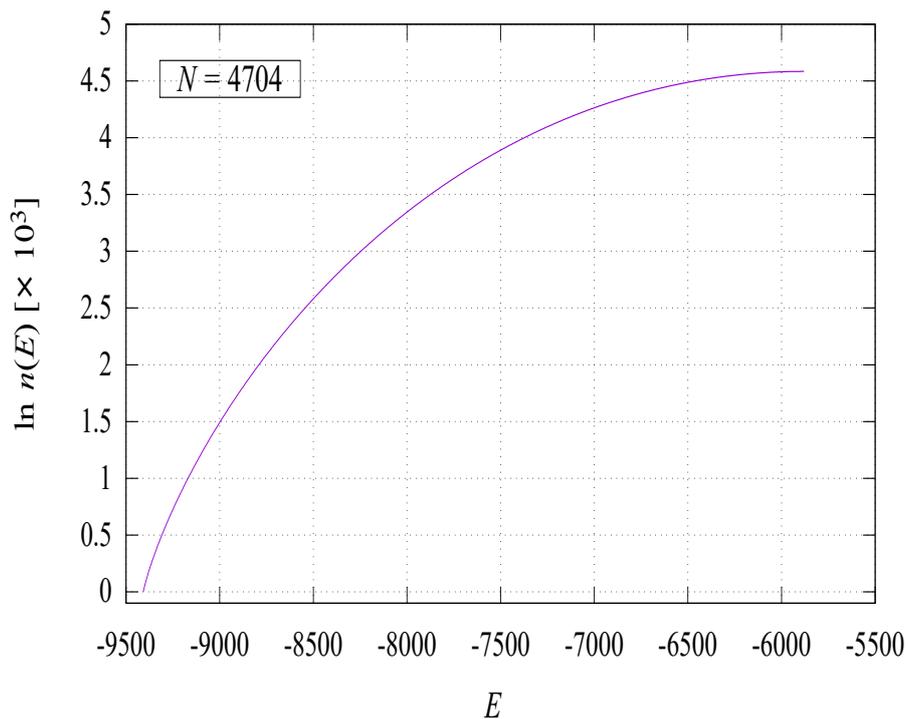


Fig. 4.9: 全水分子数 4704 個の REWL-MUCAREM シミュレーションによって得られた無次元のエントロピー $\ln n(E)$ の値。 $E = -9408$ において $\ln n(E)$ の値を 0 とした。

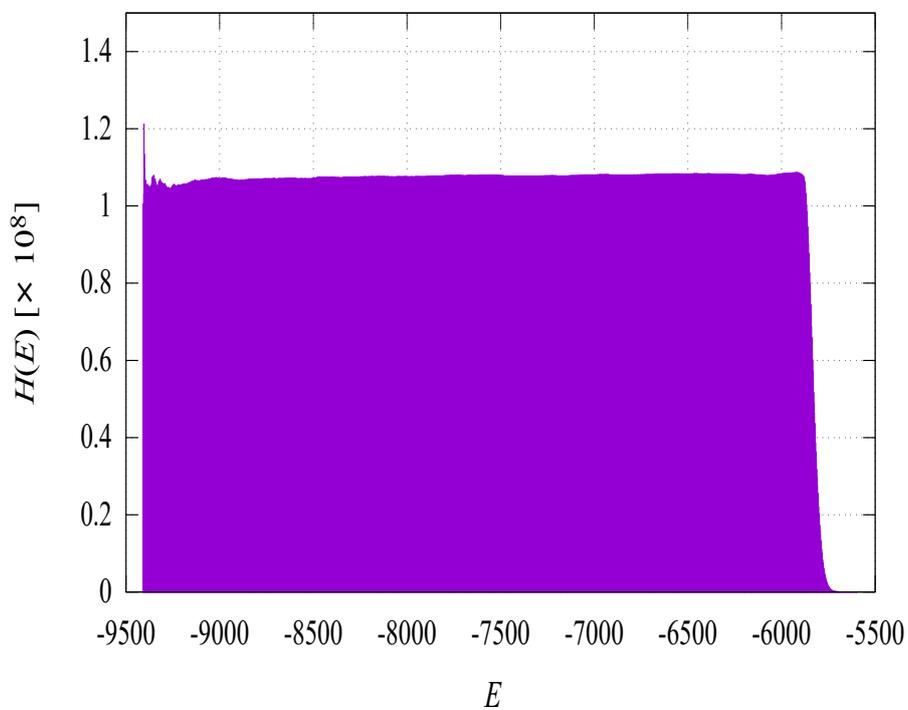


Fig. 4.10: 全水分子数 4704 における MUC A シミュレーションによって得られたエネルギーのヒストグラム。 Fig. 4.9 の $n(E)$ の値が MUCA 重み因子として使用された。

4.4.2 残余エントロピー

最終的な基底状態の状態数 W_0 の推定値も Table 4.2 に示した。図 4.11 には、系の大きさに対応した W_0 推定値と、その外挿をプロットした。外挿式は次の形式を使用した。

$$W_0\left(\frac{1}{N}\right) = W_0(0) + a\left(\frac{1}{N}\right)^\theta. \quad (4.13)$$

本研究での、 W_0^{ThisWork} の最終的な推定値 (図 4.11 の $W_0(0)$) は、Table 4.2 の最後から 2 番目の列に示されている。フィッティングには MATHEMATICA を用い、得られたフィッティングパラメータの値は、 $a = 1.992074 \pm 0.119366$ および $\theta = 0.933296 \pm 0.011084$ であった。本研究グループの先行研究 [6, 9] では、最も小さい格子サイズ $N = 128$ のデータ点もフィッティングに含めたが、 $N = 128$ は $N = 288$ の半分以下の水分子数であり、有限系としての影響が最も大きくなるため、本研究のフィッティングには含めなかった。

本研究における水分子数 N を ∞ としたときの状態数の見積もり (すなわち最終的な見積) は、

$$W_0^{\text{ThisWork}}(0) = 1.507472 \pm 0.000047. \quad (4.14)$$

この値による残余エントロピーは

$$S_0 = 0.815615 \pm 0.000063 \text{ [cal/(mol K)]}. \quad (4.15)$$

であった。

本研究におけるフィッティングでは、フィッティングに使用するデータ点の選び方によって系統誤差が存在した。例えば、すべての点を使用したフィッティングの場合、 $W_0(0) = 1.507401 \pm 0.000031$ であり、 $N = 128, 288$ を除いた場合は $W_0(0) = 1.507409 \pm 0.000057$ 、 $N = 128, 288, 360$ を除いた場合は $W_0(0) = 1.507480 \pm 0.000078$ であった。いずれの場合においても、それぞれの測定誤差は式 (4.14) の測定値と重なりがあった。

Table 4.3: Comparisons of the estimates by various methods.

Group	Methods	W_0	ΔW_0	S_0	ΔS_0
Nagle [5]	Series expansion	1.50685	0.00015	0.8147962	0.000198
Berg (2007) [6]	Multicanonical algorithm	1.50738	0.00016	0.81550	0.00021
Berg (2012) [9]	Multicanonical algorithm	1.507117	0.000035	0.815149	0.000046
Herrero [14]	Thermodynamic Integration	1.50786	0.00012	0.81613	0.00016
Kolafa [15]	Thermodynamic Integration	1.5074674	0.0000038	0.8156103	0.0000051
Ferreyra [17]	Wang-Landau algorithm	1.5070	0.0009	0.81478	0.00012
Vanderstraeten [18]	PEPS algorithm	1.507456		0.8155953	
This work	REWL-MUCAREM	1.5074723	0.000047	0.815615	0.000063

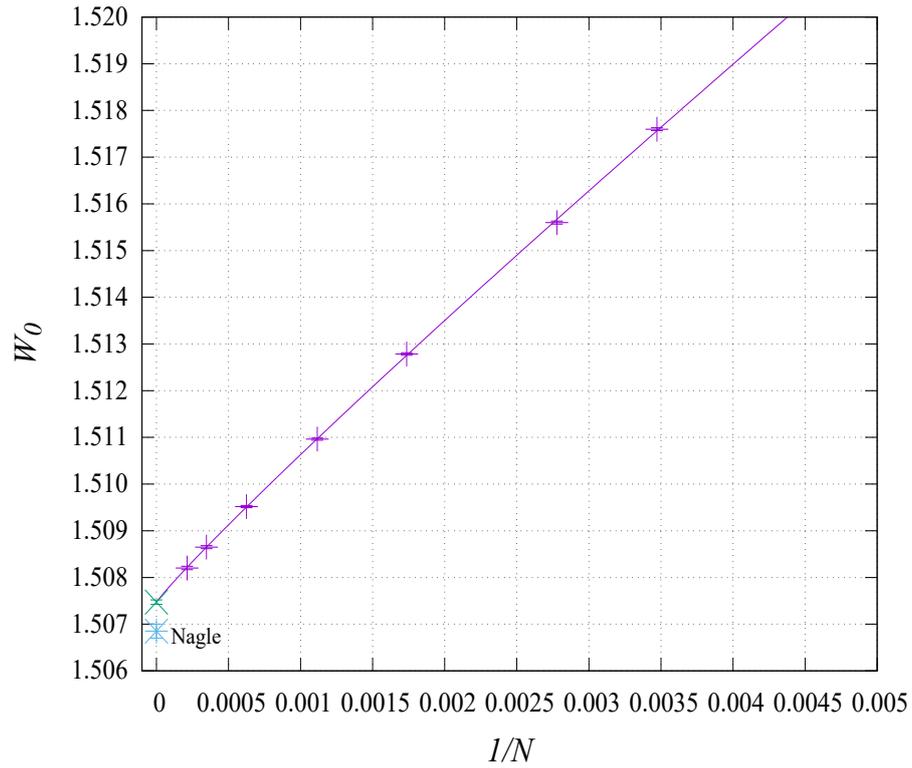


Fig. 4.11: 基底エネルギー状態における一水分子当たりの自由度 $W_0(1/N)$ 。誤差はデータ点の記号よりも十分に小さかった。

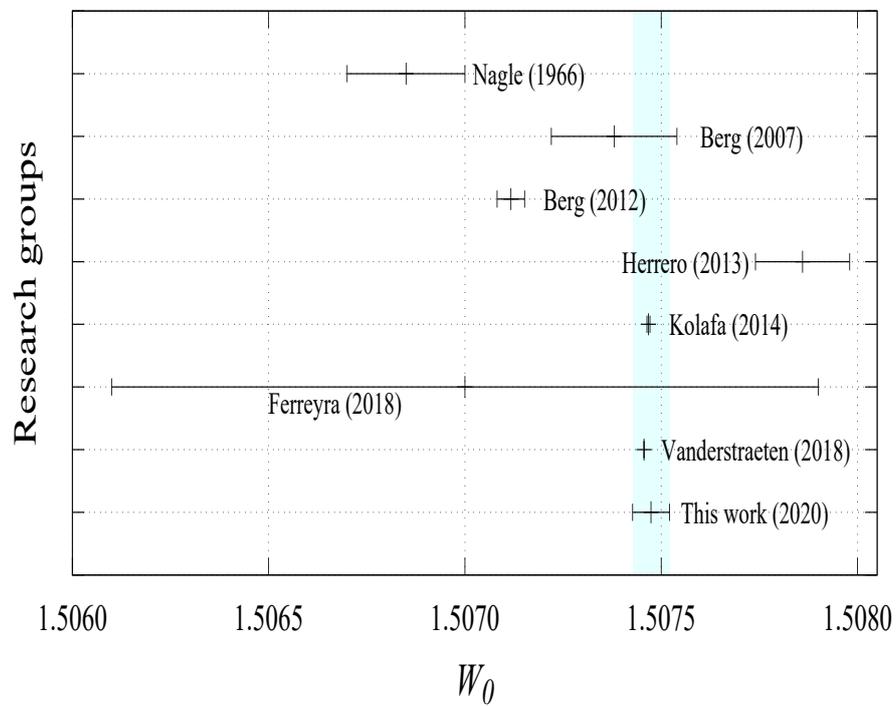


Fig. 4.12: 各研究グループによる W_0 の推定値。

本研究における W_0 の推定値を他の研究グループの結果と比較した。図 4.12 に、各グループの W_0 の推定値とその誤差をプロットした。各グループの計算方法とその結果は Table 4.3 にまとめた。Berg (2007) [6]、Kolafa (2014) [15]、Ferreyra (2018) [17]、Vanderstraeten (2018) [18] の結果と、本研究での結果は誤差の範囲内で互いに一致していた。特に、3 つの異なる計算アプローチ (PEPS アルゴリズム [18]、熱力学的積分 [15]、および REWL-MUCAREM) で、ほぼ同じ推定値を与えた。

本研究グループの MUCA による 2012 年の結果 [9] は、今回の結果と誤差の範囲で一致しなかった。これは乱数発生器の精度によるものであり、REWL-MUCAREM による結果は前回の見積もりよりも信頼性が高いと考えた。2012 年の最大の系 ($N = 2880$) では、使用した乱数発生器のため、MUCA 重み因子に基づくメトロポリス判定が正常に機能していない可能性があった。次節では、MUCA MC シミュレーションでの乱数の均一性の影響を評価する。

4.4.3 擬似乱数発生器の評価

擬似乱数発生器の性能は MC シミュレーションの結果に大きく影響する。しかし擬似乱数発生器には乱数の周期性など独自の特性がある。この節では、乱数発生器によって生成できる乱数の最小値の、MUCA MC シミュレーションへの影響について議論する。

Marsaglia 擬似乱数発生器 [24] と MersenneTwister 擬似乱数発生器 [22] という 2 つの擬似乱数発生器を比較した。Marsaglia 発生器は、本研究グループの先行研究で使用しており [6, 8, 9]、本研究では、MersenneTwister 発生器を使用した。これらの乱数発生器のソースコードは参考文献 [13, 23] のものを使用した。

乱数の精度を比較するために、これらの乱数発生器によって擬似乱数を 10^{11} 回生成した。Marsaglia 発生器 (緑色の点) と MersenneTwister 発生器 (紫色の点) によって生成された 5.0×10^{-7} 未満の値を図 4.13 にプロットした。MersenneTwister 発生器による乱数は均一な分布であるが、Marsaglia 発生器では離散的な分布が見て取れる。Marsaglia 発生器による最小乱数値は 0 であり、その次の最小値は 5.9605108×10^{-8} であった。このとき乱数の種は Seed1 = 11 および Seed2 = 20 であった。これは、使用した Marsaglia 発生器が $(0, 0.59605108 \times 10^{-8})$ までの値を、乱数として生成できないことを意味する。一方、MersenneTwister 発生器による最小乱数値 (乱数の種は 5489) は 4.9759×10^{-12} であり、Marsaglia による値 5.9605×10^{-8} よりもずっと小さい値であった。本研究で使用した MersenneTwister 発生器では 0 も乱数として生成されるが、本テスト観測されなかった。

次に 2-state model における、基底状態 X_0 から第一励起状態 X_1 への遷移確率 $w(X_0 \rightarrow X_1) = \exp[-\Delta S]$ ($\Delta S = \ln n_1 - \ln n_0$) を図 4.14 に示す。図 4.14 の挿入図は、基底状態 (エントロピー $\ln n_0$) と第一励起状態 ($\ln n_1$) の間のエントロピー差 ΔS の推定値を示す。分子数 N が増加するとエントロピー差が大きくなる。これに伴い、基底状態周辺のメトロポリス判定の受託率は小さくなる。 $n = 4704$ の場合、 $w(X_0 \rightarrow X_1)$ は $\exp[-16] \simeq 1.125 \times 10^{-7}$ であった。従って、Marsaglia 発生器では、乱数の均一性が低い (すなわち、 $(0, 0.59605108 \times 10^{-8})$ の乱数がアルゴリズム上生成できない) ため、メトロポリス判定が正常に機能しない。従って参考文献 [9] では、 $N = 2880$ の適切な推定値が得られなかった可能性があった。これが、本研究での残余エントロピーの推定値 ($S_0 = 0.815615 \pm 0.000063[\text{cal}/(\text{mol K})]$) が以前の結論 ($S_0 = 0.815148 \pm 0.000047[\text{cal}/(\text{mol K})]$) と異なる値を与えた理由である。

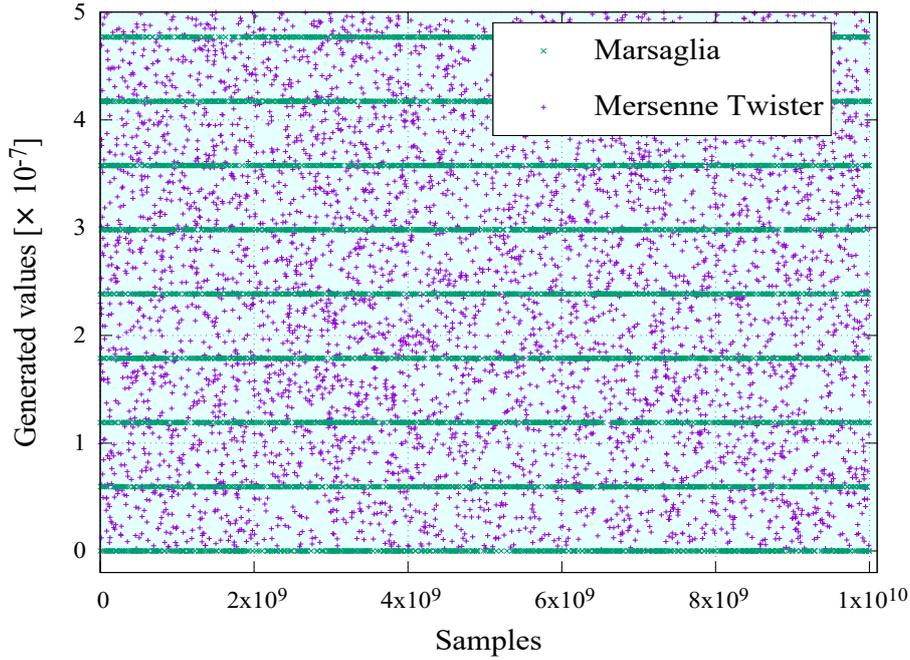


Fig. 4.13: マルサグリア乱数発生器によって得られた乱数の値 (緑) とメルセンヌツイスター乱数発生器によって得られた乱数の値 (紫)。メルセンヌツイスター乱数発生器による値は均一に分布しているが、マルサグリア乱数発生器による値は $(0, 0.59605, 1.19210, 1.78815, 2.38420, 2.98025, 3.57630, 4.17235, 4.76840 [\times 10^{-7}])$ の離散的な値をとっていた。

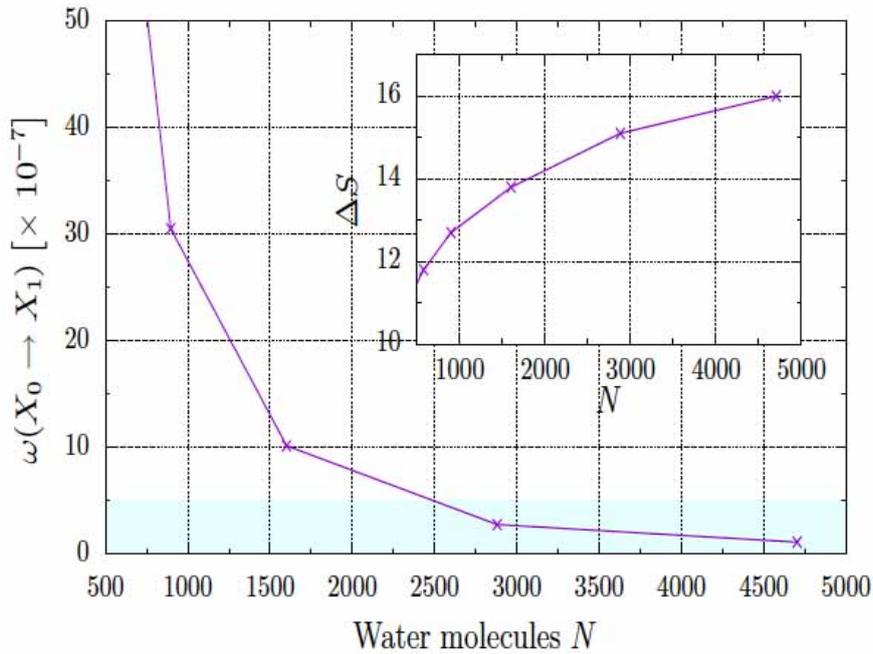


Fig. 4.14: 2-state model における基底エネルギー状態 X_0 (無次元のエントロピーの値 $\ln n_0$) から第一励起状態 X_1 ($\ln n_1$) 遷移確率 ($\omega(X_0 \rightarrow X_1) = \exp[-\Delta S]$)。ここで ΔS は $\Delta S = \ln n_1 - \ln n_0$ として定義された。挿入図は ΔS を表す。ライトブルーでハイライトされた領域は、Fig. 4.13 の領域に対応する。

最後に Marsaglia 乱数の現状として、先行研究 [9] では用いなかったが、2つの Marsaglia 乱数を1つに結合することで、この離散問題を軽減するより洗練された Marsaglia 乱数発生器は既に提案されており [13]、今後の研究では、この高性能な Marsaglia 乱数発生器を使用することが推奨される。

4.5 結論

理論的・実験的な推定は依然として困難だが、近年の急速な計算科学の進歩により、氷 I_h の残余エントロピーはシミュレーションアルゴリズムの精度をテストするための優れたベンチマークの1つになりつつある。しかし、各研究グループの計算結果には、測定誤差以上の不一致が残っている。氷 I_h の正確な残余エントロピーはまだ得られていないのである。

本研究では、REWL-MUCAREM シミュレーションによって残余エントロピーを推定した。氷 I_h の残余エントロピー S_0 の最終的な推定値は、 0.815615 ± 0.00063 [cal/deg mol] であった。もし同様の手法を用いて、本研究の最新結果よりも高い精度で残余エントロピーを推定するには、 $N = 4704$ より大きな系での W_0 の計算を行うのが妥当である。本研究における最終的な推定値は、参考文献 [9] の以前の MUCA シミュレーションの推定値を改善し、異なるシミュレーション手法を使用したグループの結果とよく一致した。その中でも本研究グループを含めた3つの異なる計算グループが誤差の範囲内で一致する推定値を示した。また擬似乱数発生器の均一性の重要性についても議論した。

References

- [1] W. F. Giaque, and M. F. Ashley, *Phys. Rev.*, **43** 81 (1933).
- [2] J. D. Bernal, and R. H. Fowler, *J. Chem. Phys.*, **1**, 515 (1933).
- [3] L. Pauling, *J. Am. Chem. Soc.*, **57**, 2680 (1935).
- [4] L. Onsager, and M. Dupuis, *Rend. Sc. Int. Fis. "Enrico Fermi"*, **10**, 294 (1960).
- [5] J. F. Nagle, *J. Math. Phys.* **7**, 1484 (1966).
- [6] B. A. Berg, C. Muguruma, and Y. Okamoto, *Phys. Rev. B*, **75**, 092202 (2007).
- [7] B. A. Berg, and W. Yang, *J. Chem. Phys.*, **127**, 224502 (2007).
- [8] C. Muguruma, Y. Okamoto, and B. A. Berg, *Phys. Rev. E*, **78**, 041113 (2008).
- [9] B. A. Berg, C. Muguruma, and Y. Okamoto, *Mol. Sim.*, **38**, 856 (2012).
- [10] B. A. Berg, and T. Neuhaus, *Phys. Lett. B*, **267**, 249 (1991).
- [11] B. A. Berg, and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [12] W. Janke, *Physica A*, **254**, 164, (1998).
- [13] B. A. Berg, *Markov Chain Monte Carlo Simulation and Their Statistical Analysis* (World Scientific, Singapore, 2004).
- [14] C. P. Herrero, and R. Ramírez, *Chem. Phys. Lett.*, **568**, 70 (2013).
- [15] J. Kolafa, *J. Chem. Phys.*, **140**, 204507, (2014).
- [16] M. V. Ferreyra, G. Giordano, R. A. Borzi, J. J. Betouras, and S. A. Grigera, *Eur. Phys. J. B*, **89**, 51 (2016).
- [17] M. V. Ferreyra, and S. A. Grigera, *Phys. Rev. E*, **98**, 042146, (2018).
- [18] L. Vanderstraeten, B. Vanhecke, and F. Verstraete, *Phys. Rev. E*, **98** 042145, (2018).
- [19] T. Hayashi and Y. Okamoto, *Phys. Rev. E*, **100**, 043304 (2019).
- [20] R. R. P. Singh, and J. Oitmaa, *Phys. Rev. B*, **85**, 144414 (2012).
- [21] B. A. Berg and T. Celik, *Phys. Rev. Lett.*, **69**, 2292 (1992).

- [22] M. Matsumoto, and T. Nishimura, *ACM Trans. Model. Comput. Simul.*, **8**, 3, (1998).
- [23] "Mersenne Twister Random Number Generator for Fortran",
[http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/VERSIONS/
FORTRAN/mt19937-64.f95](http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/VERSIONS/FORTRAN/mt19937-64.f95)
- [24] G. Marsaglia, A. Zaman, and W. W. Tsang, *Stat. Probab. Lett.*, **9**, 35, (1990).

第5章 REWL-MUCAREM with Genetic Algorithm for Protein Folding Problem and Microcanonical Inflection-Point Analysis

5.1 導入

生体を構成する物質は、それぞれが生命を維持するうえで重要な役割を果たす。特に、人体構成の約20%を占めるタンパク質は、アミノ酸の組み合わせや量の違いによって形状や働きが異なり、酵素やホルモン・免疫物質といった多岐にわたる機能を有する。従って、生命の仕組みを理解するために、タンパク質が機能を発現するメカニズムの理解が重要となる。

タンパク質は、体内で特有の立体構造を取り、機能を発現する。タンパク質が生体内で立体構造を取る過程では、20種類のアミノ酸が連なったひも状の構造(1次構造)から、 α ヘリックス、 β シートといった特徴的な立体構造(2次構造)をとり、それらが組み合わさった立体構造(3次構造)をとる。そして3次構造を持った複数のタンパク質が組み合わさって4次構造を形成し機能を発現する。すなわち、タンパク質はアミノ酸配列(1次構造)によって、立体構造と機能が決定する。物理法則にしたがったコンピュータ・シミュレーションによって、1次構造から立体構造(最安定構造)を理論的に予測する問題を、タンパク質のフォールディング問題と呼ぶ。

この立体構造予測には、自由エネルギーなどの物理量を指標として用いる。シミュレーションを行う際、これらの物理量を精度良く推定するために、より多くの構造をサンプリングする必要がある。しかしタンパク質のような複雑多体系では、温度一定の統計力学的集団であるカノニカルアンサンブルを再現するシミュレーション手法だと、ひとつの構造に留まり続け、著しく構造サンプリング効率が落ちるという問題がある。そのため、人工的な統計力学的集団に従ってサンプリングを行う拡張アンサンブル法が有効である。

本章では、本研究グループが提案した拡張アンサンブル法の計算プロトコルであるREWL-MUCAREM MC法を、全原子モデルによるアラニン重合体のフォールディングに適用した。ここでは、タンパク質の構造探索アルゴリズムである、遺伝的アルゴリズム(Genetic Algorithm:GA [1, 2])を導入したREWL-MUCAREMアルゴリズムを提案した。結果として、アラニン重合体の特徴的な振る舞いであるヘリックス-コイル転移現象を確認した。

また、コンピュータ・シミュレーションによってタンパク質の構造変化を議論する際、構造の変化に伴って変化するエネルギーや比熱、自由エネルギーといった物理量が指標となる。しかし、これらの物理量では変化量が小さく、構造変化を判別できないことがあ

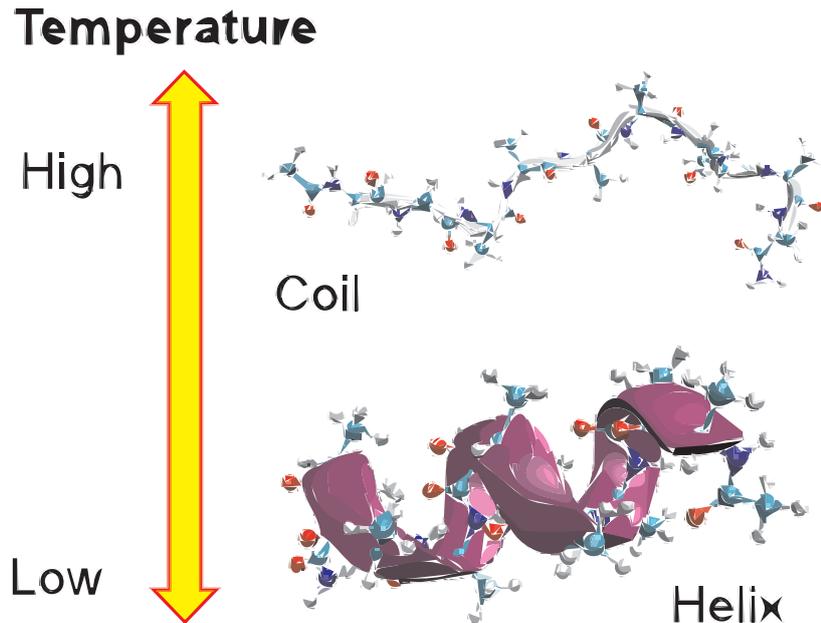


Fig. 5.1: アラニン重合体の立体構造。アラニン重合体は高温でコイル状態を取り、低温でヘリックス状態を取る。

る。そのため、本研究では、タンパク質の小さな構造変化を同定する解析手法として、マイクロカノニカルエントロピーと、そのエネルギーの導関数の概形によって物質の構造を分類する、Microcanonical Inflection-Point Analysis Method (MIPA 法) [3, 4, 5] を導入した。この手法は、イジング模型、粗視化タンパク質模型において、鋭敏な相分類が可能であったことが報告されている [4, 5]。しかし MIPA 法をタンパク質のすべての原子を詳細に取り扱った全原子模型に適用した事例はない。そのため本研究では、MIPA 法によってタンパク質のフォールディング現象を解析し、その有効性を確認した。全原子モデルを用いたアラニン重合体のヘリックス-コイル転移は、有限サイズスケリング法を用いた先行研究 [6, 7, 8] により解析されており、その臨界指数から水の存在の有無で転移の強さが変化すると予想されている [9]。文献 [9] によれば、水の影響を取り入れなかった場合 (以後、真空とする) に一次相転移のような振る舞いを、陰溶媒モデルによって水の効果を取り入れた場合 (以後、水中とする) に二次相転移のような振る舞いになることが報告されている。本研究では MIPA 法を用いた解析によって、同様の結果を得た。これにより MIPA 法は、全原子模型のタンパク質の構造転移においても、適用可能であることが示唆された。

5.2 Models

長さの異なる 2 つのアラニン重合体を考えた。アラニンの残基数 N は 10 個と 20 個とした。図 5.1 では、残基数 10 個のアラニン重合体のコイル構造とヘリックス構造を示す。アラニンの重合体は、高温領域でコイル構造、低温領域でヘリックス構造を形成する。これらの系は、ECEPP2 力場を使用した MUCA シミュレーションによってすでに研究が報告されている [6, 9, 10, 11]。

本研究では、水の効果を取り入れていない (すなわち真空中での) AMBER99 力場と、陰

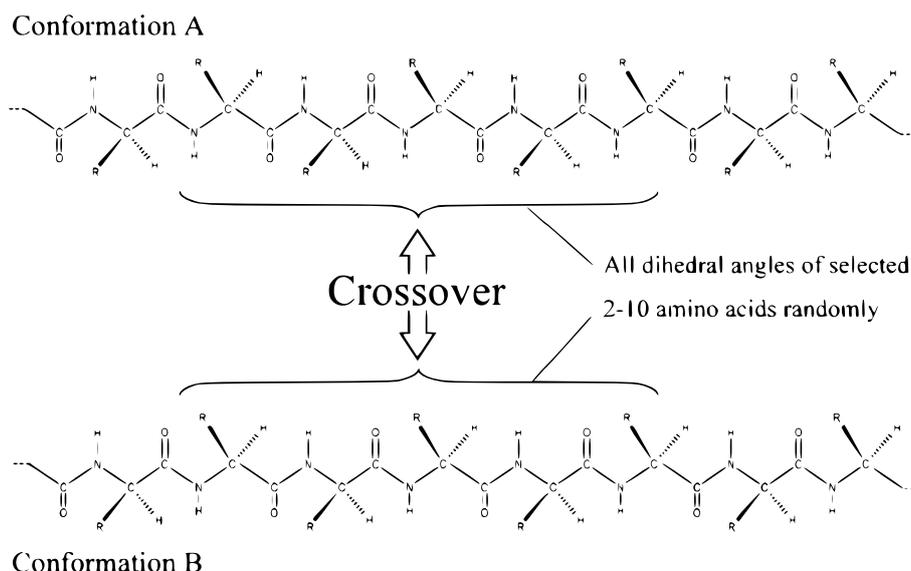


Fig. 5.2: 遺伝的アルゴリズムにおける交差の概念図。

溶媒モデルによって水の効果を取り入れた AMBER99/GBSA 力場を使用した。計算プログラムとしてオープンソースである TINKER[12] に、遺伝的アルゴリズムを導入した REWL-MUCAREM を実装した。タンパク質の末端の電荷は、タンパク質のヘリックス構造形成を阻害することが分かっているため、N 末端に中性の NH₂-基、C 末端に中性の -COOH 基を付与することで電荷を除去した。

タンパク質の立体構造は、主鎖および側鎖の 2 面角で一意に決定できる。本研究では、タンパク質がヘリックス構造を形成しているとする判断基準は、2 面角を用いて、主鎖の 2 面角 (ϕ, ψ) が ($-70 \pm 20, -37 \pm 20$) に存在した場合にその残基はヘリックス構造を形成している、と定義した。この 2 面角の基準値は、事前のレプリカ交換法によるシミュレーションによってサンプルされ、タンパク質の可視化ソフト VMD によって、ヘリックス状態にあると判断された 10,000 個のサンプルの期待値から得られたものである。ヘリックスの長さ L は、ヘリックス構造をとる連続した残基の数によって定義した。末端残基を無視したため、ヘリックスの最大長は残基数 $N = 10$ で 8、 $N = 20$ で 18 となる。

5.3 シミュレーション手法

タンパク質のひとつの立体構造は 2 面角 ϕ の集合で表現できる。従って、本研究でのモンテカルロ法の 1 MC STEP は、 $\phi \rightarrow \phi'$ として、2 面角を更新した。

5.3.1 遺伝的アルゴリズム

原理

本研究では、タンパク質の構造探索アルゴリズムである遺伝的アルゴリズム (Genetic Algorithm:GA) を使用した。遺伝的アルゴリズムは、生体高分子系を含む多くの最適化問題に適用されてきた生物の進化を模したアルゴリズムである [13, 14, 15]。以下では、モンテカルロ法における遺伝的アルゴリズムについて解説する。モンテカルロ法を使用した

遺伝的アルゴリズムと、その生体高分子のフォールディング問題への応用としては、文献 [16, 17, 18, 19, 20] などが詳しい。

モンテカルロ法における遺伝的アルゴリズム

コンピュータ・シミュレーションにおける遺伝的アルゴリズムでは、はじめに複数の初期構造を用意し、これらに対して同時に独立してシミュレーションを行う。シミュレーションを実行している間、遺伝的アルゴリズムの特徴である”交叉”という操作を、一定のシミュレーション時間間隔 (MC STEP 数) で実行する。交叉とは、同時に流れている別のシミュレーションで得られているタンパク質の立体構造情報を参考にして、新しい立体構造を提案する操作である。そして、この新たに得られた構造が元の構造より優れていれば、新しい構造を受け入れ再びシミュレーションを行う。そうでなければ、元の構造を引き継ぎシミュレーションを再開する。この操作を”選択”と呼ぶ。“交叉”と”選択”の繰り返しによって、優れた構造を探索する手法が遺伝的アルゴリズムである。シミュレーションの目的に合わせ、”優れた”という基準を設定する。

モンテカルロ法を使用する場合、第2章で紹介したレプリカ交換法のように系のコピー (レプリカ) を M 個用意し、同時に独立してシミュレーションを行う。そして、カノニカルアンサンブルに従ったモンテカルロ法を、一定の MC STEP 回実行した後、交叉の操作を行う。交叉では、別のレプリカによって得られているタンパク質の立体構造のうち、いくつかのアミノ酸残基を選択し、これを元の構造のアミノ酸残基と入れ替える。交叉のアルゴリズムを導入することで、通常 2 面角 ϕ を更新する MC STEP よりも、タンパク質の構造を大きく変化させることになり、構造空間の探索効率が高まる。この構造情報の選択の仕方には、1点交叉や2点交叉法など複数の方法が提案されている [14, 15, 17, 18] が、本研究では2点交叉法を使用した。2点交叉法では、タンパク質の残基を連続して n 個選択し、その2面角を入れ替える。この概念図を図 5.2 に示した。

本来の遺伝的アルゴリズムの選択処理では、オリジナルの構造と交叉によって生まれた構造を比較し優れた構造を次の構造とする。モンテカルロ法の場合、優れた構造を選択にメトロポリス判定を使用することで、詳細釣り合いの条件式 (2.25) を満足するため、統計的な確率分布からのサンプリングが保証される。

カノニカルアンサンブルを実現する MC シミュレーションの場合の、遺伝的交叉による新しい構造を選択する際の遷移確率 $w_{CA}(p \rightarrow c)$ は、カノニカルアンサンブルの重み因子

$$W_{CA}(E) = e^{-\beta E}. \quad (5.1)$$

より、

$$w_{CA}(p \rightarrow c) = \min \left[1, \frac{W_{CA}(E_c)}{W_{CA}(E_p)} \right] = \min (1, \exp [-\beta(E_c - E_p)]). \quad (5.2)$$

である。ここで E_p および E_c は交叉前の構造 p と交叉後の構造 c の構造のポテンシャルエネルギーを表す。式 (5.2) の遷移確率を用いることで、カノニカル分布からのサンプリングが保証される。以上がモンテカルロ法で用いられる遺伝的アルゴリズムである。

アルゴリズム

従って、モンテカルロ法における遺伝的アルゴリズムは次のように行われる。

1. M 個の MC シミュレーションを独立して同時に実行する。
2. 一定回数の MC STEP 後、ランダムに 2 つのシミュレーションのペアを作る。
3. 選ばれたペアで、タンパク質内の全アミノ酸配列から一部の連続したアミノ酸配列 n 個をランダムに選ぶ。
4. 選ばれたアミノ酸配列 n 個に含まれる 2 面角の値を、ペア間で交換する (交叉)。
5. 新しい構造を受託するか棄却するか判定する (選択)。

本研究では、このアルゴリズムを REWL-MUCAREM 法に拡張した。

5.3.2 REWL-MUCAREM with Genetic Algorithm

通常の遺伝的アルゴリズムの遺伝的交差によって得られる構造は、高いポテンシャルエネルギーを持つ不安定な構造を取ることが多い。カノニカルアンサンブルに従ったメトロポリス判定では、受託確率が小さくなりすぎてしまうことが考えられる。そのため、遺伝的選択の操作を行う前に、比較的安定な構造にするために系の緩和処理を行うなどの工夫がされることもある ([20])。本研究では、幅広いポテンシャルエネルギー領域を一度にカバーする拡張アンサンブル法、REWL-MUCAREM を使用することで、遺伝的交差アルゴリズムの受託率を確保した。

REWL-MUCAREM では、人工的な統計力学的アンサンブルであるマルチカノニカルアンサンブルを、レプリカごとにエネルギー領域を分割して用いる。はじめに、マルチカノニカルアンサンブルとレプリカ交換法でのアンサンブルにおける遺伝的アルゴリズムのメトロポリス判定の式を確認する。

マルチカノニカル法における遺伝的アルゴリズム

マルチカノニカル分布 $P_{\text{MUCA}}(E)$ を実現する、マルチカノニカル重み因子 $W_{\text{MUCA}}(E)$ は

$$W_{\text{MUCA}}(E) = \frac{1}{g(E)}. \quad (5.3)$$

であるため、遺伝的アルゴリズムのメトロポリス判定の式は、(5.2) と同様に、

$$w_{\text{MUCA}}(p \rightarrow c) = \min \left[1, \frac{W_{\text{MUCA}}(E_c)}{W_{\text{MUCA}}(E_p)} \right] = \min \left(1, \frac{g(E_p)}{g(E_c)} \right). \quad (5.4)$$

となる。

WL 法においても、式 (5.4) に従って遺伝的交差を行えば良い。この場合、 $g(E)$ はシミュレーション中に変動する値 $\tilde{g}(E)$ となる。第 2 章で述べたように、WL 法は詳細釣り合いの条件を満足しないが、遺伝的アルゴリズムを導入することでシミュレーション中のタンパク質に大きな構造変化を起こすことができ、サンプリング効率の上昇が期待できる。

注意すべき点として、式 (5.2) 及び (5.4) で紹介したカノニカルアンサンブルとマルチカノニカルアンサンブルでの遺伝的交差のメトロポリス判定では、ペアとなっている 2 つの

タンパク質のうち、片方で遺伝的交差が受託され、もう一方が棄却される可能性がある。次に導入するレプリカ交換法での遺伝的アルゴリズムでは、ペアとなっているタンパク質で同時に遺伝的交差が受託される遷移確率を導出している。すなわち、2つのタンパク質の構造を”入れ替える”アルゴリズムでの遷移確率となる。原理的にはどちらのアルゴリズムを使用してもよい。

レプリカ交換法における遺伝的アルゴリズム

M 個のカノニカルアンサンブルからなるレプリカ交換法の遺伝的アルゴリズムを考える。この重み因子は式 (2.61) より、

$$W_{\text{REM}}(X) = \prod_{i=1}^M W_{\text{CA}}(E(x^{[i]}); \beta_{m(i)}) = \prod_{m=1}^M \exp \{-\beta_m E^{[i(m)]}\}. \quad (5.5)$$

で定義される。ここで X は全レプリカでの系の状態を表す (式 (2.60))。 i ($i = 1, 2, \dots, M$) はレプリカの番号を表し、 $\beta_{m(i)}$ はレプリカ i の逆温度、 $E^{[i]}$ はレプリカ i の状態 $x^{[i]}$ でのポテンシャルエネルギーである。

いま、温度 T_m, T_n に対応する2つのレプリカ (i, j) のペアで遺伝的交叉を実行した場合、各レプリカで新しい構造 ($x'^{[i]}, x'^{[j]}$) が提案される。遺伝的交差前の系の状態を X 、後の状態を X' とすると、

$$X = \{\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots\} \rightarrow X' = \{\dots, x_m'^{[i]}, \dots, x_n'^{[j]}, \dots\}. \quad (5.6)$$

となる。

このとき、遺伝的アルゴリズムの遷移確率 $w_{\text{REM}}(X \leftrightarrow X')$ は、 $(x_m'^{[i]}, x_n'^{[j]})$ に対応するエネルギーを $(E'^{[i]}, E'^{[j]})$ とすると、

$$\begin{aligned} w_{\text{REM}}(X \rightarrow X') &= \min \left[1, \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)} \right] \\ &= \min \left[1, \frac{\exp[-\beta_m (E(x'^{[i]}))]}{\exp[-\beta_m (E(x^{[i]}))]} \times \frac{\exp[-\beta_n (E(x'^{[j]}))]}{\exp[-\beta_n (E(x^{[j]}))]} \right] \\ &= \min [1, \exp[-\beta_m (E'^{[i]} - E^{[i]}) - \beta_n (E'^{[j]} - E^{[j]})]]. \end{aligned} \quad (5.7)$$

となる。ここで仮に、レプリカ i, j のすべての2面角を入れ替える (すなわち、 $x_m'^{[i]} = x_n^{[j]}, x_n'^{[j]} = x_m^{[i]}$) と考えた場合、 $E'^{[i]} = E^{[j]}, E'^{[j]} = E^{[i]}$ となり、式 (5.7) は、

$$\begin{aligned} w_{\text{REM}}(X \leftrightarrow X') &= \min [1, \exp[-\beta_m (E^{[j]} - E^{[i]}) - \beta_n (E^{[i]} - E^{[j]})]] \\ &= \min [1, \exp[-(\beta_m - \beta_n) (E^{[j]} - E^{[i]})]]. \end{aligned} \quad (5.8)$$

となり、レプリカ交換の遷移確率 (2.65) と一致する。すなわち構造情報を”入れ替える”遺伝的アルゴリズムは、レプリカ交換法の拡張版といえる。

REWL-MUCAREM における遺伝的アルゴリズム

REWL や MUCAREM で実現される拡張アンサンブルの重み因子は、式 (2.70) あるいは (2.81) であった。REWL ではシミュレーション中に重み因子 $W_{\text{REWL}}(X)$ が変動するの

で、簡単のために MUCAREM の重み因子 W_{MUCAREM} によって、遺伝的アルゴリズムを考える。

全エネルギー領域の MUCAREM 重み因子 $W_{\text{MUCAREM}}(E)$ は、式 (2.70) より、

$$W_{\text{MUCAREM}}(X) = \prod_{i=1}^M W_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]})) . \quad (5.9)$$

であった。ここで、各エネルギー領域 m ごとレプリカの重み因子 $W_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]}))$ は、

$$W_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]})) = \exp\left[-\beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]}))\right] \quad (5.10)$$

$$\equiv \frac{1}{g^{\{m\}}(E(x_m^{[i]}))} . \quad (5.11)$$

と定義されていた(式 (5.10))。これはマルチカノニカルポテンシャルエネルギー(式 (5.12))を定義することで、

$$\varepsilon_{\text{MUCA}}^{\{m\}}(E) = \begin{cases} \frac{\partial E_{\text{MUCA}}(E_L^{\{m\}}; T_m)}{\partial E} (E - E_L^{\{m\}}) + E_{\text{MUCA}}(E_L^{\{m\}}; T_m), & \text{for } E < E_L^{\{m\}}, \\ E_{\text{MUCA}}(E; T_m) & \text{for } E_L^{\{m\}} \leq E \leq E_H^{\{m\}}, \\ \frac{\partial E_{\text{MUCA}}(E_H^{\{m\}}; T_m)}{\partial E} (E - E_H^{\{m\}}) + E_{\text{MUCA}}(E_H^{\{m\}}; T_m), & \text{for } E > E_H^{\{m\}}. \end{cases} \quad (5.12)$$

と表すことができた。

いま m 番目と n 番目のエネルギー領域に対応するレプリカ対 i と j のペアで、遺伝的交差を実行した場合、各レプリカで新しい構造 $(x'^{[i]}, x'^{[j]})$ が提案される。遺伝的交差前の系の状態を X , 後の状態を X' とすると、式 (5.6) となる。

従って、マルチカノニカルレプリカ交換用のアンサンブルにおける遺伝的アルゴリズムのメトロポリス判定式は、

$$\begin{aligned} w_{\text{MUCAREM}}(X \rightarrow X') &= \min \left[1, \frac{W_{\text{MUCAREM}}(X')}{W_{\text{MUCAREM}}(X)} \right] \\ &= \min \left[1, \frac{\exp\left[-\beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m'^{[i]}))\right]}{\exp\left[-\beta_m \varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]}))\right]} \times \frac{\exp\left[-\beta_n \varepsilon_{\text{MUCA}}^{\{n\}}(E(x_n'^{[j]}))\right]}{\exp\left[-\beta_n \varepsilon_{\text{MUCA}}^{\{n\}}(E(x_n^{[j]}))\right]} \right] \\ &= \min [1, \exp[-\Delta]] . \end{aligned} \quad (5.13)$$

となる。ここで

$$\begin{aligned} \Delta &= \beta_m \left(\varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m'^{[i]})) - \varepsilon_{\text{MUCA}}^{\{m\}}(E(x_m^{[i]})) \right) \\ &\quad - \beta_n \left(\varepsilon_{\text{MUCA}}^{\{n\}}(E(x_n^{[j]})) - \varepsilon_{\text{MUCA}}^{\{n\}}(E(x_n'^{[j]})) \right) \end{aligned} \quad (5.14)$$

である。MUCAREMにおいても、隣り合うエネルギー領域 ($n = m + 1$) における全ての立体構造情報の入れ替え ($x_m^{[i]} = x_{m+1}^{[j]}, x_{m+1}^{[j]} = x_m^{[i]}$) を仮定することにより、式 (2.78) と比較することができる。従って REM と同様に構造情報を”入れ替える”遺伝的アルゴリズムは、MUCAREM の拡張版といえる。

以上より、本研究で用いた REWL-MUCAREM のアルゴリズムでは、次の 3 つの MC update から構成されている。

1. アミノ酸残基間のランダムに選択されたひとつの 2 面角 ϕ の値を更新する。
2. 連続したいくつかのアミノ酸残基の 2 面角を、別の構造の 2 面角に更新する (遺伝的交差)。
3. シミュレーションするエネルギー領域を更新する (レプリカ交換)。

それぞれの MC update は、上から順にタンパク質の局所的な構造変化を引き起こす更新から、MUCA シミュレーションを行うエネルギー領域を更新する大局的な更新となっている。規模の異なる MC update 方法を取り入れることで、REWL-MUCAREM の構造サンプリング能力の向上が期待される。

5.3.3 Microcanonical Inflection-Point Analysis Method

アラニン重合体の構造変化を解析するために、Microcanonical Inflection-Point Analysis [4, 5] を用いた。この手法は、熱力学ポテンシャルとその微分の振る舞いによって相転移を分類する Ehrenfest の手法 [21] を有限系に拡張したものである。Microcanonical Inflection-Point Analysis は、通常の秩序パラメーターや応答関数では検出できなかった高次の相転移を検出するのに十分な感度があると報告されている [4]。

この解析手法では、ミクロカノニカルエントロピー $S(E)$ とそのエネルギーの導関数が、構造相転移を識別するために使用される。この時、ミクロカノニカルエントロピー $S(E)$ は、次の式で定義される。

$$S(E) = k_B \ln g(E). \quad (5.15)$$

ここで、 k_B はボルツマン定数であり、 $g(E)$ はポテンシャルエネルギー E に依存する系の状態密度である。

$S(E)$ とその導関数は、同一相内でのエネルギー領域では単調な振る舞いを示すが、何らかの転移が存在する場合、 $S(E)$ またはそのエネルギーの導関数に変曲点が現れる。これにより、相転移の系統的な分類が可能となる [4]。シミュレーションによる通常の相転移の分類は、無限系への熱力学的極限を考えることで、熱力学的量の発散や不連続性を検出する。しかしこの手法は、本来有限の大きさであるタンパク質系へ用いることが適切かどうかわからない。Microcanonical Inflection-Point Analysis によって、古典的な相転移分類手法の類推として、有限系であっても変曲点によって、構造変化を検出できると指摘されている [4]。

$S(E)$ に変曲点がある場合、エネルギーの一次導関数に極小値・極大値が観測される。ミクロカノニカル逆温度 $\beta(E)$ を、 $S(E)$ の導関数として、

$$\beta(E) = \frac{1}{k_B} \frac{\partial S(E)}{\partial E}. \quad (5.16)$$

Microcanonical Inflection-Point Analysis では、この種の転移を 1 次転移として分類する。同様に、2 次の転移の場合は、 $\beta(E)$ に変曲点が生じる。より高次の転移も同様の方法で定義される。本章では、これらの物理量の振る舞いを確認することで、転移の分類を行った。

5.4 初期条件

本研究でシミュレーション対象としたアラニン重合体は 10 残基と 20 残基であり、側鎖も含めた 2 面角の総数 N_d は 42 と 82 であった。1 回の MC sweep は、メトロポリス判定 N_d 回とした。擬似乱数は Mersenne Twister 擬似乱数発生器を用い、これにはオープンソースのプログラムコードを使用した。レプリカ数は 4 個とした。各レプリカは、隣接するエネルギー領域間で 80% のオーバーラップを持たせ、レプリカ交換は 100 MC sweep ごとに試行された。REWL での WL 平坦度判定は 1000 MC sweep ごとに試行された。REWL によって MUCA 重み係数を取った後、各レプリカは MUCAREM シミュレーションに移行し、状態密度 $g(E)$ を推定した。この MUCAREM の MC sweep 数の合計は 100 000 sweep であった。

5.5 結果と考察

図 5.2 は、REWL-MUCAREM シミュレーションの後に得られた DOS から (3.2), (3.3) 式によって計算された比熱 (a)(b) と、サンプルされた構造データから (2.80) によって計算されたヘリックス長 L の温度依存性 (c)(d) である。(a) と (c) はアラニン重合体の長さ $N = 10$ 、(b) と (d) は長さ $N = 20$ の結果を表す。紫の曲線は真空状態での結果であり、緑の曲線は水中での結果を示す。比熱を表す (a)(b) では、真空中においても水中においてもピークが確認できた。ピークを示す温度 T_c は、 $N = 10$ で $T_c^{\text{water}} \simeq 380[\text{K}]$ 、 $T_c^{\text{nowater}} \simeq 720[\text{K}]$ であった。また $N = 20$ で $T_c^{\text{water}} \simeq 380[\text{K}]$ 、 $T_c^{\text{nowater}} \simeq 850[\text{K}]$ であった。ここで、ヘリックス長の期待値を表す (c)(d) を確認すると、比熱の T_c と同程度の温度で、ヘリックスが形成されていることがわかる。従って、この比熱のピークはヘリックス-コイル構造転移に伴う、ポテンシャルエネルギーの変化に由来すると考えた。

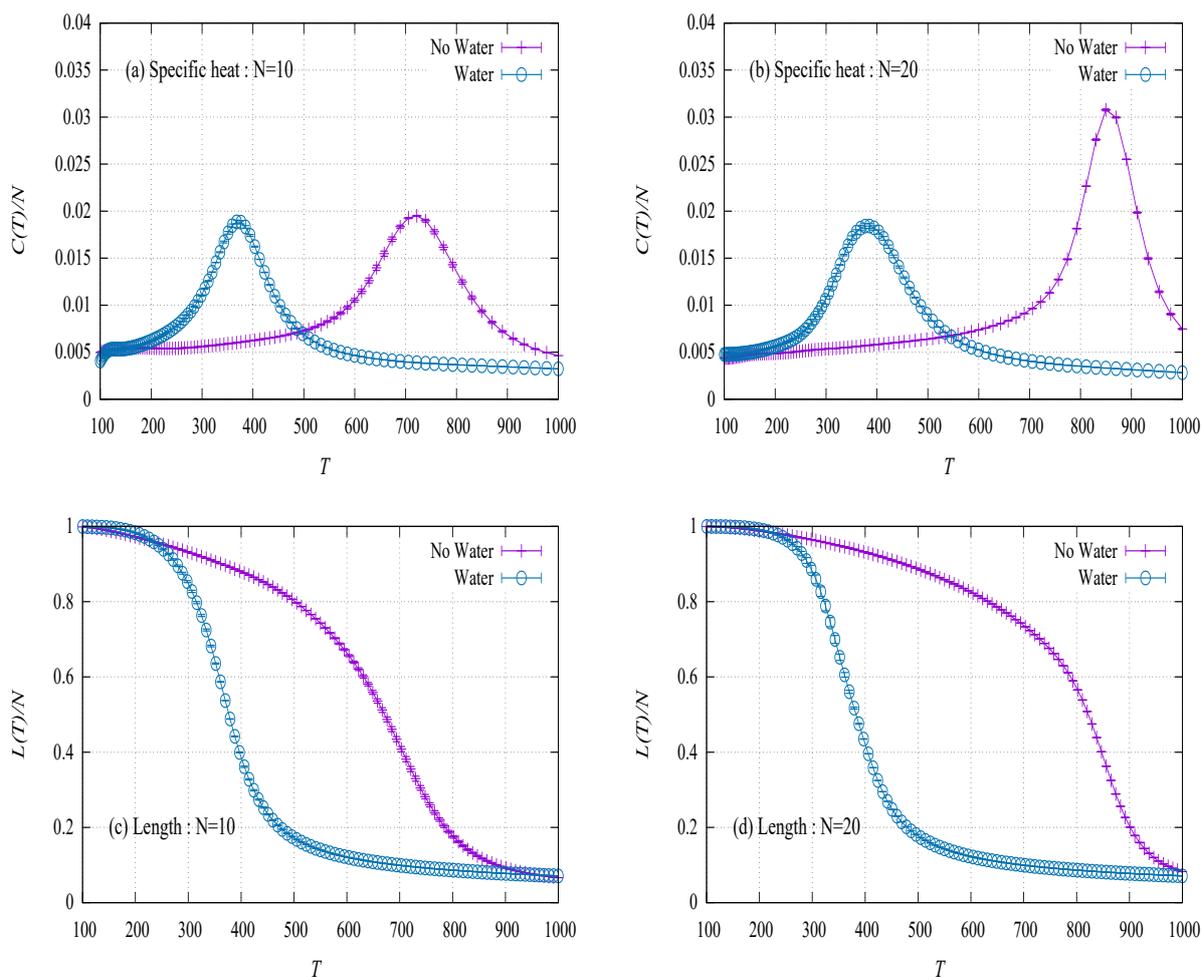


Fig. 5.3: 比熱とヘリックス長。(a)(c)は残基数 $N = 10$ で得られた期待値であり、(b)(d)は残基数 $N = 20$ で得られた期待値である。紫色の曲線は真空条件でのシミュレーションから、緑の曲線は水中条件でのシミュレーションから得られた値である。

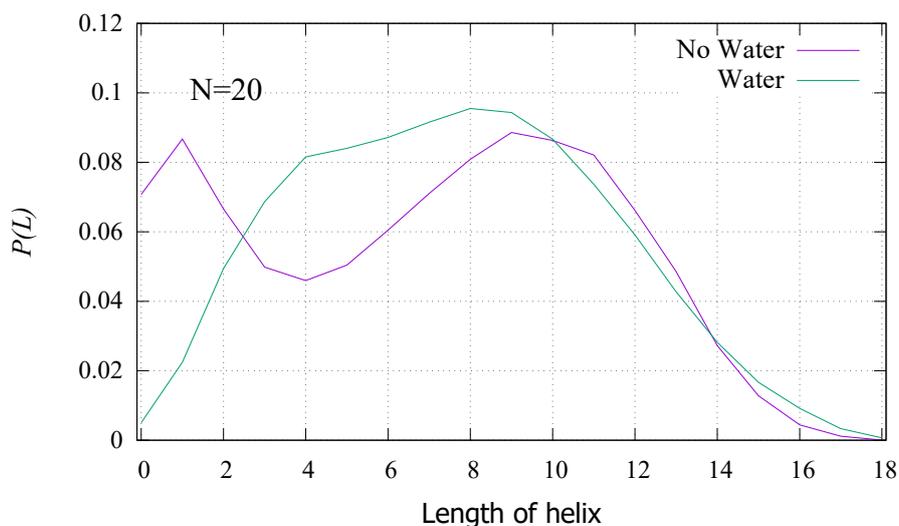


Fig. 5.4: 残基数 20 個におけるヘリックス長の確率分布。真空中の場合 (紫の曲線) は温度 $T = 720[\text{K}]$ の、水の効果を取り入れた場合 (緑の曲線) は $T = 380[\text{K}]$ での確率分布を表す。

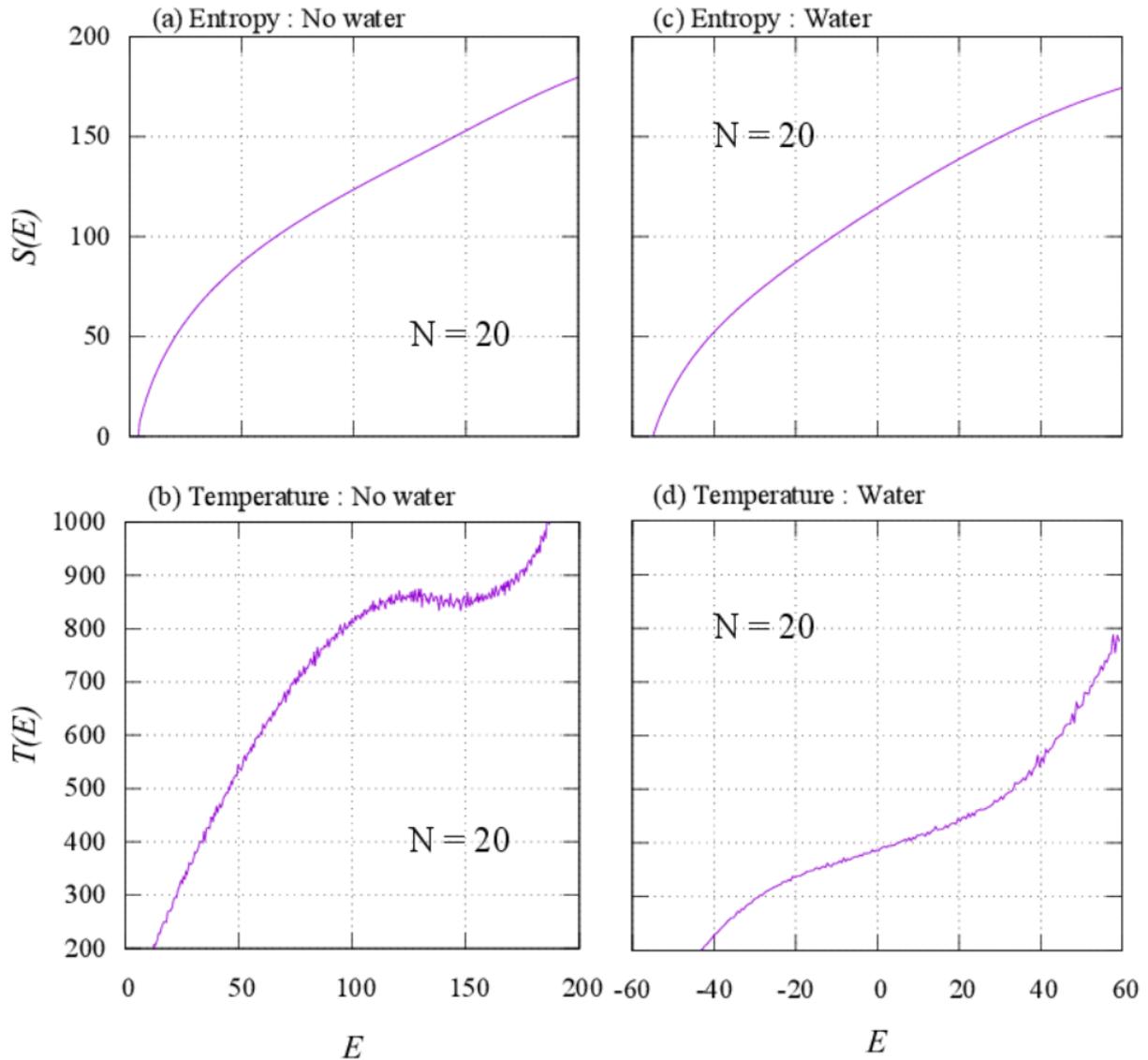


Fig. 5.5: Microcanonical inflection-point analysis によるマイクロカノニカルエントロピー (a)(c) とマイクロカノニカル温度 (b)(d) の解析結果。真空中のマイクロカノニカル逆温度の結果 (b) では極大値と極小値が観測されるが、水の効果を取り入れた場合のマイクロカノニカル逆温度の結果 (d) では極大・極小値は観測されなかった。

この構造転移を詳細に確認するために、転移温度 T_c 周辺におけるヘリックス長の確率分布を調べた (図 5.4)。図 5.4 は、残基数 20 個での、転移温度 T_c におけるヘリックス長 L 確率分布を示す。紫色の曲線は真空中での、緑色の曲線は水中の結果である。水中では、コイル状態 ($L \sim 0$) からヘリックス状態 ($L \sim 20$) まで、ひとつのピークを持って幅広く分布していることが見て取れる一方で、真空の場合は、コイル状態とヘリックス状態にふたつのピークが確認できる。これは、水の効果を取り入れるか否かによって構造転移に影響が及ぼされたことを示唆する。

Microcanonical Inflection-Point Analysis を行うため、REWL-MUCAREM によって推定された $S(E)$ とその導関数である $\beta(E)$ を計算した。図 5.5 は、 $N = 20$ の場合の $S(E)$ と $T(E) = 1/k_B\beta(E)$ を表す。(a) と (b) はそれぞれ、水を考慮しなかった場合の $S(E)$ と $T(E)$ であり、(c) と (d) は水を考慮した場合の結果である。(b) では極大値と極小値があり、これは無限系における一次相転移のような振る舞いに対応する。(d) では変曲点が確認された

が、極大値と極小値は確認できなかった。これは Microcanonical Inflection-Point Analysis Method における 2 次の相転移のような振る舞いに対応する。以上の MIPA 法による転移の分類結果は、臨界指数を用いた (無限系における) 分類結果 (文献 [9]) と一致した。

5.6 結論

本研究では、タンパク質のような複雑多体系における有効なサンプリング手法として、遺伝的アルゴリズムを導入した REWL-MUCAREM プロトコルを提案した。この REWL-MUCAREM with Genetic Algorithm を使用して、ヘリックス-コイル転移を示すアラニン重合体のモンテカルロシミュレーションを実行した。結果として、ヘリックス-コイル転移をシミュレーションできた。

またこの系におけるヘリックス-コイル転移は、水の影響を考慮に入れるか否かで、転移の振る舞いが変わることが示唆されていた。本研究では、有限系での相転移のような振る舞いを分類する手法、Microcanonical Inflection-Point Analysis Method により、ヘリックス-コイル転移を調べた。結果は、構造転移が水の存在に依存することを示唆した。水の影響を考慮しない場合のヘリックス-コイル転移は、一次相転移のような振る舞いを示したが、水の影響を陰溶媒モデルで取り入れた場合は、二次相転移のような振る舞いを意味した。これらの結果は、先行研究と結論と一致している。Microcanonical Inflection-Point Analysis Method は、有限および複雑な系にも適用できると結論付けた。

Microcanonical Inflection-Point Analysis Method では、転移の分類・検知のためにマイクロカノニカルエントロピーとそのエネルギーの導関数を使用する。数値的に導関数を得るためには、マイクロカノニカルエントロピーを高精度で推定する必要がある。REWL-MUCAREM を使用することでマイクロカノニカルエントロピーを高精度で求めることができる。そのため Microcanonical Inflection-Point Analysis Method による解析は、REWL-MUCAREM によるシミュレーションを実行することで、タンパク質のような複雑多体系においても、信頼に足る結果を得ることを期待する。

References

- [1] J. H. Holland, *Adaption in Natural and Artificial Systems*, University of Michigan Press (1975).
- [2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, (1989).
- [3] M. Bachmann, *Thermodynamics and statistical mechanics of macromolecular systems*. Cambridge University Press (2014).
- [4] K. Qi and M. Bachmann, *Phys. Rev. Lett.*, **120**, 180601 (2018).
- [5] K. Qi, B. Liewehr, T. Coci, B. Pattanasiri, M. Williams and M. Bachmann, *J. Chem. Phys.*, **150**, 054904 (2019).
- [6] Y. Okamoto, U. H. E. Hansmann, *J. Phys. Chem.*, **99**, 11276 (1995).
- [7] U. H. E. Hansmann, Y. Okamoto, *J. Phys. Chem.*, **110** 1267 (1999).
- [8] N. A. Alves, and U. H. E. Hansmann, *Physical review letters*, **84**, 1836 (2000).
- [9] Y. Peng, U. H. E. Hansmann and Nelson A. Alves, *J. Chem. Phys.*, **118**, 2374 (2003).
- [10] A. Mitsutake, and Y. Okamoto, *Chem. Phys. Lett.*, **309**, 95 (1999).
- [11] A. Mitsutake, and Y. Okamoto, *J. Chem. Phys.*, **112**, 10638 (2000).
- [12] <https://dasher.wustl.edu/tinker/>
- [13] J. Lee, H. A. Scheraga, S. Rackovsky, *J. Comput. Chem.* **18**, 1222 (1997).
- [14] T. Hiroyasu, M. Miki, M. Ogura, *Proceedings of the 44th Institute of Systems, Kyoto*, pp. 113 (2000).
- [15] T. Hiroyasu, M. Miki, M. Ogura, Y. Okamoto, *J. IPS Japan*, **43**, 70 (2002).
- [16] Y. Sakae, T. Hiroyasu, M. Miki, and Y. Okamoto, *Pac, Symp. Biocomput.*, **16**, 217 (2011).
- [17] Y. Sakae, T. Hiroyasu, M. Miki, and Y. Okamoto, *J. Comput. Chem*, **32**, 1353 (2011).
- [18] Y. Sakae, T. Hiroyasu, M. Miki, K. Ishii, Y. Okamoto, *J. Phys. Conf. Ser.* **487**, 012003 (2014).
- [19] Y. Sakae, T. Hiroyasu, M. Miki, K. Ishii, Y. Okamoto, *Mol. Sim.*, **41**, 1045 (2015).

[20] Y. Sakae, J. E. Straub, and Y. Okamoto, *J. Comput. Chem.*, **40**(2), 475 (2019).

[21] P. Ehrenfest, *Proc. Royal Acad. Amsterdam (Netherlands)*, **36**, 153 (1933).

第6章 Modified Wang-Landau Algorithm with Lorentzian Fitting Algorithm

6.1 導入

拡張アンサンブル法を実行する際、モンテカルロ法 (MC 法) と分子動力学法 (MD 法) のいずれかを用いるが、特にタンパク質のような生体高分子を取り扱う場合、MD 法が好まれることが多い。生体高分子に拡張アンサンブル法の一つであるマルチカノニカル MD 法を実行した例として、例えば参考文献 ([1, 2, 3, 4]) などがあげられるが、MC 法よりもマルチカノニカル重み因子の決定が困難になるという問題がある。このような課題に対処するため、多くの手法が提案されている [5, 6, 7] が、本章では、Lorentzian 型の Fitting Algorithm を導入した ワン-ランダウ MD 法 (Modified Wang-Landau 法) を提案する。MC シミュレーションの場合、重み因子決定法として優れているという Wang-Landau 法であったが、MD シミュレーションの場合、後述する理由によって、シミュレーションが破綻する可能性がある。Modified Wang-Landau 法では、Lorentzian Fitting Algorithm を導入することによって、この問題に対応した。はじめに、マルチカノニカル法の逐次重み決定法において提案された Lorentzian Fitting [8] について解説する。次に、この Lorentzian Fitting を導入した、Modified Wang-Landau 法のアルゴリズムを提案する。

6.2 Multicanonical method with Lorentzian Fitting Algorithm

はじめに、マルチカノニカル MD 法 (MUCA MD) の従来の逐次的重み決定法について解説した後、Lorentzian を導入した逐次的重み決定法について述べる。

6.2.1 逐次近似法

MUCA MD 法における運動方程式は、第2章の式 (2.42)-(2.45) の通りである。ここでは、重み因子 $W_{\text{MUCA}} (= 1/g(E))$ に関連して $E_{\text{MUCA}}(E; T_0)$ が重要な役割を果たす (式 2.40) が、 $E_{\text{MUCA}}(E; T_0)$ はあらかじめわかっていない。そのため、短いシミュレーションの繰り返しによって、試行錯誤しながら $E_{\text{MUCA}}(E; T_0)$ を推定する手法が逐次近似法である。逐次近似法は次のような関係式で実行される。

はじめに、 $E_{\text{MUCA}}(E; T_0) = E$ として、通常のカノニカル分布を再現するシミュレーションを実行する。このとき、温度 T_0 におけるカノニカル分布 $P_{\text{CA}}(E; T_0)$ が実現される。

$$P_{\text{CA}}(E; T_0) = \frac{1}{Z_{\text{CA}}(T_0)} g(E) \exp\left(-\frac{E}{k_{\text{B}}T_0}\right). \quad (6.1)$$

ここで $Z_{\text{CA}}(T_0)$ は系の分配関数であり、

$$Z_{\text{CA}}(T_0) = \sum_E g(E) \exp\left(-\frac{E}{k_{\text{B}}T_0}\right) \quad (6.2)$$

で定義される。カノニカル分布に従ったシミュレーション中、一定の間隔で E の測定値を保存することで、ヒストグラム $H_{\text{CA}}(E; T_0)$ を作成できる。 $H_{\text{CA}}(E; T_0)$ を全サンプル数で割ることで、シミュレーションから確率分布 $\tilde{P}_{\text{CA}}(E; T_0)$ が得られる。これは式 (6.1) と、

$$\tilde{P}_{\text{CA}}(E; T_0) = \frac{1}{Z_{\text{CA}}(T_0)} g(E) \exp\left(-\frac{E}{k_{\text{B}}T_0}\right). \quad (6.3)$$

という関係がある。従って、シミュレーションによって得られた $\tilde{P}_{\text{CA}}(E; T_0)$ と状態密度 $g(E)$ との間には、

$$g(E) = \tilde{P}_{\text{CA}}(E; T_0) Z_{\text{CA}}(T_0) \exp\left(\frac{E}{k_{\text{B}}T_0}\right). \quad (6.4)$$

という関係がある。両辺の対数を取り $k_{\text{B}}T_0$ 倍することで、(式 2.40) と定数項 ($k_{\text{B}}T_0 \ln Z_{\text{CA}}$) を除いて、

$$\begin{aligned} E_{\text{MUCA}}(E; T_0) &= k_{\text{B}}T_0 \ln g(E) \\ &= E + k_{\text{B}}T_0 \ln \tilde{P}_{\text{CA}}(E; T_0). \end{aligned} \quad (6.5)$$

という関係があり、 $E_{\text{MUCA}}(E; T_0)$ を推定することができる。 $E_{\text{MUCA}}(E; T_0)$ が、十分に平坦なマルチカノニカル分布を形成できないのであれば、 $E_{\text{MUCA}}(E; T_0)$ を重みとしたマルチカノニカルシミュレーションを再び行い、同様のプロセスによって、

$$\begin{aligned} E_{\text{MUCA}}^{i+1}(E; T_0) &= k_{\text{B}}T_0 g^i(E) \\ &= E_{\text{MUCA}}^i(E; T_0) + k_{\text{B}}T_0 \ln \tilde{P}_{\text{MUCA}}^i(E; T_0). \end{aligned} \quad (6.6)$$

$E_{\text{MUCA}}^i(E; T_0)$ を求める。ここで上付きの i はイテレーションの回数を表す。

この MUCA MD 法における逐次近似法では、ヒストグラムのサンプル数が少ない領域 (すなわち確率分布の裾) では、数値的な問題が生じやすい。これは、MD シミュレーションの運動方程式 (2.42)-(2.45) で、 $\partial E_{\text{MUCA}}^i(E; T_0)/\partial E$ を求める必要があり、式 (6.6) よりシミュレーションで求められた確率分布 $\ln \tilde{P}_{\text{MUCA}}^i(E; T_0)$ の微分につながるためである。サンプル数が少ないエネルギー領域では、 $\ln \tilde{P}_{\text{MUCA}}^i(E; T_0)$ の微分は大きく変化しやすいことと、統計的に誤差が大きくなるので、数値的な誤差が生じやすい。

6.2.2 Lorentz 型の逐次近似法

参考文献 [8] では、確率分布 $\ln \tilde{P}_{\text{MUCA}}^i(E; T_0)$ に次のような Lorentz 型の関数へのフィッティングを行うことで、数値的な問題を克服できることが指摘されている。Lorentzian Fitting 用いた場合の確率分布 $P_{\text{Lorentzian}}(E; T_0)$ は、

$$P_{\text{Lorentzian}}(E; T_0) = \frac{1}{N} \sum_{n=1}^{N_E} H(E_n) \frac{1}{2\pi\delta E} \frac{1}{1 + \left(\frac{E-E_n}{\delta E}\right)^2}. \quad (6.7)$$

$$N = \sum_{n=1}^{N_E} H(E_n). \quad (6.8)$$

で行われる。ここでは、ヒストグラム $H(E_n)$ を作成するためのエネルギーの幅を δE とした。 E_1 はヒストグラムの最小エネルギー値に、 E_{N_E} はヒストグラムの最大エネルギー値に対応する。 $H(E)$ を Lorentzian の和による重み付けで和をとることにより、エネルギーの端の領域においてもなめらかな確率分布が得られ、数値的な発散の問題が緩和される。

以上より、Lorentzian Fitting を取り入れた逐次法は次のような手順で実行される。

1. $E_{\text{MUCA}}(E; T_0) = E$ とする。
2. $E_{\text{MUCA}}(E; T_0)$ を使用した短い MD を行い、ヒストグラム $H(E)^i$ を作成する。
3. 式 (6.7),(6.8) より、 $H(E)^i$ から $P_{\text{Lorentzian}}^i(E; T_0)$ を求める。
4. $P_{\text{Lorentzian}}^i(E; T_0)$ から式 (6.5) により、 $E_{\text{MUCA}}(E; T_0)$ を推定する。
5. 手順 2 へ戻る

この過程を繰り返し、十分な精度の $E_{\text{MUCA}}(E; T_0)$ を推定する。このような逐次近似法は、実際に使用すると $E_{\text{MUCA}}(E; T_0)$ を過剰に修正してしまうなど、満足のいく $E_{\text{MUCA}}(E; T_0)$ を得るまでに時間を要することもある。そのため、MD 法の場合はレプリカ交換法を使用して $E_{\text{MUCA}}(E; T_0)$ を求める REMUCA 法などが提案されている [5, 6]。

6.3 Lorentzian 型のフィッティングを考慮したワン-ランダウ法

MC 法の場合、MUCA 重み因子を決める際には Wang-Landau 法が有効であった。Wang-Landau 法では、MC STEP ごとに状態密度 $\tilde{g}(E)$ を動的に更新 ($\tilde{g}(E) \rightarrow f \times \tilde{g}(E)$) することで、状態密度 (すなわち $E_{\text{WL}}(E; T_0) (= E_{\text{MUCA}}(E; T_0))$) を求めた。ここで、 f は重みの修正因子であり、シミュレーションが進むにつれて減少する値である (第 2 章)。MD シミュレーションにおける重み修正過程は、式 (6.5) から、

$$E_{\text{WL}}(E; T_0) = k_B T_0 \ln g(E) \rightarrow k_B T_0 (\ln g(E) + \ln f), \quad (6.9)$$

$$= k_B T_0 (\ln g(E)) + f. \quad (6.10)$$

となる。ここで f は任意の値のため、 $\ln f \rightarrow f$ とした。従って、MD は、次の手順によって実行される。

1. 運動方程式に従って 1 MD STEP だけ系の状態を変化させる。
2. $E_{\text{WL}}(E; T_0) \rightarrow E_{\text{WL}}(E; T_0) + f$ としてワン-ランダウポテンシャルエネルギーを更新する。
3. 一定の MD STEP でヒストグラム $H(E)$ を確認する。十分に平坦であれば、 $f \rightarrow f/2, H(E) = 0$ とする。
4. $f < f_{\text{final}}$ であれば、シミュレーションを終える。
5. 手順 1 へ戻る。

しかし、以上の手順で MD を実行した場合は、毎 MD STEP ごとに重みを更新することになり、サンプル数の少なさが原因で $E_{\text{WL}}(E; T_0)$ の数値微分が不安定となる。WL MD ではこの問題が課題となっており、これを回避するため、様々な手法が試みられている [9, 10, 11]

この問題に対処する方法として、本章では、MUCA MD における Lorentzian Fitting を取り入れた Modified Wang-Landau (MWL) 法を提案した。この手法における $E_{\text{WL}}(E; T_0)$ は、MUCA MD のようにある程度のサンプリングを終えた後、Lorentzian Fitting によって求められた $P_{\text{Lorentzian}}^i(E; T_0)$ から、

$$E_{\text{MUCA}}^{i+1}(E; T_0) = E_{\text{MUCA}}^i(E; T_0) + f \times k_B T_0 \ln P_{\text{Lorentzian}}^i(E; T_0). \quad (6.11)$$

によって更新される。ここでは、Lorentzian Fitting を取り入れることによって、数値微分の不安定性の問題が緩和される。また、Wang-Landau 法のように修正因子 f を取り入れ、シミュレーションの時間発展とともに f を小さくすることによって、MUCA の逐次法における $E_{\text{MUCA}}(E; T_0)$ を過剰な修正が調節できるようになっている。

従って、MWL MD は以下のようにして実行される。

1. $E_{\text{WL}}(E; T_0) = E$ とする。
2. $E_{\text{WL}}(E; T_0)$ を使用した短い MD(逐次法よりも十分短くて良い) を行い、ヒストグラム $H(E)^i$ を作成する。
3. 式 (6.7),(6.8) により、 $H(E)^i$ から $P_{\text{Lorentzian}}^i(E; T_0)$ を求める。
4. 式 (6.11) により、 $E_{\text{WL}}^i(E; T_0)$ から、 $E_{\text{WL}}^{i+1}(E; T_0)$ を推定する。
5. 一定の MD STEP でヒストグラム $H(E)$ を確認する。十分に平坦であれば、 $f \rightarrow f/2, H(E) = 0$ とする。
6. $f < f_{\text{final}}$ であれば、シミュレーションを終える。
7. 手順 2 へ戻る

このようなアルゴリズムを組むことによって、MUCA における Lorentzian Fitting 法を取り入れた逐次法と、Wang-Landau 法の利点を取り入れ、MUCA 重みの決定ができる。

6.4 結論

通常の Wang-Landau 分子動力学法アルゴリズムでは、シミュレーションの初期段階に、アルゴリズムとして非物理的な状態が生じやすく、シミュレーションが破綻することが報告されている。本章では、マルチカノニカル MD 法で提案された Lorentzian Fitting Algorithm を Wang-Landau 法に組み込むことで、アルゴリズムに由来する破綻を回避するアルゴリズム (Modified Wang-Landau 法) を提案した。これにより、タンパク質のような複雑系であっても、安定した Wang-Landau 法が期待される。現在、シミュレーションを実行中である。

References

- [1] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, *Chem. Phys. Lett.*, **259**, 321 (1996).
- [2] N. Nakajima, H. Nakamura, and A. Kidera, *J. Phys. Chem. ,* **101**, 817(1997).
- [3] J. Ikebe, D. M. Standley, H. Nakamura, and J. Higo, *Protein Sci.*, **20**, 187 (2011).
- [4] J. Higo, J. Ikebe, N. Kamiya, and H. Nakamura, *Biophys. Rev.*, **4**, 27 (2012).
- [5] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6664 (2003).
- [6] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6676 (2003).
- [7] J. G. Kim, Y. Fukunishi, and H. Nakamura, *Phys. Rev. E*, **70**, 057103 (2004).
- [8] H. Shimizu, K. Uehara, K. Yamamoto., and Y. Hiwatari, *Mol. Simul.*, **22**, 285 (1999).
- [9] J. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.*, **97**, 050601 (2006).
- [10] J. Kim, J. E. Straub, and T. Keyes, *J. Chem. Phys.*, **126**, 135101 (2007).
- [11] J. Kim, J. E. Straub, and T. Keyes, *J. Phys. Chem. B*, **116**, 8646 (2012).

第7章 本論文の結論

本研究では、状態密度を高精度で推定する計算プロトコル (REWL-MUCAREM) を提案し、これを性質の異なる3つの系に適用することで、REWL-MUCAREMの有効性を示した。

はじめに第2章で、REWL-MUCAREMを提案した。REWL-MUCAREMでは、2つの拡張アンサンブル法、すなわち Replica-Exchange Wang-Landau (REWL) 法と Multicanonical Replica-Exchange 法 (MUCAREM) を適切な順序で組み合わせ、状態密度の推定する。

第3章では、REWL-MUCAREMの有効性を検証するために、REWL、MUCAREM、REWL-MUCAREMの3つの方法で正方格子イジングモデルのMCシミュレーションを実行し、得られた状態密度を比較した。その結果、REWL-MUCAREMによって得られた状態密度は、それぞれの方法で単独に推定された状態密度よりも高精度であることが示された。

第4章では、REWL-MUCAREMプロトコルの応用として、氷 I_h 系の残余エントロピーを推定した。氷 I_h 系は、絶対零度近くで有限の残余エントロピーが実験的に観測された系である。その残余エントロピーの起源に関する理論的説明は、氷中の水分子の配向に由来する Ice rule によって提案された。氷 I_h の残余エントロピーは、理論的にも実験的にも推定するのが難しい物理量の1つであるが、コンピューターシミュレーションによる推定は、近年の計算科学の発展に伴って、従来のアプローチで得られる推定と同等以上の精度を得ることが出来るようになってきた。氷 I_h の残余エントロピーは計算アルゴリズムを検証するベンチマークのひとつになりつつある。しかし、そのシミュレーションの結果には小さな不一致が見て取れる。すなわち氷 I_h の正確な残余エントロピーはいまだ得られていないのである。本章では、REWL-MUCAREM MC プロトコルによって氷 I_h の残余エントロピーを高精度で推定した。その結果は、本研究グループの前回の推定値を改善し、他のシミュレーション手法を使用したグループと良く一致した。また、乱数の均一性の重要性についても議論した。マルチカノニカル法による状態密度の推定では、大規模複雑系に適用する場合、均一性の高い乱数を用いなければ、信頼性の低い結果につながることを示した。

第5章では、タンパク質系のフォールディング問題について、REWL-MUCAREM プロトコルを適用した。計算対象は、ヘリックス-コイル構造転移を示すとわかっているアラニン重合体とした。ここでは、タンパク質系における REWL-MUCAREM MC プロトコルのサンプリング効率を上げるために、Genetic Algorithm を追加した REWL-MUCAREM でシミュレーションを行った。結果として、タンパク質のような複雑多体系においても、REWL-MUCAREM の実用性が示された。またフォールディング過程の解析には、ミクロカノニカルエントロピーを使用した相転移分類法である Microcanonical Inflection-Point Analysis Method を使用した。ミクロカノニカルエントロピーは状態密度と密接な関係があるため、高精度で状態密度を推定する REWL-MUCAREM は Microcanonical Inflection-

Point Analysis Method と相性が良いといえる。結果として、アラニン重合体のヘリックス-コイル構造転移は、水の影響を考慮しなかった場合に 1 次転移、陰溶媒モデルによって水の影響を考慮した場合に 2 次転移のような振る舞いを持つことが示唆された。

第 6 章では、REWL-MUCAREM 法の分子動力学法への拡張を考えた。通常 **Wang-Landau** 分子動力学法アルゴリズムでは、シミュレーションの初期段階にアルゴリズムとして非物理的な状態が生じやすく、シミュレーションが破綻することが報告されている。この問題を回避するため、本章では、マルチカノニカル MD 法で提案された **Lorentzian Fitting Algorithm** を **Wang-Landau** 法に組み込むアルゴリズムを提案した。これにより、高分子系など MD 法が好まれる系においても、REWL-MUCAREM の適用が期待される。

以上のように本論文では、性質の異なる 3 つの多自由度複雑系に本研究で提案した REWL-MUCAREM を適用し、その有効性を示した。

第8章 謝辞

指導教官である岡本祐幸教授には、修士課程・博士課程の5年間を通して研究に関する数多くの議論をさせていただきました。衝突することもありましたが、どのような時でも寛大に接していただいたことで、忌憚なく意見を述べることができ、充実した議論ができました。また私が海外留学を希望した際は、快く送り出していただいたことや、海外にいても研究のサポートをしていただいたこと、大変嬉しかったです。岡本先生のもとで研究できたことを誇りに思います。ここにあらためて感謝申し上げます。

倭剛久准教授には、研究発表において厳しいながらも的確な意見を頂戴しました。先生のご意見を参考にすることで、魅力的なプレゼンテーション方など、自分では気を配らないような点に気付くことができました。感謝申し上げます。

木村明洋助教には、研究から雑談まで様々なお話させていただきました。気さくに話をさせていただけたおかげで、気軽に悩みを相談でき、安心して研究生活を送ることができました。ありがとうございました。

共同研究者である中京大学の六車千鶴教授には、修士課程1年生の頃から、氷の残余エントロピー計算の研究で大変お世話になりました。六車先生の化学的見地からのコメントは、私にとって新鮮なものでした。お忙しい中、多くの助言をしていただいたこと、感謝いたします。

University of Georgia の David P. Landau 教授には、留学の際に大変お世話になりました。拙い英語でしかコミュニケーションを取れない私と根気強く議論していただいたこと、感謝いたします。研究だけでなく、慣れない土地での生活を気にかけてくださるなど、心よりおもてなししていただいたこと、嬉しかったです。ジョージア大学で過ごした日々は、私にとってかけがえのない時間となりました。あらためてお礼申し上げます。

同じく University of Georgia の Michael Bachmann 教授には、Microcanonical Inflection-Point Analysis Method の研究を提案して頂いたとともに、タンパク質のフォールディング研究の議論していただいたこと、感謝いたします。また Bachmann 教授の授業時間の中で、私に研究発表の機会を設けてくださったことで、UGA CSP の皆様とコミュニケーションが取りやすくなりました。ありがとうございました。

同じく University of Georgia の Alfred C. K. Farris 博士には、レプリカ交換ワン-ランダウ法について、多くの議論をしていただいたこと、感謝いたします。シミュレーションにおける物理現象の見方や、最新の手法・現象について議論できたことで、大きな刺激を受けました。また同じ手法を使う研究者と出会えたことは、私にとって大きな励みとなりました。ありがとうございました。

信州大学の樋口雅彦教授、並びに志水久准教授には、信州大学卒業後も多くのアドバイスを頂戴しました。博士課程への進学に悩んでいた時も、背中を押していただいたこと、心より感謝いたします。1年間だけでしたが、先生方の指導のもと信州大学物性理論研究室で過ごせたことは、研究での信念を築く上で大事な期間となりました。あらためて

お礼申し上げます。

本研究室 OB であり、現東京大学の永井哲郎助教には、研究についてのご相談だけでなく、博士進学の際に OB としてアドバイスしていただいたこと、感謝いたします。

研究室の先輩である塚本修一郎氏には、TB 研究室の研究のみならず、幅広い分野での議論をしていただいたこと、感謝いたします。私生活の面でも遊びに誘っていただくなど、塚本氏のおかげで楽しい研究室生活を送ることができました。お礼申し上げます。

同じく先輩である松原大貴氏には、何度も研究に関する詳細な議論をさせていただきました。松原氏の物理全般にわたる広範囲の知見には幾度も驚かされるとともに、大きな刺激を受けました。感謝いたします。

同じく先輩である多知裕平氏からは、研究の大枠をとらえることの重要性を学ぶとともに、TB 研究室以外の最新の研究情報についてお話していただき、大きな刺激を受けました。また私生活においてもご相談に乗っていただくなど、多くのお時間を頂戴したこと、感謝いたします。

同じく先輩である南隆明氏には、卒業後も仲良くしていただき感謝いたします。社会人としての生活のアドバイスなど、大変参考になりました。ありがとうございました。

塚本修一郎氏、松原大貴氏、多知裕平氏、南隆明氏には、卒業後も気にかけてくださっていること、大変感謝いたします。

同期の石田兼三氏、太田匡隆氏、清水政宏氏、都竹華代氏とは、学生生活全般を通して仲良くさせていただきました。皆様と過ごした学生生活は私にとってかけがえのないものです。

後輩である野田考平氏、高田勇氣氏、和田大宗氏には、後輩ながらも多くのことを学ばせていただきました。皆様がひた向きに研究に取り組む姿を見て、何度も気を引き締めました。

秘書の加藤和歌さんには、Georgia 大学への留学の準備をはじめとして、学生生活を送る上で多くのサポートをしていただきました。ここにあらためて感謝申し上げます。

そして、これまでに関わった理論生物化学物理 (TB) 研究室及び名古屋大学生物物理グループの皆様や、博士課程への進学をサポートしていただいた父母兄、友人、知人に感謝いたします。

多くの人の支えによって、本論文を執筆することができました。心より、感謝申し上げます。